Fiber-connected Massively Distributed Antenna Systems: Energy Efficiency and Interference Management

by

Haoming Li

B.Eng., Sichuan University, 1994

M.Sc., Sichuan University, 1997

M.A.Sc., Carleton University, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Electrical and Computer Engineering)

The University Of British Columbia
(Vancouver)

October 2013

Abstract

The density of wireless access nodes keeps increasing to provide ubiquitous wireless access and meet the ever-increasing traffic demand. However, the shrinking distance among neighboring access nodes causes excessive interference and the increasing number of access nodes incurs a higher power consumption. A careful management of interference ensures a high system capacity. An improved energy efficiency in wireless access network prevents the fast growth of wireless communication systems from aggravating the global energy crisis. In this thesis, we propose a novel architecture, Fiber-connected Massively Distributed Antennas (FMDA), to address the challenges of managing interference and improving energy efficiency in wireless access networks.

A FMDA system is composed of a centralized processing system connected to a large number of antennas via optical cables. The centralized processing system processes all the radio signals and allocates all the radio resources to better manage interference; each antenna contains much simpler circuits than conventional access nodes and therefore allows a massive deployment and reduces the antenna power consumption. We first propose a novel multi-cell wireless local area network (WLAN) system based on our proposed FMDA architecture, where the centralized processing system can see the entire spectrum usage across the coverage area and control the radio signals to be sent to each antenna, thus allowing a better management of inter-cell interference. We then propose an antenna scheduling scheme in a novel cellular system composed of fiber-connected femto access nodes to manage the excessive inter-femtocell interference and reduce the energies consumed by non-sleeping access nodes, thus simultaneously improving the spectral and energy efficiency.

When the number of cooperating antennas increases, the power consumption of signal processing spikes, thus drastically degrading the overall energy efficiency due to much smaller radio transmission power levels. Focusing on two typical indoor environments, office buildings and large public venues, we propose two low-complexity downlink transmission schemes to address these energy efficiency challenges.

Preface

This thesis is based on the following publications. Dr. Victor C. M. Leung, my supervisor, co-authored all the papers and supervised all the research.

Journal papers published/submitted:

1. H. Li and V. C. M. Leung, "Low complexity zero-forcing beamforming for distributed massive MIMO systems in large public venues," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 370-382, 2013. The material is incorporated in Chapter 5.

2. H. Li and V. C. M. Leung, "Energy-efficient low-complexity zero-forcing beamforming via banded matrix inversion in indoor massively distributed antenna systems," *Submitted, under the second round of review.* The material is incorporated in Chapter 4.

3. H. Li, A. Attar, and V. C. M. Leung, "Energy conservation via antenna scheduling in fiber-connected femto base stations," *Mobile Networks and Applications*, vol. 17, no. 5, pp. 685-694, 2012. The material is incorporated in Chapter 3.

4. H. Li, J. Hajipour, A. Attar, and V. C. M. Leung, "Efficient HetNet implementation using broadband wireless access with fibre-connected massively distributed antennas architecture," *IEEE Commun. Mag.*, vol. 18, pp. 72-78, 2011. The material is incorporated in Chapter 3.

5. A. Attar, H. Li, and V. C. M. Leung, "Green last mile: How fiber-connected massively distributed antenna systems can save energy," *IEEE Commun. Mag.*, vol. 18, pp. 66-74,

2011. The material is incorporated in Chapter 3.

6. H. Li, A. Attar, V. C. M. Leung, and Q. Pang, "Collision reduction in cognitive wireless local area network over fibre," *International Journal on Advances in Internet Technology*, vol. 3, no. 1-2, pp. 1-12, 2010. The material is incorporated in Chapter 2.

Conference papers published:

1. H. Li, A. Attar, Q. Pang, and V. C. M. Leung, "Collision avoidance and mitigation in cognitive wireless local area network over fibre," *in Proc. IEEE INTERNET'09*, Cannes, France, 2009. The material is incorporated in Chapter 2.

2. H. Li, Q. Pang, and V. C. M. Leung, "Cognitive access points for dynamic radio resource management in wireless LAN over fiber," *in Proc. WWRF 20th meeting*, Ottawa, Canada, Apr. 2008. The material is incorporated in Chapter 2.

The following explains the co-authorship of each paper:

- Journal paper 3 was co-authored with Dr. Alireza Attar. I proposed the scheduling scheme, established the power consumption model, analyzed energy efficiency and conducted simulations. Dr. Attar provided inputs to the system architecture and comments on performance evaluation.

- Journal paper 4 was co-authored with Mr. Javad Hajipour and Dr. Alireza Attar. I established the system model and evaluated the system performance. Mr. Hajipour provided channel configurations and frequency-selective fading channel models. Dr. Attar proposed where Fiber-connected Massively Distributed Antennas fit in a heterogeneous network and organized the structure of the paper.

- Journal paper 5 was co-authored with Dr. Alireza Attar. I established the power consumption model, conducted simulations and wrote the simulation description and performance

evaluation in the paper. Dr. Attar organized the structure of the paper and surveyed green communication techniques in cellular networks.

- Journal paper 6 was co-authored with Dr. Alireza Attar and Dr. Qixiang Pang. I reviewed previous research in radio over fibre, proposed two interference management schemes and evaluated the system performance. Dr. Attar provided inputs to the performance evaluation. Dr. Pang provided inputs to the comparison method.

- Conference paper 1 were co-authored with Dr. Alireza Attar and Dr. Qixiang Pang. I reviewed previous research in radio over fibre, proposed two interference management schemes and evaluated the system performance. Dr. Attar provided inputs to the performance evaluation. Dr. Pang provided inputs to the comparison method and how to evaluate wireless local area network performance.

- Conference paper 2 were co-authored with Dr. Qixiang Pang. I proposed the wireless local area network over fibre architecture and described the functions of the cognitive access point. Dr. Pang provided inputs to interference scenarios in the 2.4 GHz band.

Table of Contents

List of Tables

List of Figures

List of Symbols

α	Per-wall penetration loss in linear scale, i.e., $\alpha = 10^{-L/20}$.	74
β	Linear attenuation coefficient in the unit of dB/meter	98
B	Channel bandwidth	55
$\mathcal{B}_p$	The set of banded matrices with width p	72
δ_k	The distance between the k-th user and the k-th antenna	74
Δ_{SINR}	SINR loss from using sparse precoding, defined as $\text{SINR}_{\text{dw}} - \text{SINR}_{\text{sw}}$.	82
$d_{k,b}$	The distance between the k-th user and the b-th antenna in meters	102
D	Inter-antenna distance in the unit of meter in a large public venue	116
$\mathbf{D}_p$	The difference between the true channel matrix and the banded channel matrix, given by $\mathbf{H} - \mathbf{H}_p$.	75
η	Path loss exponent	74
η_e	Energy efficiency in the unit of bit/joule	55
η_s	Average spectral efficiency per user in the unit of bps/Hz/User	55
$\mathbf{E}_p$	The difference between the sparse precoding matrix and the dense precoding matrix, given by $\mathbf{S}_p - \mathbf{W}_p$.	83
$\mathcal{ED}_\alpha$	The Toeplitz matrix with exponentially decayed off-diagonal entries at rate α, i.e., entry (i, j) is given by $\alpha^{\lvert i-j \rvert}$.	79
$\mathcal{EP}_{\alpha,\eta,\delta}$	$\mathcal{EPR}_{\alpha,\eta}$ with fixed δ, representing a Toeplitz matrix $\mathbf{H}$ where $h_{ij} = \lvert i - j \rvert^{-\eta/2}\alpha^{\lvert i-j \rvert}$ for $i \neq j$ and $\lvert\delta\rvert^{-\eta/2}\alpha^0$ for $i = j$ ($\lvert\delta\rvert < 0.5$).	79

$\mathcal{EPR}_{\alpha,\eta}$	The set of random matrices whose off-diagonal entries decay both exponentially at rate α and polynomially at rate η, i.e., entry (i,j) is given by $\lvert i-j+\delta_i\rvert^{-\eta/2}\alpha^{\lvert i-j\rvert}$, where δ_i is the normalized distance between the i-th user and the i-th antenna, subject to $\mathcal{U}(-0.5,0.5)$.	79
$\mathcal{EPRZ}_{\alpha,\eta}$	The Hadamard product of a ZMCSCG matrix and an $\mathcal{EPR}_{\alpha,\eta}$ matrix.	79
$\mathcal{EZ}_{\alpha}$	The Hadamard product of a ZMCSCG matrix and an $\mathcal{ED}_{\alpha}$ matrix.	79
$f_{k,b}$	The small-scale fading between the k-th user and the b-th antenna	115
$\mathbf{F}$	Small-scale fading matrix	74
G_r	Receiver antenna gain in dB	116
G_t	Transmitter antenna gain in dB	116
$h_{k,b}$	The channel gain between the k-th user and the b-th antenna	115
$\mathbf{H}$	Full channel matrix	16
$\mathbf{H}_p$	Banded channel matrix in Chapter 4, or multi-banded channel matrix in Chapter 5.	72
I_i	Linear-scale residual interference power of the i-th stream	78
$\bar{I}_i$	Linear-scale residual interference power of the i-th stream assuming equal stream power allocation	79
$\hat{I}_i$	Log-scale residual interference power of the i-th stream averaged over user location and fading.	80
$\hat{I}_i^{(D)}$	Log-scale residual interference power of the i-th stream averaged over user location and fading when dense precoding is used.	83
$\hat{I}_i^{(S)}$	Log-scale residual interference power of the i-th stream averaged over user location and fading when sparse precoding is used.	83
κ	Rician K-factor in linear scale	121
L	Per-wall penetration loss in dB	74

M	The number of lanes/files in an antenna grid that is deployed in a large public venue.	98
$\mathcal{MB}_{M,p}$	The set of multi-banded matrices with inter-band distance M and width p, in which entry (k,b) is zero for $\max(\lvert k\%M - b\%M\rvert, \lvert\lceil k/M\rceil - \lceil b/M\rceil\rvert) \geq p$.	100
$\mathcal{MEPR}_{M,\alpha,\eta}$	The set of multi-banded matrices defined in the context of large public venues. The matrices have exponentially and polynomially decayed off-diagonal entries and uniformly distributed user-antenna locations.	103
$\mathcal{MEPRZ}_{M,\alpha,\eta}$	The Hadamard product of an $\mathcal{MEPR}_{M,\alpha,\eta}$ matrix and a ZMCSCG matrix	103
N_c	The number of antennas in a femto-CoMP cluster	49
N_{CoMP}	Total number of femto-CoMP clusters	51
N_r	Total number of distributed antennas	49
N_s	The number of OFDM symbols per scheduling interval	87
N_{sc}	The number of subcarriers per resource block in LTE	53
N_{ts}	Total number of time slot sets in a FMDA system composed of multiple femto-CoMP clusters	49
N_U	Total number of active users in a femto-CoMP cluster	51
N_{Utotal}	Total number of active users in a FMDA system	51
O_p	Outband drop at width p, which is defined as the log-ratio of the accelerated and the normal decay in $\mathbf{W}_p$.	81
p	The width of a banded matrix where entry (k,b) is zero for $\lvert k-b\rvert \geq p$.	72
P	Per-antenna transmission power constraint	75
P_{RF}	Power consumption contributed by radio-over-fiber components	51
P_{sp}	Power consumption contributed by signal processing components	51
P_{spB}	Power consumption contributed by base amount	51
P_{spc}	Sum transmission power constraint	75

P_{spMS}	Power consumption contributed by zero-forcing beamforming and user scheduling	51
P_{tot}	Total power consumption of a FMDA system	51
P_{tx}	Transmission power per antenna	54
$\mathbf{S}_p$	Sparse precoding matrix	75
$\mathrm{SINR}_{\mathrm{dw}}$	SINR measured in dB when dense precoding is used.	82
$\mathrm{SINR}_{\mathrm{sw}}$	SINR measured in dB when sparse precoding is used.	82
$\mathrm{SIR}_{\mathrm{dw}}$	SIR measured in dB when dense precoding is used.	112
$\mathrm{SIR}_{\mathrm{sw}}$	SIR measured in dB when sparse precoding is used.	114
$\mathbf{V}$	Path loss matrix	74
$\mathbf{W}$	Precoding matrix	16
$\mathbf{W}_p$	Dense precoding matrix	75

Glossary

AP access point

BS base station

BSS Basic Service Set

CBR constant bit-rate

CDF cumulative distribution function

CogAP cognitive access point

CoMP coordinated multi-point transmissions

CPS centralized processing system

CPRI common public radio interface

C-RAN cloud-based radio access network

CSI channel state information

DAS distributed antenna system

EGC equal-gain combing

ESS Extended Service Set

FDD frequency division duplex

flop floating-point operation

FMDA Fiber-connected Massively Distributed Antennas

FTP File Transfer Protocol

GOPS giga-operation-per-second

HetNet Heterogeneous Networks

i.i.d. independent and identically distributed

IPTV IP television

ISM industrial, scientific, and medical

LTE Long-Term Evolution

MAC medium access control

MIMO multiple-input multiple-output

MRC maximum-ratio combining

OFDM orthogonal frequency-division multiplexing

PAPC per-antenna power constraint

PER packet error rate

PF proportional fairness

PF-SUS proportional fairness scheduler based on semi-orthogonal user selection

RoF radio over fiber

RRH remote radio head

RR-SUS round-robin scheduler based on semi-orthogonal user selection

SI synchronization interval

SIR signal-to-interference-ratio

SINR signal-to-interference-noise-ratio

SNR signal-to-noise-ratio

SPC sum power constraint

TCP Transmission Control Protocol

TDD time division duplex

VoIP Voice over Internet Protocol

WF waterfilling

WLAN wireless local area network

ZFBF zero-forcing beamforming

ZMCSCG zero-mean circularly symmetric complex Gaussian

Acknowledgments

My greatest thanks go to my supervisor, Dr. Victor C. M. Leung, for his guidance on my research directions, knowledge and insights in my research areas, valuable advice on my research methods, continuous encouragements to my research findings, and consistent supports on my study and research during all these years. Without his guidance and supports, this thesis would have been impossible.

I would like to thank my supervisory committee members, Dr. Lukas Chrostowski and Dr. Lutz Lampe, for their valuable advice on radio-over-fiber techniques, energy-efficiency, and user-traffic considerations. I would like to thank Dr. Vikram Krishnamurthy for his valuable comments on the analysis. I would like to thank Dr. Peyman Servati for chairing my PhD department exam. I am grateful of Dr. Claude Desset (IMEC) for clarifying the power consumption model in EARTH project D4.3.

I thank my colleagues, Dr. Alireza Attar, Dr. Qixiang Pang and Dr. Shengchun Huang, for their knowledge and wisdom that ensured the quality of our numerous research collaborations. I thank Dr. Jun-bae Seo for his insights that greatly helped my studies during my first two years in UBC. I thank Mr. SeyedAli Hosseininezhad for helps on NS-2 simulator and Linux environments, Mr. Javad Hajipour for helps on frequency-selective channel modeling, Mr. Liming Chen for sharing his knowledge on Cloud-RAN, and Dr. Hu Jin for organizing our biweekly seminars. I thank Dr. Arghavan Emami and Mr. Robert White (Radio Science Lab of UBC) for their kind helps on Wireless InSite® 3D ray-tracing simulation tool.

I would like to thank Mr. Saad Mohaboob, Mr. Chinmaya Mahapatra and Dr. Roberto Rosales,

for their knowledge and helps during the preparation of an MIMO-OFDM testbed. I thank Mr. Sebastien Maury and Mr. Maxime Dumas for many discussions over radio front-ends and digital signal processing platforms that have deepened my understanding in wireless communication systems.

Lastly, I would like to thank my father-in-law (who passed away) for his encouragement and optimism, and my mother-in-law and my wife who had to face the very difficult time while still giving me tremendous amount of supports during my last two years of PhD study. Without their supports, it would have been very difficult for me to complete this thesis.

The research in this thesis were supported in part by a grant from Bell Canada through the Bell University Laboratories program, the Canadian Natural Sciences and Engineering Research Council through grant CRDPJ 320552-04 and STPGP 396756, The University of British Columbia Four-Year-Fellowship (4YF), and The University of British Columbia PhD Tuition Scholarship.

Dedication

To my parents

Chapter 1

Introduction

Wireless communications systems need to provide a wide coverage and a large system capacity to address the ever-increasing need of ubiquitous wireless access, to support the growing traffic-intense applications including multimedia streaming and file storage in Internet, and to cater the increasing number of people on Earth. Wireless communication systems also need to accomplish these tasks in a cost-effective and energy-efficient way to meet the expectation of industries and follow the increasing global consciousness in energy crisis. Given finite radio frequency resources, techniques in the areas of radio over fiber (RoF), distributed antenna system (DAS), multiple-input multiple-output (MIMO) systems and Heterogeneous Networks (HetNet) have been developed to enhance system coverage and to increase system capacity through efficient frequency reuse and MIMO techniques. The area of green communications has also drawn attentions from telecommunication operators in reducing the infrastructure energy cost and from mobile terminal manufacturers in extending battery life. However, each of these research areas has its own challenges on achieving their goals. This motivates us to propose a novel wireless access architecture, Fiber-connected Massively Distributed Antennas (FMDA), which combines the techniques that have been developed in these research areas to increase system capacity in an energy-efficient manner. In the following we survey the advances in the aforementioned research areas, present the architecture of FMDA, state our research questions, and then summarize our main contributions. In the end of this chapter, we state the organization of this thesis.

1.1 Overview of Related Work

1.1.1 Distributed Antenna System

DAS were proposed to provide more reliable coverage by splitting the transmission power among spatially distributed antennas [1]. DAS is needed in rural areas where it is expensive to install base stations (BSs), in large buildings that wireless signals are hard to penetrate from outside (e.g., government buildings and hospitals), and in crowded large public venues where a large number of antennas are needed to provide a sufficient system capacity (e.g., stadiums and convention centers).

The DAS in the setting of outdoor cellular communication systems has evolved into coordinated multi-point transmission [2], cooperative base station [3], or multi-cell processing [4] by combining MIMO technique, which has been adopted in Long-Term Evolution (LTE) networks [5, 6]. Researchers in coordinated multi-point transmissions are investigating how BSs exchange information to smoothly handle handover and increase both uplink and downlink throughput [4, 7–14]. These studies are focused on cellular systems that use licensed bands to serve large number of users per cell through intelligent BSs. In practice, however, the cooperating BSs often communicate through a backhaul link with a finite capacity and a non-negligible delay. Therefore, many research efforts have been put into how to improve system performance under a backhaul capacity constraint [15, 16]. There are also research in deploying distributed antennas in large buildings to improve the throughput of single user through single-user MIMO techniques [17–24], or use multi-user MIMO techniques to simultaneously transmit multiple packets to multiple users [25, 26].

In hyper-dense venues such as stadiums and convention centers, wireless local area network (WLAN) operators are deploying a large number of distributed access point (AP)s to meet the increasing demand of Internet access and content delivery during live events. To mitigate the inter-AP interference that increases with the AP density, designers rely on orthogonal channels, thorough radio frequency planning and careful tuning of directional antennas. However, these measures are

limited by the number of channels available to the operator, and the time and the cost to conduct a realistic radio frequency survey. Therefore, there has been a trend of mounting the APs under-seat [27] or under-floor [28] to be closer to the mobile terminals and achieve higher inter-AP isolations by taking advantage of the heavy human body penetration loss, thus allowing more APs being deployed to provide a very-high system capacity. Under-seat or under-floor mounting also has advantages in aesthetics and ease of installation [29]. Femtocells [30] follow a similar path to improve indoor wireless coverage by locating the antennas closer to users, which will be reviewed in Section 1.1.4.

1.1.2 Radio over Fiber

A RoF system [31], composed of a central controller unit and multiple fiber-connected antennas, was originally proposed to extend system coverage by simulcasting messages through antennas. It is transparent to traffic being carried and therefore future-proof since only baseband processing units in the central controller unit need an update when wireless technology evolves. Studies in [32–35] investigated the feasibility of carrying a wide-band signal through RoF systems, specifically, multiple WLAN channels over low-cost multi-mode optical fibers. And it was shown that WLAN medium access control mechanism in RoF systems are not affected by the optical delay [36–38]. Commercial fiber-based indoor wireless networks have been deployed to deliver both WLAN and cellular communication signals inside large civil structures such as stadiums [39, 40], subways [41], hospitals [42], business buildings and shopping malls [43]. RoF has also been used in high-altitude balloons to provide radio communication and video surveillance in battlefields due to the light weight and high bandwidth of optical fibers [44]. However, most of these studies are concentrated on the design of optical-electric converter components and most of applications are only using simulcasting mode to improve coverage in hard-to-reach areas. Thus far, two abilities of the central controller unit in a RoF system have not been explored. One is to see the complete picture of radio spectrum usage in the coverage area. Another is to intelligently combine signals transmitted through antennas to make use of MIMO technique.

1.1.3 Massive MIMO

Massive array-MIMO addresses ever-increasing traffic demand by densifying radio access networks. The spectral efficiency advantages of massive array-MIMO systems have been shown in analysis under match filter precoding [45], zero-forcing beamforming [46], and demonstrated in experiments [47]. Compared with array MIMO systems, distributed MIMO systems improve link reliability by locating antennas closer to users and therefore improve spectral efficiency both indoors [1, 14, 48] and outdoors [4, 49] while requiring less transmission power. When the number of users being served, K, is smaller than the coverage radius measured in the number of wavelengths of the carrier signal, a distributed MIMO system achieves a capacity linear in K [50], demonstrating its potential to satisfy high traffic demands in large public venues such as stadiums, arenas and convention centers. While many research in the area of massively distributed MIMO systems are focused on asymptotic analysis when the number of antennas N grows to infinity, few research are conducted to reduce the demanding signal processing complexity when N is large.

1.1.4 Heterogeneous Networks

As the demand for indoor wireless connectivity increases, shrinking the cell sizes through installation of more macro-BSs is no more a sustainable solution to handle the traffic load. An emerging solution to address the increased network capacity demand is to deploy smaller scale access nodes which can form femtocells or picocells and facilitate a higher level of reuse of available spectrum, thereby shifting the cell planning model in future wireless networks towards universal frequency reuse pattern. The creation of femtocells and picocells within a macro cell shortens the communication link, especially for indoor users, which in turn reduces the co-channel interference level in the network. As the capacity of most packet-based wireless communication technologies is interference-limited, such a reduction in network interference level is directly translated into capacity enhancement of the system. On the other hand new forms of interference, mainly the mutual interference between femtocell or picocell access nodes and macro cell users arise in this new net-

work architecture paradigm. The research community in academia and industry has started to more carefully address such coexistence problems under the umbrella term *HetNet*. While HetNet previously refers to cooperation of different wireless technologies, e.g., a complementary operation of a WLAN in conjunction with a cellular network, in this section we limit our discussions HetNet based on a single wireless standard. Note that in the context of HetNet, the prefixes macro-, micro-, pico- and femto- are used to indicate the relative sizes of the cell coverage areas and do not imply any exact ratio of their coverage areas. For example, a pico-cell does not cover exactly 10^3 times larger area than a femto-cell.

HetNet aims to improve the network performance via coordinating the resource allocation and service delivery using nodes with different transmission capabilities within a given cell. In June 2009, HetNet became a study item for LTE-Advanced networks [51]. A HetNet in LTE-Advanced context is comprised of a macro cell associated with a macro-BS, and a random distribution of lower power BSs, such as micro-, pico- and femto-BSs, which form their own closed subscriber groups [51]. Thus, HetNet will not be a specific technology but will represents a network of a specific radio access technology where BSs might have different radio capabilities, ranging from macro- and micro- to femto- and pico-cell scales. As subscriber groups are closed, a user can only associate with its own BS even when nearby BSs have stronger signal strengths. The resulting independent operations among neighboring cells incur the following interference scenarios [52]. In the downlink, a given macro cell user will be interfered by nearby users that associate with micro-, pico- or femto-cells. Further, a neighboring femto-BS will interfere with users associated with another femto-BS. In the uplink, on the other hand, macro cell users will create significant interference to neighboring femto-BSs, especially when the macro cell users are near the cell edges and therefore are transmitting at the largest power level. The interference scenarios become more complicated in LTE time division duplex systems [53], where eight scenarios are described and five of them are given a high priority. Note that in the context of LTE, the above interference scenarios need to be considered in both data and control channels, and in many cases it is more important to mitigate interference in control channels, which carry synchronization and broadcasting

information that allow a user to successfully attach to the network.

The coexistence challenge in LTE has ignited the research in more intelligent BSs that self-organize and negotiate resources with neighboring BSs, termed *Self-Organizing Networks* [54, 55]. The coexistence between a macro cell and femtocells ignited both data and control channel protection schemes, which are being considered in inter-cell interference coordination scheme in LTE release 8/9 [56], enhanced inter-cell interference coordination scheme in LTE release 10 [53, 57, 58] and enhanced interference management and traffic adaptation scheme in LTE release 11/12 [59, 60]. However, these inter-cell interference management schemes are still limited by the inefficient communications between the macro-BS and the femto-BSs, which have to cross X2 interface that is implemented on a data packet network and is often susceptible to capacity bottle neck and delay.

1.1.5 Green Communications

The growing concern over the power consumption aspect of wireless and cellular networks has triggered a new research initiative in academia and industry, referred to as *green communication* [61]. A surge in wireless network power consumption can be directly translated into increasing CO_2 emission. Vodafone estimates the total gross CO_2 emissions of this firm at 1,676,949 tonnes by BSs and 615,612 tonnes by its other network equipment as compared with 251,358 tonnes by offices and 40,446 tonnes by retail stores for 2012/2013 [62], indicating that the access network is the main source of power consumption and power inefficiency.

To address the power efficiency of cellular/broadband systems, one can target component, link and/or network level power efficiency solutions [63]. Component power efficiency addresses performance enhancements of the electrical and electronic components in the system, such as power amplifiers, while sustaining power conservation. Link level power efficiency aims at delivering higher throughput to end users while maintaining or even decreasing the transmit power budget at the link level. Finally, network level solutions address the effects of static/dynamic network topologies, among other factors, on the power consumption of the network.

The main components of a typical macro cell site include a power supply unit and a cooling system, and the BS units including base band processing units and radio frequency units (power amplifier, low noise amplifier and antenna feed). The authors in [61] estimate that of the total power consumption of a typical macro cell site, 43% is consumed by the cooling system followed by 41% by the BS itself. Within the BS, the shares of the main power sinks are the feeder at 44%, radio frequency conversion and power amplifier at 15% and signal processors at 9%.

In [64] the authors present a link level power efficiency analysis of cellular networks, by developing a framework to study the effect of BS cooperation, such as coordinated multi-point transmission techniques. Further, an early deliverable in EARTH project [65, 66] develops power consumption models of various types of cellular BSs, from a component energy-efficiency point of view, in order to analyze network level green solutions based on deployment strategies in heterogeneous cellular systems. A recent deliverable of EARTH project [67, 68] presented a scalable power consumption model based on measurements in realistic LTE frequency division duplex systems including macro-BS, micro-BS, pico-BS and femto-BS. The comprehensive model scales with the number of antennas, the channel bandwidth, the coding rate, the modulation type, and the semiconductor process used in digital signal processors. The model is separately modeled for uplink and downlink. These power consumption models reveal that in femto- and pico-BSs where radio transmission power is largely reduced and active cooling units are removed, signal processing is contributing more in total power consumption.

In massive array-MIMO systems, two recent contributions analyzed the energy efficiency in the uplink [69] and downlink [70]. From the path loss aspect, however, it would be more energy-efficient to locate the antennas closer to the users, which motivates the use of a large number of antennas. In such systems, the feeder loss disappears due to the collocation of a power amplifier and the corresponding antenna. The power amplifier contribution also shrinks because a much smaller transmission power is used, which consequently allows the use of passive cooling at the distributed antennas. As a result, signal processing algorithms have a higher impact on the energy efficiency of access networks. Since the cooling is now only needed at the central unit where

signal processing units consume the most part of energies, a simpler signal processing algorithm also reduces the power consumption of cooling.

Another direction in green communication is to shut down devices when there is no traffic. Traffic-aware sleeping techniques have been proposed in femto-BSs [71] and WLAN APs [72] to take advantage of temporal and spatial traffic load variations. However, when the device wakes up, the mobile terminals have to wait for the device to boot and then synchronize to the network, thus reducing the user responsiveness. Therefore, the demand to provide anywhere connectivity at a high data rate as well as seamless mobility at any time requires that almost all BSs within the traditional last-mile networks should continue to operate regardless of temporal and spatial traffic load variations. The recently proposed cloud-based radio access network (C-RAN) [73, 74] is able to resolve the conflict between reducing power consumption and maintaining anywhere and anytime connectivity. In C-RAN, baseband processing of all the BSs are concentrated at a central location, termed *baseband unit*. The rest of each BS, termed *remote radio head*, only contains a power amplifier, a low-noise-amplifier, a set of analog-digital/digital-analog converter, and an optical transceiver that supports *common public radio interface*, *open base station architecture initiative*, or other communication protocols that vendors of baseband unit and remote radio head agree on. Compared with traditional HetNets, the energy efficiency of C-RAN is improved because of two facts. First, the power consumption of remote radio head is reduced due to its simplified design. If desired, remote radio heads can be swiftly switched on and off to adapt to temporal and spatial traffic load variations, which is possible due to the centralized processing in C-RAN. Second, the power consumption of signal processing can be reduced by the increasing utilization ratio of baseband processing units at the baseband unit. When a C-RAN covers both residential areas and business districts, the overall traffic becomes stable over day and night and therefore requires less number of baseband processing units.

The trend of C-RAN is apparent considering recent advances from major wireless equipment vendors, e.g., lightRadio from Alcatel-Lucent in 2011, Antenna Integrated Radio from Ericsson in 2011, Liquid Radio from Nokia Siemens Networks in 2011, AtomCell from Huawei in 2012 and

a successful lightRadio trial in Telefonica in 2012. Since C-RAN heavily depends on the baseband unit to accomplish the baseband processing, reducing signal processing power consumption becomes more important.

1.2 Motivation of Fiber-connected Massively Distributed Antennas

We first combine the innovations in DAS and RoF to form a fiber-connected DAS that is able to address the backhaul capacity challenge from coordinated multi-point transmission technique, exploit the centralized processing and sensing opportunities in RoF systems, and reduce the interference management difficulty in HetNets. We then employs massive MIMO technique in the fiber-connected DAS to deliver high-throughput wireless access to end users. The resulting networking paradigm is termed *Fiber-connected Massively Distributed Antennas*, which provides a high-capacity low-delay backhaul, enables centralized processing and sensing capabilities and harnesses the advantages of distributed networking architecture, thus facilitating interference management solutions and achieving a high energy efficiency. Because of the wide-band nature of optical cables and technology-transparent RoF technique, systems built on the FMDA architecture can support single wireless technology such as WLAN, LTE or any other future wireless technology, as well as HetNet which contains many combinations of existing wireless technologies. We summarized in Fig. 1.1 the motivations of FMDA architecture and the challenges that need to be addressed, in which we focus on mitigating interference to improve spectral efficiency and reducing signal processing complexity to improve energy efficiency.

1.2.1 Architecture

A FMDA system is comprised of three main components, antennas, fiber-connection medium and the centralized processing system (CPS).

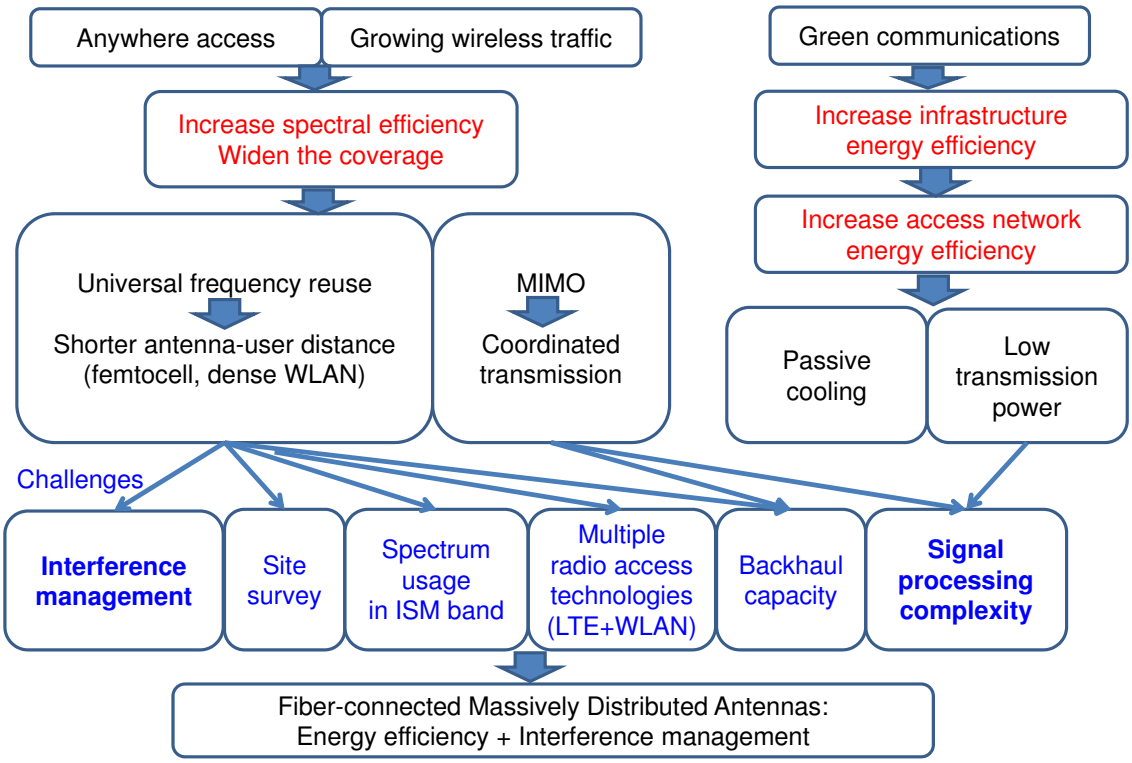


Figure 1.1: Motivation of Fiber-connected Massively Distributed Antennas. WLAN: wireless local area network. ISM band: the industrial, scientific, and medical band. LTE: Long-Term Evolution.

Antennas Antennas provide a convenient means of delivering blanket coverage for a targeted locale, while avoiding the time consuming and costly cell planning phases. Each antenna is only equipped with radio frequency components and optical-electrical converters, whereas the processing functionalities are transferred to the CPS. Each antenna forms a cell to provide wireless coverage to nearby users. Given the fast and reliable fiber-connection medium to/from the CPS, the FMDA system has the potential of scaling from a few antennas, covering tens of meters of space, to hundreds of antennas, covering a few kilometers of the targeted area. The location arrangement of individual antennas in this massive DAS is quite arbitrary, providing ease of deployment.

Centralized processing system All the processing functionalities are concentrated in the CPS which provides an opportunity to enhance the system performance from several perspectives. First, the inter-antenna interference can be easily managed as the CPS has the ability to detect the spec-

trum usage across the entire area, owing to the widely distributed antennas. Second, the signaling overhead associated with coordinated transmission and reception of data, such as through employing coordinated multi-point transmission mode in LTE context, reduces significantly as all the processing will be performed centrally. Therefore, a distributed MIMO system can be formed to increase system capacity. Third, not all antennas need to be activated at the same time. Thus, the FMDA system achieves a higher utilization ratio of processing units, which reduces the system cost and the energies consumed by signal processing.

Fiber-connection medium The backbone of FMDA is a network of optic cables connecting the CPS to each antenna. The wide-band nature and very low attenuation of optical cables allow reliable and energy-efficient delivery of wireless signals and therefore contribute to the advantages of FMDA over wireless-only solutions. The key in wide-spread deployment of FMDA is an efficient but inexpensive optical fiber backhaul, which itself is composed of two parts. First, an electrical-to-optical and optical-to-electrical converter transforms the communicated signal between optical and radio frequency domain. Second, the optical links will form a network, which can utilize passive or active optical networking protocols. If multiple antennas are fed via a shared pair of optical fibers, wavelength division multiplexing techniques can be exploited to reduce optical cable deployment cost. While the nonlinearity of the electrical-to-optical and optical-to-electrical converters incurs distortions to radio frequency signals, the resulting error vector magnitude had been shown to be as small as 1% in 16-, 64-, and 256-quadrature-amplitude-modulation signals if the input power to the fiber is smaller than 4 dBm [75]. In this thesis, we assume that radio frequency signals experience no distortions when they are transmitted over fibers.

The fiber-connection medium interconnecting the CPS and an antenna can also carry digital baseband signals in the form of high-rate in-phase/quadrature samples [76, 77], thus completely eliminating the signal distortions caused by nonlinearity while directly transmitting radio frequency signals over optical cables. In such case, however, an antenna needs a very-high-rate digital transceiver and high-speed analog/digital and digital/analog converters, which incur a higher

component cost and a higher power consumption when compared with radio frequency signaling. A comparison of radio frequency signaling and digital baseband signaling is presented in [75].

The existence of abundant optical cables inside buildings and large public venues is the key assumption we made in this thesis, which is based on the ever-decreasing cost of optical fibers and wavelength-division multiplexing components as well as the increasing fiber-to-the-home and fiber-to-the-building penetration rates. According to the European 2011-2016 forecast [78], the fiber-to-the-home/building maturity, defined as 20% household penetration of fiber-to-the-home/building, will be achieved by 2016 in eight countries in Europe and Asia. Currently, United Arab Emirates, Japan, Korea, Qatar had already achieved the maturity [79]. While the assumption is reasonable, requiring a large number of optical cables does increase the infrastructure cost of a FMDA system when compared with wireless-only solutions, and slows down the deployment speed when the required cables are not in place. In Section 6.2, we discuss how to reduce system cost by choosing different optical backbone topologies.

1.2.2 Configurations

A FMDA system can be flexibly configured as different types of communication systems, e.g., a WLAN Extended Service Set, an LTE system comprised of a large number of cooperating femtocells, a DAS that consists of a large number of antennas to cover an office building or a large public venue, or a HetNet that provisions different radio access networks over a metropolitan area. To configure FMDA as a WLAN Extended Service Set, we replace the existing APs with the antennas and convert the existing WLAN controller to the CPS by adding more functionalities. The resulting network, termed cognitive WLAN over fiber system, will be elaborated in Chapter 2. To configure FMDA as an LTE system, we replace femto-BSs with the antennas in FMDA and add the component CPS. The resulting network becomes a set of fiber-connected femto-BSs, which will be elaborated in Chapter 3. The last step of the WLAN and LTE configurations is to replace the electrical cables with optical cables. Chapter 4 and 5 elaborate the DAS configurations to cover office buildings and large public venues.

1.3 Research Questions

In each of the above configurations, there are research questions to be answered. In a cognitive WLAN over fiber system, antennas can either cooperate to improve WLAN sensing capability and reduce packet error rate, or independently operate multiple WLAN channels to form co-located WLAN Extended Service Sets. The question is which strategy brings a higher throughput and produces a lower packet error rate. We are also concerned with how the answer varies with application types, spatial traffic distribution and other network parameters. We answer these questions in Chapter 2 by considering realistic network settings and typical WLAN applications (including file transfer, Internet television, and audio streaming over Internet).

When a FMDA system is configured as an LTE system comprised of a large number of cooperating femtocells, compared with standalone femtocells, there arise one new dimension that we can explore to improve spectral and energy efficiency: each antenna can be activated or deactivated at per-slot-level, thus allowing the possibility of antenna scheduling. The sets of antennas being activated in each slot has a direct impact on spectral efficiency and signal processing complexity. Increasing the cooperation set size increases the spectral efficiency but incurs a higher power consumption in baseband processing. The question is whether it is possible to simultaneously improve spectral and energy efficiency. We answer this question in Chapter 3.

In view of the signal processing complexity challenge in massive MIMO systems and the increasing importance of low-complexity signal processing algorithms in green communications and C-RAN, we study low-complexity baseband processing schemes in our proposed FMDA architecture. As a FMDA system is equipped with a powerful, centralized signal processing capability, with traditional processing algorithms, the energy efficiency will drastically decrease when the number of cooperating antennas becomes very large. The question is to what extent the antennas should cooperate to satisfy a required system capacity. In Chapter 4 and 5, we answer this question by exploiting the distributed nature of antennas in two different network topologies.

1.4 Summary of Main Contributions

In this thesis, we propose a novel wireless access architecture, namely, FMDA, which addresses the finite-capacity backhaul challenge in the areas of DAS and HetNet, explores the opportunity of combining massive MIMO and RoF techniques, and allows the development of low-complexity beamforming schemes that are often not possible in massive array-MIMO systems. We demonstrate the advantages of FMDA in managing interference and improving energy-efficiency in both WLAN and LTE. When FMDA is employed to cover office buildings and large public venues, we carefully model the resulting channel matrix and propose two low-complexity zero-forcing beamforming (ZFBF) schemes to significantly reduce the baseband processing complexity in downlink, thus substantially improving the downlink energy efficiency. More importantly, the energy saving from our proposed ZFBF schemes increases with the network size, thus offering a great potential in future access networks. Detailed contributions are listed below.

1.4.1 Interference Management in Cognitive WLAN over Fiber

To increase the system capacity of a traditional WLAN system that is composed of a WLAN controller and multiple WLAN APs, previous research are focused on channel assignment strategies, user association schemes (or load balancing), and AP transmission power control [17]-[21]. The purpose of these strategies is to reduce co-channel and adjacent-channel interference among the APs and therefore increase the system capacity. However, as each AP typically supports only one channel, these algorithms have limited abilities to handle dynamic traffic, and become extremely complicated when channel allocation, load balancing and AP transmission power control are jointly considered.

In Chapter 2, we configure our proposed FMDA architecture as a novel WLAN Extended Service Set, where the CPS serves as a cognitive access point that is able to see the entire spectrum usage across the coverage area and control the radio frequency signals to be sent to each antenna. As a result of the wide bandwidth of optical cables, each antenna can operate multiple channels.

The antennas can also cooperate with the help of the centralized baseband signal processing at the CPS. The proposed WLAN system can exploit additional frequency channels to reduce the load in each channel and consequently the packet collisions, thus providing a much higher system capacity than the traditional WLAN system. The ability to cooperate among antennas also improved the sensing capability of the CPS, thus reducing the collisions between downlink and uplink transmissions. The two added abilities, which are only made possible by the proposed FMDA architecture, can also be combined to better accommodate dynamic traffic. Specifically, we achieve up to 62% Transmission Control Protocol throughput gain in hotspots.

This is a simulation-based research and our simulation model is based on an accurate WLAN simulation model that combines NS-2.33 simulator [80] with its *dei80211mr* WLAN rate adapter package [81]. The interference-recorded channel model incorporated in the package greatly enhances the accuracy of simulations involving channel capturing. Parts of this chapter have been included in one published journal article [82], and two published conference papers [83, 84].

1.4.2 Energy Conservation in Fiber-connected Femto Base Stations

Femtocell addresses the challenge on indoor wireless connectivity by pulling BS closer to user. The cell planning is shifted towards universal frequency reuse pattern largely due to the advantages on area spectral efficiency previously reported in cellular networks [85]. However, in very dense urban areas with a high concentration of residential and business users, inter-cell interference among femtocells using universal frequency reuse may drastically decrease system capacity. Cooperative communications among neighboring femtocells can eliminate the interference by frequency allocation, which, however, decreases the capacity gain from the use of universal frequency reuse. Meanwhile, in such hyper-dense deployments a large number of active femtocells are constantly consuming a large amount of energies even when some of them have no active users attached. While femtocell sleeping techniques can help reduce the energy waste, the wake-up time prevents anytime wireless access and therefore harms user experience.

To address these two challenges in densely deployed femtocell network, in Chapter 3 we con-

figure our proposed FMDA architecture as fiber-connected femto-BSs, where each antenna forms one femtocell. For all femtocells being formed, the CPS schedules all the attached users, allocates the frequency channel(s) in each cell, adjusts the transmission power at each antenna, and processes all the baseband signals to allow per-slot-level cooperations among the femtocells. In the formed system, we propose an antenna scheduling scheme to activate and deactivate the antennas to simultaneously improve spectral and energy efficiency. Compared with standalone femtocells, the proposed scheme is shown in a typical office building to increase energy efficiency by 64%~160% and spectral efficiency by 2%~36%.

The proposed antenna scheduling strategy relies on a large number of antennas that are able to sleep on demand while incurring negligible wake-up delay. This capability is only possible in the architecture of FMDA due to its centralized processing unit and the reduced antenna complexity. The idea of antenna scheduling also enriches network configurations on how many antennas should cooperate and how much separation distance should be kept among cooperating antenna clusters, thus offering service providers a greater flexibility on choosing different balance points between spectral and energy efficiency. Parts of this chapter have been included in three published journal articles [86–88].

1.4.3 Low-complexity Beamforming in Office Buildings

DAS have been widely used in indoor environments such as office buildings and stadiums to provide reliable coverage and a higher system capacity while using a lower radio transmission power at antennas. In Chapter 4 we consider a FMDA system that is configured as a typical DAS covering an office building. Our focus is to design a multi-user downlink transmission scheme that allows universal frequency reuse and thus provides a high system capacity.

We consider a single channel matrix $\mathbf{H}$ that includes the whole set of antennas in the entire network, and based on $\mathbf{H}$, we devise a ZFBF scheme to support multi-user transmission. When the number of antennas is large, however, it is challenging to determine the entire channel matrix $\mathbf{H}$ and power-consuming to generate the precoding matrix $\mathbf{W}$ and precode the symbols, especially when

these operations are required in each subcarrier if orthogonal frequency-division multiplexing is used. Our idea is to treat the channel matrix as a banded matrix, thus allowing a low-complexity beamforming scheme. Previously Wyner model [89] had been used to approximate a channel matrix of an one-dimensional outdoor cellular system as a banded matrix [90, 91]. However, the accuracy of Wyner model is established on a sufficiently large number of simultaneous users [92]. Banded matrices had also been applied to electromagnetic wave simulation [93], inversion of tri-diagonal matrices [94], and equalization [95]. However, in the area of distributed MIMO systems, the only literature exploiting the banded path loss structure is [96], where the authors focused on optimal allocation of channel state information rather than a low-complexity beamforming scheme.

In Chapter 4, we discover that the existence of indoor wall penetration loss allows us to discard less important off-diagonal elements in $\mathbf{H}$ and then form a banded, sparse matrix, $\mathbf{H}_p$, where p controls the sparsity of $\mathbf{H}$. The precoding $\mathbf{W}$ is then generated based on $\mathbf{H}_p$. The scheme is evaluated by applying it to a DAS that covers a dual-stripe office building floor. Compared with traditional ZFBF, our analysis and numerical evaluations show that the low-complexity beamforming scheme incurs negligible loss in signal-to-interference-noise-ratio (SINR), while offering 45%~79% gain in energy efficiency.

To the best of our knowledge, our work is the first to consider the inversion of random matrices with exponentially decayed off-diagonals or compositely decayed off-diagonals. Previously studied Wyner model, full channel matrix model and standalone femtocell model are special cases of the proposed model. In practice, the introduced parameter p can be used by system operators to obtain a fine control over the tradeoff between spectral and energy efficiency. Parts of this chapter has been included in one journal article under review [97].

1.4.4 Low-complexity Beamforming in Large Public Venues

A single outdoor micro- or macro-BS cannot provide enough capacity in large public venues such as stadiums and convention centers because of the increasing demand of Internet access and content delivery during live events. WLAN and DAS operators have been deploying a large number of

distributed antennas to offer a higher system capacity in these venues. To mitigate the interference arising from the dense deployment, the system operators still rely on a thorough radio frequency planning and careful tuning of directional antennas, which support channel allocation and transmission power control decisions. However, these measures are limited by the number of channels available to the operator; a realistic radio frequency survey is also time-consuming and very expensive (in some cases the presence of audience is even required to ensure human effects are included in the survey).

To simplify the frequency allocation difficulty and eliminate the need of a RF survey, in Chapter 5 we propose a FMDA system to cover a large public venue. Similar to Chapter 4, we design a ZFBF scheme by considering a single channel matrix $\mathbf{H}$ that includes the whole set of antennas in the venue. However, unlike an office building, the absence of walls in the venue invalidates the channel model developed in Chapter 4 and would incur a large throughput loss were our previously proposed scheme applied. To reestablish the decays in $\mathbf{H}$, we follow the under-floor antenna mounting strategy that had been adopted in hyper-dense WLAN industries [27–29] such that the heavy human body penetration loss can help increase the propagation loss among antennas. We then develop a new channel model and propose a multi-banded matrix inversion algorithm that substantially reduces the computation cost of ZFBF while incurring a negligible throughput loss.

Compared with a massive array-MIMO system located in the center of the venue, our proposed scheme provides 19 times higher energy efficiency while only incurring 6% spectral efficiency loss. We also discover that although massive array-MIMO can deliver a very high system capacity, its co-located antennas disallow a low-complexity matrix inversion. Therefore, a massively distributed MIMO system equipped with our proposed low-complexity ZFBF scheme can offer a higher energy efficiency than a massive array-MIMO system. This chapter has been included in one published journal article [98].

1.5 Thesis Organization

The rest of the thesis is organized as follows. In Chapter 2, we propose two methods that utilize the specialized capabilities of the cognitive WLAN over fiber architecture to improve system capacity by reducing packet collisions through load balancing and employing diversity to reduce the effects of packet collisions. In Chapter 3, we present how antenna scheduling can improve the energy efficiency when a FMDA system is used to form fiber-connected femto-BSs, and show its substantial advantages when compared with standalone femtocells. In Chapter 4, we present the low-complexity ZFBF scheme in a FMDA system covering an office building, analyze the residual interference due to the use of a simplified matrix inversion and then demonstrate its energy-efficiency advantage when compared with conventional ZFBF. In Chapter 5, we consider the low-complexity ZFBF scheme in a two-dimensional network and propose the use of floor-mounted directional antennas to enable the possibility of simplified matrix inversion. The main results and potential research topics are summarized in Chapter 6. Chapter 2, 3, 4 and 5 are self-contained and have been included in separate journal articles and conference papers. Each of these chapters includes its own literature survey that reviews previous solutions to the corresponding research problem.

Chapter 2

Interference Management in Cognitive Wireless Local Area Network over Fiber [1]

2.1 Introduction

Wireless local area networks (WLANs) are widely used for connecting computing equipment in homes and offices to the Internet. However, WLANs share the industrial, scientific, and medical (ISM) band with other independently-operated license-free devices such as Bluetooth radios and microwave ovens; therefore, they must tolerate interference from these devices. Cognitive radio techniques have been proposed for secondary users to exploit spectrum holes left unused in licensed frequency bands by primary users of the allocated spectrum. In this chapter, we employ Fiber-connected Massively Distributed Antennas (FMDA) in WLANs and propose a novel architecture, *cognitive WLAN over fiber*, which applies advanced cognitive radio [99] and broadband radio-over-fiber [31] technologies to an infrastructure-based IEEE 802.11 WLAN Extended Service Set (ESS) comprised of multiple access point (AP)s, each forming its own Basic Service Set (BSS). Successful simultaneous transmissions of multiple WLAN channels over low-cost multi-mode optical fibers [32–35] and clarification of WLAN medium access control (MAC) op-

[1] This chapter is based on [82, 84] co-authored with Dr. A. Attar, Dr. V. Leung and Dr. Q. Pang, and [83] co-authored with Dr. Q. Pang and Dr. V. Leung.

eration in radio-over-fiber structures [36–38] also support the proposal of cognitive WLAN over fiber as an architecture that offers huge potentials to increase system capacity and improve quality-of-service.

The ever-decreasing cost of optical fibers and wavelength-division multiplexing components has resulted in commercial fiber-based indoor wireless networks being deployed to penetrate large buildings such as stadiums [39], hospitals [42], business buildings and shopping malls [43]. It would be expensive to cover these buildings with cable-based networks due to the ever-increasing cable cost. It is also difficult to monitor and manage the radio environment within such large buildings if antennas cannot be efficiently coordinated. The success of these commercial indoor wireless networks further demonstrated potential markets for cognitive WLAN over fiber networks.

In a conventional WLAN, each AP performs carrier sensing independently and only over the channel it operates on. In contrast, a cognitive WLAN over fiber system applies cognitive radio techniques to more efficiently utilize the ISM band. The centralized architecture enables cooperative sensing and consequently reduces the interference detection time while improving the detection accuracy. Moreover, the multi-channel carrying capability of advanced broadband radio-over-fiber systems can significantly increase available radio resources at each AP. By implementing dynamic radio resource management based on accurate spectrum sensing, interference avoidance or mitigation can be easily accomplished. Effectively, the cognitive WLAN over fiber architecture enables the new concept of applying cognitive radio techniques for equal spectrum access in the ISM band.

Each AP in a conventional WLAN has an 802.11 radio modem and is digitally bridged to a distribution system, usually an 802.3 Ethernet. In a cognitive WLAN over fiber system, radio modems and bridges in the APs are moved to a centralized unit referred to as the cognitive access point (CogAP); the resulting simplified APs become antennas, which are connected to the CogAP via optical fibers that carry analog radio frequency signals. By centrally processing broadband radio frequency signals received from the antennas, the CogAP has a complete picture of the radio spectrum usage in the coverage area of the WLAN ESS. The Distributed Coordinated Function

of 802.11 MAC, which employs carrier sensing with collision avoidance, is carried out at the CogAP instead of at individual APs as in a conventional WLAN. These changes enable a cognitive WLAN over fiber system to more effectively combat packet collisions that inevitably occur over a random-access channel. A cognitive WLAN over fiber system and the structure of the CogAP are illustrated in Fig. 2.1.

In this chapter, we focus on how to reduce access collisions in a WLAN through methods made possible by the cognitive WLAN over fiber architecture, which would be difficult if not impossible to realize in a conventional WLAN. The chapter is organized as follows. In Section 2.2, we review related work on how to improve WLAN system capacity. In Section 2.3, two methods are proposed to reduce collisions among WLAN stations: load balancing to reduce collisions caused by heavy traffic, and transmitter and receiver diversity to reduce the effects of collisions. The performance of proposed methods is evaluated through Monte-Carlo simulations in Section 2.4. We summary the chapter in Section 2.5.

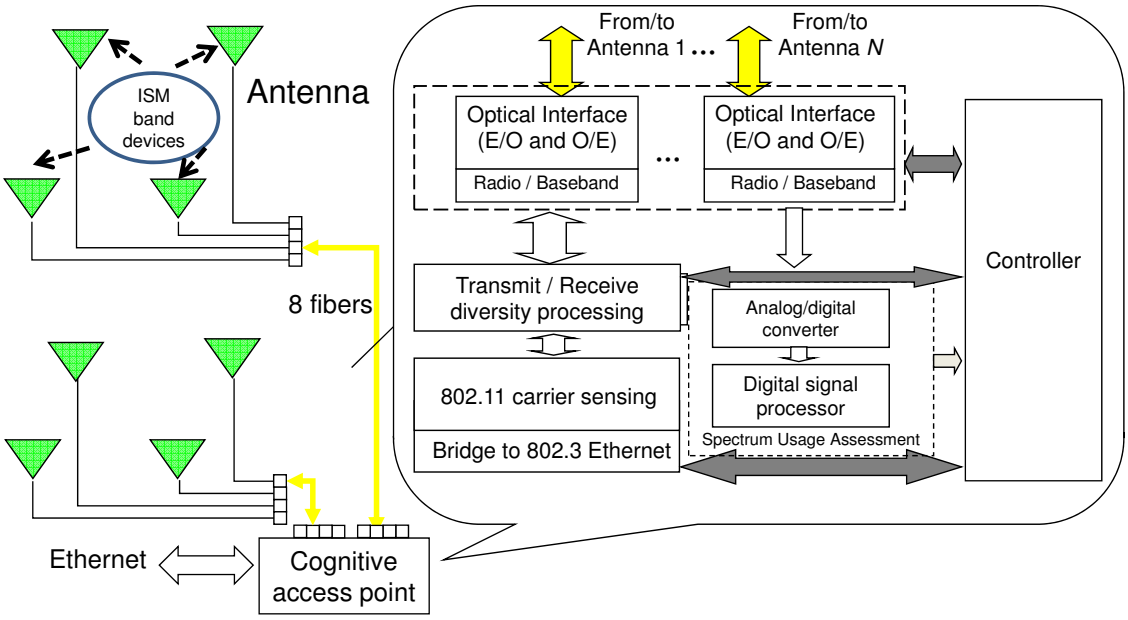


Figure 2.1: A cognitive WLAN over fiber system and the structure of the cognitive access point (CogAP). ISM band: the industrial, scientific, and medical band. E/O: electrical-optical converter. O/E: optical-electrical converter.

2.2 Related Work on WLAN

Much recent research on WLANs aims to increase system capacity of individual WLAN BSSs, and reduce co-channel and adjacent-channel interference among BSSs in a WLAN ESS.

The system capacity of a WLAN BSS can be increased through three methods: enhancing existing MAC protocols by either adjusting parameters or adding new MAC flavors to achieve a higher MAC efficiency, exploiting capture effects, and introducing multiple-input multiple-output (MIMO) to exploit spatial multiplexing. Enhancing the WLAN MAC protocol usually requires an update of station hardware or firmware. We therefore mainly review recent advances on exploiting capture effects and employing MIMO in WLAN.

The capture effect has been initially studied within the context of an ALOHA network [100]. It refers to the fact that when two packets arrive at one station at the same time, the packet with stronger signal strength will be synchronized and "captured" by the station. Luo and Ephremides [101] showed that with the capture effect, system throughput is maximized when all nodes transmit at maximum power. This conclusion, however, is based on an optimistic assumption that any packet can be successfully received as long as it has the highest power level at the receiver, regardless of how many overlapping packets are being received at lower power levels. After taking interference into account, Hadzi-Velkov and Spasenovski investigated the capture effect and its interaction with request-to-send/clear-to-send in 802.11b networks [102]. Kochut *et al.* studied the capture effect by comparing system throughput at the physical and transport layers in 802.11b networks [103]. Their comparisons showed that capture effect is magnified through variations of contention window size in the MAC layer and congestion window size in the Transmission Control Protocol (TCP) layer. Based on Bianchi's model [104], WLAN performance was derived in [105] after considering the capture effect. Capture effect and successive interference cancellation were later studied in a ZigBee network that is based on direct-sequence spread spectrum [106].

Capture effect in 802.11a networks was studied in [107, 108] through real-world experiments using commercial WLAN devices. It was shown that with an arrival time difference of up to 50 μs,

the stronger 802.11a packet can still be captured. Different from previous 802.11b capture effect studies where the stronger frame has to arrive within the preamble time of the weaker frame, this observation suggests that even when the arrival time difference of two packets is larger than the preamble length of the first packet, the stronger packet could still be captured. Such phenomena have been observed in commercial 802.11a/b/g adapters working in either direct-sequence spread spectrum mode or orthogonal frequency-division multiplexing mode.

The capture effect was further exploited in the form of "message-in-message" to increase system throughput [109, 110]. An AP sends the message with smaller channel gain first and the message with larger channel gain later such that the weaker packet's preamble can be successfully locked by one recipient and the stronger packet can also be locked by another recipient. The AP abuses carrier-sensed multiple access rule and stations use delayed ACKs. Using "message-in-message" requires the AP to update the system interference map periodically.

Exploiting diversity in WLAN is classified into micro-diversity and macro-diversity. The IEEE 802.11n standard is developed to enable micro-diversity in WLANs using MIMO. Previous work on macro-diversity includes the concept of distributed radio bridges proposed in [24] and their subsequent applications in WLAN [111, 112].

A WLAN ESS is a multi-cell WLAN system in which the WLAN controller assigns channels and sets maximum AP transmit power to different BSSs to reduce co-channel and adjacent-channel interference among them. Sub-optimal radio resource management algorithms have been extensively studied for this purpose. These algorithms address three basic problems: channel allocation [113], user association (or load balancing) [114], and transmit power control [115]. The conflict set coloring method jointly optimizes channel allocation and load balancing [116]. Measurement-driven guidelines in [117] provide a heuristic method to jointly address the three basic problems. However, due to the limited number (usually one) of channels that each BSS can support, these algorithms have limited abilities to handle dynamic traffic, and become extremely complicated when channel allocation, load balancing and transmit power control are jointly considered. Authors of [118] investigated how to coordinate medium access across multiple APs in an ESS by switching

from contention-based access to time-slotted access when the ESS is heavily loaded with audio and video streams. The switching reduces packet collisions and thus provides better quality-of-service for multimedia streams. However, the signaling protocol required by the AP coordination was not given in [33].

2.3 Collision Reduction

A collision happens when two stations access the channel at the same time, or when one station fails to sense an on-going packet transmission due to fading or hidden terminal problem and starts a new transmission. Based on the cognitive WLAN over fiber architecture, in this chapter we propose a load-balancing method to reduce collisions caused by heavy traffic, and a transmitter and receiver diversity method to reduce the impact of packet collisions by increasing the chance of successful reception. The two methods used to reduce collisions in cognitive WLAN over fiber are illustrated in Fig. 2.2.

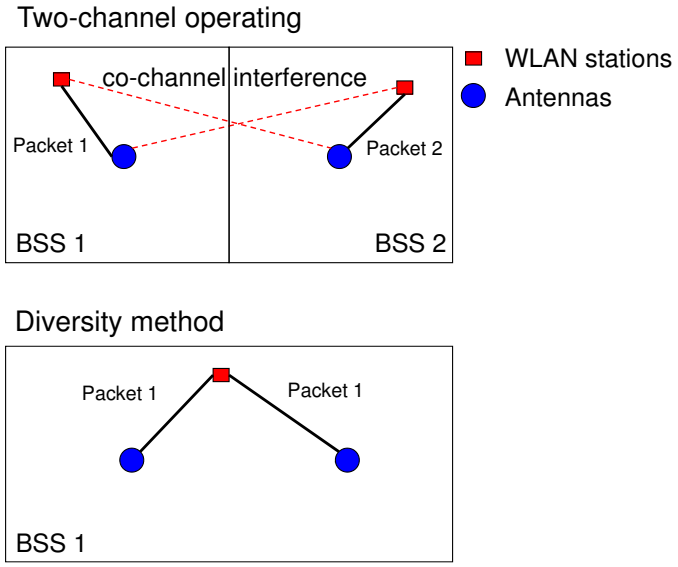


Figure 2.2: Collision reduction: Diversity and two-channel-operation. BSS: Basic Service Set.

2.3.1 Load Balancing Method

A practical load balancing technique facilitated by the cognitive WLAN over fiber architecture is to distribute the total traffic load in the frequency domain. The broadband radio-over-fiber connection between each antenna and the CogAP allows multiple channels to be allocated to any antenna. Consider the case of two antennas covering a given area: antenna A_1 operates on the channel f_1 and antenna A_2 operates on f_2. When the collision rate on f_1 is higher than a target threshold, the CogAP can use the "disassociation" process to force some of the stations to be dissociated from this channel, while simultaneously sending beacons on a different channel f_3. Stations dissociated from f_1 will then have two options. If a dissociated station receives beacons on channel f_2 from A_2, it can request to associate with A_2 on this channel. This effectively transfers a portion of the traffic load of A_1 to A_2, creating a distributed load balancing solution among antennas. Alternately, a dissociated station will receive beacons on channel f_3 from A_1, and request to associate with A_1 over this channel. In this case, load balancing occurs over the frequency domain within the same antenna, where a portion of the traffic at A_1 is switched from overloaded channel f_1 to channel f_3. The second case is particularly made possible by the broadband radio-over-fiber connections between antennas and the CogAP. In contrast, conventional WLAN APs are generally not equipped for multi-channel operations.

The gain in system throughput in the above example is two-fold: one from increased medium access efficiency due to decreased contention among stations accessing the same channel, and another from the use of three channels instead of two. We are more interested in the latter owing to its potential of linearly increasing system throughput. However, to fairly compare a cognitive WLAN over fiber with a conventional WLAN, we investigate the worst case where the new channel assigned to A_1 is the same as that assigned to A_2, i.e., f_2. We shall examine the throughput gain that can be achieved in the presence of co-channel-interference on f_2.

WLAN operations on f_1 and f_2 can be independent and as such we refer to this load-balancing method as multiple-independent-channel-operation. Let us compare a two-AP conventional WLAN,

where AP_1 operates on f_1 and AP_2 operates on f_2, with a two-antenna cognitive WLAN over fiber system, where A_1 operates on f_1 and f_2 and A_2 operates on f_2. We can certainly focus on the throughput on f_2. It is clear that the cognitive WLAN over fiber system provides the worst throughput on f_2 when all stations associated on f_2 can perfectly hear each other. We now argue that even in such a situation, the cognitive WLAN over fiber system could provide a higher throughout than a conventional WLAN. In the conventional WLAN, AP_2 can only send one data packet at a time. In the cognitive WLAN over fiber system with A_1 and A_2 independently operated, they might simultaneously send two data packets on f_2. Owing to capture effects, the two packets may both survive from the collision, thus generating a throughput gain.

2.3.2 Transmitter and Receiver Diversity Method

Besides operating channels independently, the CogAP can also manage channels to exploit macro-diversity since signals received from widely separated antennas tend to be uncorrelated. If each antenna is also equipped with multiple antenna elements, we can further implement micro-diversity in conjunction with macro-diversity. However, as only two fibers are used to connect each antenna to the CogAP (one to transmit and one to receive), wavelength division multiplexing would then be required to deliver radio frequency signals from/to different antenna elements attached to the same antenna. Here we focus on macro-diversity enabled by distributed antennas.

Receiver diversity

Consider an area covered with two antennas, using the same set of frequencies to serve a group of stations. When maximum-ratio combining (MRC) is used at the CogAP for uplink signals, not only do we achieve an array gain of 3 dB due to the coherent combining at the receiver, but also obtain a diversity order of 2 if the two signal paths from the station to the two antennas experience independent fading. The array gain and the diversity order reduce the effects of packet collisions, resulting in increased throughput and reduced packet error rate (PER). When the number of antennas increases to four, we expect a higher performance improvement due to 3 dB more in array

gain and a larger diversity order.

An immediate effect of receiver diversity is an improvement in sensing capability at the CogAP, and hence a reduction in WLAN packet collisions between downlink and uplink packets. Another effect of diversity gain is to reduce unfairness among stations in terms of their chances to access the channel due to their different distances from the antennas.

Transmitter diversity

For the downlink, we can use transmitter diversity to improve signal-to-noise-ratio at the stations without requiring them to have additional capabilities. Multiple copies of each packet are distributed to antennas and then to the destination such that when some copies are largely attenuated due to poor channel conditions, other copies can still reach the destination; hence, transmitter diversity. By reciprocity of the channel, transmitter diversity at the CogAP through multiple antennas achieves the same signal-to-noise-ratio gain as receiver diversity, subject to a total transmit power constraint on all antennas.

We investigate equal-gain combing (EGC) and MRC using transmitter diversity. In EGC scheme, each antenna is subject to a given per-antenna transmit power constraint, which reduces distortions due to the nonlinearity at optical-electrical converters. In MRC scheme, antennas are only subject to a total transmission power constraint, and therefore have a larger freedom on transmission power allocation across antennas, providing a larger signal-to-noise-ratio gain than EGC.

Both EGC and MRC require that signals from different antennas can be added coherently at the receiving station. Therefore, the CogAP must have exact channel state information from all participating antennas to the receiving station right before a packet is sent, such that signal phases can be properly shifted at the different antennas. This makes estimating channel state information for transmitter diversity more difficult than receiver diversity, where the CogAP can always rely on the physical-layer header of WLAN packets to estimate channel state information.

2.4 Performance Evaluations

We utilize the NS-2.33 simulator [80] with its *dei80211mr* WLAN rate adapter package [81] to evaluate the performance of the proposed methods based on Monte-Carlo method. The interference-recorded channel model incorporated in this package is used to accurately simulate capture effect.

2.4.1 Simulation Model

The simulation model includes two antennas connected to one CogAP, which is then connected to a fixed host computer. Single-antenna stations are either uniformly or non-uniformly placed in a $60 \times 30m^2$ area. Two antennas (or APs in the conventional WLAN) are fixed at locations (15, 15) and (45, 15) in units of meters. When no diversity is used, the CogAP communicates with stations through their closest antennas. Traffic streams only flow between stations and the fixed host. The wireless propagation model is a simplified path loss model [119] with shadowing and Rayleigh fading.

WLAN parameters follow IEEE 802.11g and two non-overlapping channels are used. Each AP in the baseline conventional WLAN operates on one channel only, while the CogAP in the cognitive WLAN over fiber system operates on both channels through the two antennas either co-operatively for macro-diversity or independently. Data mode used by each station and the CogAP is determined by the signal-to-noise-ratio-based dynamic rate adaptor in dei80211mr package. No request-to-send/clear-to-send is used. Perfect channel state information is assumed to be available at the CogAP.

The frequency plan used in the simulations is shown in Fig. 2.3 and simulation parameters are listed in Table 2.1. The synchronization interval (SI) is used to model the capture effect. When the arrival time difference of two packets is smaller than the SI, it is assumed that the receiver is able to synchronize to the packet with the stronger received power.

The simulations employ two types of traffic that represent increasingly popular Internet applications: File Transfer Protocol (FTP) over TCP in downlink representing traffic from file down-

loading applications, and constant bit-rate (CBR) traffic in both uplink and downlink representing Voice over Internet Protocol (VoIP) and IP television (IPTV). FTP over TCP traffic is saturated, i.e., stations always have packets to send. VoIP and IPTV, as multimedia traffic, have the same fixed packet interval yet different packet length due to different amount of information contained in their packets. We are mainly interested in file downloading speed and voice and video quality; thus, throughput of TCP downlink traffic and packet error rate of CBR traffic are chosen as our main performance evaluation criteria. To evaluate proposed methods for file sharing applications, we also evaluate TCP uplink throughput in some simulation scenarios. Detailed traffic parameters are listed in Table 2.1.

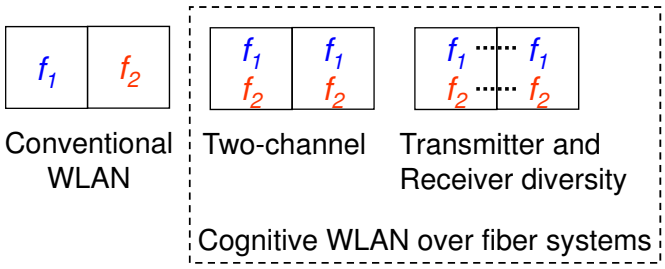


Figure 2.3: Frequency plan in simulations.

Table 2.1: Simulation parameters in cognitive WLAN over fiber systems

Propagation	Path loss exponent = 2.5 Reference distance, d_0 = 2 m Standard deviation of shadowing = 3.5 dB
Physical layer	Transmission power, P_t = 10 mW Carrier-sensing threshold = -70 dBm
MAC layer	Synchronization interval = 5 μs aSlotTime = 20 μs CWMin = 31 CWMax = 1023
FTP traffic	TCP/Reno. Packet size = 1000 bytes.
CBR traffic	Packet interval 20 ms. 1000-byte packets for IPTV; 40-byte packets for VoIP.

2.4.2 Effects of Receiver Diversity

We first investigate the effect of receiver diversity, i.e., using multiple antennas to receive a packet transmitted by a single-antenna station. Assuming the channel gain of each path is Rayleigh distributed, we know that the received signal power at the i-th antenna, $P_r^{(i)}(x,y)$, is an exponentially distributed random variable with the probability density function

$$f_{i,x,y}(P_r^{(i)}) = \frac{1}{P_{avg}^{(i)}(x,y)} \cdot e^{-\frac{P_r^{(i)}}{P_{avg}^{(i)}(x,y)}}, \tag{2.1}$$

where $P_{avg}^{(i)}(x,y)$ is the averaged power of signals received by antenna A_i from a station located at (x,y) and reflects the path loss between them. Thus, at the CogAP that receives the signals from both A_i and A_k, the p.d.f. of the total received signal power is given by

$$f_{cap,x,y}(P_r^{(cap)}) = \frac{e^{-\frac{P_r^{(cap)}}{P_{avg}^{(i)}(x,y)}} - e^{-\frac{P_r^{(cap)}}{P_{avg}^{(k)}(x,y)}}}{P_{avg}^{(i)}(x,y) - P_{avg}^{(k)}(x,y)}, \tag{2.2}$$

where $P_r^{(cap)}$ is the received signal power at the CogAP after coherently combining signals from A_i and A_k and the notation $f_{cap,x,y}(P_r^{(cap)})$ implies that the p.d.f. of $P_r^{(cap)}$ is a function of (x,y), the geographical coordinate of the transmitter. When the station has the same average path loss to the two antennas, $f_{cap,x,y}(P_r^{(cap)})$ becomes an Erlang distribution with the shape factor $N = 2$.

Given different locations of stations, the CogAP and individual antennas exhibit different outage probabilities over the whole coverage area, as shown in Fig. 2.4, where Prob($P_r \leq$threshold) represents the probability of the received signal power being smaller than or equal to the carrier-sensing threshold, commonly set at -70 dBm in WLANs. Denote the location at $x = x_0$ and $y = y_0$ by (x_0, y_0). The location $(15,15)$ and $(45,15)$ correspond to the locations of two antennas (or APs). Four edges of the coverage area correspond to $x = 0$, $x = 60$, $y = 0$ and $y = 30$. Fig. 2.5 shows the reduction on the outage probability when MRC or EGC is used, compared with the case where diversity is not employed. In both figures the dotted blue surfaces correspond to EGC and the dash

red ones correspond to MRC. The solid black surface corresponds to conventional WLAN in Fig. 2.4 and serves as the zero-reference plane in Fig. 2.5.

We observe from Fig. 2.4 and 2.5 that MRC always reduces outage probabilities more than EGC, especially for stations at corner areas where the path losses to the two antennas largely differ. To have a clearer observation, we plotted four cross sections of Fig. 2.4 and 2.5 respectively in Fig. 2.6 and 2.7, where each sub-figure shows the outage probability of a station when it is placed on a cross section (indicated by a fixed y value). For symmetry, we only plotted cross sections at y=0, 5, 10 and 15. In Fig. 2.6 and 2.7, the locations $(0,0)$ and $(60,0)$ correspond to two corners and a smaller y value indicates that the station is closer to the edge $y = 0$. We clearly observe that MRC reduces outage probabilities more than EGC, especially for stations at corner areas.

From Fig. 2.4 and 2.6, we also observe that compared with EGC, MRC has a flatter distribution of outage probabilities across the area, so that stations located farther away from the antennas will still be heard by the CogAP. Therefore, their uplink packets will less likely collide with downlink packets, and consequently their contention windows will not suffer as much from exponential increases. With receiver diversity at the CogAP, stations farther away from the antennas still have good chances to access the channel compared with those closer to the antennas.

2.4.3 Spatially Uniformly Distributed Traffic

We place stations uniformly over the whole area to represent spatially uniformly distributed traffic. For a given number of active stations, 20 different sets of station locations are randomly generated. Simulated results under these scenarios are then averaged for evaluations. From Fig. 2.8 and Fig. 2.9, we observe that under two types of spatially uniformly distributed traffic, operating two channels in both antennas increases TCP throughput by 14%~18% when only 8 stations are active. When the number of active stations increases, the TCP throughput gain also increases, reaching 36% for VoIP uplink/downlink plus FTP downlink traffic and 20% for IPTV downlink plus FTP downlink traffic when the number of active stations is 32. The packet error rate of CBR downlink traffic is also reduced by 50%. The packet error rate of CBR uplink is zero and thus not presented

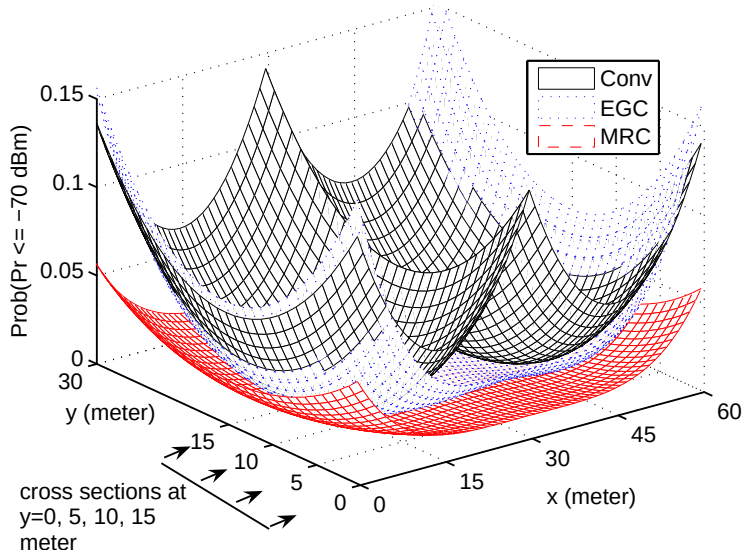


Figure 2.4: Outage probabilities in a WLAN Extended Service Set. Two antennas (or access points) are located at $(15,15)$ and $(45,15)$. Carrier-sensing threshold = -70 dBm. Conv: conventional WLAN.

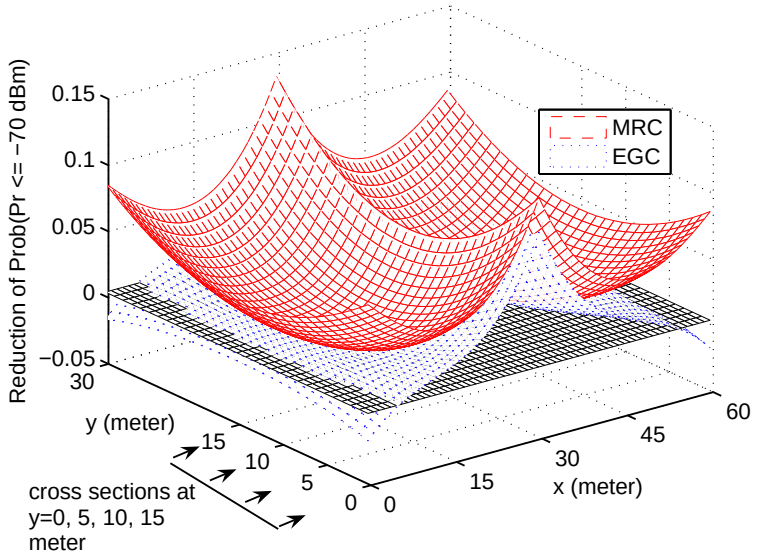


Figure 2.5: Reduction in outage probability with MRC and EGC. Carrier-sensing threshold = -70 dBm.

here. The reason is that uplink is not heavily loaded without the presence of saturated FTP/TCP traffic. The performance gain of two-channel-operation can be attributed to channel capture effects on downlink data packets in the cognitive WLAN over fiber system (See Section 2.3.1).

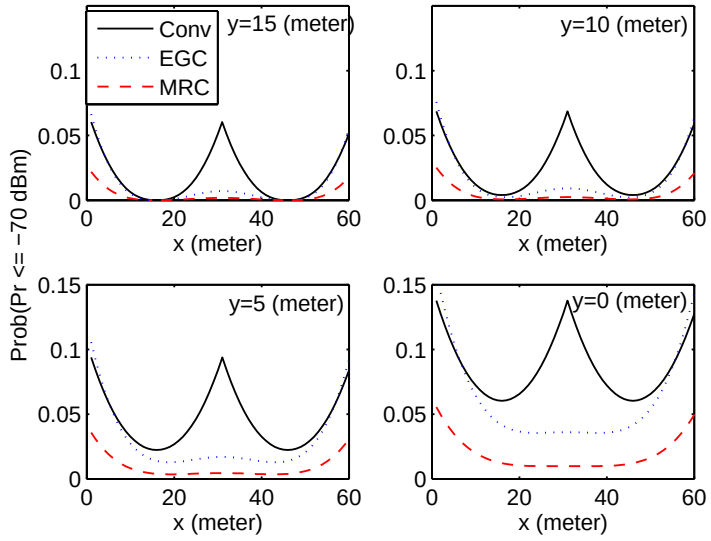


Figure 2.6: Outage probabilities in a WLAN Extended Service Set. Each sub-figure shows the outage probability of a station when it is placed on a cross section (indicated by a fixed y value). Carrier-sensing threshold = -70 dBm. Conv: conventional WLAN.

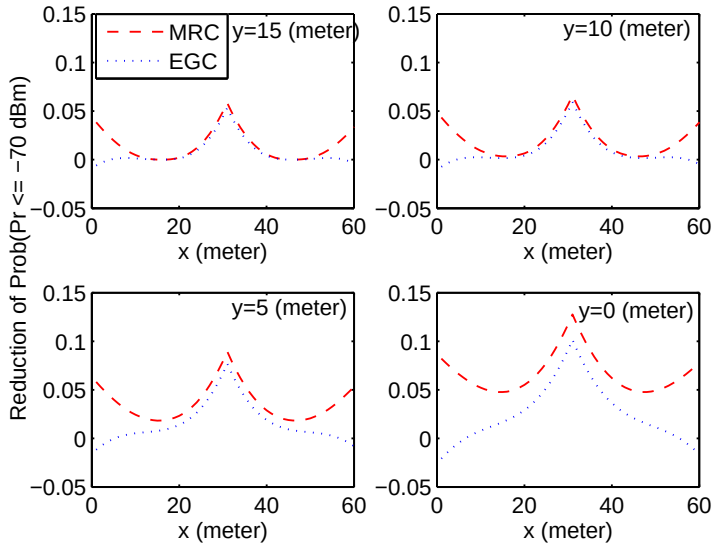
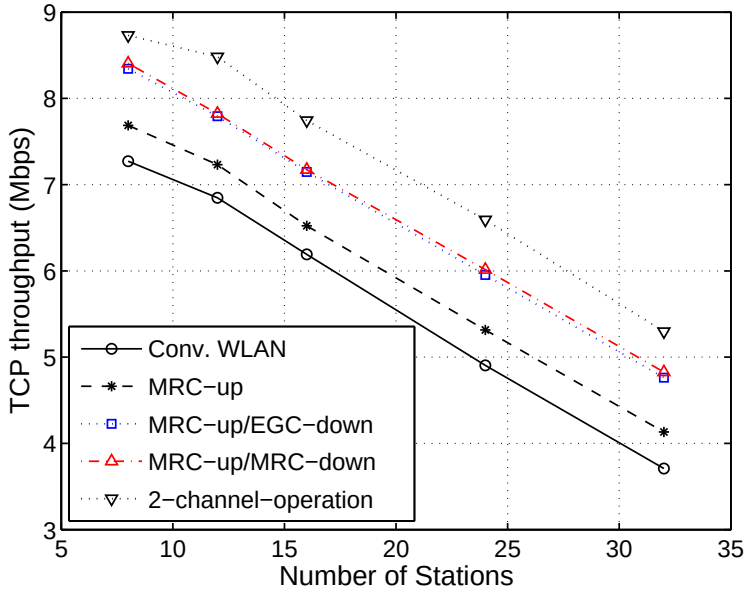


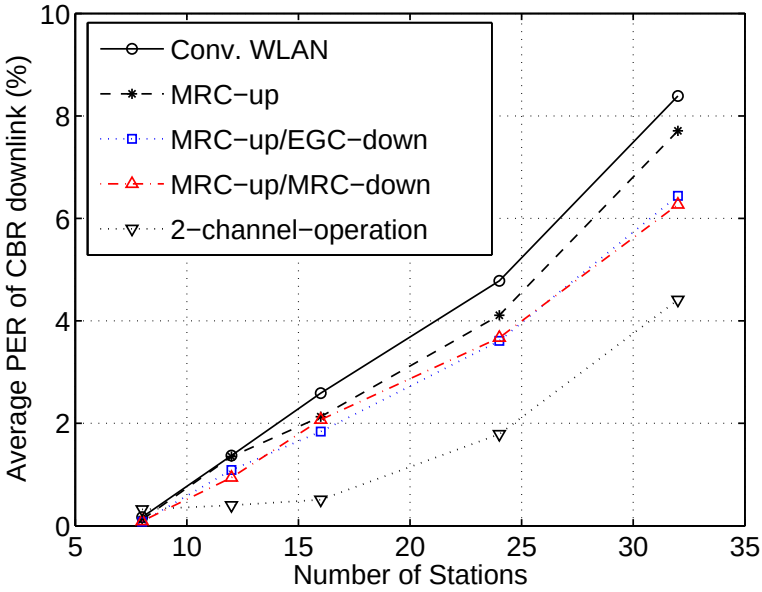
Figure 2.7: Reduction in outage probability with MRC and EGC. Carrier-sensing threshold = -70 dBm.

Note that we use the number of stations to indicate the intensity of traffic since all of stations have always-on CBR and FTP/TCP traffic.

The confidence interval of obtained average packet error rates can be estimated by *berconfint*

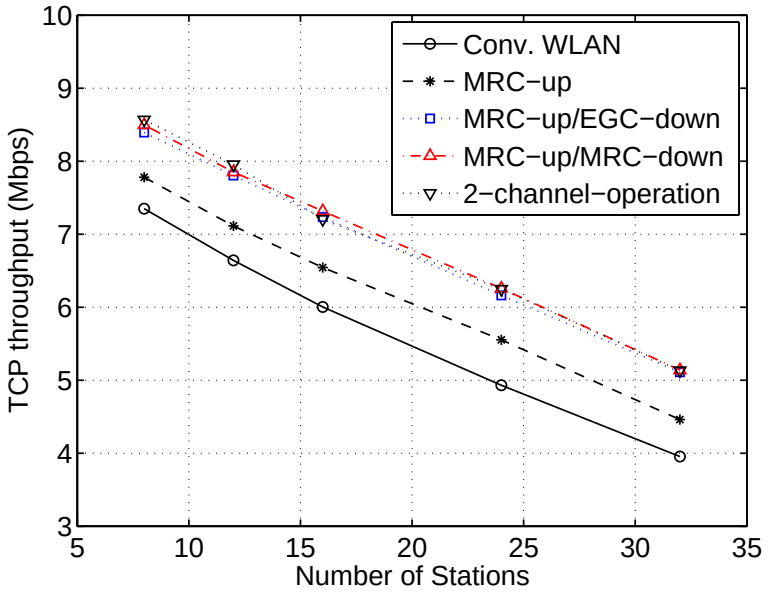


(a)

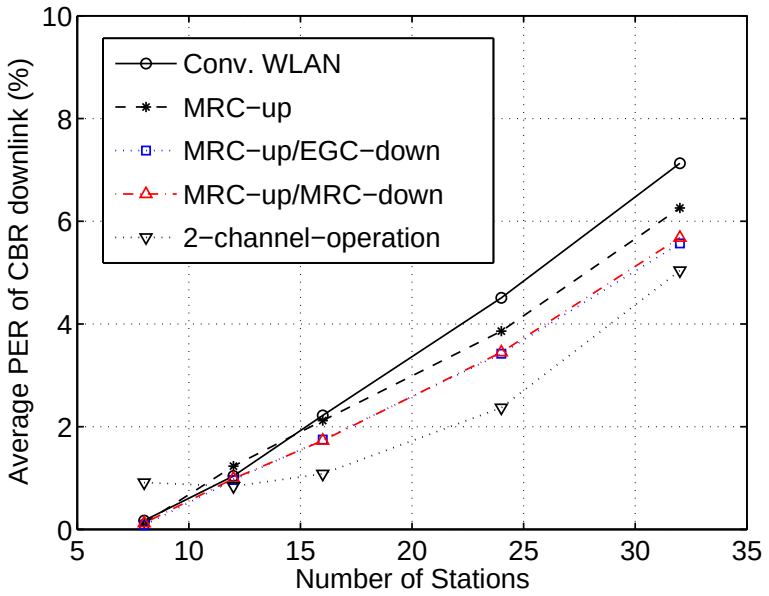


(b)

Figure 2.8: TCP throughput and average packet error rate of CBR downlink vs. Number of stations. CBR: constant bit-rate traffic. PER: packet error rate. Traffic: VoIP uplink/-downlink + FTP downlink. MRC-up: MRC is used for uplink diversity. EGC-down: EGC is used for downlink diversity. Conv. WLAN: conventional WLAN.



(a)



(b)

Figure 2.9: TCP throughput and average packet error rate of CBR traffic vs. Number of stations. Traffic: IPTV downlink + FTP downlink. MRC-down: MRC is used for downlink diversity.

function in MATLAB, provided that the number of packet errors follows binomial distribution. Each simulation run lasts 120 seconds with 20 ms packet interval. Therefore, in 12-station case, 36,000 downlink and 36,000 uplink packets are generated, resulting the following 95% confidence intervals: [0.93%, 1.08%] at 1% average error rate and [0.08%, 0.13%] at 0.1% average error rate. A higher number of stations or a higher average packet error rate generates a tighter confidence interval. To avoid clutter, we do not superpose the confidence intervals on the figures of packet error rates.

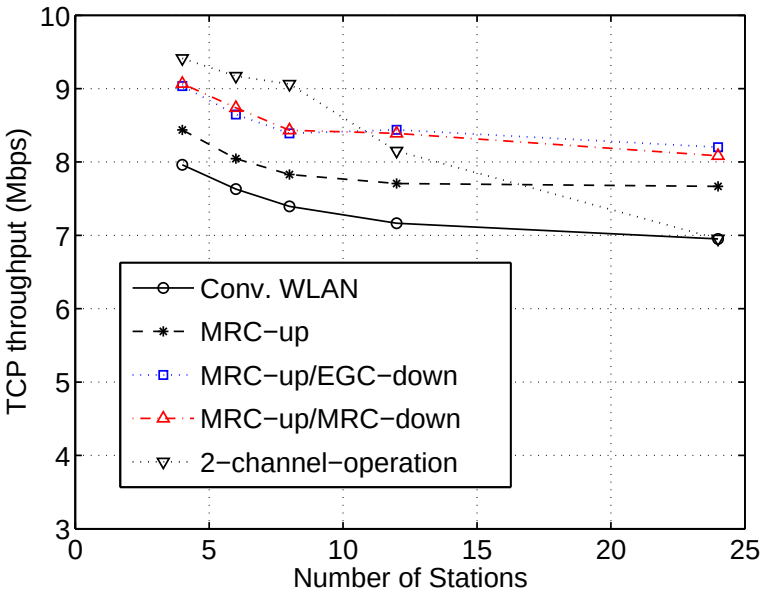
Comparing the diversity methods we have investigated, using only MRC for uplink diversity slightly increases TCP throughput and reduces downlink packet error rate for CBR traffic, while engaging additionally downlink EGC or MRC transmit diversity further improves performance by providing a higher TCP throughput gain and lower packet error rate for CBR traffic. The results also show that two-channel-operation always outperforms the diversity methods in both TCP throughout and downlink packet error rate for CBR traffic. The advantage of multi-channel operation originates from additional operation channels, which can linearly increase system capacity (assuming no co-channel interference), while diversity methods we investigate here only logarithmically increase system capacity. When the always-on VoIP traffic is not present, our simulation results showed similar TCP throughput gains from both multi-channel operation and diversity methods. The results are not presented here to avoid repetition.

As shown in Fig. 2.10, heavy traffic streams like VoIP uplink/downlink and FTP uplink/downlink largely increase the packet error rate of VoIP traffic. We observe that diversity methods still consistently improve TCP throughput at different number of active stations. The packet error rate of VoIP traffic, however, is only slightly affected, and we regard the small difference of the packet error rate between conventional WLAN and diversity methods as random effects in the simulations. In fact, our simulations show little difference among diversity methods; therefore, only MRC-uplink/MRC-downlink method is plotted in the packet error rate figure to avoid clutter. The two-channel-operation method improves TCP throughput and outperforms diversity methods when the number of active stations is less than 10. When the network contains more than 10 active sta-

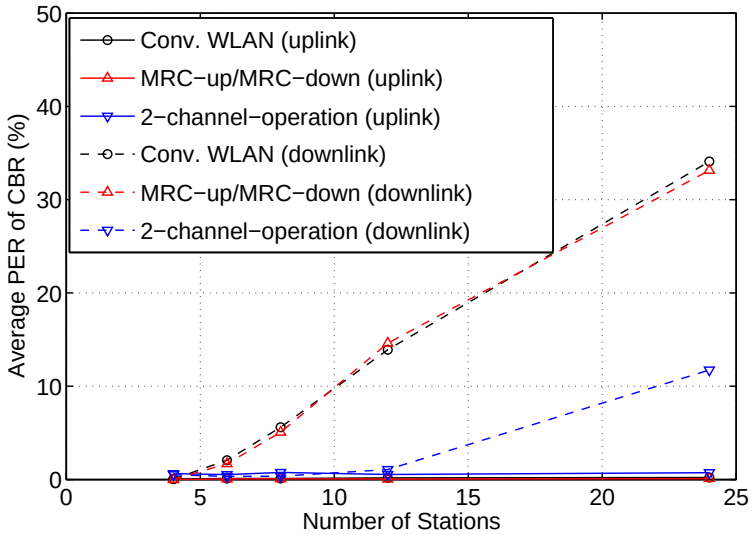
tions, TCP throughput of two-channel-operation method decreases and even becomes worse than conventional WLAN when there are 24 active stations.

To identify the reason of TCP throughput degradation of two-channel-operation, we plotted the TCP uplink and downlink throughput separately in Fig. 2.11. We observe that two-channel-operation generates the highest TCP downlink throughput but the lowest TCP uplink throughput, which taken together causes the lowest total TCP throughput. To explain the reason of TCP uplink throughput degradation, we notice that when two-channel-operation method is used, stations being served in one channel are found in areas twice as large as those in the conventional WLAN, and therefore suffer more packet collisions due to the hidden terminal problem. The above observation suggests that when there are too many active stations in the ESS formed by the cognitive WLAN over fiber system, enough number of channels should be operated to ensure proper file sharing efficiency.

Compared with diversity methods and conventional WLAN, the two-channel-operation method generates the highest packet error rate of CBR traffic when the number of active stations is less than 12, and the lowest packet error rate when the number of active stations is larger than 12. This phenomenon is not easy to see in scenarios under VoIP traffic. To see it more clearly, observe scenarios under IPTV downlink and FTP downlink traffic, enlarged in Fig. 2.12. When two-channel-operation method is used in lightly loaded networks, the packet error rate of CBR traffic is increased because capturing a new packet causes a loss of previously being received packet; in heavily loaded networks, however, the reduced packet collisions due to extra channels outweighs the disadvantage of CBR packet loss due to channel capture effect, causing a lower overall packet error rate of CBR traffic than diversity methods and conventional WLAN. Although two-channel-operation caused a bit higher packet error rate for uplink CBR traffic, the packet error rate in CBR downlink is largely decreased. For VoIP application, such balanced packet error rates provided by two-channel-operation would be very useful.



(a)



(b)

Figure 2.10: TCP throughput and average packet error rate of CBR traffic vs. Number of stations. Traffic: VoIP uplink/downlink + FTP uplink/downlink.

Explanations on the performance improvement

In diversity-based cognitive WLAN over fiber systems, the TCP throughput gains and the packet error rate reductions come from independent channel fading and more antennas being involved at

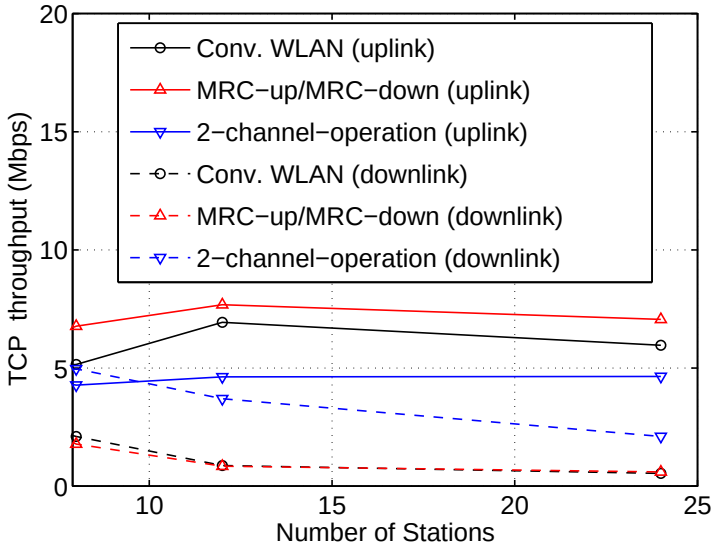


Figure 2.11: TCP throughput degradation of two-channel-operation method in heavily loaded networks. Traffic: VoIP uplink/downlink + FTP uplink/downlink.

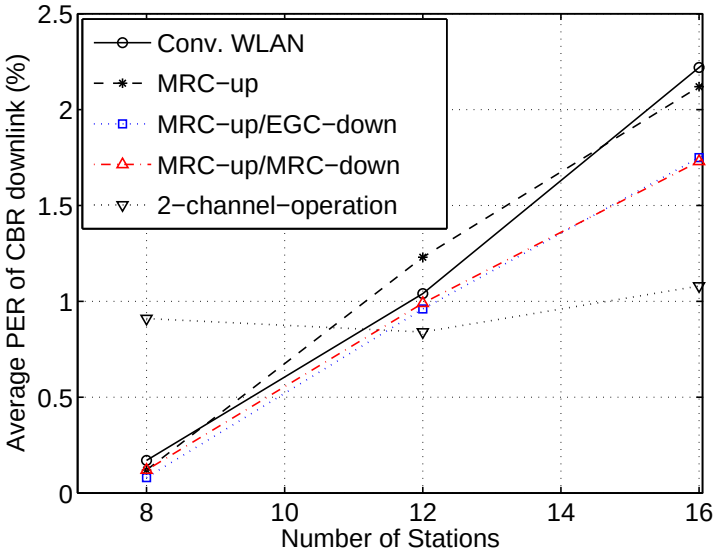


Figure 2.12: Packet error rate degradation of CBR traffic (two-channel-operation in lightly loaded networks). Traffic: IPTV downlink + FTP downlink.

the receiver or transmitter. Gains of two-channel-operation in cognitive WLAN over fiber systems come from channel capture and channel fading effects. We now further explain where these gains come from.

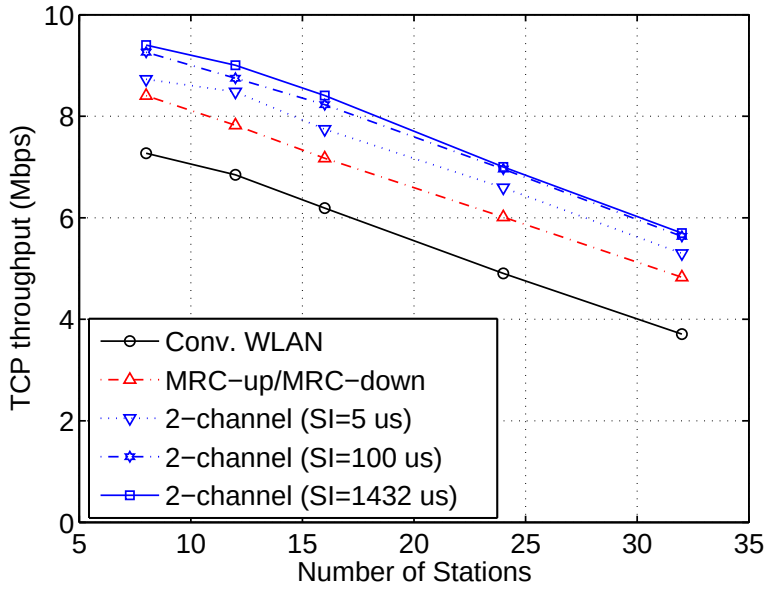
Suppose a conventional WLAN ESS serves 10 stations on channel 1 of AP_1 and another 10 stations on channel 2 of AP_2. Assume these stations are associated to the AP closer to them. Two-channel-operation actually splits the 20 stations into 4 groups, each assigned to one channel through one antenna. The resulting cognitive WLAN over fiber system can be viewed as four independently-operated conventional BSSs. Although co-channel interferences exist between these BSSs, we still gain system capacity due to the linearly increased bandwidth while the signal-to-interference-noise ratio is only logarithmically degraded. On the other hand, a diversity-based cognitive WLAN over fiber system controls antennas and forms BSSs with distributed antennas, serving stations that spread out in the whole area.

An intuitive example can observed from Fig. 2.2. Apparently, stations located in the middle of the area will favor diversity technique since they have similar average path losses to the two antennas, while stations at corner areas will favor multiple-independent-channel-operation method since there will be less co-channel interferences between corners. This observation reminds us that when the location information of stations is available at the CogAP, advanced location-aware channel management techniques can provide even higher system capacity.

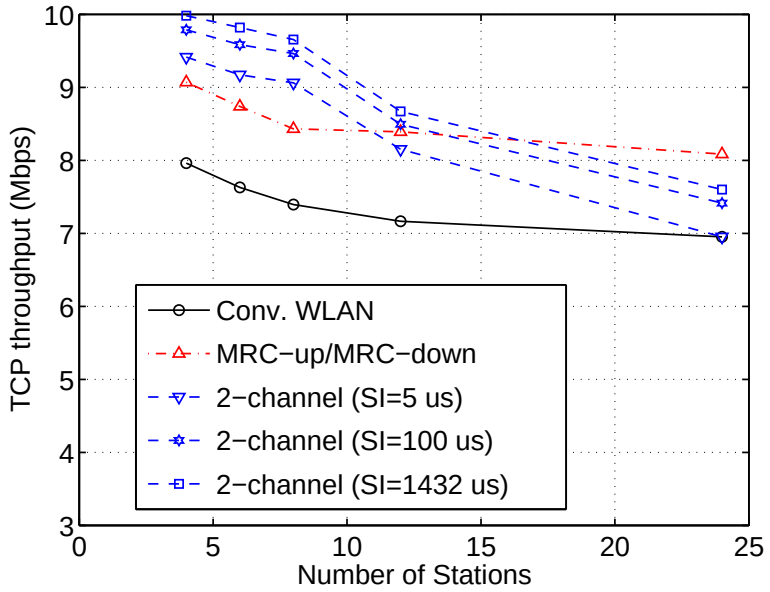
Effects of synchronization interval

We examine effects of different SI values on TCP throughput under three types of traffic, as shown in Fig. 2.13. Effects of SI on the packet error rate of CBR traffic is very small and thus omitted. Since different SI values only affect two-channel-operation; only one diversity method (MRC in uplink and downlink) is plotted for comparison purpose.

We observe that by using SI=1432 μs, TCP throughput can be increased by 5.6%~10% when compared with SI = 5 μs, owning to the fact that larger SI values cause more packet capturing than smaller SI values. However, we also notice that there is little difference between SI=1432 μs and SI=100 μs, indicating that by only looking for the strongest signal during SI=100 μs, a WLAN receiver can achieve most of the throughput gains due to capture effect. For the rest of this chapter, 100 μs is used as the SI value.



(a) Traffic: VoIP uplink/downlink + FTP downlink



(b) Traffic: VoIP uplink/downlink + FTP uplink/downlink

Figure 2.13: Effects of synchronization interval (SI).

2.4.4 Spatially Non-uniformly Distributed Traffic

When a hotspot area has much larger traffic demand than other areas, we face a spatially non-uniformly distributed traffic. We split the $60 \times 30m^2$ area into 6×3 sub-areas and place the hotspot

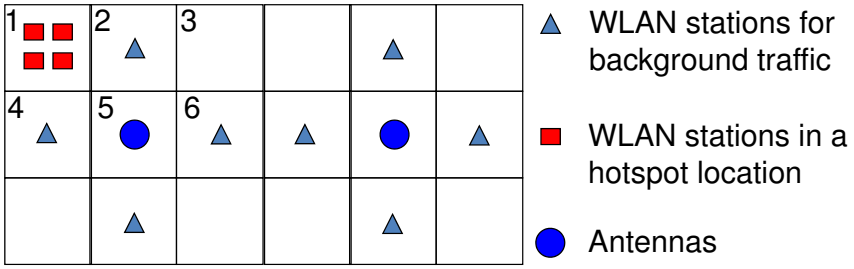
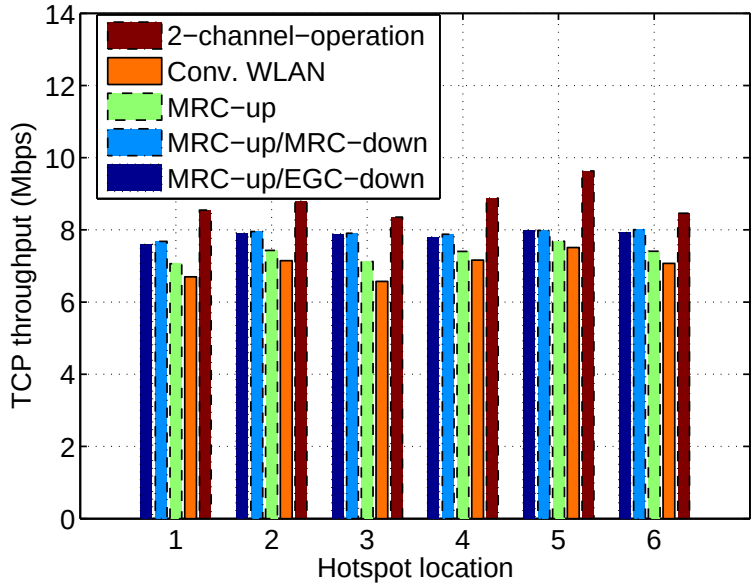


Figure 2.14: Spatially non-uniformly distributed traffic. Hotspot locations are numbered from 1 to 6.

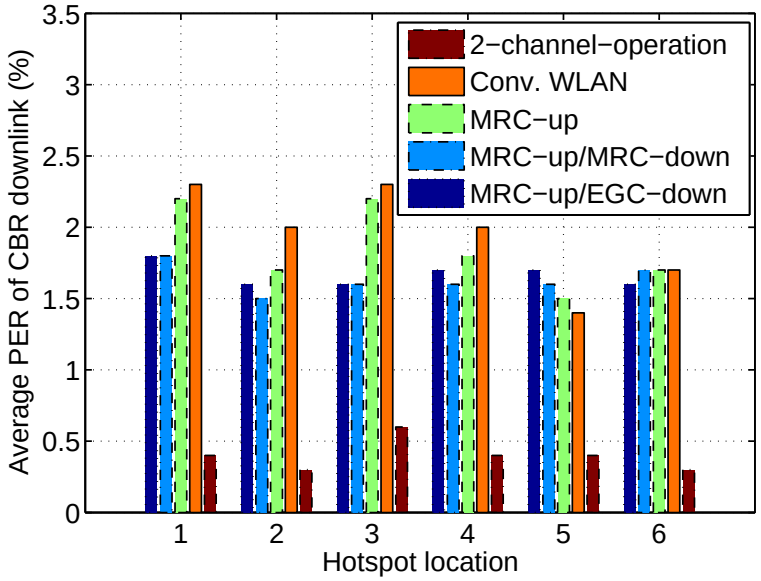
into one of these sub-areas to simulate non-uniformly distributed traffic. Totally 8 stations are used for background traffic and 4 other stations are placed in certain hotspot location, as shown in Fig. 2.14. By geometric symmetry, we only need study hotspot locations from 1 to 6. To concentrate on studying the effects of dynamic traffic, we fix the locations of background-traffic stations at the centers of sub-areas 2, 4, 6, etc. Stations that generate hotspot traffic are also fixed in the center part of their respective sub-area. Only one set of station locations is used for simulations followed.

As shown in Fig. 2.15 and Fig. 2.16, compared with conventional WLAN, diversity methods in the cognitive WLAN over fiber system achieve 10%~62% higher TCP throughput and two-channel-operation achieves 17%~48% higher TCP throughput, whereas only 14%~36% gain is achieved when the traffic is spatially uniformly distributed (comparing Fig. 2.8 to Fig. 2.10). The packet error rate of CBR downlink is also largely reduced. This demonstrates the enhanced capability of handling dynamic traffic in cognitive WLAN over fiber systems. Not surprisingly, two-channel-operation achieves a larger throughput gain when the hotspot is in the corners (e.g., location 1), while diversity methods achieve larger gains when the hotspot is in the overlapping area of antenna A_1 and A_2 (e.g., location 3 and 6). In fact in such areas, MRC in both uplink and downlink achieves higher TCP throughput than two-channel-operation when the standard deviation of shadowing increases to 10 dB.

When the hotspot moves to location 5, stations at the hotspot are closer to A_1. Therefore, co-channel interference from stations in BSS2 to those in BSS1 is less likely due to the capture effect. Thus, we observe a larger TCP throughput gain in the two-channel-operation method, as shown in



(a)



(b)

Figure 2.15: TCP throughput and average packet error rate of CBR downlink vs. Hotspot location. Traffic: VoIP uplink/downlink + FTP downlink. Antenna-distance = 30 m. Shadowing standard deviation = 3.5 dB.

Fig. 2.15 and Fig. 2.16.

We further study antenna-distance effects under spatially non-uniformly distributed traffic.

Effects of antenna-distance

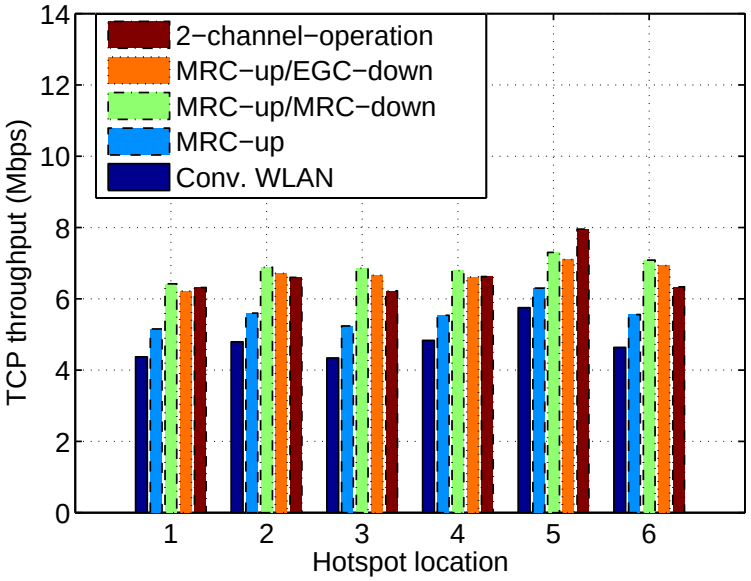
The antenna-distance is also the physical size of a BSS in our simulations. Comparing Fig. 2.17 with Fig. 2.16, we observe that both diversity and two-channel-operation methods provide larger TCP throughput gains when the antenna-distance is increased to 45 or 60 meters. Signal-to-noise-ratio gains generated by diversity methods have larger effects on throughput due to the increased antenna-distance and consequently increased path loss. Two-channel-operation method provides higher throughput gains due to the increased antenna-distance and consequently reduced co-channel interference, especially at hotspot location 5 where stations are less susceptible to co-channel interferences from stations being served by the neighboring antenna.

2.5 Summary

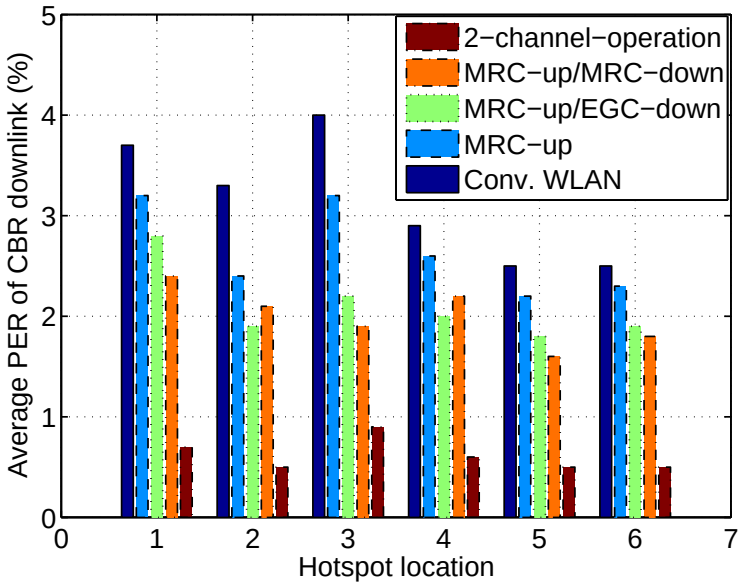
A cognitive WLAN over fiber system, as an application scenario of FMDA systems, can provide a cost-effective and efficient method for devices to equally share the ISM band by taking advantage of cognitive radio capabilities. In this chapter, we have proposed two methods that utilize the specialized capabilities of the cognitive WLAN over fiber system to improve system capacity by reducing packet collisions through load balancing and employing diversity to reduce the effects of packet collisions.

By exploiting the wide-band radio-over-fiber connections between antennas and the cognitive access point, multiple-independent-channel-operation at each antenna has been proposed to reduce the collision probability in each channel by moving stations to different channels. By exploiting the distributed antennas in a cognitive WLAN over fiber system, we have demonstrated the use of macro-diversity to increase the sensing capability of the cognitive access point. Simulation results show that both methods can achieve 14%~18% TCP throughput gain and 10%~50% packet error rate reduction in constant bit-rate traffic for spatially uniform traffic in an IEEE 802.11g network, and up to 62% TCP throughput gain when hotspots exist.

We also studied effects of synchronization interval and antenna-distance of cooperating antennas. Similar TCP throughput gain and packet error rate reduction are observed in all scenarios.

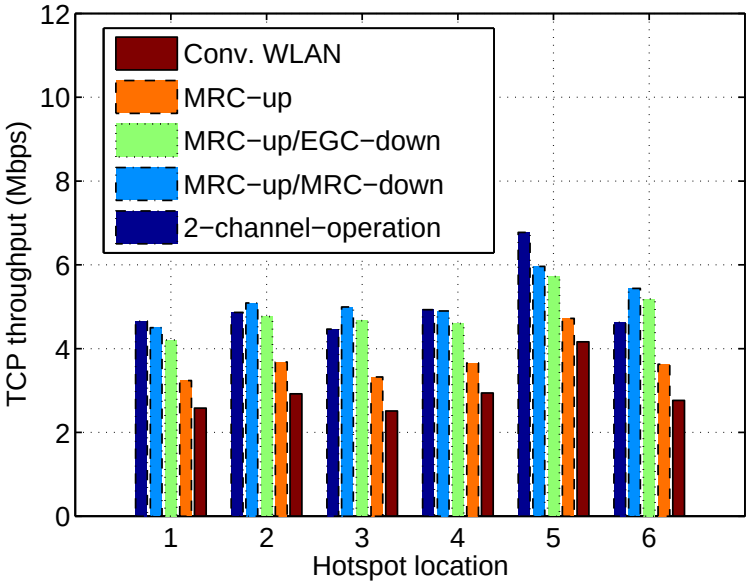


(a)

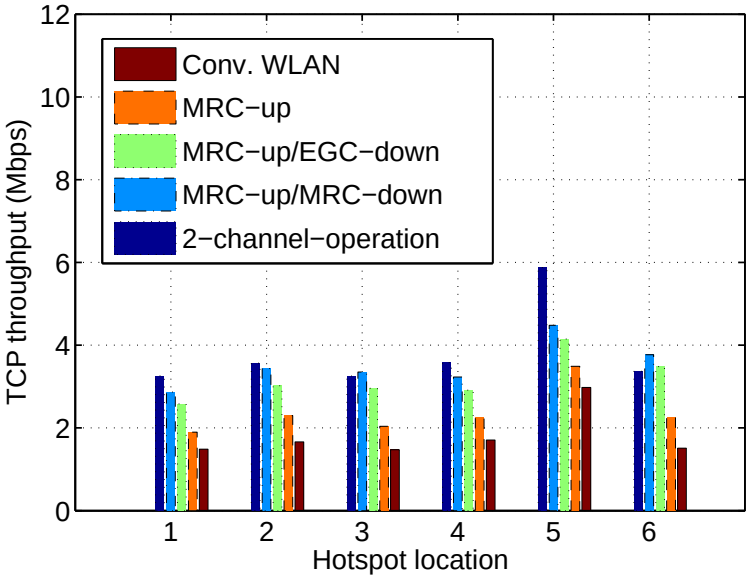


(b)

Figure 2.16: TCP throughput and average packet error rate of CBR downlink vs. Hotspot location. Traffic: VoIP uplink/downlink + FTP downlink. Antenna-distance = 30 m. Shadowing standard deviation = 10 dB.



(a) Antenna-distance = 45 m.



(b) Antenna-distance = 60 m.

Figure 2.17: Effects of antenna-distance. Traffic: VoIP uplink/downlink + FTP downlink. Antenna-distance $\in\{45\text{m}, 60\text{m}\}$. Shadowing standard deviation = 10 dB.

Chapter 3

Energy Conservation via Antenna Scheduling in Fiber-connected Femto Base Stations [1]

3.1 Introduction

Femtocell addresses the challenge on indoor wireless connectivity by pulling base station (BS) closer to users. While the cell planning is shifted towards universal frequency reuse pattern largely due to the advantages on area spectral efficiency previously reported in cellular networks [85], in very dense urban areas with a high concentration of residential and business users, inter-cell interference among femtocells using universal frequency reuse may drastically decrease system capacity. Cooperative communications among femtocells are therefore needed to eliminate the interference in the expense of decreasing femtocells' backhaul capacity. To reduce the cooperation expense, motivated by advances in radio over fiber (RoF) technique [31] that improve indoor coverage of wireless local area network [32, 33] and cellular networks [120], in this chapter we propose to configure Fiber-connected Massively Distributed Antennas (FMDA) to form localized

[1] This chapter is based on [86] co-authored with Dr. A. Attar and Dr. V. Leung, [87] co-authored with Mr. J. Hajipour, Dr. A. Attar and Dr. V. Leung, and [88] co-authored with Dr. A. Attar and Dr. V. Leung.

cooperation among femtocells. Focusing on coordinated multi-point transmissions (CoMP) for femtocells in Long-Term Evolution (LTE), to which we referred as *femto-CoMP*, we demonstrate its considerable spectral and energy efficiency enhancement over standalone femtocells.

We first formulate the power consumption under a generalized framework that takes into account various contributing factors including orthogonal frequency-division multiplexing (OFDM), error control coding, and different packet schedulers that consider user fairness. After analyzing the computation cost of both user scheduling and multiple-input multiple-output (MIMO) linear beamforming, we developed an optimization tool in a single femto-CoMP cluster for service providers to maximize energy conservation by adjusting the number of transmission antennas and controlling wireless transmission power in linear beamforming. Based on the analysis results, we use 2-MIMO and 3-MIMO femto-CoMPs as building blocks in a new group of network configurations, which employs antenna scheduling to simultaneously improve spectral and energy efficiency.

The proposed configurations are denoted as FMDA-N_r-N_c-N_{ts}: N_r ceiling-mounted fiber-connected antennas are deployed to provide wireless access in a building, every N_c of which form one femto-CoMP cluster that serves N_c single-antenna users at a time. All antennas are equally split into N_{ts} sets, each being chosen in its assigned time slot while other antennas are left unused. The idea of antenna scheduling is to selectively use a large number of antennas which are only made possible in FMDA systems such that downlink throughput can be enhanced due to reduced interference and energy efficiency can be improved due to the reduced amount of calculations in collecting channel state information (CSI) and linear beamforming. Antenna scheduling also enables a broad range of network configuration choices which offers LTE service providers the flexibility of choosing different balance points between spectral and energy efficiency.

The rest of the chapter is organized as follows. We first briefly review the architecture of FMDA and present the power consumption framework. The optimization tool in single femto-CoMP network is then developed. Based on the analysis results, we propose a group of fixed antenna scheduling strategies and verify its performance improvement in Section 3.5. In the end we summary the chapter.

3.2 Power Consumption in Fiber-connected Massively Distributed Antennas

3.2.1 Architecture

The architecture of FMDA is depicted in Fig. 3.1, composed of three components, namely the antennas, the fiber-connection medium and the centralized processing system. The antennas in FMDA transmit and receive radio frequency signals over the air. There is no processing capability embedded in the antennas, except for radio frequency signal amplification and electrical-optical/optical-electrical conversion. The antennas will be excited from the optical link, which can be OM2 or laser-optimized-OM3 multi-mode fibers for low deployment costs, or single-mode fibers for larger scale deployments. Owing to low propagation loss of optical fibers, the distance between the antennas and the centralized processing system, with off-the-shelf RoF equipment, can reach hundreds of meters, which suffice to cover most buildings in practice. FMDA integrates the capabilities of a distributed network, based on its massively distributed antenna topology and the centralized processing capability at the centralized processing system. Digital signal processors are pooled in the centralized processing system, which acts as femto-BS gateway in LTE-Advanced, to provide an unparalleled throughput delivery performance in an energy efficient manner.

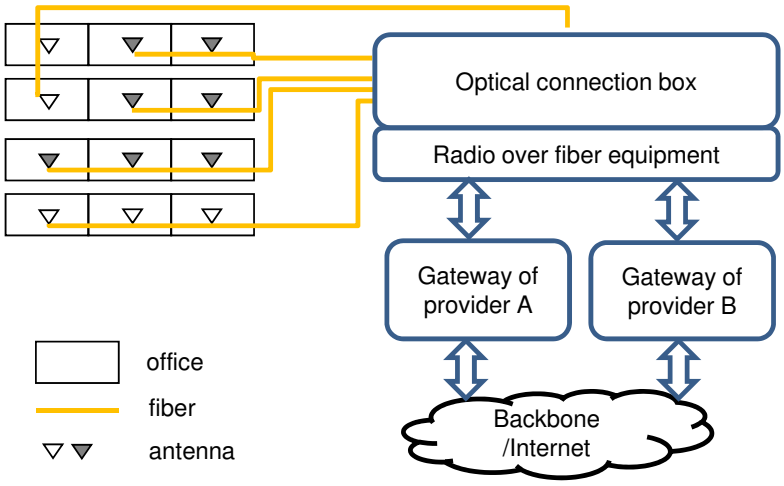


Figure 3.1: Fiber-connected femto base stations with two service providers.

In this chapter we consider a FMDA system with N_r antennas. The number of active users is N_{Utotal}. Every N_c antennas form a femto-CoMP cluster to serve N_U users. Within each femto-CoMP, a user scheduler chooses N_c users at a time to serve through $N_c \times N_c$ multi-user MIMO downlink transmission. In particular, we use zero-forcing beamforming for its simplicity and near-optimal performance at high SNRs. If each femto-CoMP contains an equal number of users, the number of femto-CoMPs, denoted by N_{CoMP}, can be given by N_{CoMP}=N_{Utotal}/N_U.

3.2.2 Power Consumption Model

The total power consumption of a FMDA network, denoted by P_{tot}, is given by

$$P_{tot} = P_{sp} + P_{RF}, \tag{3.1}$$

where P_{sp} is the signal processing component and P_{RF} is the transmission power component contributed by RoF links. In this chapter we focus on downlink as downlink traffic are much heavier than uplink traffic and thus dominate the total power consumption.

Signal processing component P_{sp}

Consider a BS with N_c antennas serves N_c users through $N_c \times N_c$ zero-forcing beamforming. We now present a general power consumption model of signal processing component, which incorporates user scheduling overhead into the micro-BS model in [121] and the cooperative BS model in [64]:

$$P_{sp} = P_{spB}\alpha_B + P_{spB}\alpha_{FO}N_{CoMP}N_c + P_{spB}\alpha_E N_{CoMP}N_c^2 + P_{spMS}, \tag{3.2}$$

$$P_{spMS} = P_{spB}\alpha_{MS}\left[N_{CoMP}\left(\frac{8N_c^3}{3} + 2N_c^2 + N_c\log_2 N_c + 5N_c + \frac{f(N_c,N_U)}{N_{sc}}\right)\right]. \tag{3.3}$$

The P_{spB} represents the power consumption contributed by base amount, which does not depend on N_c; P_{spMS} represents the power consumption contributed by zero-forcing beamforming and user scheduling. The α_B, α_{FO}, α_E and α_{MS} are positive coefficients that had been measured in

Table 3.1: Power consumption coefficients

Coefficients	α_B	α_E	α_{FO}+α_{MS}
Values	0.87	0.10	0.03

[64, 66] to quantify the power consumption amount attributed to the base amount, forward error correction and OFDM, channel estimation, and zero-forcing beamforming and user scheduling, respectively. Each coefficient establishes the relation between the floating-point operation (flop) count and the power consumption in its corresponding component; therefore, its value depends on particular signal processing hardware and channel bandwidth configuration. Table 3.1 lists the coefficients measured in the 10-MHz MIMO-OFDM reference system in [64, 66], which will be used throughout this chapter.

We explain the formulation of (3.2) as follows. The flop count of forward error correction, OFDM, channel estimation, zero-forcing beamforming and user scheduling scales linearly with N_{CoMP}. The flop count of forward error correction and OFDM modulation scales linearly with N_c. The flop count of channel estimation is proportional to the number of links and thus incurs $O(N_c^2)$ flops. The zero-forcing beamforming is dominated by matrix inversion to generate beamforming directions and one round of water-filling among N_c users to allocate power. Generating beamforming directions via matrix inversion involves $8N_c^3/3$ flops, provided Gaussian elimination with partial pivoting is used [122]; water-filling includes calculating N_c vector 2-norms (each requires $2N_c$-1 flops), sorting the norms requiring $O(N_c \log_2 N_c)$ flops, and determining the water level and power allocation requiring $5N_c$ flops. When N_c is large, the dominating term in zero-forcing beamforming is $8N_c^3/3$. For $N_c \leq 3$, however, the computation cost of matrix inversion can be reduced by taking analytic form based on Cramer rule [123], which requires 8 flops for N_c=2 and 39 flops for N_c=3. The power consumption of user scheduling is expressed as $f(N_c, N_U)$, where the function $f(\cdot)$ can be exponential in exhaustive search or polynomial $N_c^{\zeta} N_U^{\kappa}$ for schedulers with lower computation costs, where ζ and κ vary with different schemes. When the channel bandwidth changes, we remark that the power consumption of forward error correction, OFDM, channel estimation

and beamforming scales linearly with the number of subcarriers while scheduling can be done at a coarser granularity, per resource block, which results in a discounting factor $1/N_{sc}$ (N_{sc} is the number of subcarriers per resource block in LTE).

The P_{sp} in (3.2) is decomposed according to two principles. One is to group N_c-terms and thus provide the insights on how to choose N_c. Another is to group functions which are implemented in the same circuit or accomplished at the same time such that the corresponding coefficients can be measured by profiling power consumption of real products. The forward error correction and OFDM functions are usually conducted on dedicated chips; channel estimation is performed during uplink transmission; scheduling and beamforming are usually performed at digital signal processors for flexible implementation.

Computation cost of user scheduling $f(N_c, N_U)$

A proportional fairness (PF) exhaustive scheduler is extremely expensive because zero-forcing beamforming is needed at each subcarrier to estimate the instantaneous rate and the estimation is required by each of $\binom{N_U}{N_c}$ trials. Without counting PF rate update, the flop count already amounts to $\binom{N_U}{N_c}\frac{8N_c^3}{3}$.

A proportional fairness scheduler based on semi-orthogonal user selection (PF-SUS) [124] has the flop count $\frac{N_U}{2}(N_c^3 + 8N_c^2 - 10N_c + 4) - \frac{N_c}{3}(N_c^3 + 5.5N_c^2 - 10N_c - 2.5)$ and can be approximated as $\frac{N_U}{2}(N_c^3 + 8N_c^2 - 10N_c)$ when $N_U > N_c$. A round-robin scheduler based on semi-orthogonal user selection (RR-SUS) has an approximate flop count $\frac{N_U}{4}(N_c^3 + 8N_c^2 - 10N_c)$ when $N_U > N_c$. Table 3.2 lists the exact flop count of the schedulers as well as the asymptotic flop count, which only takes the highest-order of N_c terms in the exact flop count.

The simplest user scheduling to preserve fairness is round-robin. In this chapter we consider a round-robin tournament scheduler, termed *robinT*. When the set of rooms being served in two consecutive time slots is the same, the scheduler ensures that the set of users being scheduled differs. In the case that a user combination incurs a poor channel conditional number, the combination only appears once and thus has minor effects on the system throughput. Consider a typical multi-floor

office building where we deployed a FMDA system by mounting one antenna in each room. We first make robinT work for $N_c > 2$ by setting up an initial grouping in which every possible combination is equally probable. To improve spectral efficiency of each combination, we then enforce a heuristic constraint: in each round of multi-user MIMO transmission, one and only one user is scheduled from each room. Take N_U=12 and N_c=6 as an example. The constraint comes from the intuition that with only free-space path loss and wall penetration loss at present, the smallest conditional number of all possible 6×6 submatrices out of the 12×6 channel matrix results from choosing only one user per room.

Transmission power component P_{RF}

Besides the signal processing component, another power consumption contributor is the transmission power component, contributed by RoF links and given by

$$P_{RF} = P_{\text{RoFD}} \times N_{CoMP} \times N_c, \tag{3.4}$$

where P_{RoFD} is the power per RoF downlink and $N_{CoMP} \times N_c$ represents the total number of RoF links. We adopt the model presented in [125] to estimate P_{RoFD}:

$$\begin{aligned} P_{\text{RoFD}} &= \left(P_{\text{LDA}} + P_{\text{LD}} + P_{\text{PD}}\right) + P_{\text{PA}}, \\ P_{\text{LDA}} &= P_{\text{LDAout}}/\gamma_{\text{LDA}}, \;\; P_{\text{PA}} = P_{tx}/\gamma_{\text{PA}}, \\ P_{\text{LD}} &= 140 \text{ mW}, \;\; P_{\text{PD}} = 83 \text{ mW}, \end{aligned} \tag{3.5}$$

where P_{LDA} is the power consumption of one laser-diode amplifier and P_{LDAout} is its output power, set at 5 dBm in [125]; P_{LD}, P_{PD} and P_{PA} are power consumption contributed by a laser diode, a photodiode, and a power amplifier, respectively; γ_{LDA} and γ_{PA} indicate the efficiency of the laser-diode amplifier and the power amplifier, both fixed at 2.2%; P_{tx} represents the transmission power per antenna. From (3.5) we learn that when a RoF link is not excited, there still exists a fixed power consumption, $P_{\text{LDA}} + P_{\text{LD}} + P_{\text{PD}}$.

Table 3.2: Floating-point operation count of scheduling: $f(N_c, N_U)$

Scheduler	Granularity	Exact flop count	Asymptotic flop count
PF exhaustive search	Resource block	$\binom{N_U}{N_c}\frac{8N_c^3}{3}$	$\frac{8}{3}N_c^3 N_U^N$
PF-SUS	Resource block	$\frac{N_U}{2}(N_c^3 + 8N_c^2 - 10N_c + 4) - \frac{N_c}{3}(N_c^3 + 5.5N_c^2 - 10N_c - 2.5)$	$0.5N_c^3 N_U$
RR-SUS	Resource block	$\frac{N_U}{4}(N_c^3 + 8N_c^2 - 10N_c + 2) - \frac{N_c}{12}(N_c^3 - 2N_c^2 - 10N_c - 4) + \frac{2N_c^2}{N_U}$	$0.25N_c^3 N_U$
robinT	Resource block	1	1

3.3 Energy Efficiency in Single Femto-CoMP

Define energy efficiency as

$$\eta_e \triangleq \frac{\text{System throughput}}{\text{Total power consumption}} = \frac{B \times N_{Utotal} \times \eta_s}{P_{tot}}, \tag{3.6}$$

where B is the channel bandwidth, $N_{Utotal} = N_U$ for single femto-CoMP, η_s is average spectral efficiency per user under a given user scheduler and P_{tot} is total power consumption, given by (3.1), (3.2), and (3.4). The use of "average spectral efficiency per user", although not precise, can be viewed as per-user throughput averaged over channel bandwidth and collected during network operation. Our focus in this section is to find optimal N_c and P_{tx} to maximize energy efficiency in single femto-CoMP.

3.3.1 Approximate Spectral Efficiency

Average spectral efficiency per user under a given user scheduler is $\eta_s = \frac{1}{N_U}\sum_{i=1}^{N_U}\eta_{s,i}$, where $\eta_{s,i}$ is the average spectral efficiency of the i-th user. The η_s can be approximated as a time-average over possible channel variations (including user location changes, dynamic shadowing and fading):

$$\eta_s = \frac{1}{N_{slot}}\sum_{t=1}^{N_{slot}}\left\{\frac{1}{N_U}\sum_{i=1}^{N_U}\left[I_{\Omega_t}(i)\log_2\left(1 + P_{tx}\frac{\|h_{i,t}\|^2}{P_{noise}}\right)\right]\right\}, \tag{3.7}$$

where N_{slot} is the number of time slots observed; Ω_t is the scheduled user set at the t-th time slot; $I_{\Omega_t}(i)$ is an indicator function that takes value 1 when the i-th user belongs to Ω_t; P_{tx} is total transmission power constraint per MHz averaged over all antennas; P_{noise} is thermal noise power per MHz, and $\|h_{i,t}\|$ is the average frequency channel gain of the i-th user at the t-th time slot. As the system of interest is single femto-CoMP, there is no inter-CoMP interference. Note that the use of Gaussian signaling is assumed in (3.7).

The η_s in (3.7) can be periodically monitored by service providers to track traffic pattern changes and channel variations. By $\sum_{i=1}^{N_U} I_{\Omega_t}(i) = N_c$, the η_s at high SNRs can be approximated as a linear function of $\ln(P_{tx})$:

$$\eta_s = \frac{N_c}{N_U \ln 2} \ln P_{tx} + \frac{1}{N_{slot} N_U} \sum_{t=1}^{N_{slot}} \sum_{i=1}^{N_U} \left(I_{\Omega_t}(i) \log_2 \frac{\|h_{i,t}\|^2}{P_{noise}} \right), \tag{3.8}$$

where the second term can be viewed as an estimated expectation over user locations, shadowing, fading, user scheduling decisions and other random variables through Monte-Carlo simulations, provided that N_{slot} is large enough. Since the second term also scales linearly with N_c, we rewrite (3.8) as

$$\eta_s = N_c (c_1 \ln P_{tx} + c_2), \tag{3.9}$$

$$c_1 = \frac{1}{N_U \ln 2}, \quad c_2 = \frac{1}{N_{slot} N_U N_c} \sum_{t=1}^{N_{slot}} \sum_{i=1}^{N_U} \left(I_{\Omega_t}(i) \log_2 \frac{\|h_{i,t}\|^2}{P_{noise}} \right). \tag{3.10}$$

3.3.2 Maximize Energy Efficiency

Combining (3.1), (3.2), (3.4), , (3.9), (3.10) and dropping lower order N_c terms in P_{spMS}, we have

$$\eta_e = \frac{BN_U(c_1 \ln P_{tx} + c_2)}{\frac{BP_{tx}}{\gamma} + c_3}, \tag{3.11}$$

$$c_3 = P_{\text{LDA}} + P_{\text{LD}} + P_{spB} \cdot \alpha_{FO} + P_{spB} \cdot g(N_c), \tag{3.12}$$

$$g(N_c) = \frac{\alpha_B}{N_c} + \alpha_E N_c + \alpha_{MS}\left[\frac{8N_c^2}{3} + \frac{f(N_c, N_U)}{N_c N_{sc}}\right], \tag{3.13}$$

where γ_{PA} is replaced by γ for notation convenience. Since $g(N_c)$ contains all the N_c terms but no P_{tx} terms, we propose to maximize η_e by first finding optimal N_c to minimize $g(N_c)$ and then finding optimal P_{tx} under such N_c.

Optimal N_c

For PF-SUS and robinT schedulers, $g(N_c)$ is given by:

$$g_{\text{PF-SUS}} = \frac{\alpha_B}{N_c} + \alpha_E N_c + \alpha_{MS}\left[\frac{8N_c^2}{3} + \frac{N_U}{2} \cdot \frac{N_c^3 + 8N_c^2 - 10N_c}{N_c N_{sc}}\right], \tag{3.14}$$

$$g_{\text{robinT}} = \frac{\alpha_B}{N_c} + \alpha_E N_c + \alpha_{MS}\frac{8N_c^2}{3}. \tag{3.15}$$

Both $g_{\text{PF-SUS}}$ and g_{robinT} are convex as their second-order derivatives are positive.

Consider PF-SUS scheduler. Assume N_c continuous for analytical convenience. By $\arg\max_{N_c} \eta_e = \arg\min_{N_c} g(N_c)$, we find the optimal N_c, denoted by N_c^*, by solving

$$\frac{\mathrm{d}g(N_c)}{\mathrm{d}N_c} = \alpha_{MS}\left(\frac{16}{3} + \frac{N_U}{N_{sc}}\right)N_c + \alpha_E + \alpha_{MS}\frac{4N_U}{N_{sc}} - \frac{\alpha_B}{N_c^2} = 0. \tag{3.16}$$

Apparently $g_{\text{PF-SUS}}(N_c)$ is strictly increasing when $N_c > N_c^*$ and strictly decreasing when $2 \le N_c < N_c^*$. Since the value of N_{sc} is 12 in LTE, eq. (3.16) is rewritten as

$$\alpha_B = \alpha_E N_c^2 + \alpha_{MS}\left[\left(\frac{16}{3} + \frac{N_U}{12}\right)N_c^3 + \frac{N_U}{3}N_c^2\right], \tag{3.17}$$

which suggests that to double N_c^* (thus doubling the spectral efficiency), system designers need to roughly quadruple channel estimation efficiency and octuple the efficiency of beamforming and scheduling. Instead of giving involved algebraic roots of a cubic function, we list requirements on α_B, α_{MS}, α_E for some typical values of N_c^*:

$$\begin{aligned}
&\alpha_B \approx 4\alpha_E + \alpha_{MS}(43 + 2N_U), \text{for } N_c^* = 2, \\
&\alpha_B \approx 16\alpha_E + \alpha_{MS}(341 + 10.67N_U), \text{for } N_c^* = 4, \\
&\alpha_B \approx 36\alpha_E + \alpha_{MS}(1152 + 30N_U), \text{for } N_c^* = 6.
\end{aligned} \tag{3.18}$$

When a robinT scheduler is used, similarly we have,

$$\begin{aligned}
&\alpha_B \approx 4\alpha_E + 43\alpha_{MS}, \text{for } N_c^* = 2, \\
&\alpha_B \approx 16\alpha_E + 341\alpha_{MS}, \text{for } N_c^* = 4, \\
&\alpha_B \approx 36\alpha_E + 1152\alpha_{MS}, \text{for } N_c^* = 6.
\end{aligned} \tag{3.19}$$

Optimal P_{tx} under N_c^*

Let $\frac{\partial \eta_e}{\partial P_{tx}} = 0$. The optimum P_{tx}^* is found by solving

$$c_1 c_3 \gamma - BP_{tx}(c_2 - c_1) - c_1 BP_{tx} \ln P_{tx} = 0. \tag{3.20}$$

Note that P_{tx} in LTE indoor deployment is less than 1 watt/MHz. Therefore to have a solution in (3.20), we need $c_2 > c_1$; otherwise η_e would be monotonically increasing in P_{tx}. From (3.8) one can show $c_2 > c_1$ if the SNR at P_{tx}=1 is larger than $\lg e$, i.e., 4.3 dB. This condition is certainly satisfied at high SNRs.

To verify whether η_e reaches the global maximum at P_{tx}=P^*, we evaluate

$$\frac{\partial^2 \eta_e}{\partial P_{tx}^2}|_{P_{tx}=P^*} = \frac{-\gamma c_1 B N_U}{(P^*)^2 (BP^* + c_3 \gamma)} < 0. \tag{3.21}$$

So η_e is strictly concave at $P_{tx} = P^*$. From (3.20) one can show that $\frac{\partial \eta_e}{\partial P_{tx}} < 0$ when $P^* < P_{tx} < 1$ and $\frac{\partial \eta_e}{\partial P_{tx}} > 0$ when $P_{tx} < P^*$, indicating P^* is a global optimum and the convergence of a nonlinear equation solver is guaranteed.

3.4 Energy Efficiency in Multiple Femto-CoMPs

The above optimization process is based on single-CoMP configuration. In multiple femto-CoMPs, however, the parameter c_2 in (3.7) is affected by inter-CoMP interference and therefore depends on frequency reuse pattern and $\{N_{CoMP}, N_c\}$ configuration, making it difficult to have a reliable estimate on spectral efficiency. To optimize the overall system energy efficiency, one needs to find the optimal transmission power and the optimal number of cooperating antennas for each femto-CoMP. This combinatorial optimization problem is generally difficult to be solved in real time. In this section, we propose an antenna scheduling approach to heuristically solve this problem.

3.4.1 Motivation of Antenna Scheduling

Our approach exploits one major advantage of FMDA architecture, abundant antennas available in a system. Compared with an outdoor cellular network, an indoor FMDA system typically has much less users nearby each antenna; therefore, universal frequency reuse may not provide the largest system capacity. Instead of using all antennas and trying to reduce inter-CoMP interference, we could selectively use antennas in time slots and as the result of less interference and less number of femto-CoMPs per slot, spectral and energy efficiency are expected to be concurrently improved.

In this joint scheduling problem, we need to find an optimal user-antenna matching. To reduce complexity, we decouple it into two scheduling problems: user scheduling and antenna scheduling. For user scheduling, we use round-robin tournament algorithm for its simplicity, fairness and good

performance in spectral efficiency when the number of users per femto-CoMP is small. Consider a multi-floor office building where we deployed a FMDA system by mounting one antenna in each room. Once the user set is given, the antennas in corresponding rooms have to be chosen to ensure beamforming performance. Therefore, our first problem is to determine the number of users being scheduled for each time slot. Scheduling too many users causes excessive inter-CoMP interference; scheduling too few wastes transmission opportunities. Also, the users being scheduled should be spread in space to reduce inter-CoMP interference.

Once the user set is determined, CSI should be collected to enable antenna scheduling and beamforming. The second problem is to determine the amount of CSI needed for scheduling purpose. Since CSI varies over space, time and frequency, its collection is an expensive operation in MIMO systems, as indicated in (3.2), and needs to be carefully planned.

Given collected CSI, the third problem is finding the optimal clustering pattern to maximize an efficiency criterion subject to certain constraint, as required by the service provider. Examples of such balance include maximizing spectral efficiency while maintaining minimal energy efficiency in network peak hours, and maximizing the energy efficiency while maintaining minimal level of service at midnight.

3.4.2 Antenna Scheduling Based on Two-user MIMO

In the energy efficiency analysis of single CoMP cluster, after plugging $\alpha_B = 0.87$, $\alpha_E = 0.10$ (as used in [64]) and $\alpha_{MS} = 0.02$ into (3.19) to check the optimum condition of N_c^*=2, we found $\alpha_B < 4\alpha_E + 43\alpha_{MS}$, indicating that two-user MIMO provides the best energy efficiency among all N_c choices. By solving (3.20) we found that the optimal transmission power is 5.1 mW/MHz for a 10-MHz channel. The optimal transmission power is under the output power limit regulated in LTE, 100 mW per single-antenna home BS [126]. Note that we consider each antenna equivalent to one home BS (or femto-BS). Motivated by this finding, we attempt to solve the antenna scheduling problem by posing two constraints: each femto-CoMP has the same N_c; N_c is fixed at 2. The P_{tx} is fixed as transmission power control is not the focus of this chapter.

To avoid expensive CSI collection, we base the scheduling strategy on geographical information that includes dual-stripe floor plan and the existence of wall/floor penetration loss. Specifically, cooperation of two antennas only occur when they are located in neighbor rooms and when possible, the cooperation between two antennas on one side of corridor is preferred because there is only one wall penetration and thus more signal energies can be harvested in two-user MIMO operation.

Under the above constraints, we denote possible configurations as FMDA-N_r-2-N_{ts}, which indicates that the system is equipped with N_r ceiling-mounted antennas, every two of which form one femto-CoMP cluster, and all antennas are equally split into N_{ts} sets. Each set of antennas is used in its assigned time slot while the rest of antennas are left unused. Apparently the energy consumption of both signal processing and CSI collection is reduced by N_{ts} times.

In the ideal case, antenna scheduling should track channel variations on per-slot basis. However, extra CSI between each scheduled user and its neighboring antennas are needed when we want to reconfigure femto-CoMPs. To reduce the CSI collection overhead, we could increase the time interval of scheduling update to the channel coherence time or even the shadowing coherence time; the scheduling granularity in frequency can also be increased to the coherence bandwidth rather than one resource block. In this chapter, we propose multiple promising static antenna scheduling and demonstrate their potentials to simultaneously improve energy and spectral efficiency.

When standalone femtocells are operated, antenna scheduling in time is impossible as switching off any femto-BS for a long time will affect short-term user fairness while a frequent switching is impractical due to the required femto BS boot-up and network synchronization time.

3.5 Numerical Results

We consider a four-floor office building with six rooms per floor. The rooms are divided into two stripes by a two-meter wide corridor. The dimension of each room is $12 \times 6 \times 3\ m^3$. We assume two users with fixed separation d_{U}=12 meters, as depicted in Fig. 3.2, and focus on saturated down-

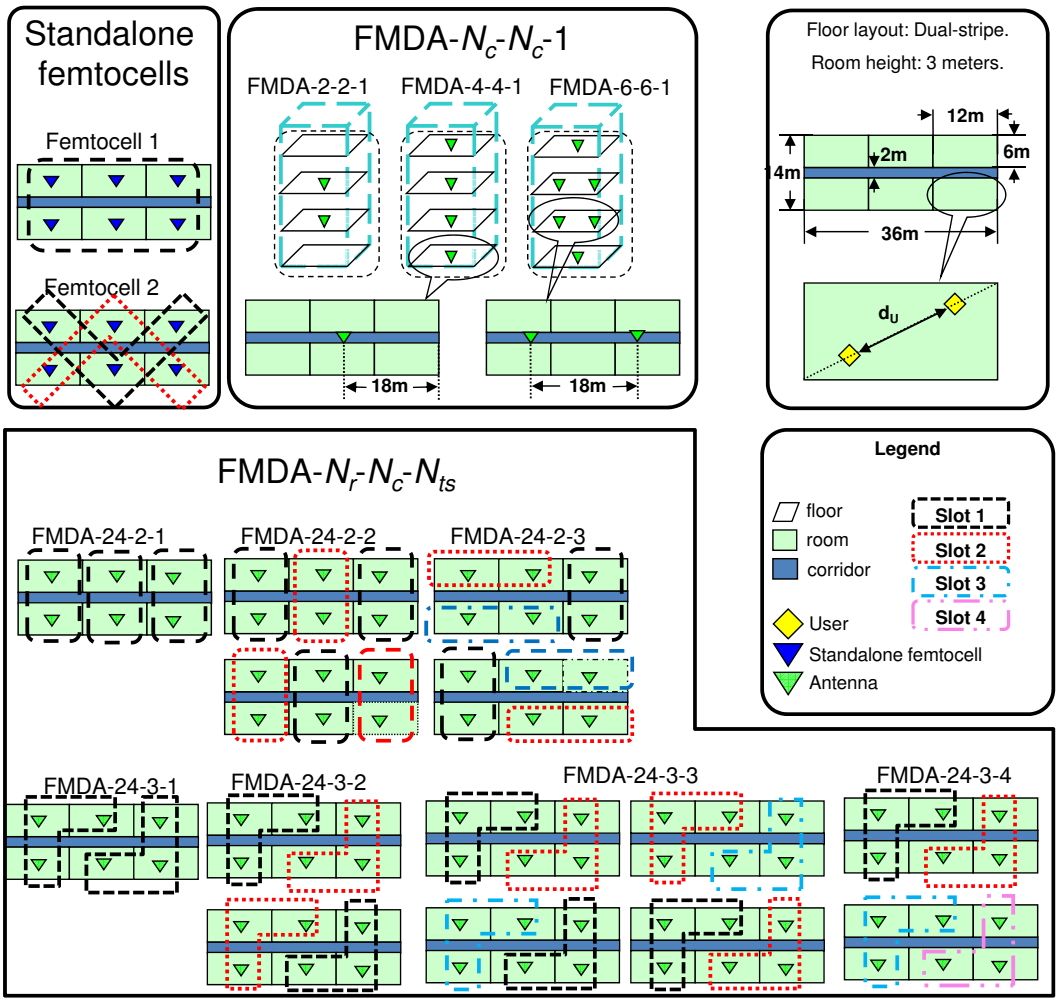


Figure 3.2: Simulation scenarios for femtocell and FMDA. Antennas of FMDA-N_r-N_c-N_{ts} are scheduled in up to four time slots.

link traffic. To provide wireless access to users in the building, we have two options: standalone femtocell and FMDA. The following describes antenna or femto-BS placement within each configuration. Antenna scheduling and clustering are detailed in corresponding FMDA configurations.

3.5.1 Simulation Model

Standalone femtocell: Each femto-BS forms a closed subscriber group and therefore operates independently. We assume that there exists one femto-BS, equipped with one antenna in each room.

In *femtocell-1* configuration, all femto BS units transmit in each slot; in *femtocell-2* configuration, all femto-BS units are equally split into two sets that transmit in alternate time slots to reduce inter-cell interference. As explained earlier, regardless the transmitting state of a femto-BS, its power consumption is assumed constant at 18 watts, which is a typical power consumption value in commercial femto-BS units.

FMDA-N_c-N_c-1: Denotes a FMDA network that is equipped with N_c antennas, where $N_c \in \{2,4,6\}$. The antennas are placed on the ceiling of the corridors and form a single femto-CoMP cluster.

FMDA-N_r-N_c-N_{ts}: Denotes a FMDA network that is equipped with N_r ceiling-mounted antennas, every N_c of which forms one CoMP cluster. All antennas are equally split into N_{ts} sets, each set being active in its assigned time slot while other antennas are left unused. Besides N_c=2 configuration, we also verified N_c=3 configuration as potential candidates for energy saving. As opposed to FMDA-N_c-N_c-1, the antennas are placed within the rooms. Water-filling was conducted within each femto-CoMP as if there is no inter-CoMP interference.

Following the air interface of LTE Release 8, we chose to use 10 MHz channel bandwidth with 9 MHz allocated for data communications. Every device is attached to only one channel in the 2.5 GHz band. The channel model of each link is composed of static path loss, spatially correlated static shadowing, and time-varying frequency-selective fading.

Shadowing and fading: While fading channels in the three studied access technologies are modeled as tapped delay lines with parameters from ITU-R M.1225, path loss and shadowing models vary with systems and different antenna placements. Link path loss follows M.1225 with the penetration loss model from COST231, i.e., 18.3 dB loss per floor and 6.9 dB loss per wall. In the femtocells, there are two links from two users in each room to their corresponding femto-BS. We model the shadowing in this case as spatially correlated with an exponential correlation function of the user separation. The shadowing is assumed to be independent for two users in different rooms. Given two links from two femto-BS units to one user, we always assume that these two links have independent shadowing due to the large femtocell separation distance.

Further in the FMDA network, when antennas are placed in corridors, the indoor hotzone model in M.2135-1 [127] is used to describe the path loss and shadowing. When an antenna is placed in each room, as in FMDA-N_r-N_c-N_{ts}, the path loss and shadowing models follow those in standalone femtocells. The noise figure at each user's receiver is assumed to be 7 dB. All other channel propagation parameters follow M.2135-1 unless otherwise noted.

User scheduling: A robinT scheduler is used in FMDA. For femtocells, robinT scheduling only suffers minor spectral efficiency loss when compared with computationally intensive PF exhaustive search. As such, only robinT is used to evaluate energy efficiency of femtocells.

Antenna scheduling in FMDA-N_r-N_c-N_{ts}: When N_{ts} = 1, neighboring antennas on one floor are clustered into 2- or 3-MIMO setting, as shown in the bottom part of Fig. 3.2, and other floors follow the same configuration. When $N_{ts} > 2$, femto-CoMPs assigned with the same time slot are placed on alternating sides of the building to reduce inter-CoMP interference. In all configurations in Fig. 3.2 except FMDA-24-3-3, only two neighboring floors are shown and other two floors follow the same pattern.

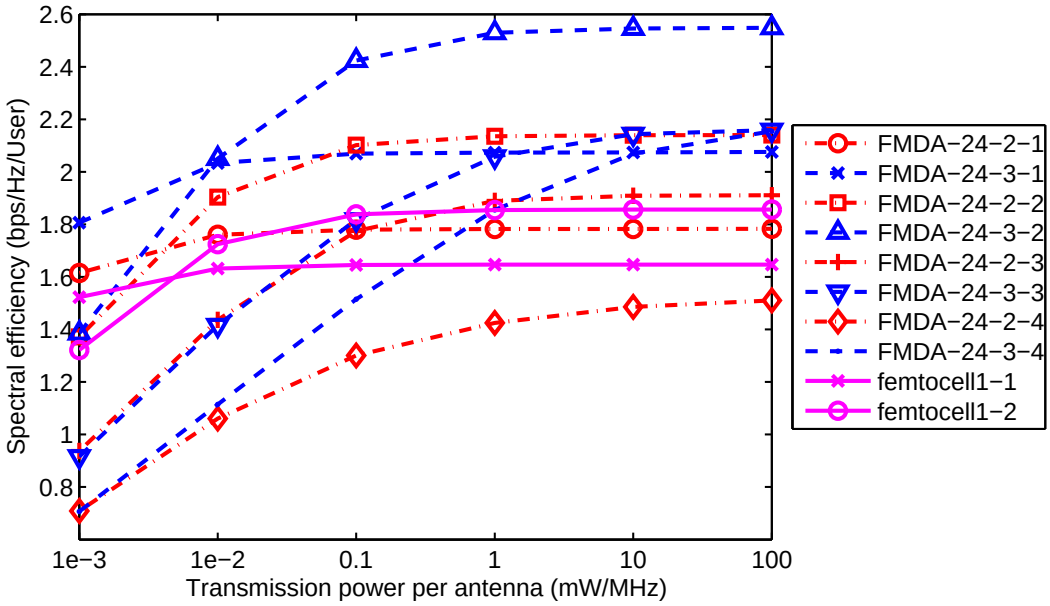
Power consumption coefficients: The α_B is set as 0.87 and α_E=0.10, as listed in Table 3.1. The α_{FO}, α_{MS} are chosen as 0.01 and 0.02, respectively; their sum matches the value 0.03 in Table 3.1.

3.5.2 Spectral Efficiency

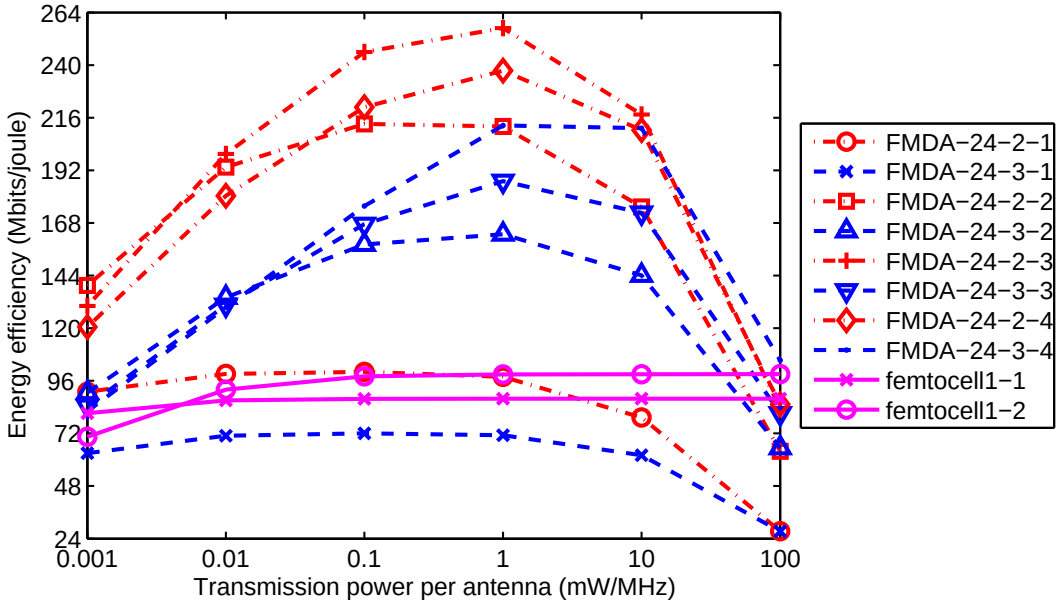
We now demonstrate the advantages of the proposed FMDA antenna scheduling in both energy efficiency and spectral efficiency in comparison with standalone femtocells. The throughput of a given user at a given time slot is the sum of Shannon capacities of all subcarriers used by this user. All packets are transmitted in a slotted manner, where each slot lasts 1 millisecond and each simulation lasts 5 seconds.

In both FMDA-24-2-N_{ts} and FMDA-24-3-N_{ts} configurations, as shown in Fig. 3.3a, the system gradually enters the interference-limited state when P_{tx} increases. Those configurations with a larger N_{ts} enter the state at higher P_{tx} values due to the increased inter-CoMP separation distance. The η_s in interference-limited state increases when N_{ts} increases from 1 to 2. Specifically, the

20% and 23% gain in η_s are achieved in FMDA-24-2-2 and FMDA-24-3-2, respectively. The η_s, however, drops when N_{ts} further increases to 3 or larger due to wasted transmission opportunities. In fact, FMDA-24-2-4 and FMDA-24-3-4 did not even enter the interference-limited state.



(a) Spectral efficiency vs. Transmission power per antenna.



(b) Energy efficiency vs. Transmission power per antenna.

Figure 3.3: Spectral and energy efficiency vs. Transmission power per antenna.

3.5.3 Energy Efficiency vs. Spectral Efficiency

Fig. 3.3b presents the energy efficiency η_e at different levels of P_{tx}, which range from 10^{-3} to 100 mW/MHz. Driving P_{tx} up to 100 mW/MHz is to show a complete picture of energy efficiency although the transmission power level is over the limit regulated in LTE [126]. As expected, the configuration group FMDA-24-2-N_{ts} provides the best energy efficiency due to the use of two-user MIMO and they reach the optimal efficiency when $P_{tx} \approx 1$ mW/MHz, approximately corresponding to the P_{tx} level when the systems become interference-limited. The phenomenon suggests when P_{tx} is less than 1 mW/MHz, energies consumed by wireless transmission are much less than those consumed by signal processing. Therefore the best strategy to be energy efficient is increasing P_{tx} until the FMDA system becomes interference-limited.

The tradeoff between energy and spectral efficiency is shown in Fig. 3.4. In all configurations we observe the bell-shaped curves because, as we observed in Fig. 3.3b, the energy efficiency first increases when the transmission power increases from a very low value to certain point where the system enters the interference-limited state. After the transmission power surpasses this point, much more energies are consumed by radio frequency power amplifiers but little throughput gain is achieved, thus causing a drop in energy efficiency. The drop is steep because of the extremely poor power amplifier efficiency, 2.2%, as modeled in Section 3.2.2.

The square area enclosed by gray dotted lines represents a preferable operation region where any operating point has advantages over standalone femtocells in both spectral and energy efficiency. Among single femto-CoMP configurations, only FMDA-4-4-1 can enter the region when P_{tx} reaches 100 mW/MHz.

In multiple femto-CoMPs, four configurations are able to simultaneously improve both energy and spectral efficiency, demonstrating the potential of antenna scheduling in achieving a desired balance between spectral and energy efficiency. Compared with femtocell-2, FMDA-24-2-3 improves energy efficiency by 160% and spectral efficiency by 2%; FMDA-24-3-2 improves energy efficiency by 64% and spectral efficiency by 36%. Compared with FMDA-4-4-1 at P_{tx}=10

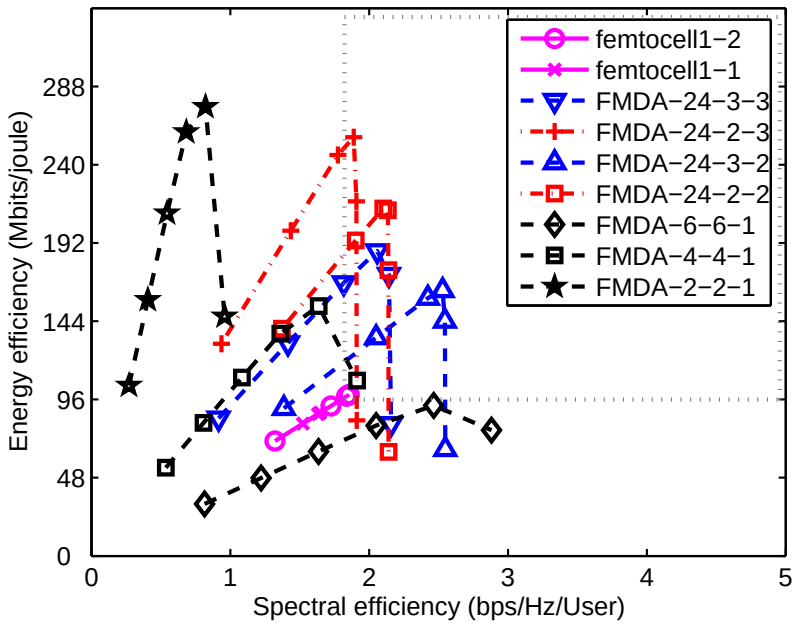


Figure 3.4: Energy efficiency vs. Spectral efficiency under different P_{tx} ($10^{-3} \sim 10^{2}$ mW/MHz). Mbits/joule: 10^6 bits/joule.

mW/MHz, FMDA-24-2-3 improves energy efficiency by 68% and spectral efficiency by 15%; FMDA-24-3-2 improves energy efficiency by 6% and spectral efficiency by 55%. However, one should also be aware of the higher infrastructure cost associated with FMDA-24-2-3 and FMDA-24-3-2 since a much higher number of ceiling-mounted antennas and more optical fibers need to be installed when compared with FMDA-4-4-1. Notice that since systems in multiple femto-CoMP setting are able to reach interference-limited state, their optimal operation points are reached at a P_{tx} level significantly lower than systems in single femto-CoMP setting.

3.6 Summary

We proposed a novel group of network configurations that employs antenna scheduling in an FMDA network to simultaneously improve both spectral and energy efficiency. Denoted as FMDA-N_r-N_c-N_{ts}, the proposed scheme selects N_r/N_{ts} antennas out of N_r ceiling-mounted fiber-connected antenna in each time slot, every N_c neighboring antennas forming a femto-CoMP cluster that serves N_c users at a time. Compared with standalone femtocells, the proposed scheme is shown in a typi-

cal office building to increase energy efficiency by 64%~160% and spectral efficiency by 2%~36%. The exact gain depends on network configurations and transmission power levels because a higher energy efficiency gain implies a lower spectral efficiency gain.

The proposed antenna scheduling strategy relies on a large number of antennas, which are only made possible in FMDA system. Downlink throughput is enhanced due to reduced interference; energy efficiency is improved as leaving antennas unused reduces energy consumed by collecting channel state information and zero-forcing beamforming. Antenna scheduling enriches network configurations on how many antennas should cooperate and how much separation distance should be kept among cooperating clusters, thus offering service providers a greater flexibility on choosing different balance points between spectral and energy efficiency.

Chapter 4

Energy-efficient Low-complexity Zero-forcing Beamforming via Banded Matrix Inversion in Indoor Fiber-connected Massively Distributed Antenna Systems[1]

4.1 Introduction

Massive array multiple-input multiple-output (MIMO) addresses ever-increasing traffic demand by densifying radio access networks. The spectral efficiency advantages of massive array-MIMO systems have been shown in analysis under match filter precoding [45], zero-forcing beamforming [46], and demonstrated in experiments [47]. Two recent contributions analyzed the energy efficiency of massive array-MIMO in the uplink [69] and downlink [70]. From the path loss aspect, however, it would be more energy-efficient to locate the antennas closer to the mobile terminals, which motivates the use of a large number of femtocells. While adopting universal frequency reuse in such networks to enjoy the advantages on area spectral efficiency [85], one needs to mitigate

[1]This chapter is based on [97] co-authored with Dr. V. Leung.

the resulting inter-cell interference. Multi-cell processing can mitigate the inter-cell interference through multi-user MIMO, which however requires channel state information (CSI) from participating femto base stations (BSs) and therefore decreases the backhaul capacity. Pooling baseband signal processing at a centralized processing system (CPS) can reduce the cost of cooperation, in which case each BS is reduced to an antenna that contains a power amplifier and a low-noise amplifier, and when a large number of such antennas are deployed, the system constitutes a Fiber-connected Massively Distributed Antennas (FMDA) system. The optical or coaxial wireline facility interconnecting the CPS and an antenna can be analog or digital. An analog facility carries radio frequency signals (e.g., by radio over fiber (RoF) technology in an optical facility), thus allowing antennas that are simpler and more transparent to standards (e.g., [87]). A digital facility carries high-rate in-phase/quadrature samples, thus eliminating signal distortions caused by nonlinearity while transmitting analog signals [76]. In such case, however, an antenna needs a very-high-rate digital transceiver and high-speed analog/digital and digital/analog converters.

In this chapter we study how to improve the energy efficiency and reduce the computation cost of FMDA systems. Locating antennas closer to mobile terminals cause the power consumption of digital signal processing rather than radio frequency transmissions to be emphasized in the overall energy consumption, which has not been widely considered in previous literature. Note that baseband signal processing is a significant contributor of the overall energy consumption, especially in femtocells where the transmission power is low and no air conditioning is needed [63]. Our research question is how to make beamforming feasible for practical implementation and energy efficient in the presence of a large number of antennas. We consider an indoor office environment where each room is equipped with one antenna. The antennas cooperate by zero-forcing beamforming (ZFBF) due to its simplicity and near-optimal performance at a high signal-to-noise-ratio (SNR) [128].

We propose a low-complexity ZFBF algorithm via banded matrix inversion that aims to reduce signal processing complexity in indoor FMDA systems. The idea is motivated by the phenomenon that in massively distributed antenna systems, only neighboring antennas need to cooperate. Large

wall-penetration losses in indoor environments strengthen such a phenomenon and allow the use of banded inversion, which offers a substantial reduction in the computation cost of matrix inversion, CSI collection, and MIMO precoding.

4.1.1 Related Work

The Wyner model [89] is a widely accepted inter-cell interference model that involves a tri-diagonal matrix, a special case of banded matrix. Specifically, BSs are linearly placed and the interference between the k-th BS and b-th BS (denoted by $v_{k,b}$) is 1 for $k = b$, α for $|k-b| = 1$, and 0 otherwise. While convenient to analyze in multi-cell processing [90, 91, 96], the Wyner model is only accurate with a sufficiently large number of simultaneous users and perfect BS power control [92]. Banded matrices had been applied to electromagnetic wave simulation [93], inversion of tri-diagonal matrices [94], and equalization [95]. Specifically, the equalization scenario resembles distributed MIMO in that a farther distance in the frequency domain implies less inter-carrier interference. With regard to ZFBF in distributed MIMO systems, the only work to exploit banded path loss structure is [96], where under a given backhaul constraint, unequal CSI in distributed antenna systems and optimal CSI allocation are discussed.

Approximated matrix inversion have been used to find a suitable starting point for iterative inversion of sparse matrices, such as Jacobi/Gauss-Siedel and conjugate-gradient methods. Given a finite banded matrix $\mathbf{V}$, Demko showed in [129] that the inverse $\mathbf{V}^{-1}$ has *exponentially decayed off-diagonals* (EDODs), i.e., $v_{k,b}^{-1} \leq K\alpha^{|k-b|}$, where K is a constant, $0 < \alpha < 1$, and both K and α depend on the width of $\mathbf{V}$. It was later shown [130, 131] that given $\mathbf{V}$ with EDODs, $\mathbf{V}^{-1}$ also has EDODs yet at a higher rate, i.e., $v_{k,b}^{-1} \leq \beta^{|k-b|}(0 < \beta < \alpha)$. When $\mathbf{V}$ has *polynomially decayed off-diagonals*, i.e., $v_{k,b} \leq K(1+|k-b|)^{-\eta}(\eta > 1)$, Jaffard's theorem [132] states that $\mathbf{V}^{-1}$ also has polynomially decayed off-diagonals at the same decay rate, i.e., $v_{k,b}^{-1} \leq K(1+|k-b|)^{-\eta}$. A good summary of Demko and Jaffard's work along with other findings can be found in [133]. The decay rate bound in [130] is for general banded matrices and increases with the width p. Such bound is too loose when applied to estimate the signal-to-interference-noise-ratio (SINR). Besides, we are

more interested in random channel matrices with a path loss matrix having *compositely decayed off-diagonals* (CDODs) that comprises both exponential and polynomial decays.

4.1.2 Contributions

In a massively distributed antenna system, it is challenging to determine the entire channel matrix $\mathbf{H}$ and power-consuming to both generate the precoding matrix $\mathbf{W}$ and precode the symbols, especially when these operations are required in each subcarrier if orthogonal frequency-division multiplexing (OFDM) is used. Our proposed low-complexity ZFBF scheme meets these challenges by discarding less important off-diagonal elements in $\mathbf{H}$ to form a banded, sparse matrix, $\mathbf{H}_p$, where p controls the sparsity of $\mathbf{H}$. The precoding $\mathbf{W}$ is then generated based on $\mathbf{H}_p$. The scheme is evaluated by applying it to a distributed antenna system that covers a dual-stripe office building floor. Compared with traditional ZFBF, the proposed scheme significantly reduces the amount of signal processing and hence improves the energy efficiency due to the reduced complexities in CSI collection, matrix inversion and precoding.

The proposed scheme contains two versions: dense $\mathbf{W}$ and sparse $\mathbf{W}$. The dense $\mathbf{W}$ version uses the full inversion of $\mathbf{H}_p$ as the precoding matrix, incurring quadratic computation cost, relative to the number of antennas. The sparse $\mathbf{W}$ version further bands $\mathbf{W}$ to achieve linear computation cost of precoding. We analyze the performance impact of banding random channel matrices, in which the variances of the entries have EDODs or CDODs. Both the SINR analysis and numerical evaluations indicate that compared with a dense $\mathbf{W}$, a sparse $\mathbf{W}$ incurs a negligible SINR loss under CDOD but significant loss under EDOD. To the best of our knowledge, our work is the first to consider the inversion of random matrices with EDODs or CDODs.

Notations: A bold capital $\mathbf{A}$ indicates a matrix $\mathbf{A}$; a bold lowercase $\mathbf{a}$ indicates a column vector $\mathbf{a}$. The l_2-norm of $\mathbf{a}$ is $\|\mathbf{a}\|$. The entry on the i-th row and j-th column of $\mathbf{H}$ is denoted by $\mathbf{H}(i,j)$ or $h_{i,j}$; $\mathbf{H}_{a:b,c:d}$ denotes the submatrix of $\mathbf{H}$ formed by rows $[a, a+1, \cdots, b]$ and columns $[c, c+1, \cdots, d]$. $\mathcal{B}_p$ denotes the set of banded matrices with *width* p, i.e., entry (k,b) is zero for $|k-b| \geq p$. We consider banded matrices with equal upper- and lower-width $p-1$, which is the number of non-

zero sub-diagonals above and below the main diagonal, respectively. Given a matrix $\mathbf{H}$, the banded $\mathbf{H}$ with width p is obtained by keeping the main diagonal and $p-1$ off-diagonals on both sides, denoted by $\mathbf{H}_p$. $\mathbf{H}^T$, $\mathbf{H}^H$, $\mathbf{H}^{-1}$, $|\mathbf{H}|$ denote the transpose, Hermitian transpose, inversion, and determinant of matrix $\mathbf{H}$, respectively. A diagonal matrix $\mathbf{D}$ is denoted by $\text{diag}[d_{1,1}, d_{2,2}, ..., d_{n,n}]$. The operator $\circ$ denotes the Hadamard product. $\mathbf{I}$ denotes identity matrix. The largest integer not larger than x is denoted by $\lfloor x \rfloor$. The variance of a random variable x is denoted by $\text{var}(x)$. zero-mean circularly symmetric complex Gaussian (ZMCSCG) distribution with variance σ^2 is denoted by $\mathcal{CN}(0, \sigma^2)$. The uniform distribution between a and b is denoted by $\mathcal{U}(a, b)$. The functions $\ln(\cdot)$ and $\lg(\cdot)$ represent natural and base-10 logarithm, respectively. RHS means "right hand side". $\mathbb{E}_{\mathbf{X}}[\cdot]$ denotes the expectation over the random variable $\mathbf{X}$.

The rest of this chapter is organized as follows. We introduce the system model in Section 4.2 and propose the banded inversion algorithms in Section 4.3. The SINR loss from using a dense and sparse precoding matrix is analyzed in Sections 4.4 and 4.5, respectively. The downlink energy efficiency is modeled and analyzed in Section 4.6. We evaluate the performance in Section 4.7 and summary this chapter in Section 4.8.

4.2 System Model

Consider a massively distributed antennas system deployed on one floor of an office building with N rooms arranged in a single-stripe topology as depicted in Fig. 4.1. The system is composed of three components, namely the antennas, the optical cables and the CPS. Each room has one user and one ceiling-mounted fiber-connected antenna, which transmits and receives radio frequency signals over the air while all the baseband signal processing is concentrated at the CPS. Optical fibers have a low propagation loss and high bandwidth, and are well-suited to carry radio frequency signals between antennas and the CPS. The fiber runs can reach hundreds of meters, thus it is feasible to implement a massive array of distributed antennas to cover most buildings in practice.

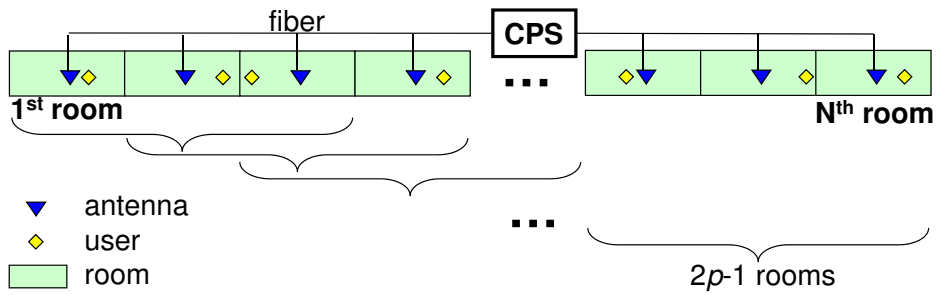


Figure 4.1: A Fiber-connected Massively Distributed Antennas system. CPS: centralized processing system.

4.2.1 Channel Matrix H

The $N \times N$ channel matrix $\mathbf{H}$ is given by $\mathbf{V} \circ \mathbf{F}$, where $\mathbf{V}$ is the path loss matrix and $\mathbf{F}$ is a frequency-flat block-fading matrix, modeled as independent and identically distributed (i.i.d.) ZMCSCG. In this chapter, we ignored static and dynamic shadowing for analytic convenience.

Let the k-th room enclose the k-th user and the k-th antenna. The path loss between the k-th user and the b-th antenna ($1 \leq k, b \leq N$) is modeled as the free space propagation loss with path loss exponent η, plus the wall penetration loss which is given by $L(k-b)$ dB, where L is 6.9 dB for concrete walls and 3.4 dB for plasterboard walls according to the COST231 indoor path loss model in ITU M.1225 [127]. The (k,b) entry in $\mathbf{V}$ is then given by

$$v_{k,b} = (|k-b| - \delta_k)^{-\eta/2} \alpha^{|k-b|} \approx (|k-b|)^{-\eta/2} \alpha^{|k-b|}, \tag{4.1}$$

where δ_k is the distance between the k-th user and the k-th antenna and $\alpha = 10^{-L/20}$ (α=0.45 for L=6.9 dB). The term $\alpha^{|k-b|}$ indicates exponentially decayed off-diagonals in $\mathbf{V}$, attributed to the wall-penetration loss. The term $(|k-b|)^{-\eta/2}$ indicates polynomially decayed off-diagonals, which are caused by free space propagation losses. The inter-antenna distance is normalized to 1.

4.2.2 Precoding

Assume perfect CSI at the transmitter [134] and perfect synchronization [135, 136]. We consider $N \times 1 \times N$ downlink transmissions: N pre-scheduled single-antenna users are simultaneously served

by N antennas through $N \times N$ multi-user MIMO downlink transmissions. After precoding, the received signal vector is $\mathbf{y} = \mathbf{HWs} + \mathbf{n}$, where $\mathbf{y}$ is the $N \times 1$ received symbols, $\mathbf{H} = [\mathbf{h}_1^T; \mathbf{h}_2^T; \cdots; \mathbf{h}_N^T]$, $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \cdots \mathbf{w}_N]$ is the $N \times N$ precoding matrix, $\mathbf{s} = [s_1 s_2 \cdots s_N]^T$ is the $N \times 1$ transmitted symbol vector, and $\mathbf{n}$ is the $N \times 1$ i.i.d. additive Gaussian noise vector in which each component has zero mean and variance σ^2. Considering ZFBF, we have $\mathbf{W} = \mathbf{H}^{-1}$. Given a stream power allocation $\{P_k\}(1 \le k \le N)$, we assume $\mathbf{s} \sim \mathcal{CN}(0, \mathrm{diag}[P_1, \cdots, P_N])$.

4.2.3 Stream Power Allocation

Stream power allocation employs waterfilling (WF) algorithm under sum power constraint (SPC), denoted by P_{spc}, or per-antenna power constraint (PAPC), denoted by P. WF-SPC establishes an analytic solution of the convex optimization problem [128]. WF-PAPC uses the interior point method to iteratively solve the problem [137]; however, the average number of iterations is hard to predict. To control nonlinearity in each RoF link while maintaining a low complexity, we use a relaxed version of WF-SPC, termed "WF-SPCr", which first achieves a lower bound of WF-PAPC by running WF with $P_{spc} = P$, as proposed in [138]. Once an initial allocation $\{P_k\}(1 \le k \le N)$ is obtained, the second step is to tighten the lower bound by scaling up all streams' power with a single scaling factor until P is first reached at one antenna. The scaling factor is found by first locating the antenna being assigned with the maximal power and then dividing its output power by P. Our benchmark is conventional WF-SPC subject to $P_{spc} = NP$ with the precoding matrix being $\mathbf{W}_p$, termed "WF-SPC-dW".

4.3 Banded Inversion of Matrices with Compositely Decayed Off-Diagonals

We now develop banded matrix inversion algorithms for matrices with CDODs. Consider an $N \times N$ matrix $\mathbf{H}$. Let $\mathbf{H} = \mathbf{H}_p + \mathbf{D}_p$ where $\mathbf{D}_p$ is the difference between the banded $\mathbf{H}$ and the true $\mathbf{H}$. Denote by $\mathbf{W}$ the full inversion of $\mathbf{H}$; $\mathbf{W}_p$ the full inversion of $\mathbf{H}_p$, also referred to as dense $\mathbf{W}$; $\mathbf{S}_p$ the banded inversion of $\mathbf{H}_p$, also referred to as sparse $\mathbf{W}$.

4.3.1 Banded Inversion

Consider a nonsingular $\mathbf{H} \in \mathcal{B}_p$ and assume $\mathbf{H}$ can be factorized into $\mathbf{LU}$ where $\mathbf{L}$ is lower-triangular, $\mathbf{U}$ is upper-triangular, and $\mathbf{L}, \mathbf{U} \in \mathcal{B}_p$ [27]. Therefore, as depicted in Fig. 4.2, we can obtain $\mathbf{W}_p$ by factorizing $\mathbf{H}_p$ into $\mathbf{L}_p\mathbf{U}_p$ via Gaussian-elimination without pivoting, followed by conventional forward and backward substitution as $\mathbf{L}_p$ and $\mathbf{U}_p$, although banded, still have dense inverses. Disabling pivoting does not cause numerical instability if $\mathbf{H}_p$ is strictly diagonally dominated, i.e., the magnitude of the diagonal entry in a row is larger than or equal to the magnitude sum of other entries in the same row. Owing to the fast decaying off-diagonals in $\mathbf{V}$, we expect that most instances of $\mathbf{H}_p$ possess this property.

To avoid the expensive multiplication $\mathbf{U}_p^{-1}\mathbf{L}_p^{-1}$, we truncate $\mathbf{L}_p^{-1}$ and $\mathbf{U}_p^{-1}$ during forward and backward substitution to obtain $\mathbf{S}_p$. For each column of $\mathbf{L}_p^{-1}$ and $\mathbf{U}_p^{-1}$, we keep only $2p$-1 entries that are closest to the main diagonal. The resulting algorithm consists of a banded LU factorization to factorize $\mathbf{H}_p$ into $\mathbf{L}_p$ and $\mathbf{U}_p$ [139], and a banded forward/backward substitution to solve $\mathbf{L}_p\mathbf{x} = \mathbf{b}_i$ and $\mathbf{U}_p\mathbf{y} = \mathbf{x}$, where $\mathbf{b}_i$ is the i-th column of $\mathbf{I}$. The banded forward/backward substitution algorithms are listed in Table 4.1.

Table 4.1: Algorithm: Banded forward/backward substitution

	Dense-**W**	Sparse-**W**

Banded Forward Substitution	for r=i+1:N
	  idx=max(i,r-p+1):r-1
	  b(r)=b(r)-L(r,idx)*b(idx)
	end
		for r=i+1: min(i+p-1,N)
		  idx=max(i,r-p+1):r-1
		  b(r)=b(r)-L(r,idx)*b(idx)
		end

Banded Backward Substitution	b(N)=b(N)/U(N,N)
	for r=N-1:-1:1
	  idx=r+1:min(r+p-1,N)
	  b(r)=(b(r)-U(r,idx)*b(idx))/U(r,r)
	end
		es=min(i+p-1,N)
		ee=max(1,i-p+1)
		b(es)=b(es)/U(es,es)
		for r=es-1:-1:ee
		  idx=r+1:min(r+p-1,es)
		  b(r)=(b(r)-U(r,idx)*b(idx))/U(r,r)
		end

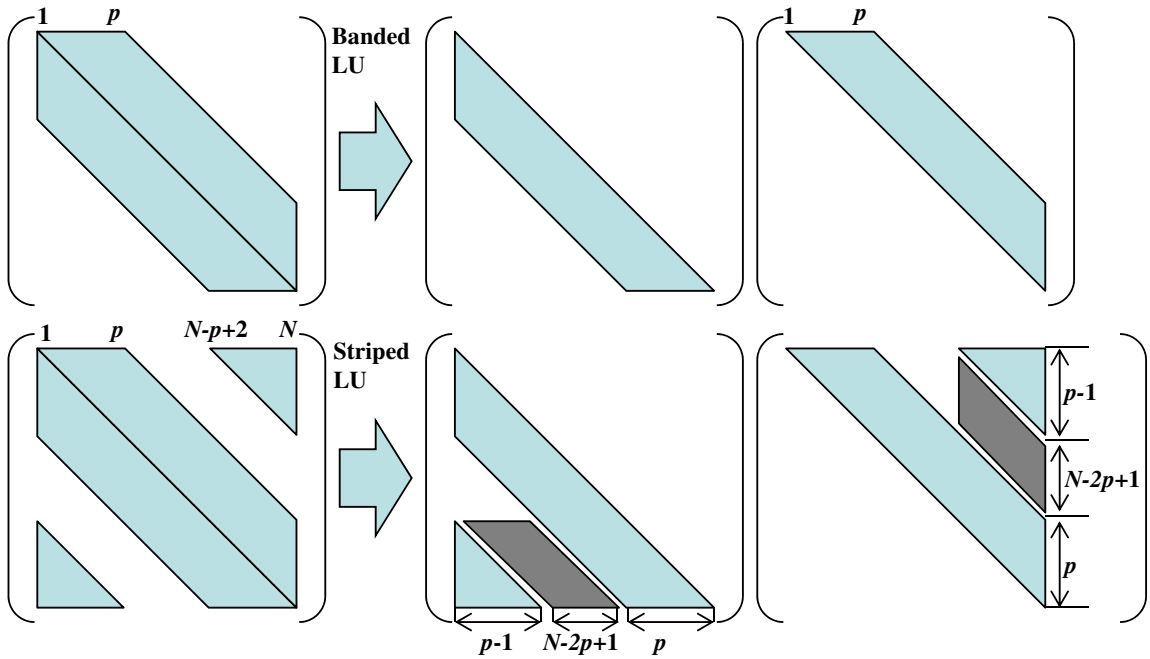


Figure 4.2: LU factorization of banded and striped matrices. Blank areas are zeros.

4.3.2 Striped Inversion

Wrapping around the topology in Fig. 4.1 leads to a circular topology that corresponds to a striped channel matrix, named after two stripes formed by the zero entries in the upper-right and the lower-left area, as shown in the bottom part of Fig. 4.2. Unlike the banded inversion, the striped inversion needs to generate the lower-left and upper-right corner areas in the inverse. However, the increased indexing complexity is negligible when $p \ll N$. Note that the valid range of p is $1 \leq p \leq N$ in the banded inversion and $1 \leq p \leq \lfloor (N+1)/2 \rfloor$ in the striped inversion.

The striped LU factorization, as listed in Table 4.2, still uses vector-matrix operations for the first and last p-1 rows, which densify the last (p-1) rows in $\mathbf{L}$ and the last (p-1) columns in $\mathbf{U}$, shown as the dark areas in Fig. 4.2. Those entries, however, are not accessed by the striped forward/backward substitution and thus, equivalently, zeros.

Table 4.2: Algorithm: Striped LU factorization of **H**

Striped LU	for k=1:N-1
	  if k<p; rl=[k+1:k+p-1 N-p+k+1:N]
	  else rl=k+1:min(k+p-1,N)
	  end
	  H(rl,k) = H(rl,k)/H(k,k)
	  H(rl,rl) = H(rl,rl) - H(rl,k)*H(k,rl)
	end

4.4 SINR Loss Using Dense W

To analyze the residual interference when the dense precoding matrix $\mathbf{W}_p$ is used, we start with path loss matrices with CDODs and then move on to random matrices.

Given an instantaneous full channel matrix $\mathbf{H}$ and the dense precoding matrix $\mathbf{W}_p$, the received signal $\mathbf{y}$ is given by:

$$\mathbf{y} = \mathbf{H}\mathbf{W}_p\mathbf{s} + \mathbf{n} = (\mathbf{H}_p + \mathbf{D}_p)\mathbf{W}_p\mathbf{s} + \mathbf{n} = (\mathbf{I} + \mathbf{D}_p\mathbf{W}_p)\mathbf{s} + \mathbf{n} \approx \mathbf{I}\mathbf{s} + \mathbf{D}_p\mathbf{W}_p\mathbf{s}, \tag{4.2}$$

where the last approximation is possible under the high SNR assumption. Let $\mathbf{A} = \mathbf{D}_p\mathbf{W}_p$, which represents the residual interference after the banded inversion. Although the main diagonal of $\mathbf{A}$ still contributes to the signal power, they are relatively small when being added by the identity matrix $\mathbf{I}$. Thus we redefine $\mathbf{A}$ as $\mathbf{D}_p\mathbf{W}_p$ with the main diagonal cleared. By $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathrm{diag}[P_1, \cdots, P_N]$, the residual interference power of the i-th stream, denoted by I_i, is given by:

$$I_i = \sum_{j=1, j\neq i}^{N} [\|a_{i,j}\|^2 P_j] \approx P \sum_{j=1, j\neq i}^{N} [\|a_{i,j}\|^2], \tag{4.3}$$

where P is the sum power constraint divided by N and P_j is the power of the j-th stream. The approximation is to decouple the analysis of residual interference from stream power allocation. Notice that the RHS in (4.3), denoted by $\bar{I}_i$, is reached under equal stream power allocation.

In the following, we analyze $\bar{I}_i$ for different types of $\mathbf{H}$, as listed in Table 4.3.

Table 4.3: Types of channel matrix $\mathbf{H}$ $(0 < \alpha < 1, 2 \le \eta \le 4)$

$\mathcal{ED}_\alpha$	The Toeplitz matrix $\mathbf{H}$ with exponentially decayed off-diagonal entries at rate α, i.e., $h_{ij} = \alpha^{\lvert i-j \rvert}$.
$\mathcal{EPR}_{\alpha,\eta}$	The set of random matrices whose off-diagonal entries decay both exponentially at rate α and polynomially at rate η, i.e., $h_{ij} = \lvert i-j+\delta_i \rvert^{-\eta/2}\alpha^{\lvert i-j \rvert}$, where δ_i is the normalized distance between the i-th user and the i-th antenna, subject to $\mathcal{U}(-0.5, 0.5)$.
$\mathcal{EP}_{\alpha,\eta,\delta}$	$\mathcal{EPR}_{\alpha,\eta}$ with fixed δ, representing a Toeplitz matrix $\mathbf{H}$ where $h_{ij} = \lvert i-j \rvert^{-\eta/2}\alpha^{\lvert i-j \rvert}$ for $i \neq j$ and $\lvert \delta \rvert^{-\eta/2}\alpha^0$ for $i = j$ $(\lvert \delta \rvert < 0.5)$.
$\mathcal{EZ}_\alpha$	The Hadamard product of a ZMCSCG matrix and an $\mathcal{ED}_\alpha$ matrix.
$\mathcal{EPRZ}_{\alpha,\eta}$	The Hadamard product of a ZMCSCG matrix and an $\mathcal{EPR}_{\alpha,\eta}$ matrix.

4.4.1 Path Loss Matrix $\mathcal{ED}_\alpha$

Given $\mathbf{H} \in \mathcal{ED}_\alpha$, we now develop how $\lg(\bar{I}_i)$ varies with p. Let $\mathbf{H} = \mathbf{H}_p + \mathbf{D}_p$ and $\mathbf{A} = \mathbf{D}_p\mathbf{W}_p$. Consider $p+1 \le i \le N-p$. The transpose of the i-th row of $\mathbf{D}_p$, denoted by $\mathbf{d}_i$, is composed of three sections: $[\alpha^{i-1};\cdots;\alpha^p]$, $[0;\cdots;0]$ and $[\alpha^p;\cdots;\alpha^{N-i}]$. Each entry in the j-th column of $\mathbf{W}_p$ is bounded by the corresponding entry in $\mathbf{w}_j = [\beta^{j-1};\cdots;\beta^{N-j}]$ $(1 \le j \le N)$, according to $\mathbf{W}_p \in \mathcal{ED}_\beta(\alpha < \beta < 1)$ which we learn from [130]. All of the multiplicative constants are dropped as our interest is to examine how $\bar{I}_i$ varies with p. By $a_{i,j} = \mathbf{d}_i^T\mathbf{w}_j$, the $a_{i,j}$ reaches its maximum when α^p in $\mathbf{d}_i$ meets 1 in $\mathbf{w}_j$. When the entry "1" in $\mathbf{w}_j$ falls in one of three sections in $\mathbf{d}_i$, there is one ruling item for each j:

$$a_{i,j} \approx \begin{cases} \frac{\alpha}{\beta-\alpha}\beta^{i-j}(\frac{\alpha}{\beta})^p, & j \le i-p \\ \frac{\alpha^p(\beta^{j-i+p}+\beta^{i-j+p})}{1-\alpha\beta}, & i-p < j < i+p-1 \\ \frac{\alpha}{\beta-\alpha}\beta^{j-i}(\frac{\alpha}{\beta})^p, & j \ge i+p-1 \end{cases} \tag{4.4}$$

After some algebraic manipulations, the $\bar{I}_i$ is approximated by

$$\bar{I}_i \approx \frac{2P\alpha^{2p}}{1-\beta^2}\left(\frac{\alpha}{\beta-\alpha}\right)^2 \propto \alpha^{2p}, \tag{4.5}$$

where $i \in [p+1, N-p]$ and $p \geq 2$. For $i \in [1, p]$ and $i \in [N-p+1, N]$, there is only one side of interference; hence, $\bar{I}_i \propto \alpha^{2p}$. By $\alpha = 10^{-L/20}$, we have $\lg(\bar{I}_i)$ decreases linearly with p $(1 < p \ll N)$.

4.4.2 Path Loss Matrix $\mathcal{EP}_{\alpha,\eta,\delta}$

Given $\mathbf{H} \in \mathcal{EP}_{\alpha,\eta,\delta}$, we can show that the interference term $\lg(\bar{I}_i)$ approximately decreases by $L + 10\eta \lg(1 + 1/p)$ when the width increases from p to p+1 $(1 < p \ll N)$. The derivation process is similar to the case of $\mathcal{ED}_\alpha$, except we encounter power series rather than geometric series. Details can be found in Appendix A.

4.4.3 Random Matrices $\mathcal{EZ}_\alpha$ and $\mathcal{EPRZ}_{\alpha,\eta}$

Consider $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$. Let $\mathbf{H} = \mathbf{V} \circ \mathbf{F}$, where $\mathbf{V} \in \mathcal{EPR}_{\alpha,\eta}$ and entries of $\mathbf{F}$ are i.i.d. ZMCSCG. Assume $\mathbf{F}$ as frequency flat and block-fading. Our interest is the long-term time-average of instantaneous capacity, termed "throughput capacity of slow-fading" in [140] and defined as the time-average of mutual information under block-fading channels, provided that an infinite number of blocks are transmitted. Given high SNRs and equal stream power allocation, the throughput capacity of the i-th stream, denoted by $\hat{C}_i$, can be approximated by $c_0 - \mathbb{E}_{\mathbf{F},\Delta}[\log_2(P/\bar{I}_i)]$, which is proportional to $\hat{I}_i$, defined as

$$\hat{I}_i \triangleq \mathbb{E}_{\mathbf{F},\Delta}\left[\lg\left(\sum_{j=1, j\neq i}^{N} \|a_{i,j}\|^2\right)\right]. \tag{4.6}$$

Following the procedure in Section 4.4.1, we know that $a_{i,j}$ is a sum of the series $\{\alpha^{|i-k|} f_{i,k} w_{k,j}\}$, where $|i-k| \geq p$, $f_{i,k} \sim \mathcal{CN}(0,1)$, and $w_{k,j}$ is the (k,j) entry in $\mathbf{W}_p$. We now investigate the statistical properties of $w_{k,j}$, in particular, $\mathbb{E}_{\mathbf{F},\Delta}[\lg(\|w_{i,j}\|^2)]$ since from the perspective of channel capacity, we are more interested in the average of log-scale interference.

Consider $\mathbf{H} \in \mathcal{EZ}_\alpha$ in which $h_{i,j}$ is given by $\alpha^{|i-j|} f_{i,j}$; $f_{i,j} \sim \mathcal{CN}(0,1)$. It can be shown that $\mathbb{E}_{\mathbf{F}}[\lg(\|w_{k,j}\|^2)]$ linearly decays with k $(3 \leq k \leq p+1)$ and when α approaches zero, the decay rate is approximated by

$$r_{EZ} = 2\lg\alpha + \lg e. \tag{4.7}$$

The derivation details are given in Appendix B, where we also show that for $k > p+1$, $\mathbb{E}_{\mathbf{F}}[\lg(\|w_{k,j}\|^2)]$ linearly decays with k at a rate smaller than (4.7). By symmetry of $\mathbf{H}_p$, the decay rate in (4.7) also applies to any row of $\mathbf{W}_p$.

Now consider $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$. We use $\mathbb{E}[\cdot]$ instead of $\mathbb{E}_{\mathbf{F},\Delta}[\cdot]$ when this causes no ambiguity. It can be shown that $\mathbb{E}[\lg(\|w_{k,1}\|^2)](k \le p)$ linearly decays with k and the decay rate is approximated by

$$r_{EPRZ} = (2k-2)\lg\alpha - \eta\lg(k-1). \tag{4.8}$$

See Appendix C for the derivation details. We list below the important findings.

The simplest estimate of $\mathbb{E}[\lg(\|w_{k,1}\|^2)]$, denoted by $LW_k^{(1)}$, is given by

$$LW_k^{(1)} = (2k-2)\lg\alpha - \eta\lg(k-1) - 2\eta\lg(2e) + \gamma/\ln 10. \tag{4.9}$$

A better estimate, denoted by $LW_k^{(3)}$, introduces an adjustment ρ_η (listed in Table C.1):

$$LW_k^{(3)} = LW_k^{(1)} + \frac{\rho_\eta}{\ln 10}\left(\frac{k-1}{k-2}\right)^{\eta/2}. \tag{4.10}$$

Comparing (4.8) (where $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$) with (4.7) (where $\mathbf{H} \in \mathcal{EZ}_\alpha$), we notice that the $\lg e$ penalty in the decay rate disappears due to the presence of the polynomial decay, which indicates that for $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$, $\mathbb{E}[\lg(\|w_{k,j}\|^2)]$ has the same decay rate as $\mathbb{E}[\lg(\|h_{i,k}\|^2)]$.

We are also interested in $\mathbb{E}[\lg(\|w_{p+1,1}\|^2)]$ as it is the largest term being dropped when $\mathbf{S}_p$ is used. Unlike $\mathcal{EZ}_\alpha$ case, $\mathbb{E}[\lg(\|w_{p+1,1}\|^2)]$ in $\mathcal{EPRZ}_{\alpha,\eta}$ drops significantly. To quantify the behavior, we define the *outband drop* as the log-ratio of the accelerated and the normal decay, i.e., $O_p \triangleq \lg(|\mathbf{W}_p(p+1,1)/\mathbf{W}_{p+1}(p+1,1)|)$. It is shown in Appendix C that O_p is negligible in $\mathcal{EZ}_\alpha$, and is given by $\gamma/\ln 10 + \psi_\eta$ in $\mathcal{EPRZ}_{\alpha,\eta}$, where ψ_η is listed in Table C.1. As will be shown later, it is the outband drop that justifies the use of $\mathbf{S}_p$ for $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$.

We now analyze $\hat{I}_i$ based on the statistics of $\mathbf{W}_p$. Applying Jensen's inequality on (4.6) produces an upper-bound of $\hat{I}_i$: $2\lg(\sum_{j=1,j\neq i}^{N}\mathbb{E}_{\mathbf{F}}[\|a_{i,j}\|])$. However, the exponential decay in $\mathbf{D}_p$ and

$\mathbf{W}_p$, as revealed in (4.7) and (4.8), suggests highly dynamic values among $\|a_{i,j}\|$ and therefore, a tighter estimation of $\hat{I}_i$ shall be possible by considering only the dominating terms in $\|a_{i,j}\|$.

Consider $p \le i \le N-p$. By (4.7) and (4.8), we approximate $a_{i,j} = \mathbf{d}_i^T \mathbf{w}_j$ by

$$a_{i,j} \approx \begin{cases} \sum_{k=0}^{j-1}\left[\frac{\alpha^{i-j+k} f_{i,j-k} w_{j-k,j}}{(i-j+k)^{\eta/2}}\right] + \sum_{k=1}^{i-j-p}\left[\frac{\alpha^{i-j-k} f_{i,j+k} w_{j+k,j}}{(i-j-k)^{\eta/2}}\right], & 1 \le j \le i-p \\ \sum_{k=p}^{i-1}\left[\frac{\alpha^{k} f_{i,i-k} w_{i-k,j}}{k^{\eta/2}}\right] + \sum_{k=p}^{N-i}\left[\frac{\alpha^{k} f_{i,i+k} w_{i+k,j}}{k^{\eta/2}}\right], & i-p+1 \le j \le i+p-2, j \ne i \\ \sum_{k=0}^{N-j}\left[\frac{\alpha^{j-i+k} f_{i,j+k} w_{j+k,j}}{(j-i+k)^{\eta/2}}\right] + \sum_{k=1}^{j-i-p}\left[\frac{\alpha^{j-i-k} f_{i,j-k} w_{j-k,j}}{(j-i-k)^{\eta/2}}\right]. & i+p-1 \le j \le N \end{cases} \tag{4.11}$$

Inspecting (4.11) reveals that the variance of $a_{i,j}$ is tightly upper-bounded by that of $\alpha(1+1/p)^{-\eta/2} a_{i,j+1}$ for $1 \le j \le i-p$. From (4.8) we know the dominating term corresponds to $a_{i,i-p}$. Similarly, we obtain dominating terms in other intervals of j: $a_{i,i-p+1}$ for $i-p+1 \le j < i$; $a_{i,i+p}$ for $i+1 \le j \le i+p-2$; $a_{i,i+p-1}$ for $i+p-1 \le j \le N$. Since the variance of each of the four terms decreases by the ratio $\alpha(1+1/p)^{-\eta/2}$ when p increases by 1, we estimate $\hat{I}_i$ by recursion:

$$\hat{I}_i|_{p=p^*} \approx \lg[\alpha^2(1+1/p)^{-\eta}] + \hat{I}_i|_{p=p^*-1} = \cdots = -pL/10 - \eta \lg p + \hat{I}_i|_{p=1}, \tag{4.12}$$

which also applies to $1 \le i \le p-1$ and $N-p+1 \le i \le N$. The above equation indicates that given $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$, the $\hat{I}_i$ decreases by $L + 10\eta \lg(1+1/p)$ dB when the width increases from p to $p+1$ ($1 < p \ll N$). Denote by $\mathrm{SINR}_{\mathrm{dw}}$ the SINR measured in dB when dense precoding matrices are used. It immediately follows that $\mathrm{SINR}_{\mathrm{dw}}$ increases by $L + 10\eta \lg(1+1/p)$ dB when the width increases from p to $p+1$ ($1 < p \ll N$).

4.5 SINR Loss Using Sparse W

Denote by $\mathrm{SINR}_{\mathrm{sw}}$ the SINR measured in dB when sparse precoding matrices are used. We now investigate the SINR loss from using $\mathbf{S}_p$, defined as $\Delta_{\mathrm{SINR}} \triangleq \mathrm{SINR}_{\mathrm{dw}} - \mathrm{SINR}_{\mathrm{sw}}$.

4.5.1 SINR Loss in $\mathcal{EZ}_\alpha$

Let $\mathbf{H} = \mathbf{H}_p + \mathbf{D}_p$. Decompose $\mathbf{S}_p$ by $\mathbf{S}_p = \mathbf{W}_p + \mathbf{E}_p$. Rewrite (4.2) as

$$\mathbf{y} \approx \mathbf{Is} + (\mathbf{D}_p\mathbf{W}_p + \mathbf{H}_p\mathbf{E}_p + \mathbf{D}_p\mathbf{E}_p)\mathbf{s}. \tag{4.13}$$

Let $\mathbf{A} = \mathbf{D}_p\mathbf{W}_p$ and $\mathbf{B} = \mathbf{H}_p\mathbf{E}_p$. We omit $\mathbf{D}_p\mathbf{E}_p$ as its dominating term's variance is much smaller than those in $\mathbf{A}$ or $\mathbf{B}$. Consider $p \le i \le N-p$. Denote by $\hat{I}_i^{(S)}$ the residual interference power of the i-th stream when $\mathbf{S}_p$ is used, and by $\hat{I}_i^{(D)}$ when $\mathbf{W}_p$ is used. We have

$$\hat{I}_i^{(S)} = \mathbb{E}_{\mathbf{F}}\left[\lg\left(\sum_{j=1,j\neq i}^{N} \|a_{i,j} + b_{i,j}\|^2\right)\right], \tag{4.14}$$

$$\hat{I}_i^{(D)} = \mathbb{E}_{\mathbf{F}}\left[\lg\left(\sum_{j=1,j\neq i}^{N} \|a_{i,j}\|^2\right)\right], \tag{4.15}$$

where $a_{i,j} = \mathbf{d}_i^T\mathbf{w}_j$ and $b_{i,j} = \mathbf{h}_i^T\mathbf{e}_j$. The $\mathbf{d}_i^T$ and $\mathbf{h}_i^T$ are the i-th row of $\mathbf{D}_p$ and $\mathbf{H}_p$, respectively; $\mathbf{w}_j$ and $\mathbf{e}_j$ are the j-th column of $\mathbf{W}_p$ and $\mathbf{E}_p$, respectively. By $\alpha \to 0$, we invoke (4.7) to take only the terms that have the maximal variance. The (4.14) and (4.15) are then simplified as

$$\hat{I}_i^{(S)} = \mathbb{E}_{\mathbf{F}}[\lg(\|A_{11} - A_{22}\|^2 + \|A_{12} - A_{21}\|^2)], \tag{4.16}$$

$$\hat{I}_i^{(D)} = \mathbb{E}_{\mathbf{F}}[\lg(\|A_{11}\|^2 + \|A_{12}\|^2)], \tag{4.17}$$

where $A_{11} = \alpha^p f_{i,i-p}/f_{i-p,i-p}$ and $A_{12} = \alpha^p f_{i,i+p}/f_{i+p,i+p}$, representing the residual interference from using $\mathbf{W}_p$; $A_{21} = \alpha^0 f_{i,i}w_{i,i-p}$ and $A_{22} = \alpha^0 f_{i,i}w_{i,i+p}$, representing the residual interference from using $\mathbf{S}_p$. We consider $\{A_{21}, A_{22}\}$ dominate $\{A_{11}, A_{12}\}$ since $\mathbb{E}_{\mathbf{F}}[\lg(\|\alpha^{i-k}f_{i,k}\|)] < \mathbb{E}_{\mathbf{F}}[\lg(\|\alpha^0 w_{i,k}\|)]$. The SINR loss of the i-th stream is then given by

$$\Delta_{\text{SINR}} \approx \mathbb{E}_{\mathbf{F}}[10\lg(\|w_{i,i-p}\|^2)] + 10\varepsilon_p - \gamma - 20p\lg\alpha - \mathbb{E}_{\mathbf{F}}[10\lg(\|f_{i,i-p}\|^2 + \|f_{i,i+p}\|^2)], \tag{4.18}$$

where $\varepsilon_p = \mathbb{E}_{\mathbf{F}}[\lg(1 + \|w_{i,i+p}/w_{i,i-p}\|^2)]$.

We now show that $(\varepsilon_{p+1} - \varepsilon_p)$ approaches 0 as $p \to +\infty$ and $\alpha \to 0$. Details can be found in Appendix D. We only list below an outline of the derivation. We first derive a set of recursive equations:

$$n_p = \sum_{k=0}^{p-1} [(-1)^k f_{i+k,i+p} n_k] / f_{i+p,i+p}, \quad d_p = \sum_{k=0}^{p-1} [(-1)^k f_{i-k,i-p} d_k] / f_{i-p,i-p}, \tag{4.19}$$

$$z_p \triangleq \|w_{i,i+p}/w_{i,i-p}\|^2 \approx \|n_p/d_p\|^2. \tag{4.20}$$

We can then evaluate ε_p in a closed-form,

$$\varepsilon_p \approx \lg e \sqrt{\frac{p\pi}{12}} [1 + \frac{3}{2\pi^2 p}] + 0.2826, \tag{4.21}$$

from which we establish that as $p \to +\infty$ and $\alpha \to 0$, the term $(\varepsilon_{p+1} - \varepsilon_p)$ approaches 0.

Considering $\mathbb{E}_{\mathbf{F}}[\ln(\|f_{i,i-p}\|^2 + \|f_{i,i+p}\|^2)] = 1 - \gamma$ and $\mathbb{E}_{\mathbf{F}}[\lg(\|w_{i,i-p}\|^2)] \propto p(2\lg\alpha + \lg e)$ from (4.7), we obtain the approximate rate at which the Δ_{SINR} linearly increases with p: $10\lg e$.

The above process also applies to $1 \le i \le p-1$ and $N-p+1 \le i \le N$, where only one dominating term is present in $\hat{I}_i^{(S)}$. For $\mathbf{H} \in \mathcal{ED}_\alpha$, however, there is no similar conclusion. Its inverse can be shown as a tri-diagonal matrix, where the main diagonal entry is $\frac{1}{1-\alpha^2}$ at the first and last rows, and $\frac{1+\alpha^2}{1-\alpha^2}$ at other rows. The off-diagonal entries of $\mathbf{H}^{-1}$ take the same value: $\frac{-\alpha}{1-\alpha^2}$. See Appendix E for details.

4.5.2 SINR Loss in $\mathcal{EPRZ}_{\alpha,\eta}$

Given $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$. The four terms in (4.16) and (4.17) become:

$$A_{11} = \alpha^p p^{-\eta/2} f_{i,i-p}/w_{i-p,i-p}, \qquad A_{12} = \alpha^p p^{-\eta/2} f_{i,i+p}/w_{i+p,i+p}, \tag{4.22}$$

$$A_{21} = \alpha^0 \delta_i^{-\eta/2} f_{i,i}/w_{i,i-p}, \qquad A_{22} = \alpha^0 \delta_i^{-\eta/2} f_{i,i}/w_{i,i+p}. \tag{4.23}$$

During the derivation of (4.8), we learn that unlike $\mathcal{EZ}_\alpha$ case, an $\mathcal{EPRZ}_{\alpha,\eta}$ matrix has a significant drop in $\mathbb{E}[\lg(\|w_{i,i-p}\|^2)]$ and $\mathbb{E}[\lg(\|w_{i,i+p}\|^2)]$. Therefore, $\{A_{11}, A_{12}\}$ dominates $\{A_{21}, A_{22}\}$ for $\eta > 2$ and $\{A_{21}, A_{22}\}$ dominates $\{A_{11}, A_{12}\}$ for $\eta < 2$. We now rewrite (4.16) and (4.17) as:

$$\hat{I}_i^{(S)} \approx \begin{cases} \mathbb{E}[\lg(\|A_{11} - A_{21}\|^2)] + \mathbb{E}[\lg(1 + \|A_{12}/A_{11}\|^2)], & \eta > 2, \\ \mathbb{E}[\lg(\|A_{11} - A_{21}\|^2)] + \mathbb{E}[\lg(1 + \|A_{22}/A_{21}\|^2)], & \eta < 2, \end{cases} \tag{4.24}$$

$$\hat{I}_i^{(D)} \approx \mathbb{E}[\lg(\|A_{11}\|^2)] + \mathbb{E}[\lg(1 + \|A_{12}/A_{11}\|^2)]. \tag{4.25}$$

Consider $\eta > 2$. We have

$$\Delta_{\text{SINR}}/10 = \mathbb{E}[\lg(1 - \|A_{21}/A_{11}\|^2)]. \tag{4.26}$$

Plugging into (4.26) the estimations of $w_{i,i-p}$ and $w_{i-p,i-p}$, we have

$$\Delta_{\text{SINR}}/10 \approx \mathbb{E}\left[\lg\left(\left\|1 - (-1)^p \sum_{m=i-p+1}^{i-1} \left(\frac{(m-i+p+\delta_m)(i-m+\delta_i)}{|p+\delta_i||\delta_m|}\right)^{-\eta/2} \cdot \frac{f_{m,i-p} f_{i,m}}{f_{i,i-p} f_{m,m}}\right\|^2\right)\right]. \tag{4.27}$$

Eq. (4.27) coincides with the *inband estimation error* of $LW_k^{(1)}$ at $k = p+1$, defined as $\mathbb{E}[\lg(\|w_{k,1}\|^2)] - LW_k^{(1)}$. It can be therefore estimated by the difference of $LW_{p+1}^{(1)}$ and $LW_{p+1}^{(3)}$:

$$\Delta_{\text{SINR}}/10 \approx \frac{\rho_\eta}{\ln 10}\left(\frac{p}{p-1}\right)^{\eta/2}. \tag{4.28}$$

4.6 Downlink Energy Efficiency

We study the downlink energy efficiency in a FMDA system with N antennas serving N pre-scheduled users through ZFBF. Radio frequency signaling is adopted to simplify the antenna design and reduce the amount of energy consumed in backhaul signalling between the antennas and the CPS. Smaller antenna-user distances and low attenuations of optical fibers allow a smaller

transmission power being used at antennas. Consequently, baseband processing contributes more in the total power consumption.

Define the *energy efficiency* as:

$$\eta_e \triangleq \frac{\text{System throughput}}{\text{Total power consumption}} = \frac{B \times N \times \eta_s}{P_{\text{RoF}} + P_{\text{BB}}}, \tag{4.29}$$

where B is the channel bandwidth; η_s is *per-user spectral efficiency* given by SINR/3; P_{RoF} is the power consumption of RoF components; P_{BB} is the power consumption contributed by baseband processing.

4.6.1 Power Consumption of Radio-over-Fiber and Downlink Baseband Processing

The P_{RoF} is decomposed as [125]:

$$P_{\text{RoF}} = N\left(\frac{P_{\text{LDAout}}}{\gamma_{\text{LDA}}} + \frac{P_{tx}}{\gamma_{\text{PA}}} + 0.083 + 0.140\right), \tag{4.30}$$

where P_{LDAout} is the output power of the laser-diode amplifier, set at 5 dBm in [125]; P_{tx} is the output power of the power amplifier; the efficiency of the laser-diode amplifier and the power amplifier are denoted by γ_{LDA} and γ_{PA} and set at 2.2% and 4.4% [141], respectively; the constants 0.083 and 0.140 represent the power consumption of the laser diode and the photodiode, respectively.

The model of P_{BB} follows [68], where measurements in a Long-Term Evolution (LTE) frequency division duplex system establish the relationship between the computation cost and the power consumption. The system operates in a 20-MHz channel, using 64-quadrature-amplitude-modulation. The P_{BB} is translated from the computation cost of four major downlink operations, namely, OFDM, encoding of forward error correction, mapping and beamforming, denoted by G_{ofdm}, G_{fec}, G_{map} and G_{bf}, respectively:

$$P_{\text{BB}} = (G_{\text{ofdm}} + G_{\text{fec}} + G_{\text{map}} + G_{\text{bf}})(1 + \eta_{\text{oh}})/c_{\text{gops}}. \tag{4.31}$$

According to [68], G_{ofdm} and G_{map}, measured in the unit of giga-operation-per-second (GOPS), scales linearly with N; G_{fec} scales linearly with both N and η_s; η_{oh} is 39.76%, representing the overhead due to leakage current, power conversion and active cooling at the CPS; c_{gops} is assumed 40 GOPS per watt. The G_{bf} is obtained by converting the floating-point operation (flop) count of beamforming into GOPS:

$$G_{\text{bf}} = 14000 \cdot 1200 \cdot \left[\left(F_{\text{csi}} + F_{\text{mi}} + F_{\text{spa}} \right) / N_s + F_{\text{pre}} \right], \tag{4.32}$$

where F_{csi}, F_{mi}, F_{spa}, F_{pre} represent the flop count of CSI collection, matrix inversion, stream power allocation and precoding, respectively; 14000 is the number of OFDM symbols per second in the downlink of an LTE frequency division duplex system; 1200 is the number of subcarriers in a 20-MHz channel. The down-scaling factor N_s is the number of OFDM symbols per scheduling interval, which is introduced because CSI collection, matrix inversion and stream power allocation only occur when the set of users changes, while precoding is required for each symbol. The G_{ofdm} is estimated by the computation cost of 2048-point inverse fast Fourier transform, i.e., $G_{\text{ofdm}} = N \cdot 14000 \cdot 2400 \cdot \log_2(2400)$. Based on G_{ofdm} and its relations to G_{map} and G_{fec} that were measured in [68], we have $G_{\text{map}} = G_{\text{ofdm}}/3$ and $G_{\text{fec}} = G_{\text{ofdm}}/3 \times (\eta_s/5.04)$, where the constant 5.04 is the baseline when the modulation is 64-quadrature-amplitude-modulation and the coding rate of forward error correction is close to 1.

4.6.2 Floating-point Operation Count of Beamforming

We consider four components: CSI collection, matrix inversion, stream power allocation and precoding. Two versions of the proposed scheme have the same flop count in CSI collection but different ones in matrix inversion and precoding. The flop count of CSI collection is linearly proportional to the number of $\mathbf{H}$ entries being used to calculate $\mathbf{W}_p$: N^2 for full $\mathbf{H}$, $2Np - p^2$ for banded $\mathbf{H}$, and $2Np$ for striped $\mathbf{H}$. The exact flop count of matrix inversion is determined by enumerating non-zero entries that participate in the calculation, as summarized in Table 4.4.

Table 4.4: Floating-point operation count of banded matrix inversion and stream power allocation

Components	**Banded** ($2 \le p \le N$)	**Striped** ($2 \le p \le N$)
LU factorization	$2Np^2 - \frac{4}{3}p^3$	$2Np^2 + \frac{4}{3}p^3$
Forward substitution (dense $\mathbf{W}$)	$N^2p - Np^2 + \frac{1}{3}p^3$	$N^2p - Np^2 + \frac{2}{3}p^3$
Backward substitution (dense $\mathbf{W}$)	$2N^2p - Np^2$	$2N^2p$
Forward substitution (sparse $\mathbf{W}$)	$Np^2 - \frac{2}{3}p^3$	$Np^2 + \frac{2}{3}p^3$
Backward substitution (sparse $\mathbf{W}$)	$3Np^2 - 2p^3$	$3Np^2 + 2p^3$
WF-SPC-dW	$2N^2 + N\log_2 N + 5$	$2N^2 + N\log_2 N + 5N$
WF-SPCr	$4Np + N\log_2 N + 5N - 2p^2 + 2p$	$N(4p + \log_2 N + 5)$

The flop count of precoding is linearly proportional to the number of non-zero entries in $\mathbf{W}$, each requiring one multiplication and one addition. Thus the flop count is $2N^2$ for dense $\mathbf{W}$, $4Np - 2p^2$ for the banded version of sparse $\mathbf{W}$, and $4Np$ for the striped version of sparse $\mathbf{W}$.

4.7 Numerical Results

We first examine the decaying behaviors of $\mathbf{W}_p$ and then numerically evaluate SINR, where two schemes are considered. Scheme "Dense" uses $\mathbf{W}_p$ for precoding and WF-SPC for power allocation, subject to the sum power constraint NP. Scheme "Sparse" uses $\mathbf{S}_p$ for precoding and WF-SPCr for power allocation, subject to per-antenna power constraint P. To allow a clearer observation on how SINR varies with p, we lower the noise floor to -154 dBm/MHz.

Lastly, we evaluate the downlink energy efficiency in an LTE system that operates in a 20-MHz channel. The noise floor is restored to the normal value, -114 dBm/MHz. Throughout this section, the banded matrix inversion is used to evaluate the decay behavior of $\mathbf{W}_p$; the striped matrix inversion is used to simulate SINR for its homogeneous property.

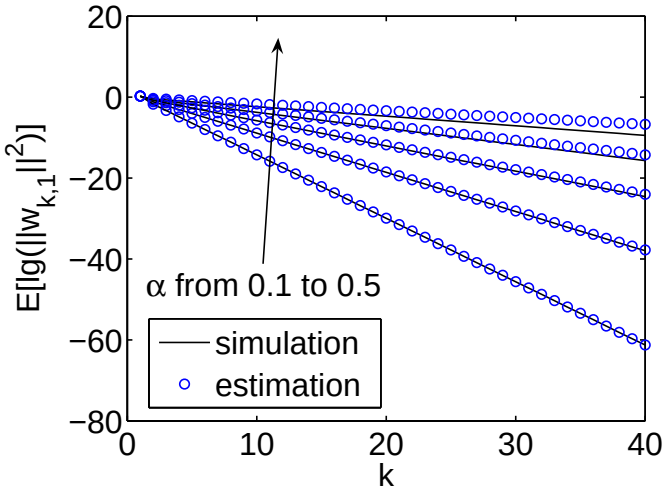


Figure 4.3: Entries of $\mathbf{W}_p$. $\mathbf{H} \in \mathcal{EZ}_\alpha$. $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. N=50. p=40.

4.7.1 SINR Loss in $\mathcal{EZ}_\alpha$

We first observe the normal decay for $\mathbf{H} \in \mathcal{EZ}_\alpha$. 20000 independent fading samples are generated according to the definition of $\mathcal{EZ}_\alpha$ in Table 4.3. From Fig. 4.3 we observe that the decay rate given in (4.7) is accurate when $\alpha \leq 0.3$ and becomes an upper-bound when $\alpha > 0.3$.

Fig. 4.4 shows SINR and Δ_{SINR} for $\alpha \in \{0.82, 0.45, 0.20, 0.09\}$, which correspond to a quarter, one, two and three walls, respectively. The SINR loss from using $\mathbf{S}_p$ linearly increases with p until p reaches a threshold, at which point the system transits from the interference-limited state to the noise-limited state. As stated in Section 4.5.1, the increment rate approaches $10p\lg e$ when α decreases and p increases. Fig. 4.5 shows that compared with $10p\lg e$, the ε_p is becoming negligible when p increases. At small α values, (4.19) and (4.20) match simulations and (4.21) well tracks the variation of ε_p. We also observed that ε_p decreases with α.

4.7.2 SINR Loss in $\mathcal{EPRZ}_{\alpha,\eta}$

We simulate SINR in the network shown in Fig. 4.1 with parameters listed in Table 4.5. In each simulation scenario, 20000 independent user-location drops and fading samples are generated. We first observe normal decay and the outband drop. In Fig. 4.6, we plotted $\mathbb{E}[\lg(\|w_{k,1}\|^2)]$ minus

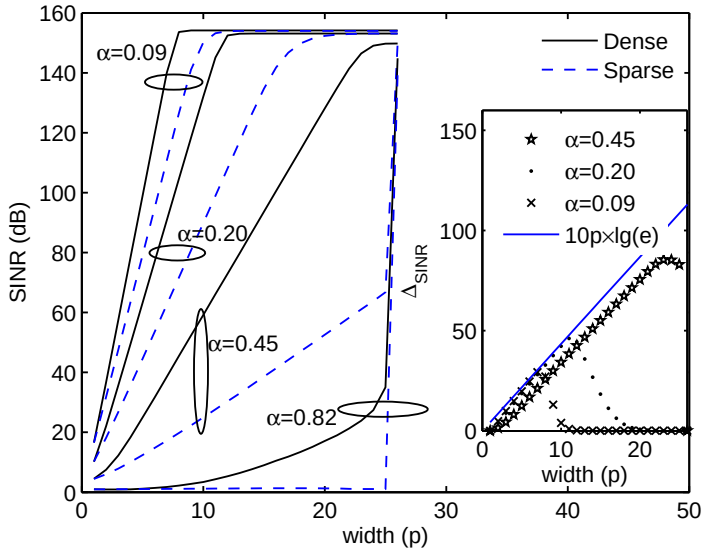


Figure 4.4: SINR using dense and sparse $\mathbf{W}$. $\mathbf{H} \in \mathcal{EZ}_\alpha$. $\alpha \in \{0.82, 0.45, 0.20, 0.09\}$. N=50.

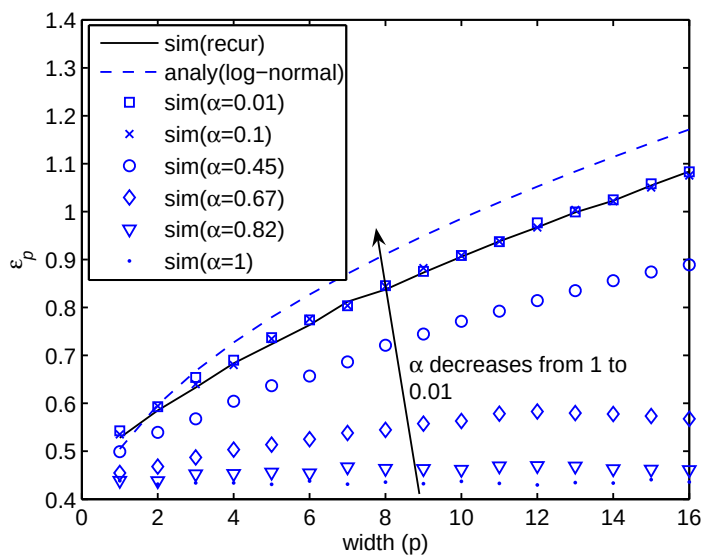


Figure 4.5: The ε_p slowly increases with p. $\mathbf{H} \in \mathcal{EZ}_\alpha$. N=50. Sim(recur): Simulation based on recursive equations. Analy(log-normal): Asymptotic analysis based on log-normal assumption. Sim($\alpha \in \{0.01, 0.1, 0.45, 0.62, 0.82, 1\}$): Simulation of ε_p.

$(2k-2)\lg\alpha$ to remove the effect of α. It can be observed that both $LW_k^{(1)}$ in (4.9) and $LW_k^{(3)}$ in (4.10) fit the simulation for $\eta \in \{2,3,4\}$ and $\alpha \in \{0.1, 0.5, 0.9, 1\}$. The gap between the analysis and the simulation decreases with η because when η becomes smaller than 2, the channel matrix

Table 4.5: Simulation parameters in office buildings

Frequency	2.5 GHz	P (per-antenna power constraint)	10 dBm/MHz
Wall penetration loss	6.9 dB	Thermal noise floor	-114,-154 dBm/MHz
D (inter-office distance)	3 meters	Noise figure at user	7 dB
Antenna gain	0 dB	η (path loss exponent)	$\{2,3,4\}$

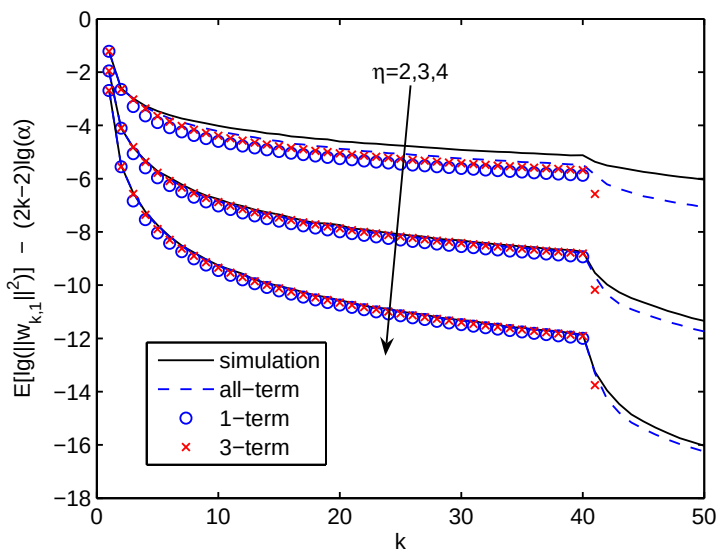


Figure 4.6: Entries of $\mathbf{W}_p$. $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$. $\eta \in \{2,3,4\}$. $\alpha \in \{0.1, 0.5, 0.9, 1\}$. N=50. p=40. 1-term: $LW_k^{(1)}$. 3-term: $LW_k^{(3)}$. All-term: the simulations based on $\{c_0, c_1, \cdots, c_{k-2}\}$ (See Appendix C for details). As the effect of α is already deducted from the Y-axis, all α values produce the same curve for a given η value.

$\mathbf{H}$ is shifting towards $\mathcal{EZ}_\alpha$, which makes the estimations given by (4.8) less accurate.

We observe that using $\mathbf{S}_p$ causes less than 4 dB SINR loss in Fig. 4.7, where to avoid cluttering, only Δ_{SINR} at $\alpha \in \{0.45, 0.20, 0.09\}$ are shown. Simulated SINRs match the SINR_{dw} analysis in Section 4.4, except at $\alpha = 1$, in which case $\mathbf{H}$ has only polynomial decay that invalidates the analysis based on dominating terms. One can notice that the existence of exponential decay is important to the proposed scheme.

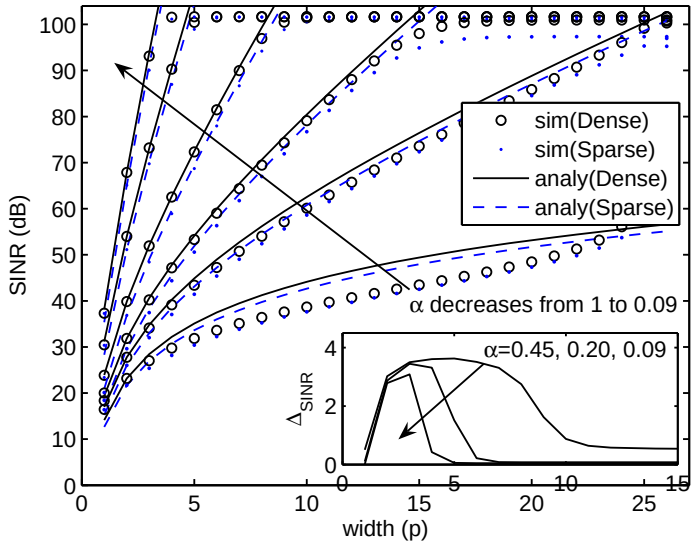


Figure 4.7: SINR using dense and sparse $\mathbf{W}$. $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$. sim: simulation; analy: analysis. $\eta = 3$. $\alpha \in \{1, 0.82, 0.67, 0.45, 0.20, 0.09\}$. N=50.

4.7.3 Downlink Energy Efficiency

We compare the energy efficiency of the dense and sparse schemes at three per-antenna power constraint levels, $P \in \{0.1, 1, 10\}$ mW/MHz, in the same LTE system as in [68]. The transmission power of all antennas in each transmission is recorded to calculate the RoF power consumption. The energy efficiency is obtained from (4.29), (4.30), (4.31), (4.32) and Table 4.4. Although the system operates in a 20-MHz frequency-selective fading channel that consists of 1200 subcarriers, we obtain the spectral efficiency by simulating a single subcarrier through 20000 independent user-location drops and fading samples. As our purpose is to compare the spectral efficiency obtained from the traditional beamforming and the proposed one, the difference evaluated at one subcarrier can represent the results evaluated in the 20-MHz channel. This choice is also made to greatly increase the simulation speed. In practical LTE systems, the proposed beamforming ought to be performed at each subcarrier. However, not each subcarrier or time slot is allocated with training symbols. So there is an opportunity to jointly design the proposed beamforming scheme and the interpolation-based channel estimation, which are often used in practical systems like LTE. This topic is further discussed in Chapter 6.

Fig. 4.8 shows how energy and spectral efficiency change when p increases, where the scheduling interval is assumed to be seven OFDM symbols. At each power level, there exists an optimal value $p = p^*$ that maximizes the energy efficiency in both dense and sparse precoding schemes. At these optimal operation points, the sparse scheme offers 22%~59% higher energy efficiency than the dense scheme due to the use of sparse precoding, only with small loss in the spectral efficiency after the system becomes noise-limited. When p reaches its maximum (equal to 26 for N=50), both schemes converge to the full inversion that degrades energy efficiency. When the power level becomes lower, such degradations worsen because the beamforming contributes more in the total power consumption. Fig. 4.9 shows the power consumption contributed by RoF and beamforming in the sparse scheme, where the RoF contribution is calculated by (4.30) and the beamforming contribution is translated from (4.32). When the transmission power level drops, the power consumption contributed by RoF drops significantly, whereas the power consumption contributed by beamforming remains the same. The observation indicates the importance of low-complexity beamforming scheme when lower transmission power levels are used.

We observe in Fig. 4.8 that the sparse scheme maximizes energy efficiency at $p^* = 4, 5, 6$ for the power level 0.1, 1 and 10 mW/MHz, respectively. The p^* is reached when the residual interference becomes commensurate with the noise floor, in which case further increasing p produces few gains in spectral efficiency but decreases the energy efficiency. Therefore, it is expected that a higher power level leads to a larger p^*.

Fig. 4.10 shows $N_s = 1$ case, where we observe similar behaviors as in $N_s = 7$. The sparse scheme is more attractive in energy efficiency because the energies consumed by beamforming are more significant at a lower N_s value, which may happen in uplink zero-forcing equalization in fast-fading channels. In such cases, the equalizer needs to track the channel variations by extracting the reference symbols that are regularly inserted by users.

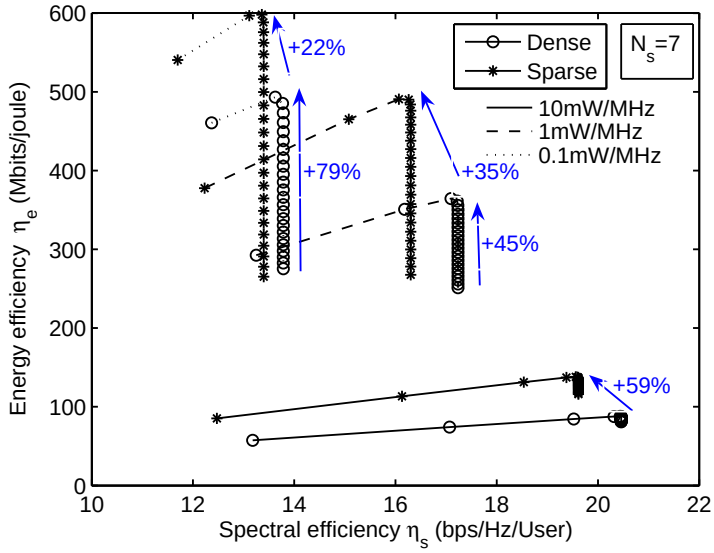


Figure 4.8: Energy efficiency η_e vs. Spectral efficiency η_s ($N_s = 7$). $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$. $\eta = 3$. $\alpha = 0.45$. N=50. The width p ranges from 2 to 26. Mbits/joule: 10^6 bits/joule.

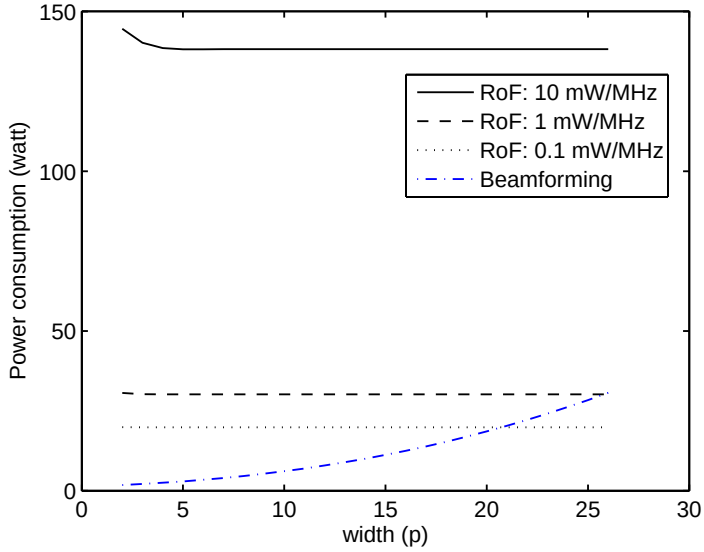


Figure 4.9: Power consumption contributed by radio-over-fiber and beamforming ($N_s = 7$). RoF: radio-over-fiber. The sparse scheme is used. $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$. $\eta = 3$. $\alpha = 0.45$. N=50. The width p ranges from 2 to 26.

4.7.4 Practical Considerations

In this subsection, we briefly discuss how to apply the proposed scheme in practice. In a typical dual-stripe office environment, users residing on two sides of the stripe are scheduled alternatively

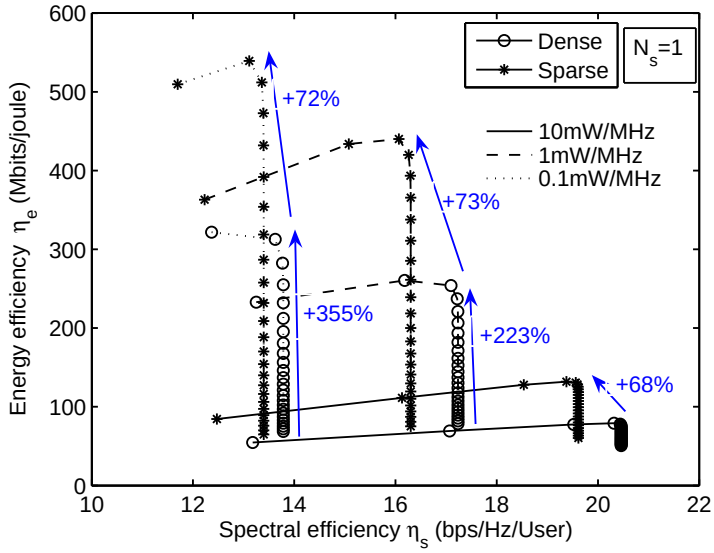


Figure 4.10: Energy efficiency η_e vs. Spectral efficiency η_s ($N_s = 1$). $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$. $\eta = 3$. $\alpha = 0.45$. N=50. The width p ranges from 2 to 26.

to form a one-dimensional network for each transmission. In each room, only one user is scheduled per transmission such that the resulting channel matrix has desired CDODs. If a room has multiple active users, the system serves the users one at a time because each room is only equipped with one antenna. If a room has no active users, the corresponding antenna does not join the beamforming and the resulting channel matrix still has CDODs.

In ensuring the desired CDOD structure, another practical issue is the corridor propagation path. Unlike the direct path crossing multiple walls, the corridor path crosses only two walls and reflects upon one wall. To attenuate this passage, the antennas can be side-mounted on the corridor-side wall such that the front-back ratio of the antennas (typically 10 dB in a directional antenna) can further attenuate the corridor path. Another benefit of such antenna mounting is to increase the isolation between neighboring rooms by taking advantage of the antenna directionality.

In a two-dimensional hyper-dense network (e.g., those deployed in stadiums and airports), the extension of the proposed scheme will be addressed in the next chapter where we study the structure of multi-banded channel matrix with the help of floor-mounted directional antennas. In a three-dimensional network (e.g., a highrise apartment), the concrete floor typically introduces

18.3 dB penetration loss [127] and therefore largely increases the off-diagonal decays when the scheduled users are distributed across multiple floors. However, one should caution to enumerate antenna/user pairs to form a channel matrix that has the desired CDOD structure.

4.8 Summary

We have proposed a low-complexity zero-forcing beamforming scheme in an indoor massively distributed antenna system, where the channel matrix can be modeled to have exponentially and polynomially decayed off-diagonals. Since only neighboring antennas need to cooperate, we band the channel matrix to reduce the computation cost of CSI collection, matrix inversion and precoding. Two versions of the scheme are presented: the dense precoding version provides quadratic computation cost in matrix inversion and precoding; the sparse precoding version provides linear computation cost. The resulting SINRs in both versions are analyzed for random channel matrices. Compared with beamforming based on the full matrix inversion, our analysis and numerical evaluations show that both versions incur negligible loss in SINR, while offering 45%~79% gain in energy efficiency at lower transmission power levels. At a higher transmission power level, although the energy saving from the proposed scheme is less noticeable due to the dominance of the radio transmission power consumption, the sparse precoding scheme still provides 22%~59% higher downlink energy efficiency than the dense precoding scheme. While this chapter is focused on beamforming, the proposed scheme can be easily extended to uplink equalization, where matrix inversion may occur more frequently than in downlink as the receiver can track the channel from regularly transmitted reference symbols by the users. Therefore in uplink scenarios, the energy saving from the proposed scheme would be more significant.

Chapter 5

Low Complexity Zero-forcing Beamforming for Massively Distributed Antenna Systems in Large Public Venues[1]

5.1 Introduction

The spectral efficiency advantages of massive array multiple-input multiple-output (MIMO) systems have been shown in analysis [46] and demonstrated in experiments [47]. Compared with array MIMO systems, distributed MIMO systems improve link reliability by locating antennas closer to users and therefore improve spectral efficiency both indoors [1, 14, 48] and outdoors [4, 49] while requiring less transmission power. When the number of users being served, K, is smaller than the coverage radius measured in the number of wavelengths of the carrier signal, a distributed MIMO system achieves a capacity linear in K [50], demonstrating its potential to satisfy high traffic demands in large public venues such as stadiums, arenas and convention centers. In practice, operators are also deploying a large number of distributed antennas or access points to cover the hyper-dense venues. Particularly in hyper-dense wireless local area networks, access points are

[1]This chapter is based on [98] co-authored with Dr. V. Leung.

preferably mounted under-seat [27] or under-floor [28] to be closer to users and achieve higher inter-access-point isolations by taking advantage of heavy human body penetration loss, thus allowing more access points being deployed to provide a very-high system capacity. Under-seat or under-floor mounting also has advantages in aesthetics and ease of installation [29].

When antennas are mounted under-floor in a distributed massive MIMO system, the channel gain matrix $\mathbf{H}$ can be modeled as a multi-banded matrix with off-diagonal entries that decay both exponentially due to heavy human body penetration loss and polynomially due to free space propagation loss. The human body penetration loss in an antenna-user link is characterized as βd ($\beta > 0$) by following the widely used linear attenuation model [142, 143], where the unit of the linear attenuation coefficient β is dB/meter and d represents the antenna-user distance in meters. Realizing that entries in the matrix are not equally important, we propose a simpler zero-forcing beamforming (ZFBF) scheme based on sparse matrix inversion. Specifically, to send a symbol to a given user, only the channel gains from its nearby antennas matter the most in calculating its beamforming vector, as also observed in [96]. While ignoring other entries causes residual inter-stream interference, the resulting throughput loss is small if the residual interference is commensurate with the inaccuracy of channel estimation, the external interference, or the interference tolerable for the highest level modulation and coding scheme.

The proposed low-complexity ZFBF scheme is designed in a two-dimensional network where a large number of antennas are uniformly deployed in an $M \times M$ grid to cover a large public venue. We study how to relate the residual interference level to the number of channel entries participating in the matrix inversion. To the best of our knowledge, these issues had only recently been addressed in [144, 145].

In [144], the matrix inversion in a linear minimum mean-square-error detector, $(\mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I})^{-1}$, was expanded as a sum of polynomials, $\sum_{l=0}^{L-1} w_l(\mathbf{H}\mathbf{H}^H)^l$, where $(\mathbf{H}\mathbf{H}^H)^l$ forms the l-th base of Krylov subspace, w_l is the corresponding coefficient, σ^2 is the noise power and $\mathbf{I}$ is an identity matrix. When the dimension of $\mathbf{H}$ grows to infinity, the coefficient set $\{w_l\}$ converges and therefore can be pre-calculated. Thus, the polynomial expansion renders quadratic computation cost, relative

to the size of $\mathbf{H}$. Note that a larger L leads to a better approximation at the expense of more matrix multiplications to find the bases. The Krylov subspace method was also explored in low-complexity code division multiple access receivers [146, 147]. In [145], a general approximate sparse matrix inverter was used to reduce the computation cost of minimum mean-square-error beamforming. In Chapter 4 of this thesis, a banded matrix inverter was proposed to enable a low-complexity ZFBF scheme by targeting a one-dimensional network that contains a stripe of indoor offices.

Approximated matrix inversion had also been extensively studied to find a suitable starting point for iterative inversion of sparse matrices. Given a finite banded matrix $\mathbf{V}$, Demko showed in [129] that the inverse $\mathbf{V}^{-1}$ has *exponentially decayed off-diagonals* (EDOD), i.e., $v_{k,b}^{-1} \leq K\alpha^{|k-b|}$, where K is a constant, $0 < \alpha < 1$, and both K and α depend on the width of $\mathbf{V}$. It was later shown [130, 131] that given $\mathbf{V}$ with EDOD, $\mathbf{V}^{-1}$ also has EDOD yet at a higher rate, i.e., $v_{k,b}^{-1} \leq \beta^{|k-b|}(0 < \beta < \alpha)$. In Chapter 4 of this thesis, the decay rate increment was obtained in the inversion of $\mathbf{V} \circ \mathbf{Z}$, where the elements of $\mathbf{Z}$ are independent and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian (ZMCSCG) random variables and the operator $\circ$ denotes the Hadamard product.

Banded matrices had been used in electromagnetic wave simulations [93], inversion of tri-diagonal matrices [94], equalization [95] and circuit design [148]. Multi-banded matrices had only been studied in graphics where neighboring pixels are more relevant [149], in inter-carrier-interference mitigation [150] and in circuit design [151].

The contribution of this chapter is that given a large public venue deployed with a distributed massive MIMO system with $M \times M$ antennas mounted under-floor, we propose a multi-banded matrix inversion algorithm that substantially reduces the computation cost of ZFBF while incurring a negligible throughput loss by keeping the most significant entries in $\mathbf{H}$ and the precoding matrix $\mathbf{W}$. By introducing a parameter p to control the sparsity of $\mathbf{H}$ and $\mathbf{W}$, we can control the computation cost and establish the relationship between the residual interference level and p. A larger p value requires more channel entries to be collected and more calculations in inverting $\mathbf{H}$ yet offer-

ing a higher throughput. The proposed algorithm includes dense and sparse precoding versions, providing quadratic and linear computation cost in M^2, respectively. The difference between this chapter and Chapter 4 lies in the network dimension and channel propagation environment: This chapter considers a two-dimensional network without inner walls, while Chapter 4 considers a one-dimensional network with inner walls which can strengthen inter-antenna isolation.

We also show that polynomially decayed off-diagonals are not enough to offer the opportunity of reducing the computation cost of matrix inversion. Instead, EDOD is needed. However, the distance-based exponential decay model (which will be made clear in Section II.A) weakens the main diagonal dominance of $\mathbf{H}$ and therefore causes the off-diagonal entries in $\mathbf{W}$ to decay slower than $\mathbf{H}$, regardless of how large β is. This phenomenon drastically reduces the throughput in sparse precoding and motivates us to examine the benefit of using directional antennas. As demonstrated by both analysis and simulations, when the directional antenna gain increases, the resulting signal-to-interference-ratio (SIR) increment in sparse precoding increases linearly with p, while the SIR of dense precoding is much less sensitive to p.

Notations: A bold capital $\mathbf{A}$ indicates a matrix $\mathbf{A}$; a bold lowercase $\mathbf{a}$ indicates a vector $\mathbf{a}$. The l_2-norm of $\mathbf{a}$ is $\|\mathbf{a}\|$. $\mathbf{H}^T$, $\mathbf{H}^H$, $\mathbf{H}^{-1}$, $|\mathbf{H}|$ denote the transpose, Hermitian transpose, inversion, and determinant of matrix $\mathbf{H}$, respectively. The entry on the i-th row and j-th column of $\mathbf{H}$ is denoted by $\mathbf{H}(i,j)$ or $h_{i,j}$; the i-th row or column of $\mathbf{H}$ is $\mathbf{h}_i$; $\mathbf{H}_{a:b,c:d}$ denotes the submatrix of $\mathbf{H}$ formed by rows $[a, a+1, \cdots, b]$ and columns $[c, c+1, \cdots, d]$. The smallest integer not less than x is denoted by $\lceil x \rceil$; the largest integer not larger than x is denoted by $\lfloor x \rfloor$; the remainder of a/b is denoted by $a\%b$. $\mathcal{B}_p$ denotes the set of banded matrices with *width* p, i.e., entry (k,b) is zero for $|k-b| \geq p$. $\mathcal{MB}_{M,p}$ denotes the set of multi-banded matrices with inter-band distance M and width p, in which entry (k,b) is zero for $\max(|k\%M - b\%M|, |\lceil k/M \rceil - \lceil b/M \rceil|) \geq p$. A diagonal matrix with the diagonal $\mathbf{a}$ is denoted by diag[$\mathbf{a}$]. ZMCSCG distribution is denoted by $\mathcal{CN}(0, \sigma^2)$, where σ^2 is the variance and the mean is zero. The uniform distribution between a and b is denoted by $\mathcal{U}(a,b)$. The functions $\ln(\cdot)$ and $\lg(\cdot)$ represent natural and base-10 logarithm, respectively. $\mathbb{E}_{\mathbf{X}}[\cdot]$ denotes the expectation over the random variable $\mathbf{X}$; $\mathbb{E}[\cdot]$ is used when this causes no ambiguity.

The rest of this chapter is organized as follows. The system model is introduced in Section 5.2. Section 5.3 presents the proposed multi-banded inversion algorithm. We analyze the SIR in dense and sparse precoding in Section 5.4 and present the numerical results in Section 5.5. Section 5.6 concludes this chapter.

5.2 System Model

Consider an occupied large public venue uniformly deployed with M^2 under-floor mounted antennas arranged in an $M \times M$ grid, as shown in Fig. 5.1. Each of the M^2 squares in the grid has one pre-scheduled user and one centrally located antenna, which transmits and receives radio frequency signals over the air while all the baseband signal processing is concentrated at the centralized processing system through optical cables. Under-floor propagation paths are blocked by concrete and steel structures, which are often present in large public venues [28]. Blocking under-floor propagation paths is to force inter-antenna interference to be attenuated by heavy human body penetration loss, thus allowing the channel matrix to be modeled as a sparse matrix.

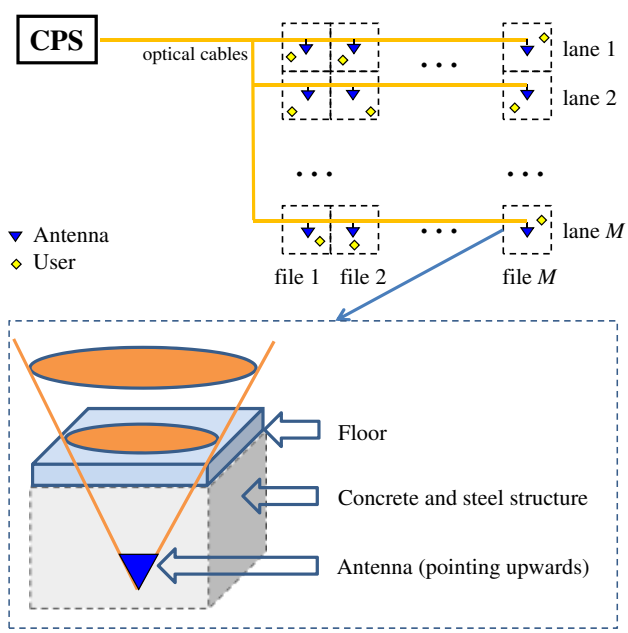


Figure 5.1: A large public venue with M^2 under-floor mounted antennas uniformly deployed over an $M \times M$ grid. CPS: centralized processing system.

5.2.1 Channel Matrix H

Enumerating the antennas and users lane by lane and letting $N = M^2$, we obtain an $N \times N$ channel matrix $\mathbf{H}$, given by $\mathbf{V} \circ \mathbf{F} = \mathbf{V} \circ (\mathbf{Z} + \mathbf{K})$, where $\mathbf{V}$ is the path loss matrix, $\mathbf{F}$ is the fading matrix (frequency-flat block-fading is assumed), $\mathbf{Z}$ is i.i.d. ZMCSCG and $\mathbf{K}$ is the matrix of Rician K-factors representing the ratios of the energy in the specular path to the energy in the scattered paths in the corresponding links. The antenna-user pair at the i-th lane across the j-th file is labeled as $(i-1)M + j$ $(1 \le i, j \le M)$ and the corresponding link is called home antenna-user link. Equivalently, the coordinate of the k-th antenna-user $(1 \le k \le N)$ is given by

$$i_k = \lceil k/M \rceil, j_k = (k-1)\%M + 1. \tag{5.1}$$

To enhance clarity of this section, we reserve (i, j) to index physical coordinates, while (k, b) is used to index rows and columns of $\mathbf{H}$, which correspond to users and antennas, respectively. We ignored shadowing for analytic convenience. Only the home antenna-user links are assumed to experience Rician fading; therefore, $\mathbf{K}$ is a diagonal matrix with each diagonal element fixed at κ.

The path loss between the k-th user and the b-th antenna is modeled as the free space propagation loss with the path loss exponent η $(\eta \ge 2)$ plus the linear attenuation loss, given by $\beta d_{k,b}$ $(\beta > 0)$, which had been widely used in the previous literature [142, 143] and is used here to characterize heavy loss induced by human body penetrations. The unit of β is dB/meter and $d_{k,b}$ represents the distance between the k-th user and the b-th antenna in meters. The (k, b) entry in $\mathbf{V}$ is then given by

$$v_{k,b} = d_{k,b}^{-\eta/2} 10^{-\beta d_{k,b}/20} = d_{k,b}^{-\eta/2} \alpha^{d_{k,b}}, \alpha = 10^{-\beta/20} \tag{5.2}$$

$$d_{k,b} = \sqrt{(i_k - i_b + x_k)^2 + (j_k - j_b + y_k)^2 + h^2} \tag{5.3}$$

where (x_k, y_k) determines the distance between the k-th user and the k-th antenna; h is the height

of all users. The inter-antenna distance is normalized to 1; $x_k, y_k \sim \mathcal{U}(-0.5, 0.5)$. We assume $h \ll 1$ such that $d_{k,b}$ in (3) can be simplified as the Euclidean distance on a plane rather than in a three-dimensional space. The set of matrices that possess the properties (5.1)~(5.3) is denoted by $\mathcal{MEPR}_{M,\alpha,\eta}$, which has a multi-banded structure, exponentially and polynomially decayed off-diagonal entries, and uniformly distributed x_k and y_k. We further denote by $\mathcal{MEPRZ}_{M,\alpha,\eta}$ the Hadamard product of an $\mathcal{MEPR}_{M,\alpha,\eta}$ matrix and a ZMCSCG matrix.

$\mathbf{V}$ can be viewed as an $M \times M$ block matrix, where the (i, j)-block, denoted by $B_{i,j}$, contains the path loss information between the users at the i-th lane and the antennas at the j-th lane. The block-diagonal that contains $B_{i,i}$ $(1 \le i \le M)$, $B_{i,i+l}$ $(1 \le i \le M - l)$, $B_{i-l,i}$ $(l + 1 \le i \le M)$ are called the main band (or 0-th band), the l-th sideband and the (-l)-th sideband, respectively $(0 \le l \le M - 1)$. The way we enumerate the antenna-users determines that each band resembles a ridge that has a peak with both sides gradually decaying, as shown by the contour pattern in Fig. 5.2. The main band has the highest peak and the sharpest decay; a sideband has a smaller peak and a slower decay when it is farther away from the main band. The term $\alpha^{d_{k,b}}$ in (5.2) indicates exponentially decayed off-diagonals within each band and also across the peaks of sidebands; the terms $d_{k,b}^{-\eta/2}$ indicate polynomially decayed off-diagonals. The value of β depends on the crowdedness of the venue, ranging from 0.1 to 0.3 dB/meter in previous measurements [142, 143].

5.2.2 Compositely Decayed Multi-banded Path Loss Matrix

The composite decay within each sideband and across the peaks of sidebands in $\mathbf{V}$ motivates us to form a sparser $\mathbf{V}$, denoted by $\mathbf{V}_p$, by keeping only the main band and p-1 sidebands on both sides, each block in these bands being a banded matrix with width p. An example is shown for p=3 in Fig. 5.2. Define the p-multi-banded $\mathbf{H}$ by $\mathbf{H}_p = \mathbf{V}_p \circ F$. The particular structure in $\mathbf{V}_p$ allows a simple approximate matrix inversion that delivers a desired spectral efficiency at a much lower computation cost than a full inversion.

The choice of p is determined by the throughput demanded by users and the highest level modulation and coding scheme in practical systems. Therefore, values of p of interest to us would

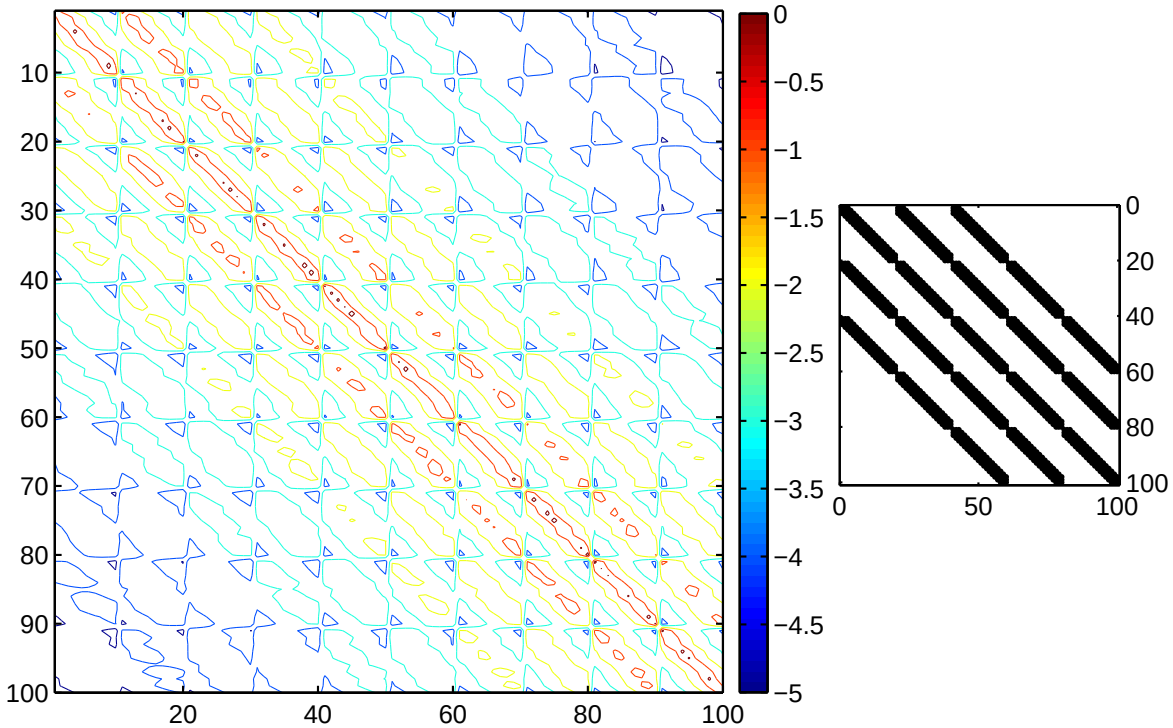


Figure 5.2: Log-scale contour of path loss matrix. Parameters: N=100, M=10, β=0.3 dB/m, η=2, κ=10. Black pattern: the mask to obtain $\mathbf{V}_3$.

be much smaller than M when the distributed massive MIMO system covers a very large area. And in such cases, obtaining $\mathbf{H}_p$ requires much less training than obtaining $\mathbf{H}$. Since links being estimated are grouped locally, it is now possible for antennas that are far away from each other to use non-orthogonal (or to reuse) training sequences, thus mitigating the system capacity reduction caused by the use of long globally orthogonal training sequences.

5.2.3 Precoding

Assuming perfect channel state information (CSI) at the transmitter and perfectly synchronized communications, we consider $N \times 1 \times N$ downlink transmissions: N pre-scheduled users are simultaneously served by N antennas through ZFBF, which is adopted for simplicity and near-optimal performance at high signal-to-noise-ratio (SNR)s [128]. The $N \times 1$ received symbol is

given by $\mathbf{y} = \mathbf{HWs} + \mathbf{n}$, where $\mathbf{H} = [\mathbf{h}_1^H \mathbf{h}_2^H \cdots \mathbf{h}_N^H]^H$, $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \cdots \mathbf{w}_N]$ is the $N \times N$ precoding matrix, $\mathbf{s} = [s_1 s_2 \cdots s_N]^T$ is the $N \times 1$ transmitted symbol vector, and $\mathbf{n}$ is the $N \times 1$ i.i.d. additive Gaussian noise vector in which each component has zero mean and variance σ^2. Thus, $\mathbf{W} = \mathbf{H}^{-1}$. Given a stream power allocation $\{P_k\}(1 \le k \le N)$, we assume $\mathbf{s} \sim \mathcal{CN}(0, \mathrm{diag}[P_1; \cdots; P_N])$.

Our focus in this chapter is hyper-dense deployment scenarios where there will always be active users. Therefore, we consider $N \times N$ channel matrices. If some users are inactive, we have a $Q \times N$ channel matrix $\mathbf{H}$ ($Q < N$). In this case, calculating the right inverse of $\mathbf{H}$ destroys the particular structure of $\mathbf{H}$ while the proposed algorithm is built on that very structure to achieve a low computation cost. A simple adaptation is to first calculate $\mathbf{H}^{-1}$ using the proposed algorithm and then remove the columns that correspond to inactive users. This adaption is suboptimal in that stream power levels should be adjusted after removing the columns. When $Q \ll N$, it is possible to devise a proper user scheduling scheme that forms a single-band channel matrix, thus further reducing computation cost.

5.2.4 Stream Power Allocation

Stream power allocation is accomplished through the waterfilling (WF) algorithm under sum power constraint (SPC), denoted as P_{spc}, or per-antenna power constraint (PAPC), denoted as P_{papc} and simplified as P for notation convenience. WF-PAPC uses the interior point method to iteratively solve the problem [137]; however, the average number of iterations is hard to predict. To maintain a low computation cost, in this chapter we continue to use the relaxed version of WF-SPC, termed "WF-SPCr" in Chapter 4, which first achieves a lower bound of WF-PAPC by running WF with $P_{spc} = P$. Once an initial allocation $\{P_k\}(1 \le k \le N)$ is obtained, all streams' power is scaled up with the same factor until any antenna's output power reaches P. In practice, satisfying the per-antenna power constraint reduces the dynamic rage of output signals at antennas and therefore the cost of power amplifiers.

5.3 Multi-banded Matrix Inversion in Two-dimensional Networks

In this section we present an approximate inversion to an $N \times N$ matrix $\mathbf{H} \in \mathcal{MEPRZ}_{M,\alpha,\eta}$. Let $\mathbf{H} = \mathbf{H}_p + \mathbf{D}_p$, where $\mathbf{D}_p$ is the difference between $\mathbf{H}_p$ and the true $\mathbf{H}$. Denote by $\mathbf{W}$ the full inversion of $\mathbf{H}$; $\mathbf{W}_p$, or dense $\mathbf{W}$, is the full inversion of $\mathbf{H}_p$; $\mathbf{S}_p$, or sparse $\mathbf{W}$, is the p-multi-banded inversion of $\mathbf{H}_p$.

The approximate solution can be obtained directly or iteratively. Since an iterative solver usually converges slower than a direct one [152], we choose a direct solver and in particular, focus on sequential methods for lower computation cost. Parallel methods in direct solvers are mostly based on the divide-and-conquer strategy, offering higher levels of parallelism while incurring higher complexities [153]. When orthogonal frequency division multiplexing is used, different subcarriers or sub-bands can be distributed to the baseband processing cores, thus achieving parallelism without suffering the overhead usually associated with divide-and-conquer methods. Since the sparsity pattern of $\mathbf{H}$ is known beforehand, we are able to design an efficient multi-banded matrix inversion algorithm by following the idea of incomplete LU factorization and incomplete forward/backward substitution, as used in Chapter 4.

5.3.1 Algorithm

Consider a nonsingular $\mathbf{H} \in \mathcal{MB}_{M,p}$ and assume that $\mathbf{H}$ can be factorized into $\mathbf{LU}$ where $\mathbf{L}$ is lower-triangular and $\mathbf{U}$ is upper-triangular. By $\mathcal{MB}_{M,p} \in \mathcal{B}_{(p-1)M+p}$, we have $\mathbf{L} \in \mathcal{B}_{(p-1)M+p}$ and $\mathbf{U} \in \mathcal{B}_{(p-1)M+p}$ [139]. Thus $\mathbf{S}_p$ can be obtained by an incomplete LU factorization followed by banded forward/backward substitution, which however incurs a floating-point operation (flop) count in the order of Np^2M^2. Considering $M \gg p$ in massive MIMO systems, we further simplify this algorithm.

Although $\mathbf{L}, \mathbf{U} \notin \mathcal{MB}_{M,p}$ due to the "fill-in" during LU factorization (i.e., some zero entries in $\mathbf{H}$ become non-zero), the most significant entries in $\mathbf{L}$ and $\mathbf{U}$ still reside in $2p-1$ sidebands. Therefore

we propose a multi-banded LU factorization based on Gauss elimination without pivoting, where for each pivot we eliminate $p-1$ sidebands independently and in parallel. And in each sideband, only $2p-1$ entries are eliminated. The process is illustrated in Fig. 5.3 and the pseudo codes are listed in Table 5.1.

The multi-banded forward and backward substitution algorithms are similarly devised, as detailed in Table 5.2. The process to find $\mathbf{S}_p$ is equivalent to masking $\mathbf{W}_p$ by keeping only $2p-1$ block diagonals, including the main band and $2p-2$ sidebands, and keeping only $2p-1$ side-diagonals within each band. The algorithm can be easily extended to topologies with unequal number of lanes and files.

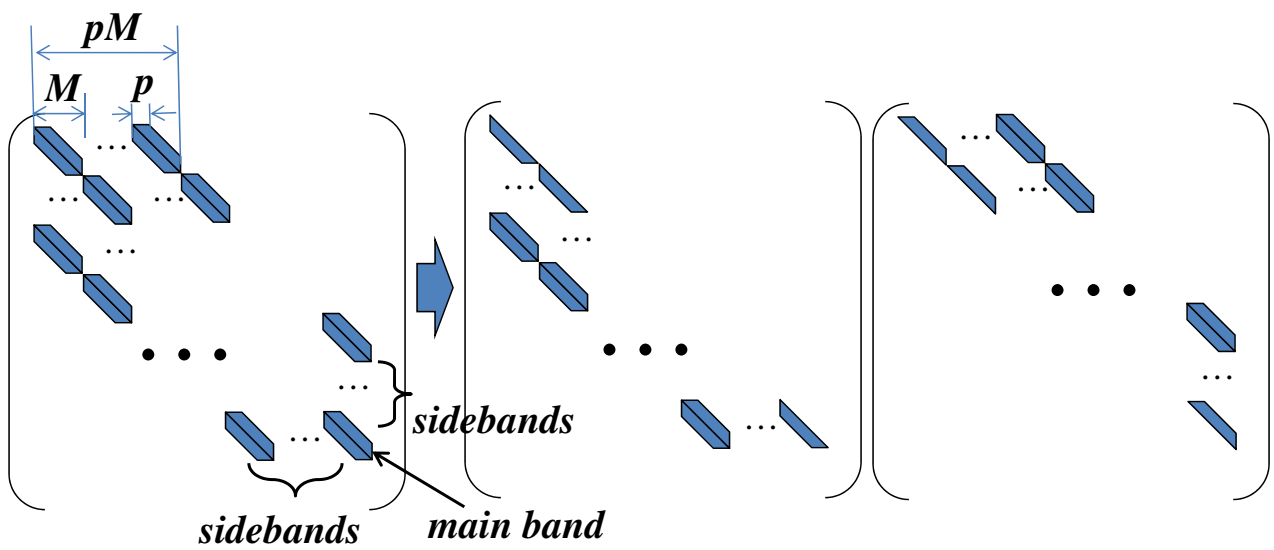


Figure 5.3: Multi-banded LU factorization. Blank areas are zeros.

5.3.2 Computation Cost

We consider four components: CSI collection, matrix inversion, power allocation, and precoding. The flop count of CSI collection is linearly proportional to the number of non-zero entries in the channel matrix: N^2 for a full $\mathbf{H}$, $N(2\sqrt{N}p-p^2)$ for a p-multi-banded $\mathbf{H}$. The flop count of matrix inversion is determined by enumerating non-zero entries that participate in the calculation and listed in Table 5.3, along with the flop count of stream power allocation schemes. The flop count

Table 5.1: Algorithm: Multi-banded LU factorization

Input	$\mathbf{H}, M, p(p < M)$
Output	$\mathbf{H}$: $\mathbf{L}$ and $\mathbf{U}$ are stored as its lower and upper triangular part, respectively.

Algorithm	for k from 1 until N-1 do
	$i_{mb} \leftarrow \lceil k/M \rceil$
	$\mathbf{r} \leftarrow [k+1 : \min(k+p-1, i_{mb}M)]$
	$k_m \leftarrow (k-1)\%M+1$
	$\mathbf{r}_s \leftarrow \max(1, k_m - p + 1) : \min(k_m + p - 1, M)$
	for i_{sb} from $i_{mb}+1$ until $\min(i_{mb}+p-1, M)$ do
	$\mathbf{r} \leftarrow [\mathbf{r} \;\; \mathbf{r}_s + (i_{sb}-1)M]$
	end (i_{sb}-loop)
	$\mathbf{H}(\mathbf{r}, k) \leftarrow \mathbf{H}(\mathbf{r}, k)/\mathbf{H}(k, k)$
	$\mathbf{H}(\mathbf{r}, \mathbf{r}) \leftarrow \mathbf{H}(\mathbf{r}, \mathbf{r}) - \mathbf{H}(\mathbf{r}, k)\mathbf{H}(k, \mathbf{r})$
	end (k-loop)

of precoding is linearly proportional to the number of non-zero entries in $\mathbf{W}$, each requiring one multiplication and one addition. Thus the flop count is $2N^2$ for using $\mathbf{W}_p$ and $2N(2\sqrt{N}p - p^2)$ for using $\mathbf{S}_p$.

As summarized in Table 5.3, the total flop counts of using $\mathbf{W}$, $\mathbf{W}_p$ and $\mathbf{S}_p$ are in the order of $2N^3$, $3N^2p^2 - p^6$ and $6Np^4 - 4p^6$, respectively, which show significant reduction in computation cost from using the proposed multi-banded inversion algorithm.

5.4 Residual Interference

In this section we analyze the residual interference when $\mathbf{W}_p$ and $\mathbf{S}_p$ are used for precoding. We start with a simple analysis on a path loss matrix where all antenna-user distances are fixed at a small constant and fading is absent. Once the matrix structure is understood, we move on to the analysis of stochastic $\mathbf{H}$.

Given an instantaneous full channel matrix $\mathbf{H}$ and the precoding matrix $\mathbf{W}_p$, the received signal $\mathbf{y}$ is:

$$\mathbf{y} = \mathbf{H}\mathbf{W}_p\mathbf{s} + \mathbf{n} = (\mathbf{H}_p + \mathbf{D}_p)\mathbf{W}_p\mathbf{s} + \mathbf{n} = (\mathbf{I} + \mathbf{D}_p\mathbf{W}_p)\mathbf{s} + \mathbf{n} \approx \mathbf{I}\mathbf{s} + \mathbf{D}_p\mathbf{W}_p\mathbf{s} \tag{5.4}$$

Table 5.2: Algorithm: Multi-banded forward/backward substitution

Forward	Backward
Input: $\mathbf{L}, M, p(p<M)$, $\mathbf{e}_i$(the i-th column of $\mathbf{I}$)	Input: $\mathbf{U}, \mathbf{b}, M, p(p<M)$
Output: $\mathbf{b} \approx \mathbf{L}^{-1}\mathbf{e}_i$	Output: $\mathbf{b} \approx \mathbf{U}^{-1}\mathbf{b}$

$b \leftarrow \mathbf{e}_i$
$B_s \leftarrow \lceil i/M \rceil - 1$
$i_0 \leftarrow (i-1)\%M + 1$
$\mathbf{r}_m \leftarrow i_0 + 1 : \min(i_0 + p - 1, M)$
$\mathbf{r}_s \leftarrow \max(1, i_0 - p + 1) : \min(i_0 + p - 1, M)$
for B_t from B_s until $\min(B_s + p - 1, M - 1)$ do
  if $B_t = B_s$:
  $\mathbf{r} \leftarrow \mathbf{r}_m + MB_s$
  for k in $\mathbf{r}$
  $\mathbf{c} \leftarrow i : k - 1$
  $\mathbf{b}_k \leftarrow \mathbf{b}_k - \mathbf{L}(k, \mathbf{c})\mathbf{b}(\mathbf{c})$
  end (k-loop)
  $\mathbf{c}_p \leftarrow [i \;\; \mathbf{r}]$
  else:
  $\mathbf{r} \leftarrow \mathbf{r}_s + MB_t$
  for k in $\mathbf{r}$
  if k is the first element in $\mathbf{r}$:
  $\mathbf{c} \leftarrow \mathbf{c}_p$
  else: $\mathbf{c} \leftarrow [\mathbf{c} \;\; k-1]$
  endif
  $\mathbf{b}_k \leftarrow \mathbf{b}_k - \mathbf{L}(k, \mathbf{c})\mathbf{b}(\mathbf{c})$
  end(k-loop)
  $\mathbf{c}_p \leftarrow [\mathbf{c}_p \;\; \mathbf{r}]$
  endif
end (B_t-loop)
	$t \leftarrow \min(i_0 + p - 1, M)$
	$\mathbf{r}_m \leftarrow t - 1 : -1 : \max(1, i_0 - p + 1)$
	$\mathbf{r}_s \leftarrow [t \;\; \mathbf{r}_m]$
	$B_{ts} \leftarrow \min(B_s + p - 1, M - 1)$
	$B_{te} \leftarrow \max(B_s - p + 1, 0)$
	for B_t from B_{ts}, step size -1, until B_{te} do
	  if $B_t = B_{ts}$:
	  $a \leftarrow t + MB_{ts}$
	  $\mathbf{b}_a \leftarrow \mathbf{b}_a / \mathbf{U}_{a,a}$
	  $\mathbf{r} \leftarrow \mathbf{r}_m + MB_t$
	  for k in $\mathbf{r}$
	  $\mathbf{c} \leftarrow k + 1 : \min(k + p - 1, a)$
	  $\mathbf{b}_k \leftarrow [\mathbf{b}_k - \mathbf{U}(k, \mathbf{c})\mathbf{b}(\mathbf{c})] / \mathbf{U}_{k,k}$
	  end (k-loop)
	  $\mathbf{c}_p \leftarrow [a \;\; \mathbf{r}]$
	  else:
	  $\mathbf{r} \leftarrow \mathbf{r}_s + MB_t$
	  for k in $\mathbf{r}$
	  if k is the first element in $\mathbf{r}$:
	  $\mathbf{c} \leftarrow \mathbf{c}_p$
	  else: $\mathbf{c} \leftarrow [\mathbf{c} \;\; k+1]$
	  endif
	  $\mathbf{b}_k \leftarrow \mathbf{b}_k - \mathbf{U}(k, \mathbf{c})\mathbf{b}(\mathbf{c}) / \mathbf{U}_{k,k}$
	  end(k-loop)
	  $\mathbf{c}_p \leftarrow [\mathbf{c}_p \;\; \mathbf{r}]$
	  endif
	end (B_t-loop)

where $\mathbf{D}_p\mathbf{W}_p$ is the residual interference after the multi-banded inversion. The last approximation is due to the high SNR assumption in our network of interest, where antennas are densely deployed to increase the system capacity; therefore, dramatically reduced antenna-user distances allow the system to operate in the high SNR regions. Although the main diagonal of $\mathbf{D}_p\mathbf{W}_p$ still contributes to the signal power, the contribution is much smaller when added by the identity matrix $\mathbf{I}$ in the term $\mathbf{I} + \mathbf{D}_p\mathbf{W}_p$. We thus define $\mathbf{A}$ as $\mathbf{D}_p\mathbf{W}_p$ with the main diagonal set to zeros. By $\mathbb{E}[\mathbf{s}\mathbf{s}^H] =$

Table 5.3: Complexity comparison of the proposed and conventional ZFBF ($2 \le p \le \sqrt{N}$)

Components	**Proposed ZFBF** (dense **W** version)	**Proposed ZFBF** (sparse **W** version)	**Conventional ZFBF**
LU factorization	$2Np^4 - \frac{4}{3}p^6$	$2Np^4 - \frac{4}{3}p^6$	$\frac{2}{3}N^3$
Forward substitution	$N^2p^2 - Np^4 + \frac{1}{3}p^6$	$Np^4 - \frac{2}{3}p^6$	$\frac{1}{3}N^3$
Backward substitution	$2N^2p^2 - Np^4$	$3Np^4 - 2p^6$	N^3
CSI collection	$N(2\sqrt{N}p - p^2)$	$N(2\sqrt{N}p - p^2)$	N^2
Precoding	$2N^2$	$2N(2\sqrt{N}p - p^2)$	$2N^2$
Stream power allocation: WF-SPC	$2N^2 + N\log_2 N$	/	$2N^2 + N\log_2 N$
Stream power allocation: WF-SPCr	/	$2Np^2 + N\log_2 N$	/
Total	$3N^2p^2 - p^6$	$6Np^4 - 4p^6$	$2N^3$

$\mathrm{diag}[P_1;\cdots;P_N]$, the residual interference power of the k-th user, denoted by $I_k^{(D)}$, is given by:

$$I_k^{(D)} = \sum_{b=1,b\neq k}^{N} \left[|a_{k,b}|^2 P_b\right] \approx P \sum_{b=1,b\neq k}^{N} |a_{k,b}|^2 \tag{5.5}$$

where $a_{k,b}$ is entry (k,b) in $\mathbf{A}$, P is the sum power constraint divided by N and P_b is the power of the b-th stream. The approximation is to decouple the analysis of residual interference from power allocation by invoking equal stream power allocation. In this section, we consider $k = (N-M)/2$, which points to the middle row that has interference coming from both sides of the k-th column.

5.4.1 Deterministic Channel Matrix

Given $\mathbf{H} \in \mathcal{MEPR}_{M,\alpha,\eta}$, we fix x_k and y_k at a small constant $c/\sqrt{2}(c \ll 1)$, simplify $d_{k,b}$ as $\sqrt{(i_k - i_b)^2 + (j_k - j_b)^2}$ for $k \neq b$, and then multiply $\mathbf{H}$ by c to normalize its main diagonal entries. The resulting matrix becomes deterministic, denoted by $\mathcal{MEPD}_{M,\alpha,\eta}$.

Lemma 1: Given $\mathbf{H} \in \mathcal{MEPD}_{M,\alpha,\eta}$. (a) $\mathbf{H}^T = \mathbf{H}$. (b) $\mathbf{H} = \mathbf{P}_\pi \mathbf{H} \mathbf{P}_\pi$, where $\mathbf{P}_\pi$ is a permutation matrix according to $\pi = (\pi_1 \pi_2 \cdots \pi_M)$ and π_i is given by

$$\begin{pmatrix} M(i-1)+1 & M(i-1)+2 & M(i-1)+3 & \cdots & iM \\ i & i+M & i+2M & \cdots & i+M(M-1) \end{pmatrix}. \tag{5.6}$$

Proof: (a) and (b) immediately follow from the construction process of $\mathbf{H}$. ■

The first row of π_i corresponds to antennas in the i-th lane while the second row corresponds to those in the i-th file. Thus, the result of applying $\mathbf{P}_\pi$ is enumerating antennas file by file instead of lane by lane.

Lemma 2: Given $\mathbf{H}_p$ as the p-multi-banded version of $\mathcal{MEPD}_{M,\alpha,\eta}$. The following properties hold:

(a) Decays within the main band: The $|\mathbf{W}_p(k,k+b)|$ decays at the same rate as $|h_{k,k+b}|$ when $|b|$ increases from 1 to $p-1$, provided that $\alpha \to 0$.

(b) Decays within sidebands: The $|\mathbf{W}_p(k,k+mM+b)|$ decays at the same rate as $|h_{k,k+mM+b}|$ when $|b|$ increases from 1 to $p-1$, provided that $\alpha \to 0$; this holds for $1 \le |m| \le p-1$.

(c) Peaks in sidebands: The $|\mathbf{W}_p(k,k+mM)| = |\mathbf{W}_p(k,k+m)|$ for $0 \le |m| \le (M+1)/2$.

Proof: (a) and (b) Consider $1 \le b \le p-1$. We have $|\mathbf{W}_p(k,k+b)| = (-1)^b M_{k+b,k}/|\mathbf{H}_p|$, where $M_{k+b,k}$ is the $(k+b,k)$ minor of $\mathbf{H}_p$. Viewing $\mathbf{H}_p$ as an $M \times M$ block matrix and denoting the (i,j)-block by $B_{i,j}$, we obtain the block index of $h_{k+b,k}$ as $i = j = \lceil k/M \rceil$. By $p \ll M$ and $\alpha \to 0$, $M_{k+b,k}$ approaches the product of three determinants: $|\mathbf{H}_{1:l,1:l}| \times |\mathbf{H}_{l+1:r-1,l+1:r-1}| \times |\mathbf{H}_{r:N,r:N}|$, where $l = (i-1)M$ and $r = (i+1)M$. We notice that only $|\mathbf{H}_{l+1:r-1,l+1:r-1}|$ varies with b and when $\alpha \to 0$, it approaches $|\mathbf{H}_{l+1:k-1,l+1:k-1}| \times |\mathbf{H}_{k:k+b-1,k+1:k+b}| \times |\mathbf{H}_{k+b+1:r-1,k+b+1:r-1}|$. Since the first and third term approach 1 when $\alpha \to 0$, the $b \times b$ matrix $\mathbf{H}_{k:k+b-1,k+1:k+b}$, denoted by $\mathbf{J}$, is of our main interest. When $\alpha \to 0$, $|\mathbf{J}|$ approaches $c\alpha^b b^{-\eta/2}$. It follows that $|\mathbf{W}_p(k,k+b)| \to c\alpha^b b^{-\eta/2}$. By $\mathbf{H}^T = \mathbf{H}$, we complete the proof of (a). The above process also holds for $1-p \le b \le -1$. The proof of (b) similarly follows.

(c) By $\mathbf{H} = \mathbf{P}_\pi \mathbf{H} \mathbf{P}_\pi$, we have $\mathbf{H}_p = \mathbf{P}_\pi \mathbf{H}_p \mathbf{P}_\pi$. So the $(k,k+m)$-minor of $\mathbf{H}_p$ is the $(k,k+m)$-minor of $\mathbf{P}_\pi \mathbf{H}_p \mathbf{P}_\pi$, which is the $(k,k+mM)$-minor of $\mathbf{H}_p$ by (5.6). Thus, $|\mathbf{W}_p(k,k+mM)| = |\mathbf{W}_p(k,k+m)|$ for $0 \le |m| \le (M+1)/2$. ■

SIR when using $\mathbf{W}_p$

From Lemma 2 we know $\mathbf{W}_p$ decays at the same rate as $\mathbf{H}$ and the main diagonal of $\mathbf{W}_p$ is constantly 1. The residual interference is therefore governed by the dominating entries in $\mathbf{A}$, each being generated when $\mathbf{W}_p(b,b)$ meets a non-zero entry in $\mathbf{D}_p$. Considering that each non-zero entry in the l-th sideband of $\mathbf{D}_p$ ($-p+1 \le l \le p-1$) has a counterpart outside of the $2p-1$ sidebands, we rewrite (5.5) as:

$$I_k^{(D)} \approx 2P \sum_{l=-p+1}^{p-1} [\sum_{b=p}^{M\lceil k/M\rceil-k} |h_{k,k+lM+b}|^2 + \sum_{b=p}^{M\lceil k/M\rceil-k+1} |h_{k,k+lM-b}|^2] \le P(8p-4)\alpha^{2p}/p^2. \tag{5.7}$$

The residual interference given in (5.7) is mainly contributed by $8p$-4 interfering antennas that geographically form a square around the k-th user, $2p$-1 interfering antennas residing at each side. The SIR in dB when using $\mathbf{W}_p$, denoted by $\mathrm{SIR}_{\mathrm{dw}}$, is then given by

$$\mathrm{SIR}_{\mathrm{dw}} \approx -20p\lg(\alpha) + 10\lg(p) - 9 \tag{5.8}$$

which increases by $-20\lg(\alpha) + 10\lg(1+1/p)$ dB when the width increases from p to p+1 ($1 < p \ll M$).

Outband drop

Lemma 2 indicates that $\mathbf{W}_p$ decays at the same rate as $\mathbf{H}_p$. However, once $|b|$ passes $p-1$, the current dominating term in $|\mathbf{J}|$ disappears due to the multi-banding of $\mathbf{H}$, which would accelerate the decay of $|\mathbf{W}_p(k,k+b)|$. To quantify the behavior, we define the *outband drop* of $\mathbf{W}_p$, denoted by O_l, as the log-ratio of the accelerated and the normal decay in the l-th sideband:

$$O_l = \lg(|\mathbf{W}_p(k,k+lM+p)/\mathbf{W}_{p+1}(k,k+lM+p)|) \tag{5.9}$$

where $0 \le l < p$. We first evaluate O_0, the outband drop in the main band. The $p \times p$ matrix $\mathbf{J}$ is given by

$$\mathbf{J} = c^p \begin{pmatrix} \alpha \cdot 1^{-\eta/2} & \alpha^2 2^{-\eta/2} & \cdots & \alpha^{p-1}(p-1)^{-\eta/2} & 0 \\ 1/c & \alpha \cdot 1^{-\eta/2} & \cdots & \cdots & \alpha^{p-1}(p-1)^{-\eta/2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha^{p-3}(p-3)^{-\eta/2} & \cdots & \cdots & \alpha \cdot 1^{-\eta/2} & \alpha^2 2^{-\eta/2} \\ \alpha^{p-2}(p-2)^{-\eta/2} & \alpha^{p-3}(p-3)^{-\eta/2} & \cdots & 1/c & \alpha \cdot 1^{-\eta/2} \end{pmatrix}. \quad (5.10)$$

Considering that $1/c$ is much larger than other entries in $\mathbf{J}$, we approximate $|\mathbf{J}|$ by taking as many $1/c$-entries as possible. Specifically, we can have $p-1$ terms in the Leibniz expansion of $|\mathbf{J}|$, where each term has p-2 $1/c$-entries with the pair $\{\mathbf{J}(b+1,b),0\}$ being replaced by $\{\mathbf{J}(1,b+1),\mathbf{J}(b+1,1)\}$ $(2 \le b \le p)$. Note that each term is only one exchange away from the permutation $\{p,1,2,\cdots,p\text{-}1\}$ and therefore has the same sign. So we have

$$|\mathbf{J}| \approx c^2 \sum_{b=1}^{p-1} [\alpha^p b^{-\eta/2}(p-b)^{-\eta/2}] \approx c^2[2\ln(p-1)+1]\alpha^p/p \quad (5.11)$$

where the second approximation is from setting $\eta = 2$ and applying Euler-Maclaurin formula. Therefore, $O_0 \approx \lg(2c\ln p)$.

Now consider $l > 0$. The $|\mathbf{J}|$ in the l-th sideband is approximated by $\sum_{m=0}^{l}\sum_{b=1}^{p-1}\chi_{m,b}$, where $\chi_{m,b} = \frac{\alpha^{\sqrt{b^2+m^2}+\sqrt{(p-b)^2+(l-m)^2}}}{\sqrt{b^2+m^2}\cdot\sqrt{(p-b)^2+(l-m)^2}}$. This summation is visualized in Fig. 5.4. To estimate $|\mathbf{J}|$, we first examine how $\chi_{m,b}$ varies with m and b. According to Minkowski inequality, the exponent part in $\chi_{m,b}$, $\sqrt{b^2+m^2}+\sqrt{(p-b)^2+(l-m)^2}$, has the minimum $\sqrt{p^2+l^2}$, which can be shown achieved at $b^* = pm/l$. The denominator of $\chi_{m,b}$, however, reaches its minimum at $b = 1$ for $0 \le m \le l/2$ and at $b = p-1$ for $l/2 \le m \le l$. Therefore, $\chi_{m,b}$ reaches its maximum at $\chi_{0,1}$ and $\chi_{l,p-1}$. For $0 < m < l$, exponential decay dominates polynomial decay; therefore, other dominating elements of $\chi_{m,b}$ reside at $\{m,\lfloor pm/l \rfloor\}$ and $\{m,\lceil pm/l \rceil\}$. In fact, $\chi_{m,b}$ can be visualized as a saddle with the central

line connecting two peaks $\chi_{0,1}$ and $\chi_{l,p-1}$. Approximating $|\mathbf{J}|$ by $\chi_{0,1}+\chi_{l,p-1}+2\sum_{m=1}^{l-1}\chi_{m,pm/l}$, we have

$$O_l = \lg\left(\frac{2\alpha^{\sqrt{p^2+l^2}}}{\sqrt{(p-1)^2+l^2}} \Big/ \frac{c\alpha^{\sqrt{p^2+l^2}}}{\sqrt{p^2+l^2}}\right) \approx \lg(2c) \tag{5.12}$$

which also holds for the (-l)-th sideband. And when O_l in (5.9) is defined at another side of the sideband, i.e., $\lg(|\mathbf{W}_p(k,k+lM-p)/\mathbf{W}_{p+1}(k,k+lM-p)|)$, (5.11) and (5.12) also hold. For notation simplicity, we only invoke the definition in (5.9). By the same argument in proving Lemma 2(c), the accelerated decay at the peak of the p-th sideband of $\mathbf{W}_p$, given by $\lg(|\mathbf{W}_{p+1}(k,k+pM)| - |\mathbf{W}_p(k,k+pM)|)$ can be also represented by O_0. Note that by symmetry of $\mathbf{H}_p$, the above results also hold for any column of $\mathbf{W}_p$.

SIR when using $\mathbf{S}_p$

After learning the structure of $\mathbf{W}_p$, we can now evaluate the SIR in dB when using $\mathbf{S}_p$, denoted by $\mathrm{SIR_{sw}}$. First, decompose $\mathbf{S}_p$ by $\mathbf{S}_p = \mathbf{W}_p + \mathbf{E}_p$ and rewrite (5.4) as

$$\mathbf{y} = (\mathbf{H}_p+\mathbf{D}_p)(\mathbf{W}_p+\mathbf{E}_p)\mathbf{s}+\mathbf{n} \approx \mathbf{I}\mathbf{s}+(\mathbf{D}_p\mathbf{W}_p+\mathbf{H}_p\mathbf{E}_p+\mathbf{D}_p\mathbf{E}_p)\mathbf{s}. \tag{5.13}$$

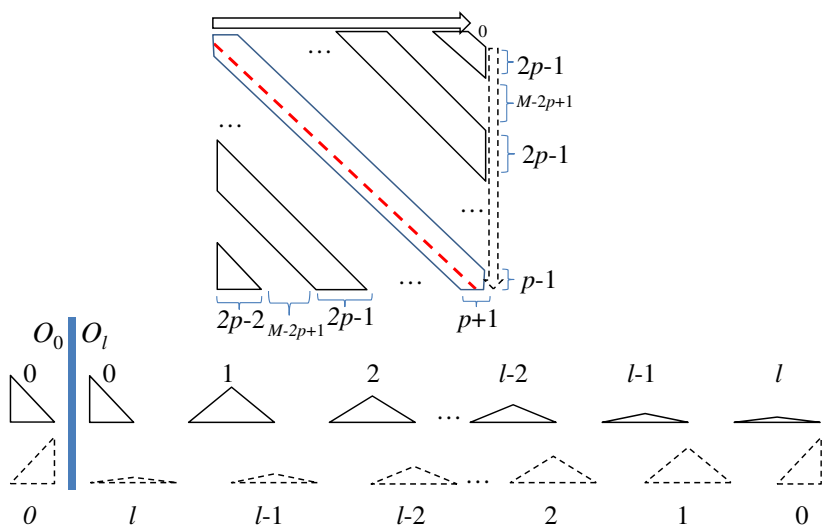


Figure 5.4: Outband drop estimation. Red dash line: the main diagonal containing dominating entries. O_0: outband drop in the main band. O_l: outband drop in the l-th side band.

Let $\mathbf{A} = \mathbf{D}_p\mathbf{W}_p$ and $\mathbf{B} = \mathbf{H}_p\mathbf{E}_p$. We omit $\mathbf{D}_p\mathbf{E}_p$ in estimating the residual interference. Denoting by $I_k^{(S)}$ the residual interference power of the k-th user, we have

$$I_k^{(S)} \approx P \sum_{b=1, b\neq k}^{N} |a_{k,b} + b_{k,b}|^2. \tag{5.14}$$

From Lemma 2 we know that $\mathbf{W}_p$ decays at the same rate as $\mathbf{H}$ and from Section 5.4.1 we learn that the outband drop in $\mathbf{W}_p$ is significant. Therefore, $|b_{k,b}|$ is much smaller than $|a_{k,b}|$, from which we conclude that $I_k^{(S)} \approx I_k^{(D)}$ and $\mathrm{SIR}_{\mathrm{sw}} \approx \mathrm{SIR}_{\mathrm{dw}}$.

5.4.2 Stochastic Channel Matrix

Consider $\mathbf{H} \in \mathcal{MEPRZ}_{M,\alpha,\eta}$ $(0 < \alpha < 1, \eta = 2)$, in which $h_{k,b} = f_{k,b}\alpha^{d_{k,b}}/d_{k,b}$.

Structure of $\mathbf{W}_p$

In an indoor office environment where each room is equipped with one antenna and has one user, we have shown in Chapter 4 that in the presence of both exponential and polynomial decay, the off-diagonal entries in $\mathbf{W}_p$ has the same decay rate as in $\mathbf{H}$ and there exists a noticeable outband drop in $\mathbf{W}_p$, which is why by using $\mathbf{S}_p$ we were able to achieve similar performance as using $\mathbf{W}_p$. The reason of having equal decay rates in $\mathbf{W}_p$ is the use of discrete exponential decay model, in which the attenuation between a user and an antenna depends on the number of walls separating them rather than the distance. The main diagonal dominance is therefore strengthened by office walls that enclose an antenna-user pair and provide isolation from its neighboring antenna-user pairs.

In a large open space, however, the exponential decay depends on the antenna-user distance and the decay model becomes continuous. When $\mathbf{H}$ is deterministic, the home antenna-user link distance is set at a small constant c to ensure the main diagonal dominance in $\mathbf{H}$ and therefore allows the proof of Lemma 2. When $\mathbf{H}$ is stochastic, however, the dominance is disrupted by randomized antenna-user distance and particularly, weakened by the two diagonals next to the

main diagonal in the main band. The first upper and lower sideband may also be comparable to the main diagonal, according to Lemma 1(b). We term this dominance structure as *tri-band tri-diagonal dominance*.

Fig. 5.5 presents simulation results for the *differential decay* in the main band, defined as $\mathbb{E}[\lg(|h_{k,k+b}/h_{k,k+b-1}|^2)]$ for $\mathbf{H}$ and $\mathbb{E}[\lg(|w_{k+b,k}/w_{k+b-1,k}|^2)]$ for $\mathbf{W}$ to signify how much decay is incurred when an entry is moving away from the main diagonal. Following the setting in Table 5.4, we compare two systems: $M \times M$ lanes and $M \times 1$ lanes. The differential decay of $\mathbf{W}$ in the $M \times 1$ system, which has only one band, can be obtained by following the same procedure in Section 5.4.1 and numerically evaluating the determinant of the random matrix $\mathbf{H}_{k+1:k+b,k:k+b-1}$, denoted by $\mathbf{J}$, at two consecutive b values. Although among $b!$ terms of $|\mathbf{J}|$, the term $x_0 = \frac{\alpha^{d_{k+b,k}}}{d_{k+b,k}} f_{k+b,k} \prod_{m=1}^{b-1} (\frac{\alpha^{d_{k+m,k+m}}}{d_{k+m,k+m}} f_{k+m,k+m})$ is the largest contributor, the evaluation should also include the main diagonal and the second upper diagonal in $\mathbf{J}$ as they are comparable to the first upper diagonal. In the $M \times M$ system, the presence of tri-band dominance further complicates the process and only evaluating $|\mathbf{J}|$ is not enough to accurately estimate $\mathbb{E}[\lg(|w_{k+b,k}|^2)]$. Simulation results in Fig. 5.5 show that the tri-band dominance further decelerates the decays in $\mathbf{W}$ even after β reaches 2.4 dB/m.

Table 5.4: Simulation parameters

Parameter	**Value**
Frequency	2.5 GHz
D (inter-antenna distance)	20 meters
Venue size	$400 \times 400 m^2$
N (number of antennas)	400
η (path loss exponent)	2
β (linear attenuation factor)	0.3 dB/meter
P (per-antenna power constraint)	0 dBm/MHz
G_t (transmitter antenna gain)	$0 \sim 5$ dB
G_r (receiver antenna gain)	0 dB
Thermal noise floor	-114 dBm/MHz
User noise figure	7 dB

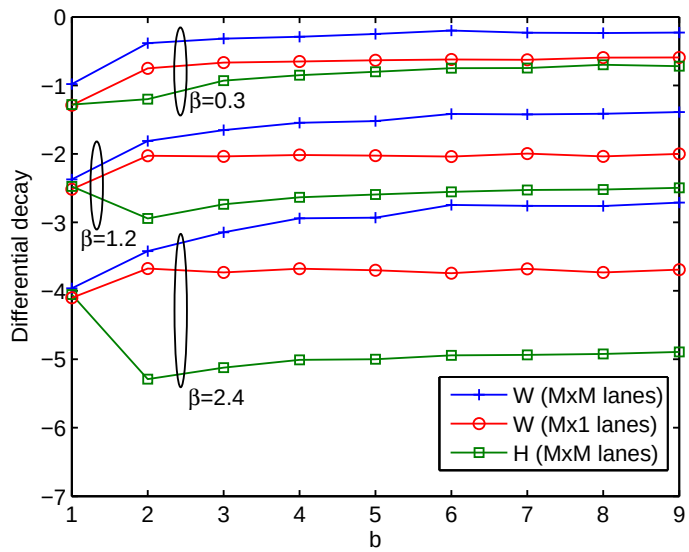


Figure 5.5: Differential decay of $\mathbf{H}$ and $\mathbf{W}$. $\beta \in \{0.3, 1.2, 2.4\}$. The decay is defined as $\mathbb{E}[\lg(|h_{k,k+b}/h_{k,k+b-1}|^2)]$ for $\mathbf{H}$ and $\mathbb{E}[\lg(|w_{k+b,k}/w_{k+b-1,k}|^2)]$ for $\mathbf{W}$.

SIR when using $\mathbf{W}_p$

Although off-diagonal entries of $\mathbf{W}_p$ decay more slowly than $\mathbf{H}$, the main diagonal entries in $\mathbf{W}_p$ are still dominating and can be expressed as:

$$\begin{aligned} |\frac{1}{w_{k,k}}| &= \left| h_{k,k} + \sum_{b \neq k} \frac{h_{k,b}(-1)^{k+b} M_{k,b}}{M_{k,k}} \right| \\ &\approx \left| h_{k,k} - \sum_{l=\pm 1} \frac{h_{k,k+l} M_{k,k+l}}{M_{k,k}} + \sum_{i=\pm M} \sum_{l=\pm 1} \frac{(-1)^{i+l} h_{k,k+i+l} M_{k,k+i+l}}{M_{k,k}} \right| \end{aligned} \tag{5.15}$$

where the approximation follows from the tri-band tri-diagonal dominance. When directional antennas are used, we are able to boost the dominance of the main diagonal and then simplify $|w_{k,k}|$ as $1/|h_{k,k}|$. In such cases, the residual interference is still governed by the dominating entries in $\mathbf{A} = \mathbf{D}_p \mathbf{W}_p$, each being generated when $w_{k+b,k+b}$ meets $h_{k,k+b}$ in $\mathbf{D}_p$. As revealed in (5.7), we consider the $8p$-4 major interfering antennas and denote each resulting interference term by $r_{e,i}$ ($1 \leq e \leq 4$, $-p+1 \leq i \leq p-1$), where e is the index of the direction of the interfering antenna (corre-

sponding to east, west, north and south, respectively). Rewrite (5.5) as:

$$I_k^{(D)}/P \approx \mathbb{E}\left[\lg\left(\sum_{e=1}^{4}\sum_{i=-p+1}^{p-1} r_{e,i}\right)\right]. \tag{5.16}$$

The evaluation of (5.16) depends on the distribution of $h_{k,k}$. For $h_{k,k}$ subject to ZMCSCG, $r_{e,i}$ is subject to a ratio distribution, i.e., $r_{e,i} = \frac{\alpha^{2\sqrt{p^2+i^2}}}{p^2+i^2}|\frac{\delta_{e,i}f_{e,i}}{g_{e,i}}|^2$, where $f_{e,i}, g_{e,i} \sim \mathcal{CN}(0,1)$ and $\delta_{e,i} \sim \mathcal{U}(-0.5, 0.5)$. Since each of $8p-4$ major interfering terms involves a different entry in $\mathbf{D}_p$ and a different main diagonal entry in $\mathbf{W}_p$, and we have simplified $|w_{k,k}|$ as $1/|h_{k,k}|$, the independence among $f_{e,i}$, $g_{e,i}$ and $\delta_{e,i}$ can be established.

As ratio distributions have infinite variance, we set $r_{e,i}$ by $\frac{\alpha^{2p}}{p^2}|\frac{\delta_{e,i}f_{e,i}}{g_{e,i}}|^2$ to approximate a lower bound of $\mathrm{SIR_{dw}}$. For $h_{k,k}$ without fading (due to a strong Rician K-factor and the use of directional antennas), $r_{e,i}$ is subject to a product distribution that has a finite variance, i.e., $r_{e,i} = \frac{\alpha^{2\sqrt{p^2+i^2}}}{p^2+i^2}|\delta_{e,i}f_{e,i}|^2$, which generates an upper bound of $\mathrm{SIR_{dw}}$. Both bounds are given as follows:

$$-10\mathbb{E}\left[\lg\left(\sum_{e=1}^{e=4}\sum_{i=-p+1}^{p-1}\frac{\alpha^{2p}}{p^2}\left|\frac{\delta_{e,i}f_{e,i}}{g_{e,i}}\right|^2\right)\right] \le \mathrm{SIR_{dw}} \le -10\mathbb{E}\left[\lg\left(\sum_{e=1}^{e=4}\sum_{i=-p+1}^{p-1}\frac{\alpha^{2\sqrt{p^2+i^2}}}{p^2+i^2}|\delta_{e,i}f_{e,i}|^2\right)\right]. \tag{5.17}$$

In evaluating the upper bound, the Central Limit Theorem (CLT) can be applied to model $\sum_{e=1}^{4}\sum_{i=-p+1}^{p-1} r_{e,i}$ as a real Gaussian random variable. However, its mean and variance have to be evaluated numerically because the close-form of $\sum_{i=-p+1}^{p-1}\frac{\alpha^{2\sqrt{p^2+i^2}}}{p^2+i^2}$ is unknown. The expectation of the logarithm of positive samples of a real Gaussian random variable also requires a numerical evaluation. To avoid numerical evaluation and gain some insights on how SIR varies with p, we propose an approximation of (5.16) by using the median of $\delta_{e,i}$ to approximate $r_{e,i}$ as $\frac{\alpha^{2p}}{p^2}|\frac{1}{4}\cdot\frac{f_{e,i}}{g_{e,i}}|^2$ such that $\delta_{e,i}$ is decoupled from $f_{e,i}/g_{e,i}$. The decoupling is motivated by the infinite variance of $f_{e,i}/g_{e,i}$ versus the finite variance of $\delta_{e,i}$; another reason of decoupling is that the cumulative distribution function (CDF) of $X_{e,i} \equiv |\delta_{e,i}f_{e,i}/g_{e,i}|^2$ can be shown as $F_{X_{e,i}}(x) = 2\sqrt{x}\tan^{-1}(\sqrt{0.25/x})$,

which complicates subsequent derivations. We now rewrite (5.16) as:

$$I_k^{(D)}/P \approx \mathbb{E}\left[\lg\left(\frac{\alpha^{2p}}{16p^2}\sum_{e=1}^{e=4}\sum_{i=-p+1}^{p-1}\left|\frac{f_{e,i}}{g_{e,i}}\right|^2\right)\right] \approx \mathbb{E}\left[\lg\left(\frac{\alpha^{2p}}{16p^2}\max_{e,i}\left|\frac{f_{e,i}}{g_{e,i}}\right|^2\right)\right] \tag{5.18}$$

where the second approximation is from $\ln(e^a+e^b)=\ln[e^a(1+e^{b-a})]=a+\ln(1+e^{b-a})\approx a$ $(0<b<a)$. Such an approximation is reasonable because of the infinite variance of $|\frac{f_{e,i}}{g_{e,i}}|$. By showing that $\mathbb{E}\{\ln[\max_{1\le i\le N}(|f_i/g_i|^2)]\}$ is the (N-1)-th harmonic number for $N>1$ (See Appendix F for details), we estimate $\mathrm{SIR_{dw}}$ as:

$$\mathrm{SIR_{dw}} \approx -10\lg\left(\frac{\alpha^{2p}}{16p^2}\right) - 10\lg e\sum_{b=1}^{8p-5}\left(\frac{1}{b}\right) = -20p\lg\alpha + 10\lg\left(\frac{16p^2}{8p-5}\right) - 5\lg e. \tag{5.19}$$

Eq. (5.19) suggests that in the case of stochastic $\mathbf{H}$, $\mathrm{SIR_{dw}}$ still increases by $-20\lg\alpha+10\lg(1+1/p)$ dB when the width increases from p to p+1 ($1<p\ll M$).

SIR when using $\mathbf{S}_p$

There are two factors increasing the SIR loss due to $\mathbf{S}_p$, when compared with Chapter 4. One is a much smaller outband drop in $\mathbf{W}_p$ due to the tri-band tri-diagonal dominance structure. Another factor is that entries in $\mathbf{W}_p$ cannot decay as fast as those in $\mathbf{H}$; therefore, banding $\mathbf{W}_p$ would incur a larger loss of information, or equivalently, the contribution from $\mathbf{E}_p\mathbf{H}_p$ will be much larger than from $\mathbf{D}_p\mathbf{W}_p$. In this section we show that with the use of directional antennas, the main diagonal dominance can be re-established and the SIR loss from using $\mathbf{S}_p$ can be largely reduced.

In practice, the under-floor or under-seat mounted antennas are reversed downtilt antennas that are omni-directional in the horizontal plane and directional in the vertical plane with typically 5 dB gain variation between 0 and 180 degree, as shown in Fig. 5.1. We use a simplified directional antenna model by assuming the channel gains of home antenna-user links are multiplied by $10^{G_t/20}$ while other links remain the same, where G_t is the transmitter antenna gain in dB. Consequently, the transmission power should be reduced by G_t to comply with radio regulations. The combined

effect is all entries in $\mathbf{H}$ being suppressed by G_t except those in the main diagonal, while the transmission power constraint remains the same.

When $p = 1$, using directional antennas has no effect on SIR_{sw}. When $p > 1$, the decrement in $|\mathbf{J}|$ due to G_t increases with b (the size of $\mathbf{J}$) in that approximately, we have $|\mathbf{J}|_{G_t=x} = |\mathbf{J}|_{G_t=0}/10^{bG_t/20}$. This indicates the effect of G_t on suppressing $\mathbf{W}_p$ off-diagonal growth is amplified by width p. Since SIR_{sw} is governed by $\mathbf{B} = \mathbf{E}_p\mathbf{H}_p$, a system operating at a larger p would benefit more from G_t, e.g., when G_t increases from 0 to 1 dB, SIR_{sw} would roughly increase by pG_t. But when G_t increases, the x_0 term in $|\mathbf{J}|$ is becoming more significant, causing a tapering SIR increment when G_t becomes higher.

Compared with SIR_{sw}, the increment of SIR_{dw} from a higher G_t is expected to be much less sensitive to changes in p because the governing residual interference term, $\mathbf{A} = \mathbf{D}_p\mathbf{W}_p$, is more sensitive to the main diagonal of $\mathbf{W}_p$ rather than its off-diagonal entries.

5.5 Numerical Results

We evaluate the signal-to-interference-noise-ratio (SINR) in a $400 \times 400m^2$ public venue with a uniformly deployed 20×20 antenna grid, i.e., N=400 and M=20. The height of all users is set at h_d=0.7 meters. We compare the proposed system, denoted by SDU ("D" stands for distributed MIMO and "U" for under-floor mounted), with two other forms of massive MIMO systems: SDC and SAC ("A" for array-MIMO and "C" for ceiling-mounted). SDC is a distributed MIMO system with its antennas evenly spread on the ceiling to benefit from the line-of-sight propagation to all users; SAC is a massive array-MIMO system with its antennas mounted in the center of the ceiling. The ceiling height, h_c, is set at 14 meters such that the edge-center distance ratio, $200/h_c$, is the same as the one in SDU, given by $10/h_d$. The per-antenna transmission power constraint of SDU is set at 1 mW/MHz. For a fair comparison, the per-antenna power constraint of SAC is raised by 26 dB such that its edge SNR is equal to the one in SDU; the per-antenna transmission power constraint of SDC is raised by the same amount.

In this section we consider two schemes. Scheme "Dense" uses $\mathbf{W}_p$ for precoding and WF-

Table 5.5: Simulation parameters of SDU, SDC and SAC. SDU: under-floor-mounted distributed antennas; SDC: ceiling-mounted distributed antennas; SAC: ceiling-mounted array antennas.

Parameters	**SDU**	**SDC**	**SAC**
ZFBF	Dense and sparse schemes.	Dense scheme.	Conventional ZFBF.
D (inter-antenna distance)	20 meters	20 meters	0.06 meter
G_t (transmitter antenna gain)	5 dB	0 dB	0 dB
η (path loss exponent)	2	2	2
β (linear attenuation factor)	0.3 dB/meter	not present	not present
κ (Rician K-factor in linear scale)	6 for home antenna-user links; 0 otherwise.	10 for all links	10 for all links

SPC for power allocation, subject to the sum power constraint NP. Scheme "Sparse" uses $\mathbf{S}_p$ for precoding and WF-SPCr for power allocation, subject to per-antenna power constraint P. For SDU, we present both schemes; SDC employs the dense scheme; SAC employs conventional ZFBF, which corresponds to the dense scheme at the maximal p value. The path loss exponent is set at η=2 as it is the case when the venue is outdoor, or indoor and occupied [154]. The range of β is set between 0.1 and 0.3 dB/meter. Common parameters are listed in Table 5.4 and differential parameters in Table 5.5. Simulation results are averaged over 10 independent user-drops, each having 100 independent fading samples. Channel matrices are wrapped around to remove edge effects; therefore, the maximal value of p is 11.

To show the computation cost reduction of SDU, we denote *per-user spectral efficiency* by η_s (approximated by SINR/3) and define *energy efficiency* as

$$\eta_e \triangleq \frac{\text{System throughput}}{\text{Total power consumption}} = \frac{B \times N \times \eta_s}{P_{\text{RoF}} + P_{\text{BB}}}, \tag{5.20}$$

where B is the channel bandwidth; P_{RoF} and P_{BB} represent the power consumption contributed by radio over fiber (RoF) components and baseband processing, respectively. We consider the same 20-MHz frequency division duplex system as in Section 4.6 and also assume $N_s = 7$, which was

explained in (4.32). After calculating P_{RoF} and P_{BB} by (4.30) and (4.31), we examine the tradeoff between energy and spectral efficiency in Fig. 5.6. One can notice that p=6 is a good balance point for both dense and sparse schemes in SDU, where the most of inter-stream interference had been removed and the energy efficiency is still considerably higher than when p reaches its maximum. We now use p=6 as the reference point for SDU. Although the spectral efficiency provided by SDU sparse scheme is 6% lower than of SAC and 21% lower than of SDC, the energy efficiency of SDU is 19 times higher than SAC and 18 times higher than SDC. While offering the substantial energy efficiency improvement, the sparse scheme in SDU only incurs a small loss in spectral efficiency when compared with the dense scheme.

Fig. 5.7 shows the power consumption contributed by RoF and beamforming in both dense and sparse schemes in SDU, where the beamforming contribution is translated from (4.32). We observe that even at small p values, beamforming in both versions is consuming much more energies than RoF. Compared with what we observed in Fig. 4.9, the power consumption dominance of beamforming over RoF in Fig. 5.7 is more significant because the number of antennas here is 400 while only 50 antennas were considered in Fig. 4.9. These observations indicate the necessity of developing a low-complexity beamforming scheme in a massively distributed MIMO system. We also notice that at small p values, the beamforming power consumption in the sparse scheme is substantially lower than in the dense scheme, which explains the energy efficiency gain when moving from the dense to the sparse scheme, as we have observed in Fig. 5.6. Note that the sparse scheme is subject to a smaller sum power constraint than the dense scheme, thus causing a smaller RoF power consumption.

In Fig. 5.8, we observe that SINR in SDU steadily increases with p while SDC requires a full matrix inversion due to its line-of-sight propagation in all links, indicating the importance of exponentially decayed off-diagonals. Although SAC and SDC generate a slightly higher SINR than SDU, the expenses include a 26 dB higher transmission power and the inability to control the tradeoff between the computation cost and the throughput. The use of "WF-SPCr" power allocation scheme causes at most 5 dB penalty in SDU, suggesting the SINR improvement from

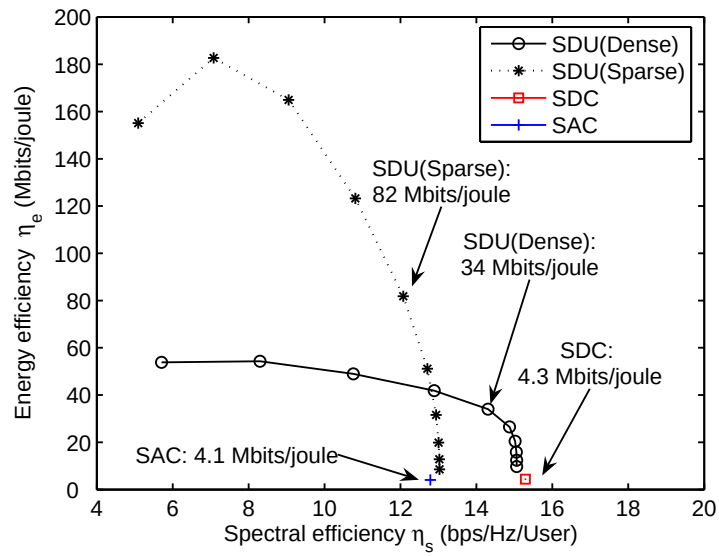


Figure 5.6: Energy efficiency η_e vs. Spectral efficiency η_s. SDU: under-floor-mounted distributed antennas; SDC: ceiling-mounted distributed antennas; SAC: ceiling-mounted array antennas. The width p increases from 2 to 11 for SDU; p=11 for SDC. η=2. N_s=7. Parameters for SDU: $\{\beta$=0.3, κ=6, G_t=5 dB$\}$.

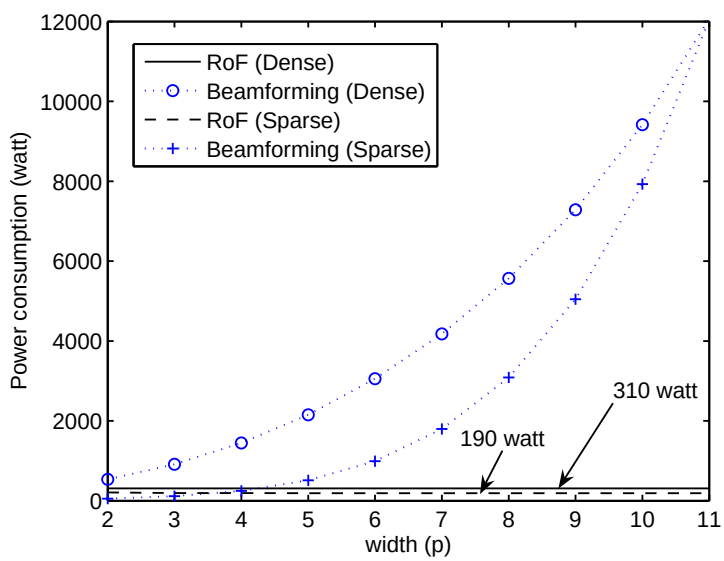


Figure 5.7: SDU: Power consumption contributed by radio-over-fiber and beamforming. RoF: radio-over-fiber. η=2. β=0.3. κ=6. G_t=5 dB. N_s=7.

any other schemes subject to a per-antenna power constraint is at most 5 dB and it shrinks with p.

To compare user fairness, we plot the CDF of SINR in Fig. 5.9, where SDC demonstrates

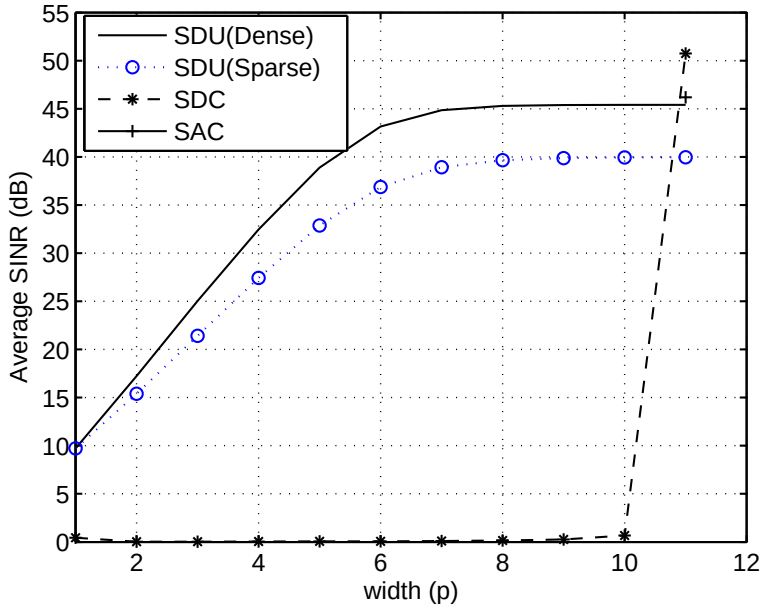


Figure 5.8: Average SINR vs. Width in SDU, SDC and SAC. 10 user-drops. 100 fading samples per user-drop.

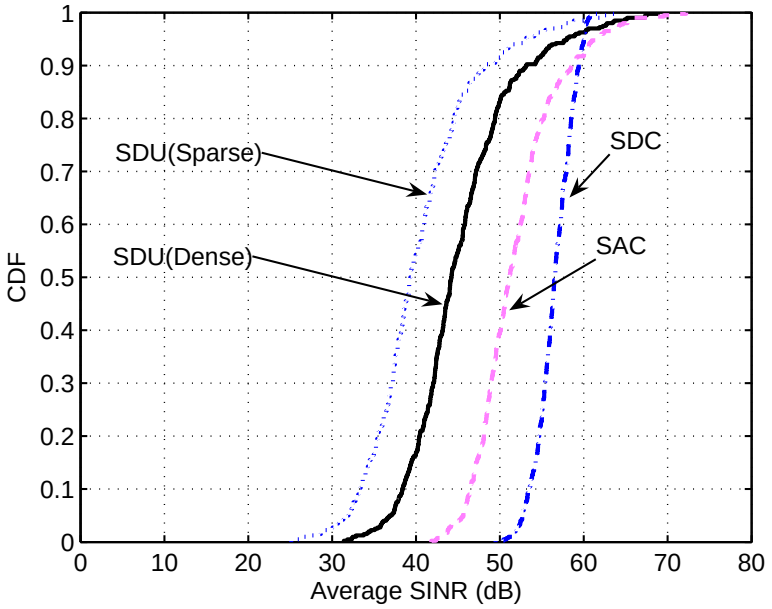


Figure 5.9: CDF of SINR in SDU, SDC and SAC. 1 user-drop. 1000 fading samples. p=11 for SDU and SDC.

the best user fairness, followed by SAC and SDU. Since SDC can be viewed as an SAC with antennas spread wider, its user fairness advantage is expected due to the better conditioned channel

matrices. The minor fairness disadvantage of SDU is attributed to the non-line-of-sight propagation caused by under-floor antenna mounting, which is the price we paid for the significantly lowered computation cost and in practice, can be compensated by user-scheduling. At other choices of p, SDU shows similar user fairness and the curves are omitted here to avoid cluttering. The CDF of antenna transmission power in SDU is presented in Fig. 5.10, where we observe that the "WF-SPCr" power allocation scheme effectively compresses the transmission power to comply with the per-antenna power constraint.

The bounds given by (5.17) are shown in Fig. 5.11, in comparison with simulated SINR when dense precoding is used. The bounds are tighter at β=0.5 than at β=0.1 because a more dominant main diagonal makes the approximations made in deriving (5.17) more accurate. The average SINRs provided by three flavors of the proposed ZFBF scheme are shown in Fig. 5.12 under different β values. The SINRs of dense and sparse precoding increase with p at a rate that is proportional to β until the system becomes noise-limited. One can observe that the $\mathrm{SIR}_{\mathrm{dw}}$ estimated by (5.19) tracks the actual SINR well.

SDU converts to full matrix inversion at p=11. We observe that when $\mathbf{S}_p$ is used, the "WF-

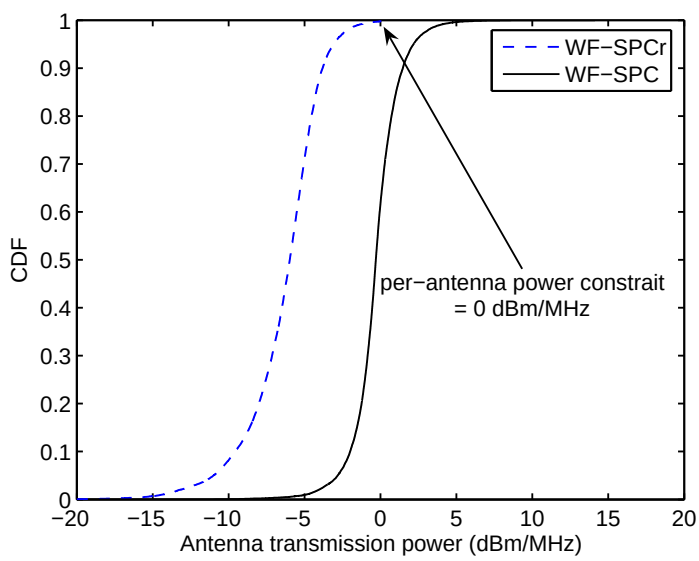


Figure 5.10: SDU: CDF of antenna transmission power.

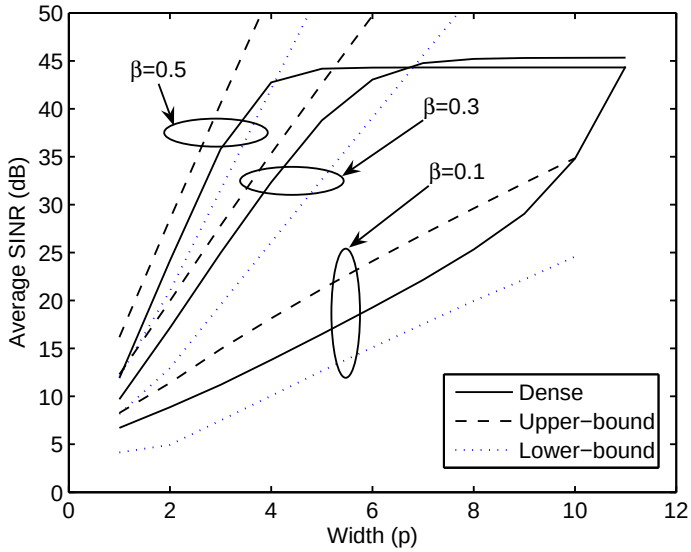


Figure 5.11: SDU: Bounds of average SINR. η=2. $\beta \in \{0.1, 0.3, 0.5\}$. κ=6. G_t=5 dB.

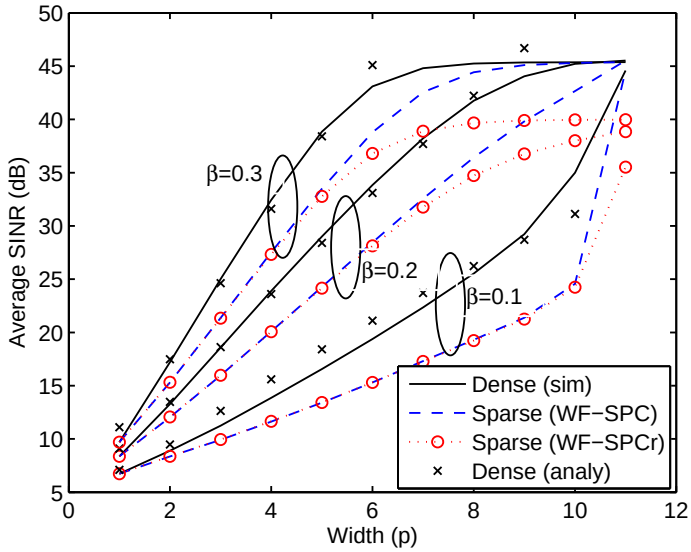


Figure 5.12: SDU: Average SINR vs. Width. η=2. $\beta \in \{0.1, 0.2, 0.3\}$. κ=6. G_t=5 dB.

SPCr" power allocation scheme only incurs minor throughput losses at a high SINR region, making it a better choice than the "WF-SPC" scheme due to its simplicity of enforcing the per-antenna power constraint. We also observe that SINR sharply decreases when β decreases to 0.1, suggesting that polynomial decay alone is not enough to offer the opportunity of reducing the computation cost of matrix inversion. Instead, exponential decay is needed, and preferably at a large attenuation

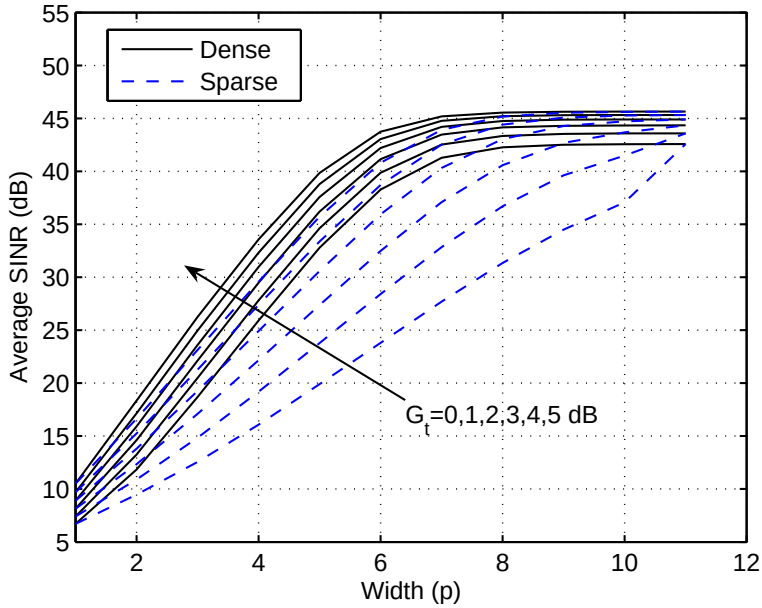
rate, to accelerate the off-diagonal decays in $\mathbf{H}$.

The effect of G_t is examined in Fig. 5.13, where β=0.3 and κ=6. When G_t increases from 0 to 5 dB, SINRs in both dense and sparse schemes increase, yet at a rate that decreases with G_t. At p=6, SINR in the sparse scheme increases by 4.6 dB when G_t increases from 0 to 1 dB, but only 4 and 3.5 dB when G_t further increases to 2 and 3 dB, respectively. The diminishing benefit of G_t on the SINR increment in the sparse scheme can be better observed in Fig. 5.13b, where the five curves correspond to the SINR increment by increasing G_t from 0 dB to 1, 2, 3, 4 and 5 dB, respectively. We also observe in Fig. 5.13b that at all G_t values, SINR in the sparse scheme linearly increases with p for $p < 7$ whereas SINR in the dense scheme remains relatively constant when p increases.

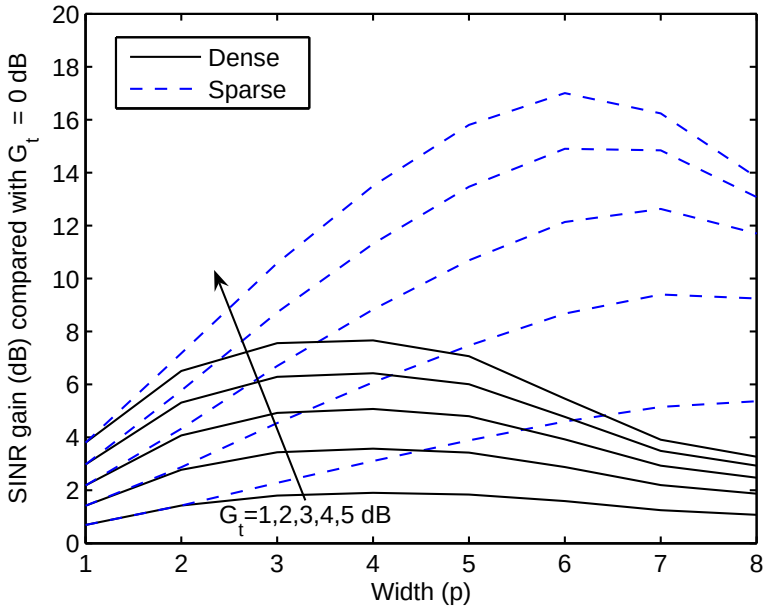
In Fig. 5.14, we compare the performance of the proposed dense scheme with Krylov subspace method [144, 146], where we set L=6 and calculate the coefficient set $\{w_l\}$ for each individual $\mathbf{H}_p$. Two different network sizes are considered: 120×120 and $400 \times 400m^2$, corresponding to 36×36 and 400×400 channel matrices, respectively. At β=0.1 and G_t=0 dB, Krylov method outperforms the dense precoding scheme. At a higher β or a larger G_t, the proposed dense precoding scheme outperforms Krylov method as the latter encounters extremely ill-conditioned matrices while solving $\{w_l\}$. The curves at higher β are omitted to avoid cluttering in the figure. The SINR of Krylov method drops for 400×400 channel matrices. The reason could be related to accumulating errors while multiplying matrices to obtain the subspace bases. We remark that although the above comparison favors the dense scheme, one should be aware that Krylov method is a general matrix inversion solver that does not depend on particular matrix structures.

5.6 Summary

Large public venues can be deployed with a distributed massive MIMO system with $M \times M$ antennas mounted under-floor, which makes the antennas closer to the users and provides a very high system capacity. Under such network configurations, the channel matrix $\mathbf{H}$ can be modeled as a multi-banded matrix with off-diagonal entries decaying both exponentially due to heavy human



(a) Average SINR vs. Width.



(b) Average SINR gain when compared with G_t=0 dB.

Figure 5.13: SDU: G_t effect on SINRs in dense and sparse scheme. η=2. β=0.3. κ=6. $G_t \in \{0, 1, \cdots, 5\}$ dB.

penetration loss and polynomially due to free space propagation loss. The composite decays motivate us to devise a multi-banded matrix inversion algorithm that, by keeping the most significant

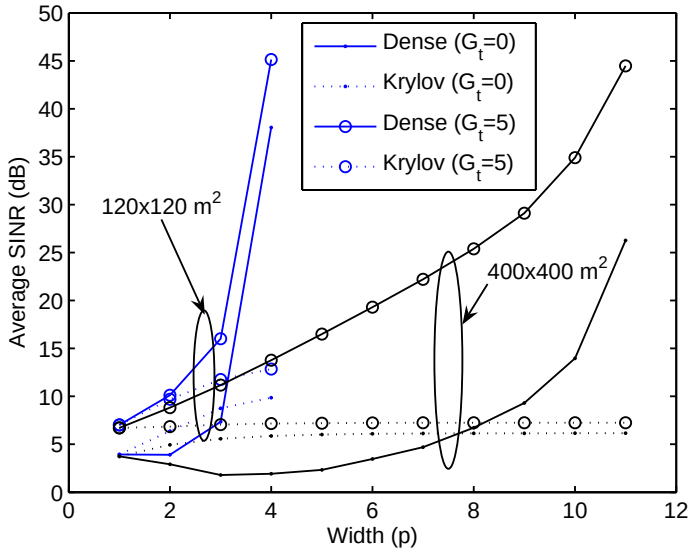


Figure 5.14: SDU: Dense scheme vs. Krylov subspace method (L=6). η=2. β=0.1. κ=6. $G_t \in \{0,5\}$ dB.

entries in **H** and the precoding matrix **W**, substantially reduces the computation cost of ZFBF while only incurring a small throughput loss. Note that the very-high system capacity is only possible when per-antenna power constraint remains fixed while the antenna density increases, in which case the signal strength at users can increase with the antenna density, thus increasing the system capacity.

The proposed algorithm includes dense and sparse precoding versions, providing quadratic and linear computation cost in M^2, respectively. We have introduced a parameter p to control the sparsity of **H** and **W** and thus achieve the tradeoff between computation cost and system throughput. We have shown both analytically and numerically how the SIR linearly increases with p in dense precoding. In sparse precoding, we have demonstrated the necessity of using directional antennas by both the analysis and simulations. When the directional antenna gain increases, the resulting SIR increment in sparse precoding linearly increases with p, while the SIR of dense precoding is much less sensitive to p.

Chapter 6

Conclusions and Future Work

In this chapter, we conclude this thesis by summarizing the main findings and point out some potential research topics.

6.1 Summary of Work Accomplished

We have proposed a novel wireless access architecture, Fiber-connected Massively Distributed Antennas (FMDA), in Chapter 1. In Chapter 2, we have demonstrated its capability to mitigate interference that often arises in wireless local area networks (WLANs) and shown its flexibility to handle spatially non-uniformly distributed traffic. The energy-efficiency of the proposed FMDA has been compared with standalone femtocells in Chapter 3, where an antenna scheduling scheme has been proposed to simultaneously improve spectral and energy efficiency in FMDA systems. When FMDA are deployed in office buildings and large public venues, we have proposed two low-complexity zero-forcing beamforming (ZFBF) schemes that significantly reduce the baseband processing complexity in downlink and therefore substantially improve the energy efficiency.

- In Chapter 2, we have applied FMDA in WLANs to form a cognitive WLAN over fiber system. The system can provide a cost-effective and efficient architecture for devices to equally share the industrial, scientific, and medical band by taking advantage of cognitive radio capabilities. In the formed system, we have proposed two methods to reduce collisions

among stations, with multiple independent channels operating at each antenna, and transmitter and receiver diversity through cooperation of adjacent antennas. Multi-channel-operation method is enabled by wide-band optical fibers and diversity method is enabled by distributed antennas in the cognitive WLAN over fiber architecture. Extensive simulations have shown substantial improvements in Transmission Control Protocol throughput and packet error rate reduction of constant-bit-rate traffic streams, especially under dynamic traffic conditions.

- In Chapter 3, we have proposed fiber-connected femto base stations based on our proposed FMDA architecture. After establishing a power consumption model of the proposed system, we have developed an optimization tool in a femtocell cluster based on coordinated multi-point transmissions to maximize energy efficiency by adjusting the number of transmission antennas and controlling transmission power in ZFBF. Based on the analysis results, we have proposed an antenna scheduling scheme to simultaneously improve spectral and energy efficiency. Compared with standalone femtocells, the proposed scheme has been shown in a typical office building to increase energy efficiency by 64%~160% and spectral efficiency by 2%~36%. The exact gain depends on network configurations and transmission power levels.

- In Chapter 4, we have considered a FMDA system that covers an office building and discovered that channel matrices in such environment can be modeled with exponentially and polynomially decayed off-diagonals. Based on this analytic channel model, we have proposed an energy-efficient low-complexity ZFBF scheme, which reduces the computation cost of channel state information collection, matrix inversion and precoding by banding the channel matrix. We have proposed two versions of the scheme. The dense precoding version provides quadratic computation cost and the sparse precoding version provides linear computation cost, relative to the number of antennas. The resulting signal-to-interference-plus-noise ratios (SINRs) in both versions have been analyzed for random channel matrices. Compared with beamforming via full matrix inversion, our analysis and numerical evaluations have shown that both versions incur negligible loss in SINR, while offering 45%~79%

gain in energy efficiency at lower transmission power levels. The sparse precoding version provides 22%~59% higher downlink energy efficiency than the dense version. We have also introduced a parameter p to control the sparsity of the channel matrix and the precoding matrix, and thus achieved a flexible control over the tradeoff between the system capacity and the energy efficiency.

- In Chapter 5, we have considered a FMDA system that covers large public venues. As the absence of wall penetrations invalidates the channel model developed in Chapter 4 and would incur a large throughput loss were our previously proposed scheme applied, we have employed the under-floor antenna mounting strategy to increase the propagation loss among antennas with the help of the heavy human body penetration loss. With this strategy, we have developed a new channel model and proposed a multi-banded matrix inversion scheme that substantially reduces the computation cost of ZFBF while incurring a negligible throughput loss. Compared with a massive array multiple-input multiple-output (MIMO) system located in the center of the venue, our proposed scheme has been shown to provide 19 times higher energy efficiency while only incurring 6% spectral efficiency loss. Our proposed scheme includes dense and sparse precoding versions, providing quadratic and linear computation cost, respectively, relative to the number of antennas. By introducing a parameter p to control the sparsity of the channel matrix, we have presented analysis and numerical evaluations to show that the signal-to-interference-ratio increases linearly with p in dense precoding. In sparse precoding, we have demonstrated the necessity of using directional antennas by both analysis and simulations. When the directional antenna gain increases, the resulting signal-to-interference-ratio increment in sparse precoding has been shown to increase linearly with p, while the signal-to-interference-ratio of dense precoding is much less sensitive to changes in p. We have also discovered that although massive array-MIMO can deliver a very high system capacity, its co-located antennas disallow a low-complexity matrix inversion.

6.2 Future Work

We list below potential research topics in how to apply the proposed FMDA architecture to manage interference and improve energy efficiency in access networks:

Interference management of time division duplex systems in Long-Term Evolution Long-Term Evolution (LTE) time division duplex (TDD) systems has many advantages over LTE frequency division duplex (FDD) systems because: (1) FDD spectrums have to allocated in pairs while TDD spectrums can be allocated one chunk at a time; (2) TDD supports asymmetric traffic often seen in access networks; (3) Reciprocated channels in TDD make it easier to have channel state information at the transmitter. However, the use of the same frequency in uplink and downlink complicates the interference scenarios in TDD systems and calls for more delicate control over timing and transmission power at both users and base stations. This topic, termed as "traffic adaptation" [59, 60], is only at the starting stage in the Third Generation Partnership Project. Considering that the centralized processing system (CPS) in a FMDA system is able to see the spectrum usage in the entire network and each antenna can be arbitrarily configured by the CPS in transmitting or receiving state, we expect to see that a rich set of interference management problems in LTE TDD traffic adaptation can be easier solved when the LTE system is built on our proposed FMDA architecture.

It is also possible to extend the idea to the area of WLAN. An increasing number of devices are equipped with an 802.11ac network adaptor to support giga-bit-per-second throughput, which is often needed in ad-hoc multimedia streaming in practice. However, such ad-hoc applications might interfere with infrastructure-based WLAN operations. When the infrastructure-based WLAN is built on FDMA, the cognitive function of the CPS can be exploited to detect the ad-hoc transmission and accordingly adjust the transmission strategies of the antennas being affected.

Joint design of channel estimation and approximated channel matrix inversion To achieve an overall lower complexity and putting engineering efforts in the most important and effective areas that improve energy efficiency, a joint design of interpolation-based channel estimation and a low-

complexity beamforming is a future direction. In this area, a potential research topic is to reduce the number of matrix inversion based on frequency and time coherence in channel state information. Previous studies have revealed linear beamforming can be interpolated in both frequency [155–157] and time [158]. Another potential topic is to investigate how the pilot contamination affects the proposed low-complexity ZFBF scheme.

In fact, a system designer can consider the overall system power consumption as a cross-layer concept, which involves many components such as pilot design, channel estimation, carrier and time synchronization, and forward error correction. Each component of practical systems often has limits that incur a SINR loss. In such cases, a perfect solution in signal processing is often not needed. The problem of improving energy efficiency then becomes a system-wide optimization problem.

Extension of the proposed low-complexity zero-forcing beamforming scheme We considered ZFBF in Chapter 4 and 5. When minimum mean-square-error beamforming is used, two matrix multiplications occur in calculating the precoding matrix. Since the width of the product of two banded matrices is the sum of their widths, the multiplications would triple the width of the precoding matrix and therefore largely increase the computation cost. How to extend the proposed scheme to minimum mean-square-error beamforming is a potential research topic.

We have only considered downlink due to its dominance in access networks. It is worth studying the extension of the proposed low-complexity scheme to uplink, where the sparse scheme is more attractive in increasing energy efficiency because matrix inversion happens more frequently in uplink zero-forcing equalization, especially in fast-fading channels. In such cases, the equalizer needs to track the channel variations by extracting the reference symbols that are regularly inserted by users.

It is also interesting to study how the presence of shadowing affects the low-complexity ZFBF schemes proposed in Chapter 4 and 5. Once the shadowing is introduced, the proposed banded and multi-banded channel matrix structures would not be good approximations due to the largely

increased variance of channel entries. The transmitter thus needs to track the most significant entries and either apply a general sparse matrix inversion solver or permute the matrix to put significant entries closer to the main diagonal. However, more entries in **H** need to be collected to make an informative decision, thus increasing the overhead of collecting channel state information.

Reduce system cost by choosing different optical backbone topologies We have proposed to connect all antennas to the CPS via point-to-multi-point optical links; therefore, each antenna consumes one strand of fiber. This choice of optical backbone largely increases the system cost in places where deploying optical cables are expensive. An alternate optical backbone is a time-division multiplexing optical ring, as used in current stages of cloud-based radio access networks, where each remote radio head (RRH) takes turns to communicate with the baseband unit. Such topology is already supported in the latest common public radio interface (CPRI) specification. Without a large buffer at RRHs, a full-scale downlink beamforming or uplink equalization would not be possible since at any time, only one RRH per ring is transmitting/receiving baseband signal samples to/from the baseband unit. Compared with a point-to-multi-point topology, a ring topology will leads to different antenna cooperation strategies. It would be also interesting to consider the impact of bit-interleaved passive optical network [159], a novel passive optical network architecture recently proposed by Alcatel-Lucent to significantly reduce the sampling rate of optical network units, thus reducing their component cost and power consumption.

Flexible configuration of RRH in cloud-based radio access networks To support a CPRI interface, commercially available chips consume 0.5 watt in serialization/deserialization [160] and 1 watt at the optical transceiver [161], which is significantly higher than what is needed to support a radio-over-fiber interface. If the design of RRH is flexible enough to support both radio-over-fiber and CPRI interface, the baseband unit can remotely configure RRHs to operate in either interface. When the traffic demand is low, radio-over-fiber signaling is advantageous because we avoided the more power-consuming signal path, which involves high-speed analog-digital/digital-analog

converters and digital transceivers. When the traffic demand is high, the system operator may want to configure the RRH as CPRI-mode to ensure a very-high order of signal modulation can be used to provide very-high-rate wireless links. Further, the CPRI circuits in a RRH should be able to be remotely turned off when there is no active user located nearby this RRH. Since a RRH is already equipped with a micro controller to support CPRI, adding the above flexibilities should not cause much increase in the RRH cost.

Bibliography

[1] A. Saleh, A. Rustako, and R. Roman, "Distributed antennas for indoor radio communications," *IEEE Trans. Commun.*, vol. 35, no. 12, pp. 1245–1251, 1987. → pages 2, 4, 97

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

[3] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE VTC 2001*, vol. 3, no. 53, 2001, pp. 1745–1749. → pages 2

[4] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing," in *Proc. IEEE ICC'08*, 2008, pp. 4033–4037. → pages 2, 4, 97

[5] *Coordinated multi-point operation for LTE physical layer aspects*, 3GPP TR36.819 Std., Rev. 11.1.0, December 2011. → pages 2

[6] 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," *IEEE Commun. Mag.*, pp. 148–155, February 2012. → pages 2

[7] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *EURASIP Journal on Wireless Communications and Networking*, no. 2, pp. 222–235, 2004. → pages 2

[8] B. Song, R. L. Cruz, and B. D. Rao, "Downlink optimization of indoor wireless networks using multiple antenna systems," in *Proc. IEEE INFOCOM 2004*, vol. 4, 2004. → pages

[9] F. Roemer, M. Fuchs, and M. Haardt, "Distributed MIMO systems with spatial reuse for high-speed indoor mobile radio access," in *Proc. WWRF 2008*, 2008. → pages

[10] J. Zhang and J. G. Andrews, "Distributed antenna systems with randomness," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3636–3646, 2008. → pages

[11] A. Tolli, M. Codreanu, and M. Juntti, "Cooperative MIMO-OFDM cellular system with soft handover between distributed base station antennas," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, 2008. → pages

[12] ——, "Linear multiuser MIMO transceiver design with quality of service and per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 7 Part 1, pp. 3049–3055, 2008. → pages

[13] N. H. Dawod, R. Hafez, and I. Marsland, "Uplink multiuser MIMO-OFDM in widely spaced antenna arrays," in *2005 Canadian Conference on Electrical and Computer Engineering*, 2005, pp. 155–159. → pages

[14] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *Proc. ACSSC'07*, 2007, pp. 83–87. → pages 2, 4, 97

[15] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Downlink multicell processing with limited-backhaul capacity," *EURASIP Journal on Advances in Signal Processing*, 2009. → pages 2

[16] P. Marsch and G. Fettweis, "On multicell cooperative transmission in backhaul-constrained cellular systems," *Annals of Telecommunications*, vol. 63, pp. 253–269, 2008. → pages 2

[17] Q. Zhou, H. Zhang, and H. Dai, "Adaptive spatial multiplexing techniques for distributed MIMO systems," in *Proc. Conference on Information Sciences and Systems 2004*, pp. 674–677. → pages 2

[18] H. Dai, "Distributed versus co-located MIMO systems with correlated fading and shadowing," in *Proc. IEEE ICASSP 2006*, vol. 4, 2006. → pages

[19] Z. Ni and D. Li, "Effect of fading correlation on capacity of distributed MIMO," in *Proc. IEEE PIMRC 2004*, vol. 3, 2004. → pages

[20] L. Dai, "Distributed antenna system: Performance analysis in multi-user scenario," in *Proc. CISS 2008*, 2008, pp. 85–89. → pages

[21] E. Song, J. Park, J. Kim, S. Hwang, and W. Sung, "Performance of cooperative transmission schemes using distributed antennas," in *Proc. IEEE CCNC 2008*, 2008, pp. 41–45. → pages

[22] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 69–73, 2007. → pages

[23] J. Zhang and J. G. Andrews, "Cellular communication with randomly placed distributed antennas," in *Proc. IEEE GLOBECOM 2007*, pp. 1400–1404. → pages

[24] V. C. M. Leung and A. W. Y. Au, "A wireless local area network employing distributed radio bridges," *Wireless Networks*, vol. 2, no. 2, pp. 97–107, 1996. → pages 2, 24

[25] Y. Li and J. Zhang, "Hot-spot wireless access exploiting shadowing diversity of distributed antennas," in *Proc. IEEE International Symp. on Broadband Multimedia Systems and Broadcasting*, 2008, pp. 1–5. → pages 2

[26] T. Zhou, Y. Yang, S. J. Eggerling, Z. Zhong, and H. Sharif, "A novel distributed MIMO aware MAC protocol design with a Markovian framework for performance evaluation," in *Proc. IEEE MILCOM 2008*, 2008, pp. 1–6. → pages 2

[27] Cisco Systems. (2011, Nov) Wireless LAN design guide for high density client environments in higher education. → pages 3, 18, 98

[28] Ruckus Wireless. (2012) Deploying very high density WiFi: Design and configuration guide for stadiums (Ver 1.0). → pages 3, 98, 101

[29] Aruba Networks. (2010, Oct) High-density wireless networks for auditoriums: Validated reference design. → pages 3, 18, 98

[30] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, 2008. → pages 3

[31] T.-S. Chu and M. J. Gans, "Fiber optic microcellular radio," *IEEE Trans. Veh. Technol.*, vol. 40, no. 3, pp. 599–606, 1991. → pages 3, 20, 48

[32] M. Sauer, A. Kobyakov, and J. George, "Radio over fiber for picocellular network architectures," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 3301–3320, 2007. → pages 3, 20, 48

[33] A. Nkansah, A. Das, C. Lethien, J.-P. Vilcot, N. J. Gomes, I. J. Garcia, J. C. Batchelor, and D. Wake, "Simultaneous dual band transmission over multimode fiber-fed indoor wireless network," *IEEE Microw. Wireless Compon. Lett.*, vol. 16, no. 11, pp. 627–629, 2006. → pages 48

[34] T. Niiho, M. Nakaso, K. Masuda, H. Sasai, K. Utsumi, and M. Fuse, "Transmission performance of multichannel wireless LAN system based on radio-over-fiber techniques," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 2, pp. 980–989, 2006. → pages

[35] A. Das, A. Nkansah, N. J. Gomes, I. J. Garcia, J. C. Batchelor, and D. Wake, "Design of low-cost multimode fiber-fed indoor wireless networks," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 8, pp. 3426–3432, 2006. → pages 3, 20

[36] K. K. Leung, B. McNair, L. J. C. Jr, and J. H. Winters, "Outdoor IEEE 802.11 cellular networks: MAC protocol design and performance," in *Proc. IEEE ICC 2002*, vol. 1, New York, NY, USA, Apr. 2002, pp. 595–599. → pages 3, 21

[37] M. G. Larrodé, A. M. J. Koonen, and P. F. M. Smulders, "Impact of radio-over-fibre links on the wireless access protocols," in *Proc. NEFERTITI Workshop*, Brussels, Belgium, Jan. 2005. → pages

[38] A. Das, M. Mjeku, A. Nkansah, and N. J. Gomes, "Effects on IEEE 802.11 MAC throughput in wireless LAN over fiber systems," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 1–8, 2007. → pages 3, 21

[39] [Online]. Available: http://www.zinwave.com → pages 3, 21

[40] (2012, June) CommScope launches specialist 18-beam antenna for stadium use. CommScope. [Online]. Available: http://news.commscope.com/news/ → pages 3

[41] (2013, June) SOLiDs Underground Reputation. SOLiD Technologies. [Online]. Available: http://solid.com/blog/?tag=in-building-wireless → pages 3

[42] [Online]. Available: http://www.innerwireless.com → pages 3, 21

[43] [Online]. Available: http://www.powerwave.com/inbuilding.asp → pages 3, 21

[44] (2010, February) FORAX-HARC: High Antennas for Radio Communications. Syntonics. [Online]. Available: http://www.syntonicscorp.com/products/datasheets/ → pages 3

[45] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010. → pages 4, 69

[46] ——, "How much training is required for multiuser MIMO?" in *Proc. IEEE ACSSC'06*, 2006, pp. 359–363. → pages 4, 69, 97

[47] C. Shepard, H. Yu, N. Anand, E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. IEEE MobiCom'12*, 2012, pp. 53–64. → pages 4, 69, 97

[48] T. Alade, H. Zhu, and H. Osma, "Performance evaluation of in-building DAS for high data rate wireless transmission," in *Proc. IEEE WCNC'12*, 2012, pp. 759–764. → pages 4, 97

[49] L. Dai, S. Zhou, and Y. Yao, "Capacity analysis in CDMA distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2613–2620, Nov 2005. → pages 4, 97

[50] A. Ozgur, O. Leveque, and D. N. Tse, "Spatial degrees of freedom of large distributed MIMO systems and wireless ad hoc networks," *IEEE JSAC*, vol. 31, no. 2, pp. 202–214, 2013. → pages 4, 97

[51] M. Kiiski, "LTE-Advanced: The mainstream in mobile broadband evolution," in *Proc. IEEE Wireless Conference (EW) European'10*, 2010, pp. 983–988. → pages 5

[52] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects*, 3GPP TR 36.814 Std., Rev. Release 9, Ver. 9.0.0, March 2010. → pages 5

[53] *Evolved Universal Terrestrial Radio Access (E-UTRA); TDD Home eNode B (HeNB) Radio Frequency (RF) requirements analysis*, 3GPP TR 36.922 Std., Rev. V11.0.0. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36922.htm → pages 5, 6

[54] *Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements*, 3GPP TS 32.500 Std. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/32500.htm → pages 6

[55] *Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions*, 3GPP TR 36.902 Std. [Online]. Available: http://www.3gpp.org/ftp/specs/html-info/36902.htm → pages 6

[56] L. Lindbom, R. Love, S. Krishnamurthy, C. Yao, N. Miki, and V. Chandrasekhar, "Enhanced inter-cell interference coordination for heterogeneous networks in LTE-Advanced: A survey," *arxiv preprint arxiv:1112.1344*, 2011. → pages 6

[57] *Evolved Universal Terrestrial Radio Access (E-UTRA); FDD Home eNode B (HeNB) Radio Frequency (RF) requirements analysis*, 3GPP TR 36.921 Std. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36921.htm → pages 6

[58] M. Huang and W. Xu, "Macro-femto inter-cell interference mitigation for 3GPP LTE-A downlink," in *Proc. IEEE WCNC'12*, 2012, pp. 75–80. → pages 6

[59] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further enhancements to LTE Time Division Duplex (TDD) for Downlink-Uplink (DL-UL) interference management and traffic adaptation*, 3GPP TR 36.828 Std. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36828.htm → pages 6, 133

[60] Z. Shen, A. Khoryaev, E. Eriksson, and X. Pan, "Dynamic uplink-downlink configuration and interference management in TD-LTE," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 51–59, 2012. → pages 6, 133

[61] G. Koutitas and P. Demestichas, "A review of energy efficiency in telecommunication networks," *Telfor journal*, vol. 2, no. 1, pp. 2–7, 2010. → pages 6, 7

[62] Vodafone Sustainability Report 2012/13. Vodafone Group Plc. [Online]. Available: http://www.vodafone.com/content/sustainability/operating_responsibly → pages 6

[63] L. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. V. der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, 2010. → pages 6, 70

[64] A. J. Fehske, P. Marsch, and G. P. Fettweis, "Bit per Joule efficiency of cooperating base stations in cellular networks," in *Proc. 3rd Workshop on Green Commun.*, Miami, Florida, Dec. 2010. → pages 7, 51, 52, 60

[65] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *Proc. Future Network and Mobile Summit*, Florence, Italy, June 2010. → pages 7

[66] EARTH deliverables D3.1. [Online]. Available: https://bscw.ict-earth.eu/pub/bscw.cgi/31474 → pages 7, 52

[67] EARTH deliverables D4.3. [Online]. Available: https://bscw.ict-earth.eu/pub/bscw.cgi/31474 → pages 7

[68] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. J. Gonzalez, H. Klessig, I. Godor, M. Olsson, M. A. Imran, A. Ambrosy, and O. Blume, "Flexible power modeling of LTE base stations," in *Proc. IEEE WCNC'12*, 2012, pp. 2858–2862. → pages 7, 86, 87, 92

[69] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *arXiv preprint arXiv:1112.3810 (2011).* → pages 7, 69

[70] D. W. K. Ng and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Commun.*, no. 9, pp. 3292–3304, 2012. → pages 7, 69

[71] L. Saker, S. E. Elayoubi, R. Combes, and T. Chahed, "Optimal control of wake up mechanisms of femtocells in heterogeneous networks," *IEEE JSAC*, vol. 30, no. 3, pp. 664–672, 2012. → pages 8

[72] A. P. Jardosh, K. Papagiannaki, E. M. Belding, K. C. Almeroth, G. Iannaccone, and B. Vinnakota, "Green WLANs: On-demand WLAN infrastructures," *Mobile Networks and Applications*, vol. 14, no. 6, pp. 798–814, 2009. → pages 8

[73] "C-RAN: The road towards green RAN (white paper, ver. 2.5.0)," China Mobile Research Institute, Tech. Rep., October 2011. → pages 8

[74] Y. Lin, Z. Zhu, Q. Wang, and R. Sabhikhi, "Wireless network cloud: Architecture and system requirement," *IBM Journal of Research and Development*, vol. 54, no. 1, pp. 4:1–4:12, Jan.-Feb. 2010. → pages 8

[75] D. Wake, A. Nkansah, and N. J. Gomes, "Radio over fiber link design for next generation wireless systems," *Journal of Lightwave Technology*, vol. 28, no. 16, pp. 2456–2464, 2010. → pages 11, 12

[76] Common Public Radio Interface. [Online]. Available: http://www.cpri.info/ → pages 11, 70

[77] Open Base Station Architecture Initiative. [Online]. Available: http://www.obsai.com/ → pages 11

[78] G. Finnie, "European FTTH Forecast, 2011-2016," Fiber-to-the-Home Council Europe, Tech. Rep., 2012. [Online]. Available: http://www.ftthcouncil.eu/documents/Reports/Market_Forecast_December_2011.pdf → pages 12

[79] "Press Release: UAE, Number One in the Global Ranking of FTTH penetration rate," Fiber-to-the-Home Council Middle East & North Africa, Tech. Rep., 2011. [Online]. Available: http://www.ftthcoucilmena.org/ → pages 12

[80] [Online]. Available: http://www.isi.edu/nsnam/ns/ → pages 15, 29

[81] [Online]. Available: http://www.dei.unipd.it/wdyn/?IDsezione=5091 → pages 15, 29

[82] H. Li, A. Attar, V. C. M. Leung, and Q. Pang, "Collision reduction in cognitive wireless local area network over fibre," *International Journal on Advances in Internet Technology*, vol. 3, no. 1-2, pp. 1–12, 2010. → pages 15, 20

[83] H. Li, Q. Pang, and V. C. M. Leung, "Cognitive access points for dynamic radio resource management in wireless LAN over fiber," in *Proc. WWRF 20th meeting*, Ottawa, Canada, Apr. 2008. → pages 15, 20

[84] H. Li, A. Attar, Q. Pang, and V. C. M. Leung, "Collision avoidance and mitigation in cognitive wireless local area network over fibre," in *Proc. IEEE INTERNET'09*, Cannes, France, 2009, pp. 133–138. → pages 15, 20

[85] M.-S. Alouini and A. J. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1047–1066, 1999. → pages 15, 48, 69

[86] H. Li, A. Attar, and V. C. M. Leung, "Energy conservation via antenna scheduling in fiber-connected femto base stations," *Mobile Networks and Applications*, vol. 17, no. 5, pp. 685–694, 2012. → pages 16, 48

[87] H. Li, J. Hajipour, A. Attar, and V. C. M. Leung, "Efficient HetNet implementation using broadband wireless access with fibre-connected massively distributed antennas architecture," *IEEE Commun. Mag.*, vol. 18, pp. 72–78, June 2011. → pages 48, 70

[88] A. Attar, H. Li, and V. C. M. Leung, "Green last mile: How fiber-connected massively distributed antenna systems can save energy," *IEEE Commun. Mag.*, vol. 18, pp. 66–74, 2011. → pages 16, 48

[89] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple access channel," *IEEE Trans. Inf. Theory*, vol. 40, pp. 1713–1727, Nov. 1994. → pages 17, 71

[90] O. Somekh, O. Simeone, Y. Bar-Ness, and A. Haimovich, "Cooperative multicell zero-forcing beamforming in cellular downlink channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3206–3219, July 2009. → pages 17, 71

[91] J. Hoydis, M. Kobayashi, and M. Debbah, "On the optimal number of cooperative base stations in network MIMO," *Arxiv preprint arXiv:0912.4595*, 2010. → pages 17, 71

[92] J. Xu, J. Zhang, and J. G. Andrews, "On the accuracy of the Wyner model in cellular networks," *IEEE Trans. Wireless Commun.*, pp. 3098–3109, Sept. 2011. → pages 17, 71

[93] L. Tsang, C. H. Chan, and H. Sangani, "A banded matrix iterative approach to Monte Carlo simulations of large-scale random rough surface scattering: TE case," *Journal of Electromagnetic Waves and Applications*, vol. 7, no. 9, pp. 1185–1200, 1993. → pages 17, 71, 99

[94] A. Kavčić and J. M. F. Moura, "Matrices with banded inverses: Inversion algorithms and factorization of Gauss-Markov processes," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1495–1509, 2000. → pages 17, 71, 99

[95] A. F. Molisch, M. Toeltsch, and S. Vermani, "Iterative methods for cancellation of intercarrier interference in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 2158–2167, 2007. → pages 17, 71, 99

[96] P. de Kerret and D. Gesbert, "Towards optimal CSI allocation in multicell MIMO channels," *arXiv preprint arXiv:1110.6778*, 2011. → pages 17, 71, 98

[97] H. Li and V. C. M. Leung, "Energy-efficient low-complexity zero-forcing beamforming via banded matrix inversion in indoor massively distributed antenna systems," *submitted to IEEE Trans. Wireless Commun.* → pages 17, 69

[98] ——, "Low complexity zero-forcing beamforming for distributed massive MIMO systems in large public venues," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 370–382, August 2013. → pages 18, 97

[99] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Personal Commun. Mag.*, vol. 6, no. 4, pp. 13–18, Aug. 1999. → pages 20

[100] J. J. Metzner, "On improving utilization in ALOHA networks," *IEEE Trans. Commun.*, vol. COM-24, p. 447C448, April 1976. → pages 23

[101] W. Luo and A. Ephremides, "Power levels and packet lengths in random multiple access," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 46–58, 2002. → pages 23

[102] Z. Hadzi-Velkov and B. Spasenovski, "Capture effect in IEEE 802.11 basic service area under influence of Rayleigh fading and near/far effect," in *Proc. IEEE PIMRC 2002*, vol. 1, 2002, pp. 172–176. → pages 23

[103] A. Kochut, A. Vasan, A. U. Shankar, and A. Agrawala, "Sniffing out the correct physical layer capture model in 802.11b," in *Proc. IEEE ICNP 2004*, 2004, pp. 252–261. → pages 23

[104] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE JSAC*, pp. 535–547, 2000. → pages 23

[105] H. Chang, V. Misra, and D. Rubenstein, "A general model and analysis of physical layer capture in 802.11 networks," in *Proc. IEEE INFOCOM 2006*, 2006, pp. 1–12. → pages 23

[106] D. Halperin, T. Anderson, and D. Wetherall, "Taking the sting out of carrier sense: interference cancellation for wireless LANs," in *Proc. ACM MobiCom 2008*, 2008, pp. 339–350. → pages 23

[107] J. Lee, W. Kim, S. J. Lee, D. Jo, J. Ryu, T. Kwon, and Y. Choi, "An experimental study on the capture effect in 802.11a networks," in *Proc. ACM WiNTECH 2007*, 2007, pp. 19–26. → pages 23

[108] J. Lee, J. Ryud, S.-J. Lee, and T. Kwon, "Improved modeling of IEEE 802.11a PHY through fine-grained measurements," *Computer Networks*, 2009. → pages 23

[109] J. Boer, H. van Bokhorst, W. J. Diepstraten, A. Kamerman, R. Mud, H. van Driest, and I. R. J. Kopmeiners (Lucent Technologies, "Wireless LAN with enhanced capture provision (US Patent 5,987,033)," 1999. → pages 24

[110] N. Santhapuri, J. Manweiler, S. Sen, R. R. Choudhury, S. Nelakuduti, and K. Munagala, "Message in message (MIM): A case for reordering transmissions in wireless networks," in *Proc. HotNets 2008*, 2008. → pages 24

[111] A. Miu, H. Balakrishnan, and C. E. Koksal, "Multi-radio diversity in wireless networks," *Wireless Networks*, vol. 13, no. 6, pp. 779–798, 2007. → pages 24

[112] K. Whitehouse, A. Woo, F. Jiang, J. Polastre, and D. Culler, "Exploiting the capture effect for collision detection and recovery," in *Proc. Embedded Networked Sensors 2005*, 2005, pp. 45–52. → pages 24

[113] B. Kauffmann, F. Baccelli, A. Chainteau, V. Mhatre, K. Papagiannaki, and C. Diot, "Measurement-based self organization of interfering 802.11 wireless access networks," in *Proc. IEEE INFOCOM 2007*, Anchorage, AK, USA, May 2007, pp. 1451–1459. → pages 24

[114] A. Kumar and V. Kumar, "Optimal association of stations and APs in an IEEE 802.11 WLAN," in *Proc. National Conference on Communications (NCC) 2005*, IIT, India, Feb. 2005. → pages 24

[115] V. Mhatre, K. Papagiannaki, and F. Baccelli, "Interference mitigation through power control in high density 802.11 WLANs," in *Proc. IEEE INFOCOM 2007*, Anchorage, AK, USA, May 2007, pp. 535–543. → pages 24

[116] A. Mishra, V. Brik, S. Banerjee, A. Srinivasan, and W. Arbaugh, "A client-driven approach for channel management in wireless LANs," in *Proc. IEEE INFOCOM 2006*, Barcelona, Spain, Apr. 2006, pp. 1–12. → pages 24

[117] I. Broustis, K. Papagiannaki, S. V. Krishnamurthy, M. Faloutsos, and V. Mhatre, "MDG: Measurement-driven guidelines for 802.11 WLAN design," in *Proc. ACM MobiCom 2007*, New York, NY, USA, 2007, pp. 254–265. → pages 24

[118] M. Abusubaih, J. Gross, and A. Wolisz, "An inter-access point coordination protocol for dynamic channel selection in IEEE 802.11 Wireless LANs," in *Proc. IEEE Workshop on Autonomic Communications and Network Management (ACNM) 2007*, Munich, Germany, May 2007, pp. 17–24. → pages 24

[119] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005. → pages 29

[120] M. J. Crisp, S. Li, A. Watts, R. V. Penty, and I. H. White, "Uplink and downlink coverage improvements of 802.11g signals using a distributed antenna network," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 3388–3395, 2007. → pages 48

[121] A. J. Fehske, F. Richter, and G. P. Fettweis, "Energy efficiency improvements through micro sites in cellular mobile radio networks," in *Proc. 2nd Workshop on Green Commun.*, Hawaii, USA, Dec. 2009. → pages 51

[122] L. N. Trefethen and D. B. III, *Numerical Linear Algebra*. Society for Industrial Mathematics, 1997. → pages 52

[123] J. Eilert, D. Wu, and D. Liu, "Efficient complex matrix inversion for MIMO software defined radio," in *Proc. IEEE ISCAS'07*, 2007, pp. 2610–2613. → pages 52

[124] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE JSAC*, vol. 24, no. 3, pp. 528–541, March 2006. → pages 53

[125] M. Crisp, R. V. Penty, I. H. White, and A. Bell, "Wideband radio over fiber distributed antenna systems for energy efficient in-building wireless communications," in *Proc. IEEE VTC'10-Spring*, Taipei, Taiwan, May 2010. → pages 54, 86

[126] *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception*, 3GPP TS36.104 Std. → pages 60, 66

[127] ITU M.1225. [Online]. Available: http://www.itu.int/rec/R-REC-M.1225/en → pages 64, 74, 96

[128] D. N. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005. → pages 70, 75, 104

[129] S. Demko, "Inverses of band matrices and local convergence of spline projections," *SIAM J. Numer. Anal.*, vol. 14, no. 4, pp. 616–619, 1977. → pages 71, 99

[130] ——, "Decay rates for inverses of band matrices," *Mathematics of Computation*, vol. 43, no. 168, pp. 491–499, 1984. → pages 71, 79, 99, 150

[131] C. de Boor, "Dichotomies for band matrices," *SIAM J. Numer. Anal.*, vol. 17, no. 6, pp. 894–907, 1980. → pages 71, 99

[132] S. Jaffard, "Propriétés des matrices 'bien localisées' près de leur diagonale et quelques applications," *Ann. Inst. H. Poincaré Anal. Non Linéaire*, vol. 7, no. 5, pp. 461–476, 1990. → pages 71

[133] K. Gröchenig and M. Leinert, "Symmetry and inverse-closedness of matrix algebras and functional calculus for infinite matrices," *Transactions of the American Mathematical Society*, vol. 358, no. 6, pp. 2695–2712, 2006. → pages 71

[134] H. V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis, and G. Caire, "Achieving high data rates in a distributed MIMO system," in *Proc. IEEE MobiCom'12*, 2012, pp. 41–52. → pages 74

[135] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE JSAC*, vol. 31, no. 2, pp. 303–313, 2013. → pages 74

[136] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE JSAC*, vol. 31, no. 2, pp. 160–171, 2013. → pages 74

[137] F. Boccardi and H. Huang, "Zero-forcing precoding for the MIMO broadcast channel under per-antenna power constraints," in *Proc. IEEE SPAWC'06*, July 2006, pp. 1–5. → pages 75, 105

[138] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, 2008. → pages 75

[139] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Johns Hopkins University Press, 1996. → pages 76, 106

[140] D. N. Tse and S. V. Hanly, "Multiaccess fading channels. I. Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, 1998. → pages 80

[141] G. Auer, V. Giannini, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, C. Desset, and O. Blume, "Cellular energy efficiency evaluation framework (Invited Paper)," in *Proc. IEEE VTC'11-Spring*, 2011. → pages 86

[142] D. M. J. Devasirvatham, C. Banerjee, M. J. Krain, and D. A. Rappaport, "Multi-frequency radiowave propagation measurements in the portable radio environment," in *Proc. IEEE ICC'90*, 1990, pp. 1334–1340. → pages 98, 102, 103

[143] J. Medbo and J.-E. Berg, "Simple and accurate path loss modeling at 5 GHz in complex indoor environment with corridors," Tech. Rep., 2000. → pages 98, 102, 103

[144] J. Hoydis, M. Debbah, and M. Kobayashi, "Asymptotic moments for interference mitigation in correlated fading channels," in *Proc. IEEE ISIT'11*, 2011, pp. 2796–2800. → pages 98, 127

[145] P. de Kerret and D. Gesbert, "Sparse precoding in multicell MIMO systems," in *Proc. IEEE WCNC'12*, 2012, pp. 958–962. → pages 98, 99

[146] S. Moshavi, E. G. Kanterakis, and D. L. Schilling, "Multistage linear receivers for DS-CDMA systems," *International Journal of Wireless Information Networks*, vol. 3, no. 1, pp. 1–17, 1996. → pages 99, 127

[147] L. Cottatellucci and R. R. Mller, "A systematic approach to multistage detectors in multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3146–3158, 2005. → pages 99

[148] S. Aslan, E. Oruklu, and J. Saniie, "Architectural design tool for low area band matrix LU factorization," in *Proc. IEEE EIT'11*, 2011, pp. 1–6. → pages 99

[149] R. Szeliski, "Locally adapted hierarchical basis preconditioning," *ACM Trans. Graphics*, vol. 25, no. 3, pp. 1135–1143, 2006. → pages 99

[150] H. G. F. et al, "Method of equalizing a multicarrier signal for intercarrier interference," Patent wo/2008/151 337. 18, Dec, 2008. → pages 99

[151] H. Li, "Efficient modeling and analysis for large scale circuits," Ph.D. dissertation, Purdue University, 2007. → pages 99

[152] Y. Saad, *Iterative methods for sparse linear systems*. PWS publishing company, 1996. → pages 106

[153] A. Cleary and J. Dongarra, "Implementation in ScaLAPACK of divide-and-conquer algorithms for banded and tridiagonal linear systems," Computer Science Dept., University of Tennessee, Tech. Rep. Technical Report CS-97-358, 1997. → pages 106

[154] F. Heereman et al, "Pathloss model and prediction of range, power and throughput for 802.11n in large conference rooms," *AEU-International Journal of Electronics and Communications*, 2011. → pages 121

[155] M. Borgmann and H. Bolcskei, "Interpolation-based efficient matrix inversion for MIMO-OFDM receivers," in *Proc. IEEE ACSSC'04*, 2004, pp. 1941–1947. → pages 134

[156] D. Cescato and H. Bolcskei, "Algorithms for interpolation-based QR decomposition in MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1719–1733, April 2011. → pages

[157] S. Aubert, J. Tournois, and F. Nouvel, "On the implementation of MIMO-OFDM schemes using perturbation of the QR decomposition: Application to 3GPP LTE-A systems," in *Proc. IEEE ICASSP'11*, May 2011, pp. 3236–3239. → pages 134

[158] W. Liu, K. Choi, and H. Liu, "Complexity-reduced channel matrix inversion for MIMO systems in time-varying channels," in *Proc. IEEE VTC'10-Spring*, Taipei, Taiwan, May 2010. → pages 134

[159] C. V. Praet, G. Torfs, Z. Li, X. Yin, D. Suvakovic, H. Chow, X.-Z. Qiu, and P. Vetter, "10 Gbit/s bit interleaving CDR for low-power PON," *Electronics Letters*, vol. 48, no. 21, October 2012. → pages 135

[160] *TLK6002: Dual Channel 0.47Gbps to 6.25Gbps Multi-Rate Transceiver (SLLSE34A Rev. A)*, Texas Instruments, September 2010. → pages 135

[161] *Digital Diagnostic SFP, 850nm 3.072/2.4576 Gb/s, RoHS OBSAI/CPRI Compatible Optical Transceiver (AFBR-57J5APZ)*, Avago Technologies, January 2013. → pages 135

[162] M. A. Stephens, "Use of the Kolmogorov-Smirnov, Cramér-Von Mises and related statistics without extensive tables," *Journal of the Royal Statistical Society, Series B (Methodological)*, pp. 115–122, 1970. → pages 154

[163] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers, Scientists and Mathematicians*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. → pages 155

Appendix A

Derivation of $\lg(\bar{I}_i)$ for $\mathbf{H} \in \mathcal{EP}_{\alpha,\eta,\delta}$

Let $\mathbf{H} = \mathbf{H}_p + \mathbf{D}_p$ and $\mathbf{A} = \mathbf{D}_p\mathbf{W}_p$. Consider $p+1 \le i \le N-p$. The transpose of the i-th row of $\mathbf{D}_p$, denoted by $\mathbf{d}_i$, is composed of three sections: $[(i-1)^{-\eta/2}\alpha^{i-1};\cdots;p^{-\eta/2}\alpha^p]$, $[0;\cdots;0]$ and $[p^{-\eta/2}\alpha^p;\cdots;(N-i)^{-\eta/2}\alpha^{N-i}]$. Each entry in the j-th column of $\mathbf{W}_p$ is bounded by the corresponding entry in $\mathbf{w}_j = [\beta^{j-1};\cdots;\beta^{N-j}]$ $(1 \le j \le N)$, according to $\mathcal{EP}_{\alpha,\eta,\delta} \in \mathcal{ED}_\alpha$ and $\mathbf{W}_p \in \mathcal{ED}_\beta (\alpha < \beta < 1)$ which we learn from [130]. All of the multiplicative constants are dropped as our interest is to examine how $\bar{I}_i$ varies with p. By $a_{i,j} = \mathbf{d}_i^T\mathbf{w}_j$, the $a_{i,j}$ reaches its maximum when $p^{-\eta/2}\alpha^p$ in $\mathbf{d}_i$ meets 1 in $\mathbf{w}_j$. When the entry "1" in $\mathbf{w}_j$ falls in each of three sections in $\mathbf{d}_i$, the corresponding $a_{i,j}$ involves a summation of power series, $\sum_{k=a}^{b} k^{-\eta/2}\alpha^k$, where $1 < a < b$. By approximating the summation by $a^{-\eta/2}\frac{\alpha^a}{1-\alpha}$, we have:

$$a_{i,j} \approx \begin{cases} p^{-\eta/2}\frac{\alpha}{\beta-\alpha}\beta^{i-j}(\frac{\alpha}{\beta})^p, & j \le i-p \\ p^{-\eta/2}\frac{\alpha^p(\beta^{j-i+p}+\beta^{i-j+p})}{1-\alpha\beta}, & i-p < j < i+p-1 \\ p^{-\eta/2}\frac{\alpha}{\beta-\alpha}\beta^{j-i}(\frac{\alpha}{\beta})^p, & j \ge i+p-1 \end{cases} \tag{A.1}$$

After some algebraic manipulations, the $\bar{I}_i$ is approximated by

$$\bar{I}_i \approx \frac{2Pp^{-\eta}\alpha^{2p}}{1-\beta^2}\left(\frac{\alpha}{\beta-\alpha}\right)^2 \propto p^{-\eta}\alpha^{2p}, \tag{A.2}$$

where $p+1 \le i \le N-p$ and $p \ge 2$. For $1 \le i \le p$ and $N-p+1 \le i \le N$, there is only one side of interference; hence, $\bar{I}_i \propto p^{-\eta}\alpha^{2p}$. When the width increases from p to p+1 $(1 < p \ll N)$, we have $\lg(\bar{I}_i)$ approximately decreases by $L+10\eta\lg(1+1/p)$.

Appendix B

Inverse of $\mathcal{EZ}_\alpha$

Suppose j=1. By Cramer rule, $w_{k,1} = \mathbf{C}_{1,k}/|\mathbf{H}_p| = (-1)^{k+1}M_{1,k}/|\mathbf{H}_p|$, where $\mathbf{C}_{1,k}$ is the (1,k) cofactor of $\mathbf{H}_p$ and $M_{1,k}$ is the (1,k) minor of $\mathbf{H}_p$. It is difficult to work with $w_{k,1}$ as it is subject to certain ratio distribution and has a heavy tail. The correlation between $M_{1,k}$ and $|\mathbf{H}_p|$ also makes the analysis intractable. However, the problem can be simplified by only counting dominating terms in $M_{1,k}$ and $|\mathbf{H}_p|$ since the variance of off-diagonal entries in $\mathbf{H}_p$ are exponentially decayed. Here by "x dominates y", we mean var(x)≫var(y). Therefore, among N! terms in the Leibniz expansion of $|\mathbf{H}_p|$, the term $h_{11}h_{22}\cdots h_{NN}$ is dominating. As for $M_{1,k}$, there will be multiple dominating terms. Take $p = 5$ and $k \in \{2,3,4,5\}$ as an example. Given $\mathbf{H} \in \mathcal{EZ}_\alpha$, where $h_{i,j} = \alpha^{|i-j|}f_{i,j}$ and $f_{i,j} \sim \mathcal{CN}(0,1)$, we observe that the variance of dominating terms in $M_{1,k}$ should have α^{k-1}. Therefore, the number of dominating terms in $M_{1,2}$, $M_{1,3}$, $M_{1,4}$, $M_{1,5}$ is 2^0, 2^1, 2^2 and 2^3, respectively. We now conjecture the number of dominating terms in $M_{1,k}$ is 2^{k-2} ($2 \le k \le p$) and prove it by induction.

Assume there are 2^{k-2} dominating terms in $M_{1,k}$. We approximate $M_{1,k}$ by $|\mathbf{A}_k|\prod_{m=k+1}^{N} f_{m,m}$,

where $\mathbf{A}_k$ is a (k-1)×(k-1) matrix given by

$$\mathbf{A}_k = \begin{pmatrix} \alpha f_{21} & f_{22} & \cdots & \alpha^{k-4} f_{2,k-2} & \alpha^{k-3} f_{2,k-1} \\ \alpha^2 f_{31} & \alpha f_{32} & \cdots & \alpha^{k-5} f_{3,k-2} & \alpha^{k-4} f_{3,k-1} \\ \alpha^3 f_{41} & \alpha^2 f_{42} & \cdots & \alpha^{k-6} f_{4,k-2} & \alpha^{k-5} f_{4,k-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha^{k-2} f_{k-1,1} & \alpha^{k-3} f_{k-1,2} & \cdots & \alpha f_{k-1,k-2} & f_{k-1,k-1} \\ \alpha^{k-1} f_{k,1} & \alpha^{k-2} f_{k,2} & \cdots & \alpha^2 f_{k,k-2} & \alpha f_{k,k-1} \end{pmatrix}. \tag{B.1}$$

The $M_{1,k+1}$ is then given by

$$\begin{aligned} M_{1,k+1} &\approx |\mathbf{A}_{k+1}| \prod_{m=k+2}^{N} f_{m,m} = \begin{vmatrix} \mathbf{A}_k & \mathbf{b}^T \\ \mathbf{c} & \alpha f_{k+1,k} \end{vmatrix} \prod_{m=k+2}^{N} f_{m,m}, \\ \mathbf{b} &= [\alpha^{k-2} f_{2,k}, \cdots, \alpha^1 f_{k-1,k}, \alpha^0 f_{k,k}], \\ \mathbf{c} &= [\alpha^k f_{k+1,1}, \cdots, \alpha^3 f_{k+1,k-2}, \alpha^2 f_{k+1,k-1}]. \end{aligned} \tag{B.2}$$

By Laplace formula, we expand $|\mathbf{A}_{k+1}|$ along the last column, $[\mathbf{b}, \alpha f_{k+1,k}]^T$. Our purpose is to find terms with variance equal to α^k. One can notice that among the k terms in the last column, only two terms expand dominating terms: $\alpha f_{k+1,k}$ and $\alpha^0 f_{k,k}$. Thus, the number of dominating terms in $M_{1,k+1}$ is 2^{k-1}. By induction, we proved that the number of dominating terms in $M_{1,k}$ is 2^{k-2}.

We now focus on evaluating $\mathbb{E}_{\mathbf{F}}[\lg(\|w_{k,1}\|^2)]$, denoted by m_k. After introducing $\alpha^{k-1} z_k$ to represent the sum of 2^{k-2} terms in $|\mathbf{A}_k|$ where each term has a variance equal to α^{k-1}, we can approximate $\|w_{k,1}\|$ as $\|\alpha^{k-1} z_k\| / \prod_{m=1}^{k} \|f_{m,m}\|$. Thus,

$$m_1 = \gamma / \ln 10, m_2 = 2 \lg \alpha + \gamma / \ln 10, \tag{B.3}$$

$$m_k = (2k-2) \lg \alpha + \mathbb{E}[\lg(\|z_k\|^2)] - \sum_{m=1}^{k} \mathbb{E}[\lg(\|f_{m,m}\|^2)], k \geq 3. \tag{B.4}$$

The γ in (B.3) is Euler constant - it comes from calculating the mean of Gumbel distribution. Note that we used $\mathbb{E}[\cdot]$ instead of $\mathbb{E}_{\mathbf{F}}[\cdot]$ when this causes no ambiguity. We now apply a recursive method to evaluate $\mathbb{E}[\lg(\|z_k\|^2)]$ in (B.4). Let $d_k = m_k - m_{k-1}$:

$$d_k = 2\lg\alpha + \gamma/\ln 10 + \mathbb{E}[\lg(\|z_k/z_{k=1}\|^2)]/\ln 10. \tag{B.5}$$

Since half of 2^{k-2} terms in z_k are expanded by $\alpha f_{k,k-1}$ and another half by $f_{k-1,k-1}$, we write

$$\alpha^{k-2} z_{k-1} = \sum_{m=1}^{k-3} [\alpha^{k-1-m} N_{k-2,m} (-1)^m (f_{k-2,m} f_{k-1,k-2} - f_{k-1,m} f_{k-2,k-2})], \tag{B.6}$$

$$\alpha^{k-1} z_k = \alpha^{k-2} z_{k-1} \cdot \alpha f_{k,k-1} + \sum_{m=1}^{k-2} [\alpha^{k-m} N_{k-1,m} (-1)^m (f_{k,m} f_{k-1,k-1})], \tag{B.7}$$

where $N_{k-1,m}$ is the (k-1,m) minor of $\mathbf{A}_{k-1}$, $N_{k-2,m}$ is the (k-2,m) minor of $\mathbf{A}_{k-2}$, and the term $(-1)^m$ is obtained by examining the signature of permutation for each term. Dividing (B.7) over (B.6) produces:

$$\alpha z_k / z_{k-1} = \alpha (f_{k,k-1} + x f_{k-1,k-1}), \tag{B.8}$$

where $x = \frac{\sum_{m=1}^{k-2} [\alpha^{k-m} N_{k-1,m} f_{k,m}]}{\sum_{m=1}^{k-2} [\alpha^{k-m} N_{k-1,m} f_{k-1,m}]}$. Owing to the group independence among $f_{k,k-1}$, $f_{k-1,k-1}$, $f_{k,m}$, $f_{k-1,m}$ and $N_{k-1,m} (1 \le m \le k-2)$, for each instance of $N_{k-1,m}$, the numerator in x is a linear combination of independent and identically distributed (i.i.d.) complex Gaussian random variables $f_{k,m}$ $(1 \le m \le k-2)$ and the denominator in x is a linear combination of i.i.d. complex Gaussian random variables $f_{k-1,m}$ with an equal set of coefficients. The distribution of x therefore does not depend on the instances of $N_{k-1,m}$. In fact, $x \sim g_1/g_2$, where g_1 and g_2 are i.i.d. zero-mean circularly symmetric complex Gaussian (ZMCSCG) subject to $\mathcal{CN}(0,1)$. We further conclude that z_k/z_{k-1} $(3 \le k \le p)$ has the same distribution as $g_3 + g_4 g_1/g_2$, where $g_i (1 \le i \le 4)$ are i.i.d. ZMCSCG subject to $\mathcal{CN}(0,1)$.

We applied two-sample Kolmogorov-Smirnov test to compare 10^6 samples from $\|z_k/z_{k-1}\| (k \in \{4,5\})$ and from $\|g_3 + g_4 g_1/g_2\|$. The P-value, defined in [162], is 0.54 for k=4 and 0.63 for k=5,

indicating that the null hypothesis, i.e., two sets of samples being consistent, cannot be rejected even at 50% significance level.

To evaluate $\mathbb{E}[\ln(\|z_k/z_{k-1}\|^2)]$ in (B.5), we write

$$\mathbb{E}[\ln(\|z_k/z_{k-1}\|^2)] = \mathbb{E}[\ln(\|g_4\|^2)] + \mathbb{E}[\ln(\|g_3/g_4 + g_1/g_2\|^2)]. \tag{B.9}$$

The second expectation in the RHS of (B.9) can be evaluated as 1 by applying the following facts: circular symmetry of g_i; $\mathbb{E}_\Theta[\ln(m + n\cos\theta)] = 2[\ln(\sqrt{m-n} + \sqrt{m+n}) - \ln 2](m > n > 0)$; the cumulative distribution function (CDF) of a Rayleigh ratio is $F_{R_0}(r_0) = r_0^2/(r_0^2 + 1)(r_0 > 0)$ [163]. Combine (B.9) and (B.4):

$$m_k = \begin{cases} (2k-2)\lg\alpha + \gamma/\ln 10, & k = 1,2 \\ (2k-2)\lg\alpha + \gamma/\ln 10 + (k-2)\lg e, & 3 \le k \le p \end{cases}. \tag{B.10}$$

For other values of j, there are only minor variations in the above process. Thus, we have shown that $\mathbb{E}_{\mathbf{F}}[\lg(\|w_{k,j}\|^2)]$ linearly decays with k for $3 \le k \le p$. The above analysis is based on dominating terms. Therefore when α approaches zero, the decay rate is approximated by

$$r_{EZ} = 2\lg\alpha + \lg e. \tag{B.11}$$

For $k > p+1$, the structure of $M_{1,k}$ can be similarly analyzed to show that z_k/z_{k-1} has the same distribution as $g_3 + c_0 g_4 g_1/g_2$, where $g_i(1 \le i \le 4)$ are i.i.d. and subject to $\mathcal{CN}(0,1)$. The $c_0(0 < c_0 < 1)$ is a constant depending on p. Thus, for $k > p$, $\mathbb{E}_{\mathbf{F}}[\lg(\|w_{k,1}\|^2)]$ also linearly decays with k yet with a rate smaller than $2\lg\alpha + \lg e$.

At $k = p+1$, the bottom-left entry of $M_{1,p+1}$ becomes zero, so we expect $\mathbb{E}[\lg(\|w_{p+1,1}\|^2)]$ decay faster.

To quantify the behavior, we define the *outband drop* as the log-ratio of the accelerated and the

normal decay, i.e., $O_p \triangleq \lg(|\mathbf{W}_p(p+1,1)/\mathbf{W}_{p+1}(p+1,1)|)$. Owing to the exponential number of dominating components in $M_{1,p+1}$, the lost component in $M_{1,p+1}$ does not affect the overall value while p grows. Therefore, (B.10) also holds for $k = p+1$.

Appendix C

Inverse of $\mathcal{EPRZ}_{\alpha,\eta}$

Consider $\mathbf{H} \in \mathcal{EPRZ}_{\alpha,\eta}$ $(0 < \alpha < 1, \eta \geq 2)$ and $h_{i,j}$ is given by $|i-j+\delta_i|^{-\eta/2}\alpha^{|i-j|}f_{i,j}$, where $f_{i,j} \sim \mathcal{CN}(0,1)$ and $\delta_i \sim \mathcal{U}(-0.5,0.5)$. We only show the case of $j = 1$; the derivation process at other values of j follows similarly.

While the variance of dominating terms in $M_{1,k}$ should still be α^{k-1} according to the same arguments in Appendix B, introducing polynomial decay reduces the number of dominating terms in $M_{1,k}$. By inspection, at $k \geq 3$ we have two choices:

$$M_{1,k}^{(1)} \triangleq (-1)^k \alpha^{k-1} \cdot x_0 \left(1 - \sum_{m=1}^{k-2} [a_m x_m]\right) \cdot \prod_{m=k+1}^{N} [|\delta_m|^{-\eta/2} f_{m,m}], \tag{C.1}$$

$$M_{1,k}^{(2)} \triangleq \alpha^{k-1} \cdot z_0 \left(1 - \sum_{m=1}^{k-2} [b_m z_m]\right) \cdot \prod_{m=k+1}^{N} [|\delta_m|^{-\eta/2} f_{m,m}], \tag{C.2}$$

where

$$x_0 = (k-1+\delta_k)^{-\eta/2} f_{k,1} \prod_{m=2}^{k-1} (|\delta_m|^{-\eta/2} f_{m,m}), \qquad z_0 = \prod_{m=1}^{k-1} [(1+\delta_{m+1})^{-\eta/2} f_{m+1,m}], \tag{C.3}$$

$$a_m = \left(\frac{(m+\delta_{m+1})(k-m-1+\delta_k)}{(k-1+\delta_k)|\delta_{m+1}|}\right)^{-\eta/2}, \qquad x_m = \frac{f_{m+1,1} f_{k,m+1}}{f_{k,1} f_{m+1,m+1}}, \tag{C.4}$$

$$b_m = \left(\frac{(2+\delta_{m+2})|\delta_{m+1}|}{(1+\delta_{m+1})(1+\delta_{m+2})}\right)^{-\eta/2}, \qquad z_m = \frac{f_{m+2,m} f_{m+1,m+1}}{f_{m+1,m} f_{m+2,m+1}}. \tag{C.5}$$

The signs in (C.1) and (C.2) are determined by the signatures of corresponding permutations; the summation $\sum_{m=1}^{k-2}[a_m x_m]$ is formed by replacing the product $h_{k,1}h_{m+1,m+1}$ in x_0 with $h_{m+1,1}h_{k,m+1}$ one at a time. Comparing $\mathbb{E}_{\mathbf{F},\Delta}[\ln\|x_0\|]$ with $\mathbb{E}_{\mathbf{F},\Delta}[\ln\|z_0\|]$ reveals that $M_{1,k}^{(1)}$ dominates over $M_{1,k}^{(2)}$. After approximating $|\mathbf{H}_p|$ as the product of its main diagonal, we write

$$\|w_{k,1}\| \approx \alpha^{k-1}\left\|\frac{(k-1+\delta_k)^{-\eta/2}f_{k,1}}{|\delta_1|^{-\eta/2}|\delta_k|^{-\eta/2}f_{1,1}f_{k,k}}\cdot(1-\sum_{m=1}^{k-2}[a_m x_m])\right\|, \tag{C.6}$$

$$\mathbb{E}[\lg(\|w_{k,1}\|^2)] = (2k-2)\lg\alpha + \eta\mathbb{E}[\lg|\delta_1\delta_k|] + 2\gamma/\ln 10 + B_k. \tag{C.7}$$

The B_k in (C.7) is defined as

$$B_k \triangleq \mathbb{E}\left[\lg\|c_0 f_{k,1} - \sum_{m=1}^{k-2}[c_m g_m]\|^2\right], \tag{C.8}$$

where $g_m = \frac{f_{m+1,1}f_{k,m+1}}{f_{m+1,m+1}}$,$c_0 = (k-1+\delta_k)^{-\eta/2}$ and $c_m = (\frac{(m+\delta_{m+1})(k-m+1+\delta_k)}{|\delta_{m+1}|})^{-\eta/2}$. Since Central Limit Theorem (CLT) cannot be applied to B_k as g_m has infinite variance, we obtain two estimations of B_k by taking only c_0-term, and taking three most significant terms, respectively:

$$B_k^{(1)} = \mathbb{E}[\lg\|c_0 f_{k,1}\|^2], \tag{C.9}$$

$$B_k^{(3)} = \mathbb{E}[\lg\|c_0 f_{k,1} - c_1 g_1 - c_{k-2}g_{k-2}\|^2]. \tag{C.10}$$

Based on $B_k^{(1)}$, we estimate $\mathbb{E}[\lg(\|w_{k,1}\|^2)]$ with $LW_k^{(1)}(3 \le k \le p)$:

$$LW_k^{(1)} = (2k-2)\lg\alpha - \eta\lg(k-1) - 2\eta\lg(2e) + \gamma/\ln 10. \tag{C.11}$$

At k=1 and 2, $LW_1^{(1)} = -\eta\lg(2e) + \gamma/\ln(10)$ and $LW_2^{(1)} = LW_1^{(1)} + 2\lg\alpha - 1.5\eta\lg 3$.

To estimate $\mathbb{E}[\lg(\|w_{k,1}\|^2)]$ based on $B_k^{(3)}$, we first conjecture

$$B_k \approx \mathbb{E}_\Delta[\lg\|c_0\|^2] + \mathbb{E}_\Delta\left[\mathbb{E}_{\mathbf{F}}\left[\lg\|f_{k,1} - a\frac{f_x f_y}{f_z}\|^2\right]\right], \tag{C.12}$$

where $a = \sqrt{c_1^2 + c_{k-2}^2}/c_0$; f_x, f_y, f_z are i.i.d. and subject to $\mathcal{CN}(0,1)$. Given i.i.d. $z_i (1 \le i \le 4)$ subject to $\mathcal{CN}(0,1)$, we know $\mathbb{E}[\lg \|z_1/z_2 + a z_3/z_4\|^2] = \frac{2a^2 \lg a}{a^2-1}$, which is tightly upper-bounded by $a/\ln 10$ for $0 < a < 1$. We also know, for $\delta_i \sim \mathcal{U}(-0.5,0.5)$ and at $k > 3$,

$$\mathbb{E}[\lg(k-1+\delta_k)] \approx \lg(k-1), \mathbb{E}[\lg|\delta_1|] = \mathbb{E}[\lg|\delta_k|] = -\lg(2e). \tag{C.13}$$

Combining (C.7), (C.12) and (C.13), we obtain the estimation of $\mathbb{E}[\lg(\|w_{k,1}\|^2)]$ based on $B_k^{(3)} (3 < k \le p)$:

$$LW_k^{(3)} = (2k-2)\lg\alpha - \eta\lg(k-1) - 2\eta\lg(2e) + \gamma/\ln 10 + \mathbb{E}_\Delta[a]/\ln 10. \tag{C.14}$$

To evaluate $\mathbb{E}_\Delta[a]$ in (C.14), we consider the following:

$$\mathbb{E}[\sqrt{c_1^2 + c_{k-2}^2}] \approx (k-2)^{-\eta/2}\mathbb{E}[\sqrt{(\frac{1+\delta_2}{|\delta_2|})^{-\eta} + (\frac{1+\delta_k}{|\delta_{k-1}|})^{-\eta}}], \tag{C.15}$$

where the expectation in the RHS, denoted by ρ_η, can be evaluated through Monte-Carlo method at different η values without depending on particular values of k (See Table C.1). Thus for $k > 3$,

$$LW_k^{(3)} \approx (2k-2)\lg\alpha - \eta\lg(k-1) - 2\eta\lg(2e) + \gamma/\ln 10 + \frac{\rho_\eta}{\ln 10}(\frac{k-1}{k-2})^{\eta/2}. \tag{C.16}$$

For $k \in \{1,2,3\}$, $LW_k^{(3)}$ can be explicitly calculated.

Unlike $\mathcal{EZ}_\alpha$ case, (C.11) does not hold at $k = p+1$ because with the bottom-left entry of $M_{1,k}$ being zero, the B_k lost the most significant component, c_0. In that case we rewrite (C.7):

$$\mathbb{E}[\lg(\|w_{p+1,1}\|^2)] \approx 2p\lg\alpha - 2\eta\lg(2e) + 2\gamma/\ln 10 - \eta\lg(p-1) + \psi_\eta. \tag{C.17}$$

When we choose $\{c_1, c_2, c_{k-2}, c_{k-1}\}$ to estimate B_k, the ψ_η can be evaluated through Monte-Carlo method, as listed in Table C.1. The outband drop is then given by the difference between (C.17) and (C.11), i.e., $\gamma/\ln 10 + \psi_\eta$.

Table C.1: Constants used in the analysis

η	ρ_η	ψ_η	$\gamma/\ln 10+\psi_\eta$	η	ρ_η	ψ_η	$\gamma/\ln 10+\psi_\eta$
0.5	0.9789	0.3389	n/a	2	0.4393	-0.6533	-0.4026
1	0.7220	-0.0222	n/a	3	0.2949	-1.2118	-0.9611
1.5	0.5548	-0.3498	n/a	4	0.2125	-1.7318	-1.4811

Appendix D

Derivation of $(\varepsilon_{p+1} - \varepsilon_p)$ for $\mathbf{H} \in \mathcal{EZ}_\alpha$

Define $z_p \triangleq \|w_{i,i+p}/w_{i,i-p}\|^2$. By $\alpha \to 0$, we can follow the procedure in Appendix B to establish $w_{i,i+p}/w_{i,i+p-1} \approx (f_{i+p-1,i+p} + f_{i+p-1,i+p-1} g_1/g_2)/f_{i+p,i+p}$, and then obtain

$$\frac{z_p}{z_{p-1}} \approx \left\| \frac{f_{i+p-1,i+p} + f_{i+p-1,i+p-1}\frac{g_1}{g_2}}{f_{i-p+1,i-p} + f_{i-p+1,i-p+1}\frac{g_3}{g_4}} \cdot \frac{f_{i-p,i-p}}{f_{i+p,i+p}} \right\|^2 . \tag{D.1}$$

A suitable model is needed to remove the correlation between g_1/g_2 and $f_{i+p,i+p}$, and between g_3/g_4 and $f_{i-p,i-p}$. First we inspect (D.1) and establish the following recursive relationship (which can be proved by induction):

$$n_p = \frac{\sum_{k=0}^{p-1}[(-1)^k f_{i+k,i+p} n_k]}{f_{i+p,i+p}}, \quad d_p = \frac{\sum_{k=0}^{p-1}[(-1)^k f_{i-k,i-p} d_k]}{f_{i-p,i-p}}, \tag{D.2}$$

$$z_p \triangleq \left\| \frac{w_{i,i+p}}{w_{i,i-p}} \right\|^2 \approx \left\| \frac{n_p}{d_p} \right\|^2 , \tag{D.3}$$

where $n_0 = p_0 = 1$. Observing the sequence of n_p and d_p reveals that their tails become heavier when p grows. So,

$$\varepsilon_p = \mathbb{E}[\lg(1+z_p)] \approx \int_1^{+\infty} \lg(1+z) f_{z_p}(z) \mathrm{d}z \approx \lg e \int_1^{+\infty} (\ln z + 1/z) f_{z_p}(z) \mathrm{d}z, \tag{D.4}$$

where the first approximation is from the heavy tail of z_p and the second is from the first order Taylor expansion. For $z_p > 1$, we model $z_p = z_{p-1}\kappa_{p-1}$, where $\kappa_{p-1} = \|\frac{f_{i+p-1,i+p}f_{i-p,i-p}}{f_{i-p+1,i-p}f_{i+p,i+p}}\|^2$ for $\kappa_{p-1} > 1$. By CLT, $\ln z_p$ can then be modeled as $\mathcal{N}(0, 2\sigma_g\sqrt{p})$, where $\sigma_g = \pi/\sqrt{6}$. Applying the asymptotic expansion of the complementary error function and $\Pr(\kappa_{p-1} > 1) = 0.5$ to (D.4), we write $\varepsilon_p \approx \frac{\lg e}{\sqrt{2\pi}}[\sigma_g\sqrt{p} + \frac{1}{4\sigma_g\sqrt{p}}]$, from which we establish that as $p \to +\infty$ and $\alpha \to 0$, the term $(\varepsilon_{p+1} - \varepsilon_p)$ approaches 0.

Appendix E

Inverse of $\mathcal{ED}_\alpha$

Denote by $\mathbf{T}^{(N)}$ an $N \times N$ matrix that belongs to $\mathcal{ED}_\alpha$. Denote the inverse of $\mathbf{T}^{(N)}$ by $\mathbf{W}^{(N)}$. We have

$$|\mathbf{T}^{(N)}| = (1-\alpha^2)|\mathbf{T}^{(N-1)}| = \cdots = (1-\alpha^2)^{N-1}, \tag{E.1}$$

from which we obtain the main diagonal entries of $\mathbf{W}^{(N)}$, $w_{i,i}^{(N)} = \frac{1}{1-\alpha^2}$ for $i = 1, N$ and $w_{i,i}^{(N)} = \frac{1+\alpha^2}{1-\alpha^2}$ otherwise.

Now we consider off-diagonal entries of $\mathbf{W}^{(N)}$. We have $w_{i,j}^{(N)} = \mathbf{C}_{j,i}^{(N)}/|\mathbf{T}^{(N)}| = (-1)^{i+j} M_{j,i}^{(N)}/|\mathbf{T}^{(N)}|$, where $\mathbf{C}_{j,i}^{(N)}$ is the (j,i) cofactor of $\mathbf{T}^{(N)}$ and $M_{j,i}^{(N)}$ is the (j,i) minor of $\mathbf{T}^{(N)}$. For $|i-j| = 1$, we have

$$M_{j,i}^{(N)} = (1-\alpha^2)^{i-1}(\alpha-\alpha^3)|\mathbf{T}^{(N-i-1)}| = \cdots = \alpha(1-\alpha^2)^{N-2}, \tag{E.2}$$

from which we establish that the off-diagonal entries of $\mathbf{W}^{(N)}$ take the same value: $\frac{-\alpha}{1-\alpha^2}$.

For $|i-j| > 1$, it is easy to show the existence of two linearly dependent rows in the submatrix formed by deleting the j-th row and i-th column of $\mathbf{T}^{(N)}$. Therefore, the inverse of $\mathbf{T}^{(N)}$ is tri-diagonal.

Appendix F

Derivation of $\mathbb{E}\{\ln[\max_{1\le i\le N}(|f_i/g_i|^2)]\}$

Theorem: Given i.i.d. $f_i, g_i \sim \mathcal{CN}(0,1)$ $(1 \le i \le N)$. $\mathbb{E}\{\ln[\max_{1\le i\le N}(|f_i/g_i|^2)]\}$ is equal to the (N-1)-th harmonic number ($N > 1$).

Proof: Let $X_i \equiv |f_i/g_i|^2$, which is a ratio of two i.i.d. exponentially distributed random variables. By $f_{X_i}(x) = 1/(x+1)^2$ and $F_{X_i}(x) = x/(x+1)$, the CDF of $\max_i X_i$ is given by $F(x) = [x/(x+1)]^N$. Denoting $\mathbb{E}\{\ln[\max_{1\le i\le N}(|f_i/g_i|^2)]\}$ by S_N, we have

$$S_N = \int_0^{+\infty} \ln x \, \mathrm{d}F(x) = (\frac{x}{x+1})^N \ln x|_0^{+\infty} - \int_0^{+\infty} \frac{x^{N-1}}{(x+1)^N} \mathrm{d}x \tag{F.1}$$

from which we have $S_N - S_{N-1} = 1/(N-1)$. ■