RICE UNIVERSITY

Client Beamforming for Rate Scalability of MU-MIMO Networks

by

Hang Yu

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

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Lin Zhong, Chair
Associate Professor of Electrical and Computer Engineering

Edward W. Knightly
Chair and Professor of Electrical and Computer Engineering

Ashutosh Sabharwal
Professor of Electrical and Computer Engineering

David B. Johnson
Professor of Computer Science

Houston, Texas

April, 2015

ABSTRACT

Client Beamforming for Rate Scalability of MU-MIMO Networks

by

Hang Yu

In a multi-user MIMO (MU-MIMO) network, an AP with multiple antennas can simultaneously serve multiple clients to increase the achievable rate. To realize such rate increase, the MU-MIMO AP leverages beamforming techniques such as zero-forcing beamforming to eliminate the intra-cell interference. However, current MU-MIMO networks suffer from two fundamental problems that limit their *rate scalability* to the number of served clients. First, for a MU-MIMO network with a single cell, as the number of simultaneous clients increases and approaches the number of antennas on the AP, the achievable rate of the cell often flattens and may even drop. Second, for a MU-MIMO network with multiple cells, the APs cannot simultaneously serve their clients due to inter-cell interference, and thus the simultaneous clients are constrained to a single cell which limits the rate scalability.

Our unique perspective to tackle the rate scalability problems of MU-MIMO networks is that modern mobile devices such as laptops, tablets, and smartphones as clients can be equipped with multiple antennas. These client antennas provide important degrees of freedom and can be leveraged for beamforming. We propose two solutions that leverage client beamforming differently. First, to tackle the rate scalability problem in a single MU-MIMO cell, we exploit client beamforming to improve the orthogonality between the channel vectors of the clients. Such channel orthogo-

nality between clients determines the SNR and rate reduction when the MU-MIMO AP cancels the intra-cell interference, and is therefore critical for the achievable rate of a MU-MIMO cell to become more scalable to the number of clients. We devise an 802.11ac-based protocol called MACCO, in which each client locally optimizes its beamforming weights based on the channel knowledge overheard from other clients. To tackle the rate scalability problem in multiple MU-MIMO cells, we exploit client beamforming together with AP beamforming to coordinately cancel the inter-cell interference. To achieve such coordinated interference cancellation in a practical way, we propose a two-step optimization including antenna usage optimization and beamforming weight optimization. We devise another 802.11ac-based protocol called CoaCa, which integrates the two-step optimization into 802.11ac with small modifications and negligible overhead.

We implement both MACCO and CoaCa on the WARP SDR platform leveraging the WARPLab framework, and experimentally evaluate them under real-world indoor wireless channels. The results demonstrate the effectiveness of MACCO and CoaCa toward achieving better rate scalability in MU-MIMO networks with a single or multiple cells.

Acknowledgments

I would like to firstly thank my advisor, Lin Zhong, who has guided me throughout my academic life at Rice University. He has made significant contributions to not only this thesis, but also the development of myself as a researcher. When I started my Ph.d. career, he spent significant effort on directing me to the right path of doing high-quality research. When I was encountered with hard problems, he constantly helped me overcome the obstacles with patience.

Ashutosh Sabharwal deserves my special thanks for his insightful and inspiring advice to my research. This thesis work would never be done successfully without his continuous involvement. I also appreciate the help from all other committee members, including Edward W. Knightly and David. B. Johnson. Their feedback is of great value.

This thesis work includes the effort from many members in the Electrical and Computer Engineering Department of Rice University. Oscar Bejarano provided help on the design and implementation of CoaCa. Ardalan Amiri Sani and Clayton Shepard offered many insightful suggestions that further improved the quality of this thesis. I also would like to thank all the group members of RECG, including Xiaozhu Lin, Zhen Wang, Siqi Zhao, Robert LiKamWa, Chao Xu, Kevin Boos, Min Hong Yun, Jie Liao, and Abeer Javed, for their help on my research and life in Houston.

Finally, I am indebted to my family, in particular my fiancee and my parents. Their continuous love and care supported me whenever I felt frustrated or depressed. I cannot imagine how I would get this thesis done without them.

Contents

Abstract ii

Acknowledgments iv

List of Illustrations viii

List of Tables xii

1 Introduction 1

1.1 Problem Statement: Rate Scalability of MU-MIMO Networks 2

1.2 Client Beamforming for Rate Scalability 4

1.2.1 Rate Scalability of a Single MU-MIMO Cell 7

1.2.2 Combating Inter-cell Interference in Multiple MU-MIMO Cells 9

1.3 802.11ac-based Realization 13

1.3.1 MACCO: Improving Rate Scalability of a Single 802.11ac Cell 13

1.3.2 CoaCa: Combatting Inter-cell Interference in Multiple 802.11ac Cells 15

1.4 Summary of Contributions 16

2 Background 18

2.1 Interference Cancellation in MU-MIMO 18

2.2 MU-MIMO in 802.11ac 23

3 Related Work 26

3.1 Rate Scalability of a MU-MIMO Cell 26

3.2 Inter-cell Interference in a MU-MIMO Network 28

3.2.1 Network-MIMO- and CoMP-based Solutions 29

3.2.2 Distributed Solutions 30

4 Rate Scalability of a Single MU-MIMO Cell 32

4.1 Problem Statement 32

4.2 Successive Beamforming Optimization 35

4.2.1 Cases 1: $\boldsymbol{M}$=2 and $\boldsymbol{K}$=2 37

4.2.2 Cases 2: $\boldsymbol{M}$=3 and $\boldsymbol{K}$=3 41

4.2.3 Cases 3: $\boldsymbol{M}$=3 and $\boldsymbol{K}$=2 43

4.2.4 General Case 44

4.2.5 Computational Complexity 45

5 Inter-cell Interference between MU-MIMO Cells 46

5.1 Problem Statement 46

5.2 Overview 47

5.3 Antenna Usage Optimization 49

5.3.1 Illustrative Example 49

5.3.2 Network of Two Cells 52

5.3.3 Network of More than Two Cells 55

5.3.4 Practical Implications 57

5.4 Antenna Usage Optimization with Fairness 58

5.4.1 Client Starvation 59

5.4.2 Energy-per-stream Fairness 61

5.5 Channel Analysis for Beamforming Weight Optimization 63

5.5.1 Illustrative Example 63

5.5.2 Network of Two Cells 65

5.5.3 Network of More than Two Cells 68

5.5.4 Practical Implications 69

6 The Design of MACCO 70

6.1 Local Beamforming Optimization 71
6.2 Implicit Channel Acquisition 74
6.3 Evaluation 75
6.3.1 WARP Implementation 75
6.3.2 Results 80
6.4 Rate and Energy Efficiency Tradeoff 90

7 The Design of CoaCa 92

7.1 Interleaved Channel Sounding 92
7.2 Channel Reporting and Overhearing 96
7.3 Cell Clustering 98
7.4 Experimental Evaluation 100
7.4.1 WARP-based Implementation 100
7.4.2 Experimental Setup 102
7.4.3 Accuracy of Interference Cancellation and Interference Alignment 102
7.4.4 Achievable Rate Improvement 106
7.4.5 Tradeoff between Achievable Rate and Fairness 117

8 Conclusion 119

Bibliography 122

Illustrations

1.1 Radio frontend power consumption of wireless transmitter prototypes reported by papers from IEEE ISSCC and IEEE JSSC. The power efficiency of wireless radios has been greatly improved in the last ten years. 6

1.2 Receive beamforming by the client creates a virtual channel vector from the AP to the client, which is a weighted combination of the physical channel vectors to the client antennas. 9

1.3 An example with two APs each serving up to three clients. Coordinating the two cells to cancel the inter-cell interference delivers the maximum number of (four) streams in the MU-MIMO network. . . 11

2.1 Intra-cell and inter-cell interference cancellation in a MU-MIMO network. The AP uses three antennas to cancel the intra-cell interference between Client1 and Client2, and the inter-cell interference to Client3. Given two additional antennas Client3 could cancel the inter-cell interference from the AP, saving one antenna for the latter. Client4 employs interference alignment to align its channel to that of Client3, so that the inter-cell interference from the AP is naturally eliminated when the AP cancels the interference to Client3. 19

2.2 The channel sounding process in 802.11ac. First the AP sends a NDP-A frame and a NDP frame for all specified clients to estimate their downlink channels, and then each client sequentially replies with a BF-R frame containing the estimated channels. 23

4.1 Representations of a client's channel vector and an AP's beamforming weight vector. Employing the beamforming weight vector $\mathbf{w}$ by the AP projects the channel vector $\mathbf{h}$ to the direction of $\mathbf{w}$, leading to a reduced SNR at the client. 33

4.2 Orthogonality between two clients' channel vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, represented by the angle between them. 33

4.3 Receive beamforming weight optimization by the clients C1 and C2 in Case 1. Depending on the channel knowledge, the client maximizes its SNR before or after the signal projection by the zero-forcing beamforming AP. 39

5.1 The optimal use of the AP and client antennas in our illustrative example. The interference from AP2 to Client1 is cancelled by Client1, while the interference from AP1 to Client3 and Client4 are cancelled by Client3 and AP1, respectively. 50

6.1 Our experimental setup in a lab room for the evaluation of MACCO. An AP with eight antennas simultaneously serves up to eight MACCO clients each with four antennas. 79

6.2 Achievable rate for Omni, AntSel, MRC, Clustered MRC and MACCO as function of the number of clients, $\boldsymbol{K}$, with $\boldsymbol{M}$=8 antennas on the AP and low client SNR. MACCO achieves better rate scalability compared to others. 81

6.3 Achievable rate for Omni, AntSel, MRC, Clustered MRC and MACCO as function of the number of clients, $\boldsymbol{K}$, with $\boldsymbol{M}$=8 antennas on the AP and medium client SNR. MACCO achieves better rate scalability compared to others. 82

6.4 Achievable rate for Omni, AntSel, MRC, Clustered MRC and MACCO as function of the number of clients, $\boldsymbol{K}$, with $\boldsymbol{M}$=8 antennas on the AP and high client SNR. MACCO achieves better rate scalability compared to others. 82

6.5 (Simulated) rate achieved by MACCO and alternative approaches for large-scale MU-MIMO cells where $\boldsymbol{M}$=$\boldsymbol{K}$=16. 84

6.6 (Simulated) rate achieved by MACCO and alternative approaches for large-scale MU-MIMO cells where $\boldsymbol{M}$=$\boldsymbol{K}$=32. 85

6.7 Achievable rate improvement from MACCO with different cell scales ($\boldsymbol{M}$ and $\boldsymbol{K}$). 87

6.8 Rate achieved by AntSel, MRC, and MACCO with different number of client antennas ($\boldsymbol{N}$). 87

7.1 Timeline of CoaCa where two APs concurrently serve their clients. To provide each AP and client the necessary information to optimize its antenna usage, an AP starts polling its clients with the optimal order only after all the APs have transmitted their NDP-A and NDP frames. To optimize its beamforming weights based on the optimal antenna usage, a client overhears the reported virtual channels from other clients in the same cell that report before it. 93

7.2 Interference reduction from interference cancellation by the AP. The number of AP antennas does not have an obvious impact on the cancellation accuracy. 104

7.3 Interference reduction from interference cancellation by the client. The number of client antennas does not have an obvious impact on the cancellation accuracy. 104

7.4 Interference reduction from interference alignment by the client. The number of client antennas does not have an obvious impact on the alignment accuracy. 105

7.5 Rate achieved by MACCO/n+ and CoaCa for four cases with different numbers of antennas on APs and clients. CoaCa outperforms MACCO/n+ when it delivers more streams. However the rate improvement from CoaCa is lower than the expected multiplexing gain increase due to imperfect channel orthogonality. . . 107

7.6 One example experiment for case 4 with two APs and eight clients. The clients are close to not only the associated but also the interfering AP. However the transmit power of the APs is properly scaled down to mimic realistic Wi-Fi signal and interference strength. 111

7.7 Measured SNR/INR of realistic APs in two example indoor locations in a campus building: (top) first floor and (bottom) third floor. . . . 112

7.8 Probability distribution of the rate improvement by CoaCa, under the client SNR of 0-20 dB (top) and 20-40 dB (bottom) respectively. . 114

7.9 Rate achieved by 802.11n+, MACCO, and CoaCa with high client SNR (top) and low client SNR (bottom). CoaCa achieves higher rate gain with higher client SNR, due to the logarithmic relationship between rate and SNR. 116

Tables

4.1 Denotations used in this thesis. 36

5.1 The optimal use of the AP and client antennas in our illustrative example, where "→" and "- - →" indicate data communication and inter-cell interference cancellation, respectively. 52

5.2 Number of antennas on the APs and clients in the three cases, for the study of the client starvation issue and energy-per-stream fairness issue in antenna usage optimization. 59

7.1 Median measured SNR, INR, and SIRin multiple interference-limited real-world scenarios. 110

7.2 Number of antennas on the APs and clients for Case 1-4. 111

Chapter 1

Introduction

Modern wireless networks desperately need higher data rates to accommodate an increasing number of clients. This is especially true for 802.11 networks since the Wi-Fi interface is ubiquitous on modern mobile devices such as laptops, tablets, smartphones, and even wearable devices. Multi-user multiple-input multiple-output (MU-MIMO) is an important technology toward solving this data rate crisis in wireless networks and is appearing in newest generation of wireless deployment, e.g., 802.11ac [1] and LTE [2]. In a MU-MIMO network, an access point (AP)* with multiple antennas can leverage beamforming techniques such as zero-forcing beamforming to cancel the intra-cell interference between multiple clients and simultaneously serve them, achieving up to proportionally increased data rate. Theoretical analysis shows that with M antennas on the AP, up to K=M clients can be simultaneously served where each client receives a single spatial stream [3]. Therefore, when the MU-MIMO network is congested with enough clients, the achievable rate improvement from MU-

*We use "AP" to refer to the infrastructure node in a MU-MIMO network, such as the access point in 802.11 networks and the base station in LTE networks.

MIMO is bounded by the number of antennas on the AP.

1.1 Problem Statement: Rate Scalability of MU-MIMO Networks

The central problem that this thesis research seeks to solve is fundamental to MU-MIMO networks under zero-forcing beamforming: *rate scalability*. It is manifested in two ways. First, for a MU-MIMO network with a single cell, the achievable rate of the cell does not scale well to the number of clients [4]: as the number of simultaneously served clients increases and approaches the number of antennas on the AP, i.e., $K \to M$, the achievable rate often flattens and may even drop. This is because as the AP uses zero-forcing beamforming to cancel the intra-cell interference between simultaneously served clients, the SNR and achievable rate of each client decreases since the channel vectors of the clients are not perfectly orthogonal to each other. As the number of served clients increases, such client SNR decrease usually leads to a trivial achievable rate improvement, or even a rate decrease when the number of clients and antennas on the AP are relatively large (e.g., eight), and henceforth the rate scalability problem.

Second, for a MU-MIMO network with multiple interfering cells, the APs cannot simultaneously serve their clients with zero-forcing beamforming. This is because zero-forcing beamforming within a cell does not care about inter-cell interference. As a result, the APs in interfering cells have to avoid the inter-cell interference by serving their clients at different time or on different frequency bands. Such interference avoidance approach significantly limits the rate scalability of MU-MIMO to clients in multiple cells. Given these two fundamental problems of MU-MIMO networks, we raise the following question in this thesis work: *can we make the achievable rate of a MU-MIMO network more scalable to (i) clients in a single cell and (ii) clients in multiple cells?*

There are two existing approaches toward solving such rate scalability problem of MU-MIMO. The first is *Massive-MIMO* [5] of which the key idea is to equip many antennas on the AP to make sure $M>>K$. The extra degrees of freedom provided by the spare M-K antennas can be leveraged in two ways: they can be used to enhance the client SNR and thus compensate for the SNR loss from poor channel orthogonality, or cancel the inter-cell interference in order to allow multiple APs to serve their clients at the same time. Such approach, however, cannot be easily adopted by the majority of real-world APs that often have form factor, cost and power constraints. Meanwhile, to build Massive-MIMO APs one needs to address a set of new challenges such as collecting channel information without overwhelming overhead, making the architecture and processing scalable to the number of antennas, etc.. Up to date, only research prototypes based on software-defined-radio (SDR) platforms, such as [6] and [7], have been publicly reported. The second approach is based on *Network-MIMO* [8]. The key idea in Network-MIMO is to interconnect multiple interfering APs in close proximity with a backbone in order to form a giant virtual AP. Given the backbone inter-cell interference is converted into intra-cell interference, which can be naturally cancelled by the virtual AP with zero-forcing beamforming. In addition, the virtual AP possesses all the antennas from the interconnected physical APs, such that a few antennas are likely spare and can be leveraged to enhance the rate scalability of a single MU-MIMO cell similar to how Massive-MIMO achieves it. Despite its effectiveness, in practice Network-MIMO has equal or even more restricted applicability compared to Massive-MIMO, due to the demanding requirement of accurate time and frequency synchronization, as well as the bandwidth-hungry backbone connection between the physical APs for sharing transmitted symbols. Consequently, similar to Massive-MIMO, up to date only research prototypes based on Network-MIMO have been publicly reported [9, 10].

1.2 Client Beamforming for Rate Scalability

Our approach to solve the rate scalability problems of MU-MIMO is radically different from existing approaches: while Massive-MIMO and Network-MIMO seek to enhance the MU-MIMO AP by providing it spare antennas, our approach instead focuses on the clients that nowadays can be easily equipped with multiple antennas and leverages these antennas for *beamforming*. Due to the continuously reduced cost and power consumption of wireless transceivers, modern mobile devices can easily afford a small number of (up to four) antennas. Figure 1.1 shows the radio frontend (RF) power consumption of wireless transmitter prototypes reported by papers [11–48] from IEEE ISSCC [49] and IEEE JSSC [50]. Although the numbers are from research prototypes, we believe their trend is representative of their commercial counterpart. One can clearly observe that the power efficiency of wireless transmitters has been significantly improved in the past ten years, despite of certain fluctuations due to limited available data. Existing work has shown the feasibility for modern mobile devices to use their antennas for beamforming [51]. When properly optimized, client beamforming can complement the limited number of antennas on practical APs that cannot afford Massive-MIMO or Network-MIMO.

There are two ways to exploit client beamforming to achieve better rate scalability of MU-MIMO. First, to tackle the rate scalability problem in a single MU-MIMO cell, client beamforming can be exploited to improve the orthogonality between the channel vectors of the clients. Such channel orthogonality between clients determines the SNR and rate reduction from the intra-cell interference cancellation by the AP,

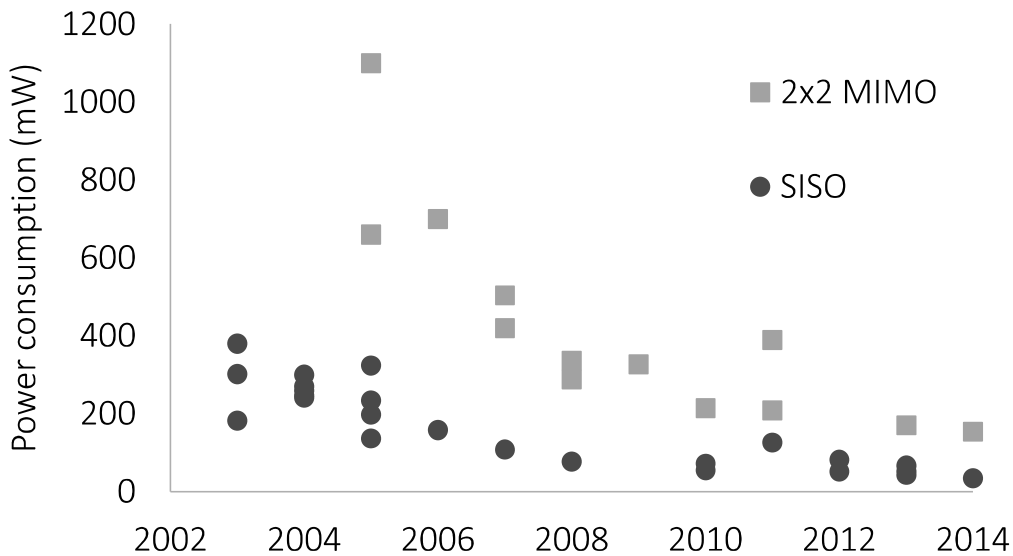


Figure 1.1 : Radio frontend power consumption of wireless transmitter prototypes reported by papers from IEEE ISSCC and IEEE JSSC. The power efficiency of wireless radios has been greatly improved in the last ten years.

and is therefore critical for the achievable rate of a MU-MIMO cell to become more scalable to the number of served clients. Second, to tackle the inter-cell interference problem in multiple MU-MIMO cells, client beamforming can be exploited together with AP beamforming to coordinately cancel the inter-cell interference. This thesis investigates both ways of using client beamforming to improve the rate scalability of MU-MIMO networks with one and multiple cells, respectively. We next introduce the two solutions resulting from this investigation.

1.2.1 Rate Scalability of a Single MU-MIMO Cell

Client beamforming can be exploited to compensate for the client SNR reduction due to the intra-cell interference cancellation by the zero-forcing beamforming AP. By combining the signals received at its multiple antennas with receive beamforming (a.k.a. post-combining), a client can leverage the physical channels between the AP and itself to create a *virtual channel*. To see how the virtual channel is created, let us consider the example in Figure 1.2 where an AP with two antennas transmits to two clients each with two antennas. We denote the physical channel vectors from the AP's antennas to each client's two antennas as $\mathbf{h}_1^1, \mathbf{h}_1^2$ and $\mathbf{h}_2^1, \mathbf{h}_2^2$, and each client's receive beamforming weight vector as $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively. The received signal after the client's beamforming is a weighted combination of the signal received at each antenna, whereas the weights are determined by the beamforming weight vector. Therefore, the signals from the AP can be considered being sent on a virtual channel from the antennas on the AP to a virtual antenna on each client. Given that the receive beamforming weight vectors of the clients are determined, the AP can perform zero-forcing beamforming simply based on their virtual channels. By creating the virtual channel, beamforming can improve the client SNR in two ways. First, by coherently combining the signals from multiple physical antennas, a stronger virtual channel with less attenuation can be produced. This is due to the combining gain from receive beamforming. More importantly, the virtual channel can be properly adjusted to be more orthogonal to the virtual channels of other clients. Such *channel orthogonality* is actually the key to make the achievable rate of a MU-MIMO cell more scalable to the number of served clients [52]. Existing solutions that leverage client antennas for receive beamforming [53–56] remain theoretical and do not consider the channel orthogonality to optimize the beamforming weight vectors. Instead, they employ either antenna selection or maximum ratio combining (MRC) to optimize the beamforming weights of the client. Though increasing the achievable rate compared to a single client antenna without beamforming, such beamforming solutions often lead to a trivial improvement of the rate scalability. The key reason is that the rate scalability of MU-MIMO is primarily determined by the channel orthogonality toward which neither antenna selection nor MRC can effectively improve.

We propose a novel solution that optimizes the client beamforming by considering the orthogonality between the client channel vectors. The solution works in a successive manner: it optimizes the clients' beamforming weight vectors one by one; for each client the optimization is based on the optimized virtual channel vectors of all previous clients. To calculate the optimal beamforming weight using given channels from previous clients, our solution balances two properties of the virtual channel for each client: the magnitude of the channel vector and the orthogonality of the chan-

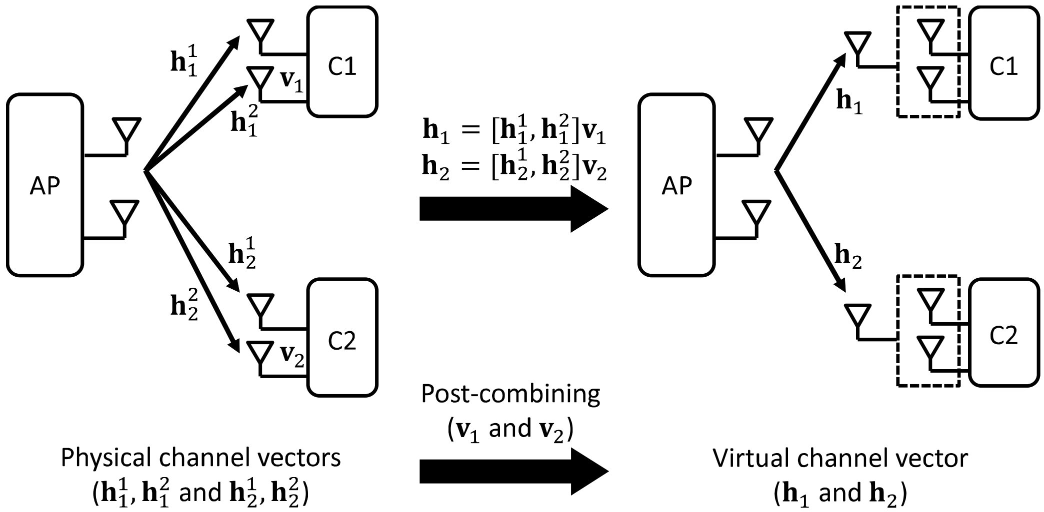


Figure 1.2 : Receive beamforming by the client creates a virtual channel vector from the AP to the client, which is a weighted combination of the physical channel vectors to the client antennas.

nel vector to previous clients' channels. Compared to antenna selection and MRC, our solution not only achieves better rate in the network but also makes the rate more scalable to the number of served clients; this is accomplished by successively leveraging the channels of the clients to improve their orthogonality between each other.

1.2.2 Combating Inter-cell Interference in Multiple MU-MIMO Cells

Client beamforming can be alternatively leveraged to assist the interfering APs to coordinately cancel inter-cell interference in a MU-MIMO network with multiple cells. When inter-cell interference is cancelled, multiple APs can simultaneously transmit to their served clients so that a greater number of spatial streams can be achieved in the MU-MIMO network that often leads higher achievable rate. To see how inter-cell interference can be cancelled by leveraging not only AP but also client beamforming, let us consider the example in Figure 1.3 with two interfering cells. With three antennas on each AP and three clients in each cell, if inter-cell interference is avoided instead of being cancelled, only a single AP (AP1 or AP2) is able to serve its clients at the same time where up to three streams in the MU-MIMO network can be enabled. Cancelling the inter-cell interference, however, can enable four streams, which is accomplished in the following way. First, each AP uses two antennas to deliver two streams to two clients (Client 1 and Client 3 in Cell 1, Client4 and Client6 in Cell 2), and the third antenna to cancel the inter-cell interference to one client in the other cell (Client1 in Cell 1, Client4 in Cell 2). Then, the other client in each cell (Client3 in Cell 1, Client6 in Cell 2) uses its three antennas to cancel the inter-cell interference from the two streams sent by each AP. As a result, we can activate four simultaneous streams in the network. Observe that both APs and clients have properly shared the responsibility of canceling the interference between them. In fact, such coordinated inter-cell interference cancellation by both APs and clients achieve the maximum number of streams in the MU-MIMO network. Then the question for us to answer is: *how to practically achieve such coordinated inter-cell interference cancellation in*

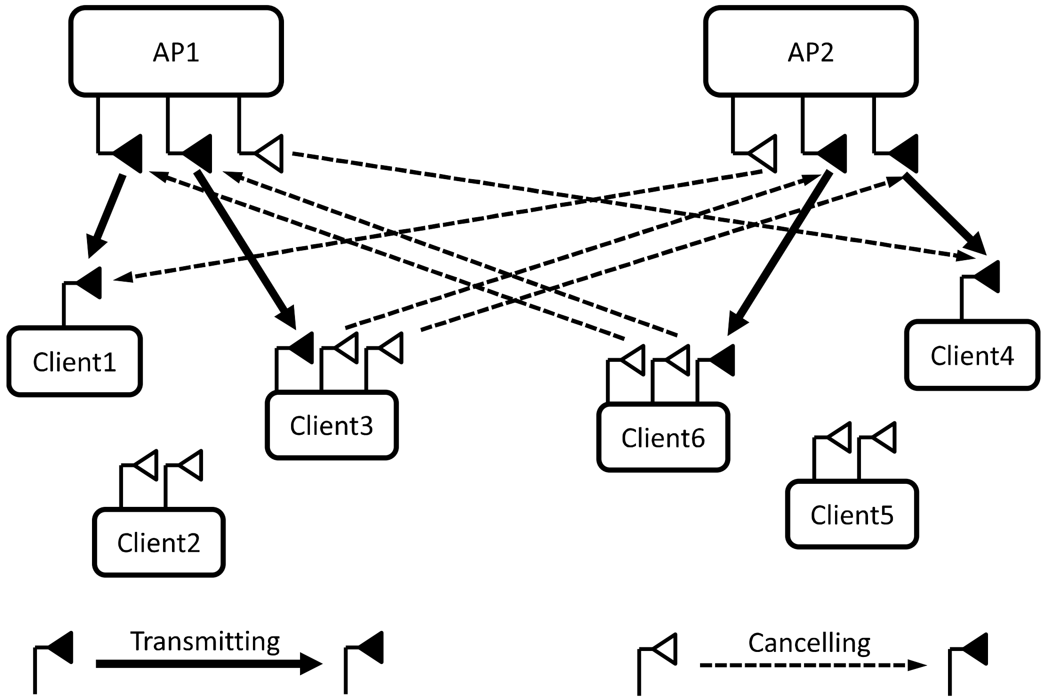


Figure 1.3 : An example with two APs each serving up to three clients. Coordinating the two cells to cancel the inter-cell interference delivers the maximum number of (four) streams in the MU-MIMO network.

a MU-MIMO network composed of multiple interfering cells?

Determining the beamforming weights for each AP and client that cancel inter-cell interference with coordination is a non-trivial process. The theoretically optimal solution that maximizes the achievable rate, or, achieves network capacity, can be empirically found by jointly optimizing the beamforming weights for all APs and clients. The optimal solution may not completely eliminate inter-cell interference since interference below the noise power is often no longer considered the capacity-limiting factor. However, identifying the optimal solution is impractical because (*i*) it is computationally intractable due to the lack of an analytical solution, and (*ii*) it has to be done in a centralized way with full channel knowledge of the entire MU-MIMO network.

To tackle these two issues in the theoretically optimal solution, we present a novel solution that allows practical cell coordination for inter-cell interference cancellation. The key idea in our solution is that the process of identifying the beamforming weights can be broken into two separate steps, namely *antenna usage optimization* and *beamforming weight optimization*. The first step determines how each AP and client antenna should be used: data communication or inter-cell interference cancellation. The optimal antenna usage of each AP and client collectively maximizes the number of streams in the MU-MIMO network. The second step determines the beamforming weights of each AP and client based on its optimized antenna usage. Given the use of antennas, it is possible to adopt practical beamforming techniques with a closed-form solution such as zero-forcing beamforming. The feasibility of such two-step optimization is based on an important heuristic we use to simplify the problem: we strive to completely eliminate inter-cell interference and maximize the multiplexing gain of MU-MIMO. Only with this heuristic the antenna usage can be reduced into a binary form including data communication and inter-cell interference cancellation, and optimized in a separate step prior to the optimization of beamforming weights.

1.3 802.11ac-based Realization

We next discuss the realization of our theoretical solutions provided in the previous chapter. Our realization is based on the 802.11ac protocol that supports MU-MIMO. Nonetheless, our theoretical solutions can be similarly applied to alternative venues such as LTE networks. In the following, we offer two 802.11ac-based protocols, MACCO and CoaCa, that solve the rate scalability problem in a single and multiple MU-MIMO cells respectively.

1.3.1 MACCO: Improving Rate Scalability of a Single 802.11ac Cell

To realize the solution in Chapter 1.2.1, we propose an 802.11ac-based protocol called MACCO. The design of MACCO is motivated by the channel sounding process used by an 802.11ac AP to acquire the downlink channels from all clients. In the channel sounding process, the MU-MIMO AP sends a sounding frame to all the clients and then the clients report their channels one by one. MACCO exploits this process in two parts to let a client optimize its beamforming weight vector. First, when a client responds to the sounding from the AP, , it reports the virtual channel that it locally calculates based on the optimized beamforming, instead of its physical channels. Second, other clients that have not yet reported their channels overhear the reported virtual channel from the client and use it to optimize their beamforming for improved channel orthogonality. Because of the distributed nature of the beamforming weight optimization in MACCO, the clients do not need to report the lengthy physical channels to the AP and the AP does not have to send the centrally optimized beamforming weight vectors back to the clients, both of which would incur considerable overhead [57]. Since channel reporting by clients is mandated by 802.11ac, MACCO does not incur any overhead by letting each client overhear the channel knowledge broadcast on the wireless medium. Notably, the successive overhearing in MACCO provides each client just enough channel knowledge to perform the beamforming weight optimization in our solution.   we realize a prototype of MACCO using the WARP platform [58], and experimentally evaluate its performance in real-world indoor wireless environments. The experimental results indicate that compared to existing solutions, on average MACCO can boost the achievable rate by 35% for a MU-MIMO cell with eight AP antennas and eight clients. While the rate improvement is modest, with only four antennas on each client, MACCO can near-proportionally increase the rate as the MU-MIMO AP serves more clients, for a majority of SNR regimes and cell topologies in our experiments.

1.3.2 CoaCa: Combatting Inter-cell Interference in Multiple 802.11ac Cells

For our theoretical solution reported in Chapter 5, the separation of the antenna usage optimization and beamforming weight optimization greatly simplifies the cell coordination effort, allowing us to practically integrate such two-step optimization into 802.11ac. Our proposed protocol, called CoaCa, leverages the channel sounding process in 802.11ac to let each AP and client locally perform the two-step optimization in a distributed way. CoaCa includes two key designs. First, CoaCa interleaves the channel sounding process from all the APs, so that each node can easily acquire the necessary global information to optimize the use of their antennas. Such information including the number of antennas on each node only needs a few bits to be represented, and therefore each AP can explicitly share it. Second, by reporting and overhearing the virtual channels in the interleaved channel sounding, each node can obtain just enough channel knowledge to optimize its beamforming weights. CoaCa incurs negligible overhead over 802.11ac and compatibly works with unmodified 802.11ac clients. While the current design of CoaCa only allows downlink MU-MIMO which is consistent to 802.11ac, the optimized beamforming weights for the APs and clients can be also used for uplink MU-MIMO leveraging channel reciprocity. However, realizing uplink MU-MIMO faces a new set of challenges such as misaligned symbol timing and clock frequency offset between clients, which are studied by other prior work, e.g., [59] and outside the scope of this thesis work. We realize a prototype of CoaCa on WARP and evaluate its performance in realistic indoor wireless environments. Our experimental results show that on average CoaCa is able to improve the achievable rate of a two-cell MU-MIMO network by 41% to 52%, with up to four antennas on each AP and client. Even though the achievable rate improvement is lower than the multiplexing gain increase (50% to 67%), CoaCa considerably outperforms existing solutions that often only allow a single cell to operate.

1.4 Summary of Contributions

This thesis makes the following contributions.

First, for the rate scalability problem of MU-MIMO, we study its manifestation in two types of MU-MIMO networks, with a single and multiple MU-MIMO cells, respectively.

Second, we provide two theoretical solutions that rely on client beamforming to tackle the rate scalability problem of MU-MIMO. We show that they outperform existing theoretical solutions and can be readily realized in practical MU-MIMO networks.

Third, we offer two protocols that realize our theoretical solutions in 802.11ac networks. Our protocol leverage the overhearing opportunity in the channel sounding process of 802.11ac, to allow local optimization of client beamforming without overhead.

Finally, we provide WARP-based implementations of our protocols, and experimentally evaluate their performance in real-world indoor environments. Our measurement results show the capability of our protocols to greatly improve the rate scalability of MU-MIMO in a single or multiple 802.11ac cells.

Chapter 2

Background

This chapter provides necessary background regarding MU-MIMO and 802.11ac.

2.1 Interference Cancellation in MU-MIMO

MU-MIMO increases the achievable rate of the network by achieving a *multiplexing gain*, defined as the number of concurrent streams or simultaneously served clients. To appreciate the multiplexing gain in a MU-MIMO network composed of multiple interfering cells, both intra-cell and inter-cell interference must be sufficiently suppressed. With multiple antennas, an AP and a client can *pre-code* and *post-combine* the signal sent or received at each antennas, by applying a complex weight to the signal. The set of weights is denoted as the *pre-coding vector* and *post-combining vector*, respectively.

Pre-coding and post-combining can be either linear or non-linear. For a single MU-MIMO cell, capacity-achieving schemes such as dirty paper coding are non-linear, and prohibitively expensive to implement in practice [60, 61]. As a result, the focus of this work is on linear pre-coding schemes that are practical. Linear pre-coding and

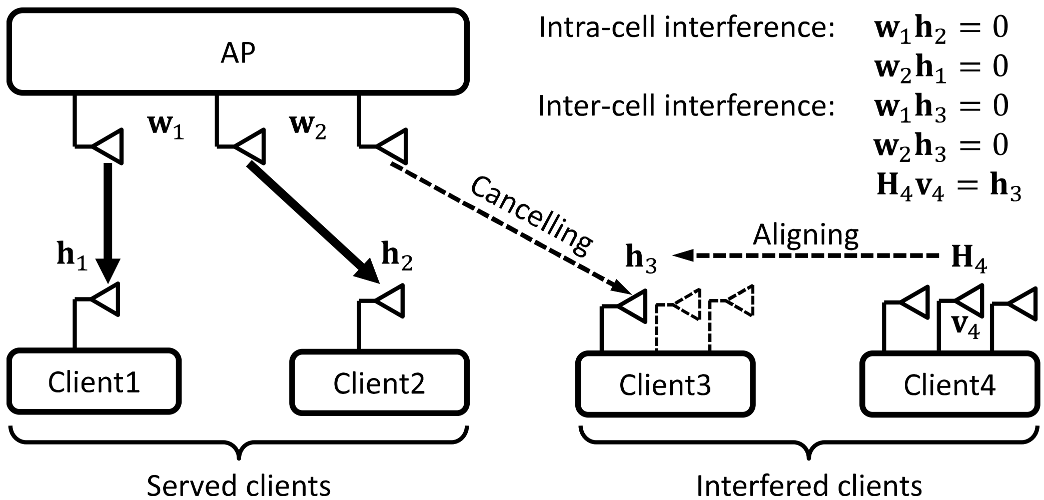


Figure 2.1 : Intra-cell and inter-cell interference cancellation in a MU-MIMO network. The AP uses three antennas to cancel the intra-cell interference between Client1 and Client2, and the inter-cell interference to Client3. Given two additional antennas Client3 could cancel the inter-cell interference from the AP, saving one antenna for the latter. Client4 employs interference alignment to align its channel to that of Client3, so that the inter-cell interference from the AP is naturally eliminated when the AP cancels the interference to Client3.

post-combining are also called *beamforming*.

Intra-cell interference cancellation. To cancel the intra-cell interference between multiple clients, an AP can leverage *zero-forcing beamforming* (ZFBF) to pre-code the signal to each individual client. One must orthogonalize the downlink channels of the simultaneously served clients such that each client only receives its own data stream without interference. Zero-forcing beamforming achieves such orthogonality in the following way: it selects the transmit beamforming weight vectors of the AP, $\mathbf{w}_j$, to make sure the channel vectors from the AP to the clients i.e., $\mathbf{h}_k$, are orthogonal, or

$$\mathbf{w}_j\mathbf{h}_k{=}0 \ (j \neq k).$$

To satisfy the orthogonality constraints, the dimension of $\mathbf{w}_j$, which is the number of antennas on the AP, must be no smaller than the number of served clients. This is the maximum multiplexing gain that can be achieved in a single MU-MIMO cell.

When the AP does not cancel intra-cell interference, it can leverage *conjugate beamforming* (ConjBF) to maximize the SNR at the served client, or

$$\mathbf{w}_k = \mathbf{h}_k^*.$$

When the client has multiple antennas and uses them for receive beamforming, the channel between the AP and client becomes a matrix, $\mathbf{H}_k$, and the AP should employ *eigen beamforming* (EigenBF) to maximize the client SNR, or

$$\mathbf{w}_k = \mathbf{v}_{max}(\mathbf{H}_k^* \mathbf{H}_k)$$

where $\mathbf{v}_{max}()$ is the eigenvector of a matrix corresponding to the largest eigenvalue. EigenBF can be conceptually considered to decompose the channel matrix $\mathbf{H}_k$ into a set of orthogonal eigen channels, whose SNR is represented by the corresponding eigenvalue. Naturally, to maximize the client SNR the AP transmits signals on the best eigen channel, $\mathbf{v}_{max}(\mathbf{H}_k^* \mathbf{H}_k)$.

Inter-cell interference cancellation. One can extend zero-forcing beamforming for the AP to cancel inter-cell interference. That is, if there are any spare antennas on the AP after cancelling the intra-cell interference, one can use them to orthogonalize the channel vectors of the clients in other cells in the same way. Note that in the proper context without introducing ambiguity, we use *antenna* to refer to the *Degree of Freedom* (DoF) provided by a physical antenna on an AP or a client. In Figure 2.1, since there are three antennas, the AP uses two of them to cancel the intra-cell interference between Client1 and Client2, and the third (spare) one to cancel the inter-cell interference to Client3 in the interfering cell.

Alternatively, a client can also cancel inter-cell interference if the client features multiple antennas for receive beamforming, a.k.a., post-combining [51]. Note that receive beamforming actually allows a client to separate and recover both the intended and interference streams. For ease of explanation we simply use the term *interference cancellation* to denote such capability of a receive beamforming client. To cancel the inter-cell interference, the client j chooses its receive beamforming weight vector, a.k.a. the post-combining vector, $\mathbf{v}_j$, such that

$$\mathbf{w}_k \mathbf{H}_j \mathbf{v}_j = 0$$

for all served clients k by the AP, where $\mathbf{H}_j$ is the channel matrix from the AP to client j. For example, in Figure 2.1 if Client3 had two spare antennas they could be used to cancel the two streams from the AP to Client1 and Client2, saving the third antenna on the AP. This would provide the AP with the flexibility to use its third antenna to potentially serve another client.

Another technique for dealing with inter-cell interference is *interference alignment* [62–65]. The key idea in interference alignment is to align the channel vectors of multiple clients, i.e.,

$$\mathbf{H}_j \mathbf{v}_j = \mathbf{H}_k \mathbf{v}_k,$$

so that the interference between the AP and these clients traverses a single aligned channel and requires fewer AP antennas to be cancelled. Such alignment can be conceptually understood as if the client used its own antennas to cancel the interference saving the antennas on the AP. Note that interference alignment is more commonly

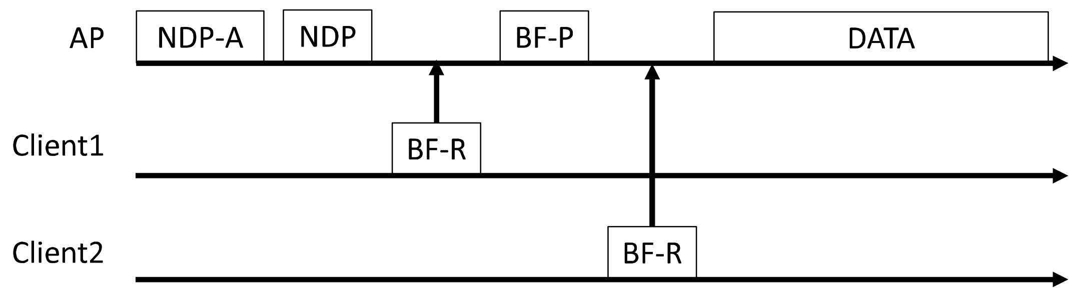


Figure 2.2 : The channel sounding process in 802.11ac. First the AP sends a NDP-A frame and a NDP frame for all specified clients to estimate their downlink channels, and then each client sequentially replies with a BF-R frame containing the estimated channels.

assumed for uplink transmission, e.g., multiple clients transmit to an AP they interfere with on an aligned channel. In this work we leverage channel reciprocity to apply interference alignment to downlink MU-MIMO in which the AP transmits to multiple clients it interferes with on the same aligned channel. In Figure 2.1, without interference alignment, the AP would need two spare antennas to cancel the inter-cell interference to both Client3 and Client4. When Client4 aligns its channel to that of Client3 by setting $\mathbf{H}_4\mathbf{v}_4=\mathbf{h}_3$, the AP only needs one antenna to cancel the inter-cell interference.

2.2 MU-MIMO in 802.11ac

We next briefly introduce the MU-MIMO feature supported by 802.11ac [1], the latest amendment to the 802.11 protocol family. 802.11ac allows the AP to use MU-MIMO techniques to simultaneously transmit downlink streams to up to four of its served clients. For APs in the interference range of each other, 802.11ac does not allow them to transmit at the same time; instead, the APs contend to access the medium using CSMA/CA.

Channel knowledge is necessary for the AP to calculate the transmit beamforming weights that cancel the intra-cell interference. To acquire channel knowledge, 802.11ac mandates an explicit channel sounding process, which we show in Figure 2.2. To sound the channel, the AP first broadcasts a *Null Data Packet Announcement* (NDP-A) frame. The purpose of the NDP-A frame is to specify the set of clients the AP is about to serve, and notify them to prepare for estimating and reporting their downlink channels. After the NDP-A frame, the AP sends a *Null Data Packet* (NDP) frame that allows the clients to estimate their downlink channels leveraging the training symbols in the frame. Then, the specified clients sequentially report their estimated channels to the AP with a *Beamforming Report* (BF-R) frame. The NDP-A frame designates a client that must immediately reply with the BF-R frame after the NDP frame, while other clients must wait for the *Beamforming Report Poll* (BF-P) frame from the AP to respond. The explicit channel sounding process in 802.11ac does not require channel reciprocity which is needed by implicit channel estimation [6]. The inter-frame interval in the channel sounding process is *SIFS* (16 μs), which is shorter than *DIFS* (34 μs) and thereby provides the APs and clients guaranteed medium access without being intervened by other 802.11 nodes.

Chapter 3

Related Work

In this chapter, we discuss related work. MU-MIMO has been demonstrated practical in real-world wireless environments, e.g., [6, 7, 9, 10, 52, 59, 66–68]. Modern wireless standards such as 802.11ac [1] have adopted MU-MIMO to increase the achievable rate of the network.

3.1 Rate Scalability of a MU-MIMO Cell

Researchers have observed the rate scalability problem in a single MU-MIMO cell from experiments. The authors in [66] observed that when a zero-forcing beamforming AP with four antennas increases its number of served clients from one to four, the achievable rate of each client drastically drops. The authors in [6] found out that even when the zero-forcing beamforming AP has sixteen antennas, the achievable rate of the network starts to decrease with more than twelve clients. The authors in [52] observed that the per-client SNR reduction from MU-MIMO is determined by the orthogonality between channels, and proposed a method for the clients to evaluate such channel orthogonality and properly choose the data rate that does not exceed the channel capacity.

There are three major approaches toward solving the rate scalability problem in a single MU-MIMO cell. The first approach is to let the AP group clients into clusters, only serve the clients in the same cluster at a time, and schedule clusters in a TDMA fashion [64,69]. However, such an approach does not fully exploit the spatial multiplexing gain of MU-MIMO: the rate improvement is bounded by the reduced number of clients in a cluster. In 802.11ac networks, the overhead from channel sounding and channel contention further amplify the inefficiency of such clustering approach, as the AP needs to contend for and sound the channel for each individual cluster.

The second approach is to increase the number of antennas on the AP, or Massive-MIMO [5]. Argos proposed in [6] features many antennas (M=64) on the base station (AP) to ensure M>>K. BigStation proposed in [7] similarly adopts up to twice as many antennas as clients. However, Massive-MIMO is not suitable for real-world 802.11ac APs that usually have a small form factor or limited cost and energy budget. The authors in [9] and [10] adopt Network-MIMO [8] to connect multiple adjacent APs as a giant virtual AP. Since the virtual AP possesses the antennas from all connected APs, it is able to serve a larger number of clients simultaneously. However, Network-MIMO requires well planned AP deployment and fine-grained time and frequency synchronization between the connected APs, and therefore cannot be applied to 802.11ac networks that are distributed and uncontrolled.

The third approach leverages multiple client antennas for receive beamforming as we do in MACCO, but existing solutions based on it [53–56, 70] remain theoretical, do not consider channel orthogonality between clients, and suffer from two practical concerns. First, the solutions are often heuristic with expensive iterations, and thus hard to implement in practice. Second, the optimization has to be centrally executed by the AP with full channel information. Unfortunately, to enable this in 802.11ac networks, one has to radically modify the 802.11ac protocol to let the AP optimize the beamforming weight vectors for the clients, and explicitly deliver the optimization results to the clients. More importantly, such centralized optimization also incurs considerable overhead to the channel sounding of 802.11ac . In contrast, by letting the clients overhear other clients' virtual channels to improve the channel orthogonality, MACCO allows local, per-client beamforming weight optimization in a distributed, 802.11ac-compliant way.

3.2 Inter-cell Interference in a MU-MIMO Network

Inter-cell interference as an important capacity limiting factor for MU-MIMO networks remains experimentally under-explored.

3.2.1 Network-MIMO- and CoMP-based Solutions

An effective but practically expensive way to tackle inter-cell interference in MU-MIMO networks is to let neighbouring APs collaborate with each other via a high-speed, low-latency connection, e.g., [9, 10, 64, 71, 72]. In LTE-based cellular networks, such technique is often referred to as coordinated multi-point (CoMP), e.g., [73–75]. Note that neighbouring base stations in cellular networks are already connected via the backbone; therefore the effort of implementing CoMP-based solutions is not as massive as that in Wi-Fi networks that are often distributed. In 802.11-based Wi-Fi networks, there are different ways of exploiting such AP connection with various implementation complexities. Network-MIMO [9, 10] allows multiple neighbouring APs to behave as a single massive AP, by synchronizing their time and frequency, and letting them share transmitted and received samples. Such technique provides similar functionality as CoMP does, but requires highly accurate inter-AP synchronization and high-speed inter-AP connection. IAC proposed in [64] only needs the APs to share the samples for joint encoding and decoding without synchronization. Robinhood proposed in [71] requires a set of APs to retransmit the received packets from the clients to the rest of APs, while cancelling the interference between the packets. OpenRF proposed in [72] uses a central controller to coordinate neighbouring APs to cancel the inter-cell interference. Due to the requirement of AP connection and a centralized solution, the applicability of these approaches is restricted to carefully planned and centrally controlled networks such as an enterprise 802.11ac network. By leveraging multi-antenna clients to assist the interfering AP to coordinately cancel the inter-cell interference, CoaCa is distributed and works in any 802.11ac networks.

NEMOx proposed in [76] seeks to enable efficient spatial reuse in distributed MIMO networks, by proposing a scalable architecture that connects a smaller number of neighbouring APs for clustered Network-MIMO. Each cluster includes a giant virtual AP formed by multiple connected APs, and contends with each other via asynchronous CSMA. Clearly, one can apply MACCO to a single cluster in NEMOx. However, within the interference range, no more than one virtual AP is allowed to operate at the same time. CoaCa similarly clusters the APs and assumes CSMA between clusters, but employs a fundamentally different approach to address the interference inside a cluster. Instead of connecting the APs for Network-MIMO, CoaCa relies on AP and client beamforming to coordinately cancel the inter-cell interference.

3.2.2 Distributed Solutions

802.11n+ presented in [65] leverages beamforming to enable concurrent streams in 802.11n networks with SU-MIMO. It allows nodes in one cell to transmit and receive first, and nodes in another cell to opportunistically transmit and receive by cancelling the interference to and from the first cell. Nodes in the second cell must overhear the ongoing transmission from nodes in the first cell, to acquire the necessary channel and use it for interference cancellation. Notably, the interference cancellation in 802.11n+ is uncoordinated and one-way: nodes in the overhearing cell must use their spare antennas to cancel the inter-cell interference. The number of spare antennas on nodes in the overhearing cell should be no smaller than that of the ongoing streams in the overheard cell. Since an 802.11n link with SU-MIMO usually includes one or two streams, each node in the overhearing cell only needs one or two spare antennas. In an 802.11ac-based MU-MIMO network where an AP can simultaneously transmit to K clients (e.g., K=4), in order to increase the number of streams by starting concurrent streams in the overhearing cell, 802.11n+ would require each AP and client in it to have at least K spare antennas to cancel the inter-cell interference, which is usually infeasible in practice. The fundamental reason why 802.11n+ does not effectively extend to 802.11ac is its uncoordinated, one-way inter-cell interference cancellation, where nodes in the overhearing cell have to contribute all the required antennas to solely carry the burden of cancellation. By achieving coordinated, two-way inter-cell interference cancellation, CoaCa can enable more streams than 802.11n+ does in 802.11ac networks, while still requiring small protocol modifications and incurring negligible overhead.

Chapter 4

Rate Scalability of a Single MU-MIMO Cell

In this chapter, we study the rate scalability problem of a single MU-MIMO cell. For MU-MIMO networks with multiple interfering cells, we assume the inter-cell interference is avoided as multiple MU-MIMO APs serve their clients either at different time or on different channels. Such interference avoidance approach toward inter-cell interference is in accordance to 802.11ac [1]. In Chapter 5, we will assume the interfering APs operate on the same channel, and seek to allow them to transmit at the same time, by cancelling the inter-cell interference instead of avoiding it.

4.1 Problem Statement

The simultaneous transmissions from a MU-MIMO AP to multiple clients often do not improve the achievable rate in the network in proportion to the number of served clients. As the number of clients approaches the number of antennas on the AP, the achievable rate often improves insignificantly, and may even decrease [6, 7, 66]. The key reason is that as the AP pre-codes the signals from its multiple antennas to eliminate the intra-cell interference, the SNR of each served client decreases. This is

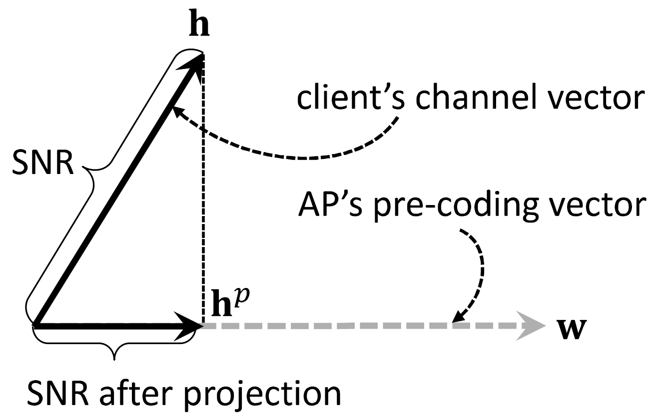


Figure 4.1 : Representations of a client's channel vector and an AP's beamforming weight vector. Employing the beamforming weight vector $\mathbf{w}$ by the AP projects the channel vector $\mathbf{h}$ to the direction of $\mathbf{w}$, leading to a reduced SNR at the client.

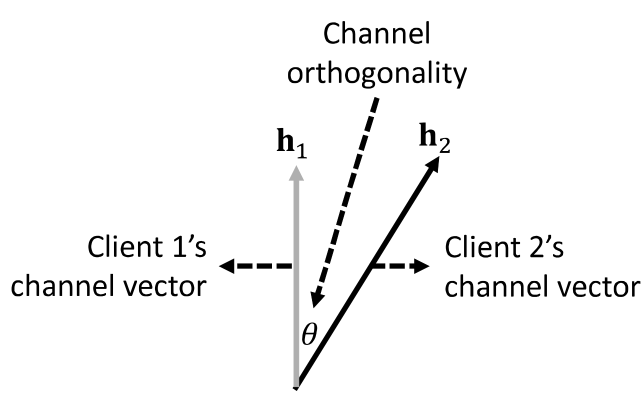


Figure 4.2 : Orthogonality between two clients' channel vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, represented by the angle between them.

because one can consider transmit beamforming by the AP as *projecting the client's channel vector* $\mathbf{h}$ *to the direction of the AP's beamforming weight vector* $\mathbf{w}$. As a result, to the client, the resultant channel is the projected channel, $\mathbf{h}^p$, as shown in Figure 4.1. In the figure,

$$\mathbf{h}^p = \mathbf{h} \cdot \cos\langle \mathbf{h}, \mathbf{w} \rangle,$$

where "$\langle \cdot \rangle$" refers to the angle between two vectors. Clearly, when $\mathbf{w}$ is aligned with $\mathbf{h}$, it preserves the maximum channel SNR; when $\mathbf{w}$ is perpendicular to $\mathbf{h}$, it completely nullifies the signal. Such relationship can be extended to that between two channel vectors, which we refer to as the *channel orthogonality.* For the channel vectors of two clients, $\mathbf{h}_1$ and $\mathbf{h}_2$, their orthogonality can be represented by the angle between them, or

$$\theta = \arccos \frac{\mathbf{h}_1 \cdot \mathbf{h}_2}{||\mathbf{h}_1|| ||\mathbf{h}_2||},$$

as shown in Figure 4.2.

The channel orthogonality between clients is critical for zero-forcing beamforming, since to cancel the intra-cell interference, the AP must set $\mathbf{w}_1 = \mathbf{h}_2^{\perp}$ and $\mathbf{w}_2 = \mathbf{h}_1^{\perp}$. If $\mathbf{h}_1$ and $\mathbf{h}_2$ are perfectly orthogonal, zero-forcing beamforming retains the maximum channel SNR and can linearly scales up the achievable rate of the MU-MIMO network.

Otherwise, as the AP serves more clients, zero-forcing beamforming reduces the rate of each client.

Realistic MU-MIMO channels usually exhibit random, unpredictable channel orthogonality. The authors in [66] reported that there is no clear correlation between the channel orthogonality and the distance between clients. The channel orthogonality can be very poor even when the clients are placed at least a few wavelengths from each other. This is because typical indoor environments have rich scattering, and the radio hardware has intrinsic randomness that changes the phase of the received signal. The authors in [52] observed that the channel orthogonality can vary from frame to frame, and consequently proposed a solution that the clients use to evaluate the channel orthogonality between each other and perform rate selection.

The central question we seek to answer in this chapter is: *how do we leverage client beamforming to improve the channel orthogonality and thus rate scalability of a single MU-MIMO cell?*

4.2 Successive Beamforming Optimization

We next discuss how clients in a single MU-MIMO cell optimize their receive beamforming weight vector.

The key idea in our solution is that *clients successively perform their beamforming*

Table 4.1 : Denotations used in this thesis.

Symbol	Definition
M	Number of antennas on the AP
K	Number of simultaneously served clients
N	Number of antennas on the client
$\mathbf{h}_k$	Virtual channel vector of the kth client
$\mathbf{h}_k^n$	Physical channel vector of the kth client's nth antenna
$\mathbf{h}_{k,j}$	The jth virtual channel vector of the kth client
$\mathbf{w}_k$	Transmit beamforming weight vector of the AP for the kth client
$\mathbf{v}_k$	Receive beamforming weight vector of the kth client
$\mathbf{H}_k$	Physical channel matrix from the AP to the kth client

optimization based on the virtual channels from previous clients. As shown in Chapter 1, with receive beamforming, a client creates a virtual channel from the AP to itself. The orthogonality between the clients' virtual channels determines the rate scalability. We provide a closed-form solution to the optimal beamforming weight vectors that maximize the client's achievable rate, given by the following equation:

$$\mathbf{v}_{k,j} = \mathbf{H}_k\mathbf{H}_L^{\perp}\mathbf{v}_{jmax}((\mathbf{H}_k\mathbf{H}_L^{\perp})^*(\mathbf{H}_k\mathbf{H}_L^{\perp})), \tag{4.1}$$

where "$*$" stands for conjugate transpose, and other denotations are provided in Table 4.1 and Chapter 4.2.4. We will derive this result by analyzing three representative cases.

4.2.1 Cases 1: M=2 and K=2

We first study a simple case where an AP with two antennas serves two clients (C1 and C2) each with two antennas. To see how multiple antennas can benefit the client, let us first assume both C1 and C2 employ antenna selection and only use a single antenna, with the corresponding channel vector $\mathbf{h}_1$ and $\mathbf{h}_2$. We denote the transmit beamforming weight vector for C1 and C2 employed by the AP as $\mathbf{w}_1$ and $\mathbf{w}_2$. Then, the SNR of C1 and C2 normalized to its noise power can be expressed as

$$\rho_1 = ||\mathbf{w}_1 \cdot \mathbf{h}_1||^2 = ||\mathbf{h}_1||^2 \cos^2\langle \mathbf{w}_1, \mathbf{h}_1 \rangle,$$

$$\rho_2 = ||\mathbf{w}_2 \cdot \mathbf{h}_2||^2 = ||\mathbf{h}_2||^2 \cos^2\langle \mathbf{w}_2, \mathbf{h}_2 \rangle,$$

where $||\mathbf{w}_1||^2$=$||\mathbf{w}_2||^2$=1 due to power normalization. To cancel the intra-cell interference between C1 and C2 with zero-forcing beamforming, the AP must employ $\mathbf{w}_1$=$\mathbf{h}_2^{\perp}$ and $\mathbf{w}_2$=$\mathbf{h}_1^{\perp}$. Such intra-cell interference cancellation reduces ρ_1 and ρ_2, since $\mathbf{w}_1$ and $\mathbf{w}_2$ are usually not aligned with $\mathbf{h}_1$ and $\mathbf{h}_2$.

When C1 and C2 use both antennas, they can beamform the received signals. Visually, such receive beamforming "stretches" and "steers" the physical channel vectors of the client to create a virtual channel vector. For clarity, when the clients use more than a single antenna, we use $\mathbf{h}_k^n$ to denote the physical channel vectors of the nth antenna at the kth client, and $\mathbf{h}_k$ the virtual channel vector of the kth client.

Before discussing the beamforming weight optimization by C1 and C2, we need to clarify that in this case the same number of streams can be achieved by letting the AP only serve a single client, say C1, with single-user MIMO (SU-MIMO). However, SU-MIMO similarly has a rate scalability problem to the number of streams, since the SNR of each stream decreases when C1 decodes the streams. Such SNR reduction depends on the orthogonality between the two physical channel vectors $\mathbf{h}_1^1$ and $\mathbf{h}_1^2$. This is also known as the "antenna correlation" that reduces the capacity of SU-MIMO. Without extra antennas, C1 cannot improve the channel orthogonality between the two streams. Therefore, we argue that the AP should employ MU-MIMO

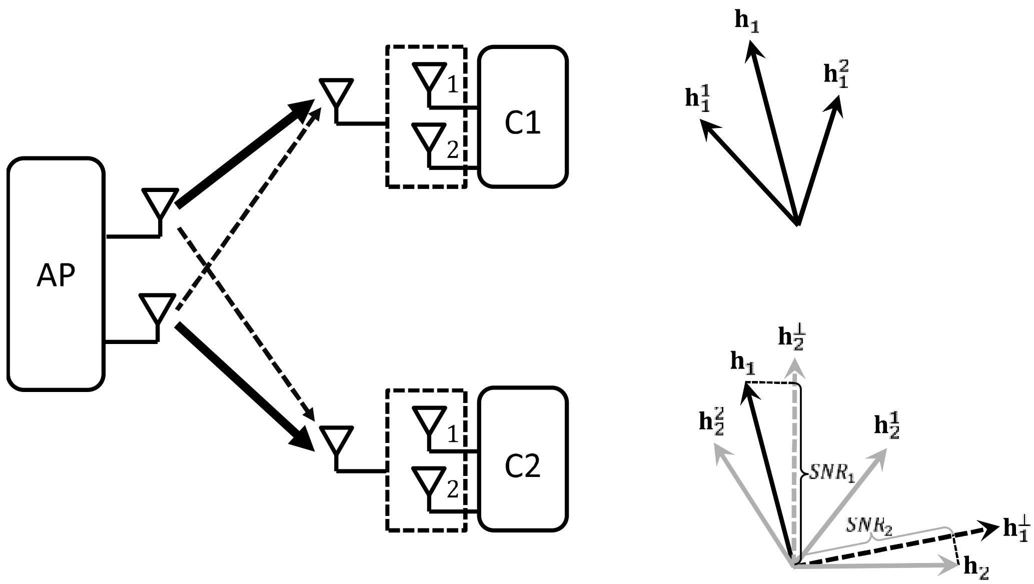


Figure 4.3 : Receive beamforming weight optimization by the clients C1 and C2 in Case 1. Depending on the channel knowledge, the client maximizes its SNR before or after the signal projection by the zero-forcing beamforming AP.

to serve both clients instead of one of them with two streams, in order to leverage the receive beamforming gain from the client antennas. Such argument is similarly proposed in [77] to exploit client diversity.

Optimization by C1. We first discuss the optimization performed by C1. We assume it first reports the virtual channel to the AP without knowing $\mathbf{h}_2$. Therefore, C1 optimizes its beamforming weight vector, $\mathbf{v}_1$, to maximize its SNR solely based on its physical channel vector $\mathbf{h}_1^1$ and $\mathbf{h}_1^2$. To achieve this, C1 employs maximum ratio combining:

$$\mathbf{v}_1 = \mathbf{H}_1 \mathbf{v}_{max}(\mathbf{H}_1^* \mathbf{H}_1),$$

where $\mathbf{H}_1$=$[\mathbf{h}_1^1, \mathbf{h}_1^2]^{\mathrm{T}}$. The resultant virtual channel vector, $\mathbf{h}_1$, is a weighted combination of $\mathbf{h}_1^1$ and $\mathbf{h}_1^2$ with maximum magnitude, as shown in Figure 4.3. From the figure, observe that the beamforming gain from maximum ratio combining creates a virtual channel that is stronger than either of the two physical channels. However, the channel orthogonality between C1 and C2 may not be improved since C1 is unaware of the direction of C2's virtual channel. Nonetheless, given its limited channel knowledge, C1 has to employ maximum ratio combining wishing a resultant SNR improvement.

Optimization by C2. Now let us switch the focus to C2. We assume it optimizes its beamforming weight vector and reports to the AP after C1. As a result, C2 can overhear the BF-R frame from C1, and acquire the knowledge of the virtual channel $\mathbf{h}_1$. Then, C2 knows the direction that the AP will project its signal onto, $\mathbf{h}_1^{\perp}$, and can maximize its SNR after such projection. To do this, C2 employs:

$$\mathbf{v}_2 = (\mathbf{h}_1^{\perp}\mathbf{H}_2)^*,$$

where $\mathbf{H}_2$=$[\mathbf{h}_2^1, \mathbf{h}_2^2]^{\mathrm{T}}$. As shown in Figure 4.3, the resultant virtual channel vector, $\mathbf{h}_2$, is closer to being orthogonal to $\mathbf{h}_1$, even though its magnitude is not maximized. However, after the AP's signal projection, $\mathbf{h}_2$ is guaranteed to have the maximum magnitude and the SNR of C2 is maximized thanks to the knowledge of $\mathbf{h}_1$. In addition, we can see the SNR increase of C1 due to improved channel orthogonality.

To summarize, in this case where M=2 and K=2, a client can leverage beamforming to adjust its virtual channel vector such that its SNR is maximized before or after the AP's signal projection, depending on whether the client has the knowledge of each other's virtual channel.

4.2.2 Cases 2: M=3 and K=3

Now let us consider the case where an AP with three antennas serves three clients each with two antennas. Consequently, the physical or virtual channel vectors reside in a three-dimensional space. Compared to the previous case we do not change the number of antennas on the clients, N, since its impact on the beamforming weight optimization is trivial as we will show in the following. We similarly assume C1, C2 and C3 successively send their BF-R frames to the AP to report their optimized virtual channels.

Optimization by C1. For C1 who does not know $\mathbf{h}_2$ and $\mathbf{h}_3$, its optimization is similar to that in the previous case, i.e., $\mathbf{v}_1$=$\mathbf{H}_1\mathbf{v}_{max}(\mathbf{H}_1^*\mathbf{H}_1)$ whereas N only changes the dimension of $\mathbf{H}_1$.

Optimization by C2. We then discuss how C2 performs the optimization. Similar to the previous case, C2 has the knowledge of $\mathbf{h}_1$ by overhearing the BF-R frame from C1. However, since the channel vectors are in a three-dimensional space, $\mathbf{h}_1^{\perp}$ becomes a two-dimensional plane, represented by a 2×2 matrix instead of a vector. After $\mathbf{h}_2$ is projected onto this plane, it will be further projected to a line that is orthogonal to the channel vector of C3, $\mathbf{h}_3$, which, however, C2 is unaware of. Therefore, without knowing $\mathbf{h}_3$, C2 should maximize the magnitude of the projection of $\mathbf{h}_2$ onto $\mathbf{h}_1^{\perp}$. The resultant optimal beamforming vector is given by

$$\mathbf{v}_2 = \mathbf{H}_2\mathbf{h}_1^{\perp}\mathbf{v}_{max}((\mathbf{H}_2\mathbf{h}_1^{\perp})^*(\mathbf{H}_2\mathbf{h}_1^{\perp})), \tag{4.2}$$

which we prove next.

We first formulate the optimization problem faced by C2:

$$\max_{\mathbf{w}_2,\mathbf{v}_2}(||\mathbf{v}_2\mathbf{H}_2\mathbf{w}_2||), \text{ s.t. } \mathbf{w}_2\mathbf{h}_1 = 0.$$

To solve this constrained optimization problem, we convert it into an unconstrained problem. This is achieved by rewriting $\mathbf{w}_2$ as $\mathbf{w}_2$=$\mathbf{h}_1^{\perp}\mathbf{u}_2$ where $\mathbf{u}_2$ is a two-dimensional vector. In other words, since $\mathbf{w}_2$ lies in the orthogonal space of $\mathbf{h}_1$, it can be represented by the linear combination of two basis vectors of $\mathbf{h}_1^{\perp}$. Therefore, the original problem is converted into the following unconstrained optimization problem:

$$\max_{\mathbf{w}_2,\mathbf{v}_2}(||\mathbf{v}_2\mathbf{H}_2\mathbf{h}_1^{\perp}\mathbf{u}_2||),$$

which is a standard eigen beamforming problem under the channel $\mathbf{H}_2\mathbf{h}_1^{\perp}$. According to Chapter 2, the optimal beamforming weight vector is given by

$$\mathbf{v}_2 = \mathbf{H}_2\mathbf{h}_1^{\perp}\mathbf{v}_{max}((\mathbf{H}_2\mathbf{h}_1^{\perp})^*(\mathbf{H}_2\mathbf{h}_1^{\perp})).$$

Optimization by C3. We finally switch to the optimization by C3. The signal of C3 is projected by the AP onto a space that is perpendicular to both $\mathbf{h}_1$ and $\mathbf{h}_2$, so that the transmit beamforming weight vector for C3 employed by the AP is $\mathbf{w}_3$=span$(\mathbf{h}_1, \mathbf{h}_2)^{\perp}$ where "span()" produces the plane spanned by two vectors. As a result, by extending Equation 4.2 we have the optimal beamforming weight vector employed by C3:

$$\mathbf{v}_3 = \mathbf{H}_3[\mathbf{h}_1, \mathbf{h}_2]^{\perp}\mathbf{v}_{max}((\mathbf{H}_3[\mathbf{h}_1, \mathbf{h}_2]^{\perp})^*(\mathbf{H}_2[\mathbf{h}_1, \mathbf{h}_2]^{\perp})). \tag{4.3}$$

4.2.3 Cases 3: M=3 and K=2

We finally study the case with more AP antennas then clients. In the previous two cases we assume a congested MU-MIMO cell where K=M, so that the AP serves M clients each with a single stream. As a result, each client uses all antennas to beamform the signals, and maximize its SNR which uniquely determines its achievable rate. We next discuss the case where M>K so that AP can send more than one stream to a client. Note, the client needs no smaller than N antennas to receive N streams from the AP. However, if extra antennas are available, the client can leverage them to improve the channel orthogonality between streams. To show this, we assume N=4 such that the client can apply our solution to up to two streams. We suppose the AP transmits one stream to C1, and two streams to C2.

Optimization by C1. Since C1 receives a single stream from the AP, its optimization does not differ from that in the previous two cases where $\mathbf{v}_1 = \mathbf{H}_1 \mathbf{v}_{max}(\mathbf{H}_1^* \mathbf{H}_1)$.

Optimization by C2. For C2, since the AP delivers two streams to it, it must report two virtual channel vectors to the AP, denoted as $\mathbf{h}_{2,1}$ and $\mathbf{h}_{2,2}$. Naturally, C2 should jointly optimize $\mathbf{h}_{2,1}$ and $\mathbf{h}_{2,2}$, based on the knowledge of $\mathbf{h}_1$ overheard from C1. The optimization can be extended from the previous case with a single stream from the AP: when there are two streams, they should be sent on the best two eigen channels. As a result, the optimal beamforming weight vectors for C2 are given by

$$\mathbf{v}_{2,1} = \mathbf{H}_2 \mathbf{h}_1^{\perp} \mathbf{v}_{max}((\mathbf{H}_2 \mathbf{h}_1^{\perp})^* (\mathbf{H}_2 \mathbf{h}_1^{\perp})),$$

$$\mathbf{v}_{2,2} = \mathbf{H}_2 \mathbf{h}_1^{\perp} \mathbf{v}_{smax}((\mathbf{H}_2 \mathbf{h}_1^{\perp})^* (\mathbf{H}_2 \mathbf{h}_1^{\perp})),$$

where $\mathbf{v}_{smax}()$ is the eigenvector of a matrix that corresponds to its second largest eigenvalue. To decode the two streams, C2 individually post-combines the received signals with $\mathbf{v}_{2,1}$ and $\mathbf{v}_{2,2}$, to eliminate the interference between them.

4.2.4 General Case

We can generalize the solutions in the previous three cases to those with arbitrary M and K where $M \geq K$. For the kth client, let use denote the set of known channel vectors of other clients as $\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_l$, and its own channel matrix as $\mathbf{H}_k$. Then, the set of optimal beamforming weight vectors are given by:

$$\mathbf{v}_{k,j} = \mathbf{H}_k \mathbf{H}_L^{\perp} \mathbf{v}_{jmax}((\mathbf{H}_k \mathbf{H}_L^{\perp})^* (\mathbf{H}_k \mathbf{H}_L^{\perp})),$$

where $\mathbf{H}_L = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_l]$ and $\mathbf{v}_{jmax}()$ stands for the eigenvector of the matrix corresponding to its jth largest eigenvalue.

4.2.5 Computational Complexity

We can modularize the computation involved in the above beamforming optimization (Equation 4.1) to achieve polynomial complexity. Calculating the optimal set of beamforming weight vectors and virtual channel vectors includes the following operations: (*i*) calculating the conjugate of a matrix; (*ii*) calculating the eigenvectors and eigenvalues of a matrix; (*iii*) inverting/pseudo-inverting a matrix; (*iv*) multiplying two matrices. We can accomplish all of them by using the following two modules that already have efficient hardware implementation: (*i*) matrix singular value decomposition (SVD); (*ii*) matrix multiplication. In addition, the complexity of Equation 4.1 is given by $O(M^2L) + O(MNL) + O(N^2M) + O(NL)$, where N, L, and M are all small numbers (e.g., $N \leq 4$, $L < 8$ and $M \leq 8$). Therefore, we suspect that the computation in our solution is affordable by modern clients.

Chapter 5

Inter-cell Interference between MU-MIMO Cells

In this chapter, we study the rate scalability problem of multiple MU-MIMO cells. In the previous chapter that deals with a single MU-MIMO cell, we assume within the interference range only one AP can transmit to its clients at the same time on the same channel, i.e., no inter-cell interference is present. This chapter abandons this assumption and studies the possibility of enabling concurrent MU-MIMO transmissions from multiple MU-MIMO cells. When more than a single AP can serve their clients at the same time, they can allow more streams in the network with improved rate scalability. To achieve this, we need to tackle the inter-cell interference between multiple MU-MIMO cells.

5.1 Problem Statement

Inter-cell interference has become a key factor that limits the capacity of multi-cell MU-MIMO networks, because it prevents the APs from serving their clients concurrently. As we show in Chapter 1, by leveraging not only the AP but also the client antennas for beamforming to coordinately cancel the inter-cell interference, one can increase the number of concurrent streams or clients in the network for improved rate scalability. However, determining the beamforming weights for each AP and client is a non-trivial process. The optimal solution that is capacity achieving has not yet been founded even theoretically, and can be only empirically identified by jointly optimizing the beamforming weights for all APs and clients. The optimal solution may not completely eliminate inter-cell interference since interference below the noise power is no longer the capacity-limiting factor. However, empirically identifying the optimal solution is not practical because (*i*) it is computationally intractable due to the lack of a theoretical solution, and (*ii*) it has to be done in a centralized way with full channel knowledge of the MU-MIMO network. The central question we seek to answer in this chapter is: *for a MU-MIMO network with interfering cells, how do we achieve coordinated inter-cell interference cancellation using both AP and client beamforming?*

5.2 Overview

To improve the achievable rate of a MU-MIMO network by allowing more streams in its cells, we propose a novel solution that relies on AP and client beamforming to practically achieve coordinated inter-cell interference cancellation. The key idea in our solution is that the process of determining the beamforming weights by each

AP and client can be broken into two steps, namely *antenna usage optimization* and *beamforming weight optimization*. We are motivated by the following important observation: when we strive to completely eliminate interference, we can execute the two optimizations sequentially. This is because the constraint of completely eliminating inter-cell interference reduces the antenna usage into a binary form. That is, one can use an antenna for either transmitting or receiving streams, or cancelling inter-cell interference. It is therefore plausible to optimize the beamforming weights solely based on the given optimized antenna usage.

Such two-step optimization significantly reduces the cell coordination effort for cancelling the inter-cell interference. First, coordinately optimizing the antenna usage by each AP and client merely requires the information of the number of antennas on all nodes in the network. Not requiring any channel knowledge, cell coordination in this step is simplified since the required information can be represented with only a few bits and each AP can explicitly share it with negligible overhead. Second, given the optimized antenna usage, an AP or a client is fully aware of its duty toward cancelling the inter-cell interference. To determine the optimal beamforming weights that fulfill this duty, the AP or client only needs a subset of channel knowledge in the network. Such reduction of the required channel knowledge further makes cell coordination easier. With much simplified cell coordination, the two-step optimization can be integrated into 802.11ac retaining the distributed nature of the protocol, which we will elaborate in Chapter 7.

5.3 Antenna Usage Optimization

In this chapter, we provide an algorithm that identifies the best antenna usage for each AP and client to maximize the multiplexing gain of the MU-MIMO network. Recall that we use an AP antenna or a client antenna for either delivering streams or cancelling inter-cell interference. Therefore, our algorithm finds the optimal allocation of the AP and client antennas for such two uses. In the following, we first use a simple but illustrative example with two APs and four clients to demonstrate the process of finding the optimal antenna usage, and then provide the algorithm that can be applied to MU-MIMO networks with arbitrary number of cells.

5.3.1 Illustrative Example

Our example shown in Figure 5.1 includes two MU-MIMO cells where each AP is equipped with two antennas and serves up to two clients simultaneously. To find the optimal antenna usage, our algorithm needs to identify the best set of clients in each cell that can be simultaneously served by their corresponding AP. In other words, the selected clients must be able to coordinately cancel the inter-cell interference with

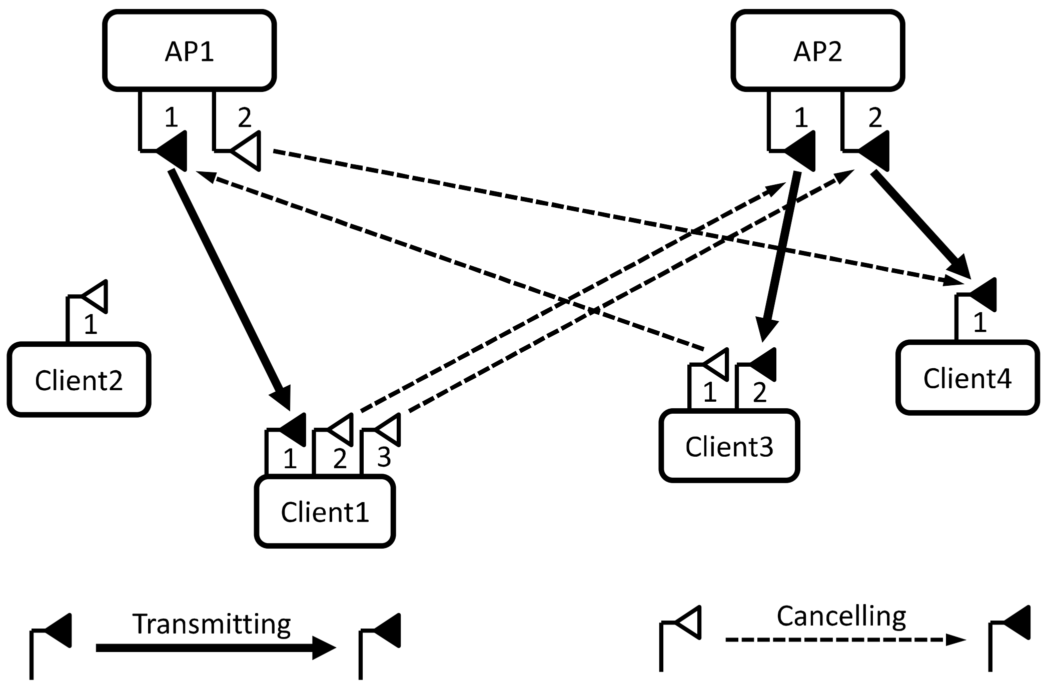


Figure 5.1 : The optimal use of the AP and client antennas in our illustrative example. The interference from AP2 to Client1 is cancelled by Client1, while the interference from AP1 to Client3 and Client4 are cancelled by Client3 and AP1, respectively.

their interfering AP.

Our algorithm starts by letting a single AP, say AP1, serve both of its clients, and tries to add concurrently served clients in the other cell similar to 802.11n+ [65]. Clearly, with only two antennas, AP2 cannot cancel the interference to both clients in Cell 1 while serving either of its own clients. Then, unlike 802.11n+ which simply stops by letting AP1 serve its two clients, our algorithm asks AP1 to serve only a single client, say Client1, and again seeks to add concurrently served clients in Cell 2. Noticeably, now AP2 can serve both of its clients if the inter-cell interference is cancelled in the following way. First, Client1 uses its three antennas to cancel the two interfering streams from AP2. Second, Client3 uses its two antennas to cancel the interfering stream from AP1. Last, while Client4 with a single antenna cannot cancel the interfering stream from AP1, observe that AP1 has a spare antenna that can be just leveraged to cancel the interference to Client4. This way, we have found the best set of clients to serve in each cell that collectively achieves a multiplexing gain of three. We illustrate the optimal use of each AP and client antenna in Table 5.1.

Table 5.1 : The optimal use of the AP and client antennas in our illustrative example, where "$\rightarrow$" and "$\dashrightarrow$" indicate data communication and inter-cell interference cancellation, respectively.

	Antenna 1	Antenna 2	Antenna 3
AP1	$\rightarrow$ Client1	$\dashrightarrow$ Client4	$\times$
AP2	$\rightarrow$ Client3	$\rightarrow$ Client4	$\times$
Client1	$\leftarrow$ AP1	$\dashleftarrow$ AP2	$\dashleftarrow$ AP2
Client3	$\dashleftarrow$ AP1	$\leftarrow$ AP2	$\times$
Client4	$\leftarrow$ AP2	$\times$	$\times$

5.3.2 Network of Two Cells

We next present our algorithm that identifies the best antenna usage for a two-cell MU-MIMO network with arbitrary number of clients and arbitrary number of antennas on each AP and client. Our algorithm is motivated by the optimization process for the illustrative example in Chapter 5.3.1.

Since we are only interested in congested MU-MIMO networks where the number of clients is no smaller than that of the AP antennas, we assume in Cell 1 AP1 has N antennas and the N clients have P_n $(n = 1, 2, \cdots, N)$ antennas, and in Cell 2 AP2 has M antennas and the M clients have Q_m $(m = 1, 2, \cdots, M)$ antennas. We further sort the clients in each cell based on their number of equipped antennas:

$$P_1 \le P_2 \le \cdots \le P_N,$$

$$Q_1 \le Q_2 \le \cdots \le Q_M.$$

Algorithm for optimizing the antenna usage. Similar to the example in Chapter 5.3.1, to obtain the optimal antenna usage, our algorithm needs to determine the optimal set of served clients in both cells given that the inter-cell interference can be coordinately cancelled. Initially, our algorithm only allows Cell 1 to operate, by letting AP1 serve its N clients. Then, it seeks to add clients in Cell 2 that can be concurrently served by AP2. The maximum number of added clients in Cell 2, denoted as L, is constrained by the inter-cell interference from AP2 to clients in Cell 1 and that from AP1 to clients in Cell 2. First, AP2 may use up to L' antennas to transmit L' streams to L' clients in Cell 2. While a few clients in Cell 1 with enough ($\geq L'+1$) antennas can cancel these L' data streams as inter-cell interference, the remaining clients in Cell 1 without enough ($\leq L'$) antennas must rely on the spare M-L' antennas on AP2 to cancel the inter-cell interference. We can keep increasing L' until those clients in Cell 1 that must cancel their inter-cell interference do not have enough antennas:

$$L' = \max(n : P_{M-n+1} \geq n + 1). \tag{5.1}$$

Second, AP2 can serve up to L'' clients in Cell 2 that must have enough ($\geq N+1$) antennas to cancel the interference from AP1. We can keep increasing L'' until these clients do not have enough antennas:

$$L'' = M - \min(m : Q_m \geq N + 1) + 1. \tag{5.2}$$

Note, we define $P_{N+1} = Q_{M+1} = +\infty$ to ensure the correctness of Equation (5.1) and (5.2) in cases where L'=0 and L''=0. The maximum number of added clients in Cell 2 is then given by

$$L = \min(L', L''). \tag{5.3}$$

In the next step, we let AP1 remove clients from its served set. When AP1 removes K ($K = 1, 2, \cdots, N$) clients, these clients should have the fewest ($P_1, P_2, \cdots, P_K$) antennas. Afterwards, AP1 has K spare antennas that can be exploited to cancel the interference to clients in Cell 2. Naturally, AP2 can serve up to K clients in Cell 2 that are not subject to the interference from AP1. Since AP1 performs such interference cancellation, we should pick clients in Cell 2 with the fewest ($Q_1, Q_2, \cdots, Q_K$) antennas to enjoy such benefit. Therefore, Cell 2 is left with M-K clients with $Q_{K+1}, \cdots, Q_M$ antennas, and Cell 1 is left with N-K clients with $P_{K+1}, \cdots, P_N$ antennas. Similar to the previous step where K=0, we have

$$L'(K) = \max(n : P_{K+M-n+1} \geq n + 1), \tag{5.4}$$

$$L''(K) = M - \min(m : Q_{K+m} \geq N - K + 1) + 1, \tag{5.5}$$

and

$$L(K) = \min(L'(K), L''(K)). \tag{5.6}$$

It is easy to verify that $L(N)$=M where only a single cell (Cell 2) is operating.

Finally, for each K=$0, 1, \cdots, N$ our algorithm calculates $L(K)$, and finds the optimal set of clients (N-K in Cell 1, L in Cell 2) that maximizes the number of streams N-K+$L(K)$. Given the optimal set of clients in each cell and their duty in cancelling the inter-cell interference, the optimal use of the AP and client antennas is meanwhile determined.

5.3.3 Network of More than Two Cells

We next extend our algorithm to T (T>2) cells. With T cells, we need to rewrite N and K as vectors, meaning that the T-1 APs serve (N_1-K_1, $\cdots$, N_{T-1}-K_{T-1}) clients respectively. Then, the maximum number of clients that can be added in the last cell is given by

$$L'(K_1, \cdots, K_{T-1}) = \max(n : P'_{M-n+1} \geq n + 1), \tag{5.7}$$

$$L''(K_1, \cdots, K_{T-1}) = |\{Q'_m : Q'_m \geq \sum_{t=1}^{T-1}(N_t - K_t) + 1\}|, \tag{5.8}$$

and

$$L(K_1, \cdots, K_{T-1}) = \min(L', L''), \tag{5.9}$$

where "$|\cdot|$" represents the cardinality of a set, P'_n the number of spare antennas on client n, and Q'_m the equivalent number of antennas on client m, determined by K'_t, the number of spare antennas on AP t. We define the spare antennas on an AP or a client as the remaining antennas after a few of them are used to cancel the inter-cell interference within the first T-1 cells. We define the equivalent number of antennas Q'_m in the following way. For a given client in the last cell, if an AP in the previous T-1 cells uses one spare antenna to cancel the interference to the client, the client can be considered to equivalently have N_t-K_t additional antennas for canceling this AP's interference. In other words, one can "transfer" the interference cancellation capability from an AP to a client it interferes with, by providing the client such equivalent antennas. Given Q'_m, the best allocation of the spare antennas on the T-1 APs is given by successively assigning those of each AP to the K'_t clients in the last cell that have the smallest Q'_m.

Given the definitions of P'_n, Q'_m and K'_t, we next explain how to derive Equation (5.7) and (5.8). For Equation (5.7), notice that the last AP can only cancel the interference to M-L' clients in all previous T-1 cells. These clients that may belong to more than one cell must have the fewest spare antennas. The remaining clients must have enough spare antennas to cancel the L' streams from the last AP. For Equation (5.8), the served clients in the last cell should not be subject to the interference from all previous T-1 APs. Since we have considered the interference cancellation from the T-1 APs with their spare antennas, we just need to find the maximum number of clients in the last cell with enough antennas, i.e., $Q'_m \geq \sum_{t=1}^{T-1}(N_t - K_t) + 1$. After calculating L for each $(K_1, \cdots, K_{T-1})$, we can find the optimal set of clients in each cell with the maximum multiplexing gain.

It is noticed that the above algorithm works in a recursive manner. Therefore, the complexity exponentially increases with the number of cells, T. To deal with this scalability issue, we leverage a cell clustering technique to convert a large-scale MU-MIMO network into a few small clusters, each of which includes up to three cells. We elaborate the cell clustering technique in Chapter 7.3.

5.3.4 Practical Implications

Observe that to identify the optimal antenna usage our proposed algorithm only requires the information about the number of antennas on each AP and client. Such information is global in the MU-MIMO network, but can be compactly represented with only a few bits. Therefore, explicitly sharing such information in the network does not incur noticeable overhead, and meanwhile significantly simplifies the cell coordination. Given such information, each AP and client can execute the same algorithm in a synchronized way to achieve coordination.

5.4 Antenna Usage Optimization with Fairness

In this chapter, we study the adaptations of the antenna usage optimization in Chapter 5.3, in order to tackle two fairness-related issues: *client starvation* and *energy-per-stream fairness*. We first discuss these two issues and then provide our adaptations.

First, our antenna usage optimization does not provide serving fairness between clients: certain clients are likely being starved, and achieving a much lower rate compared to that achieved by using 802.11ac. This is because the antenna usage optimization seeks to maximize the number of streams in the MU-MIMO network by selecting the optimal set of clients from each cell. Notably, our algorithm in Chapter 5.3 preferably selects clients with more antennas over those with fewer antennas. Consequently our algorithm may never select a client with few antennas and starve it, even with increasing number of streams and higher achievable rate of the MU-MIMO network compared to 802.11ac.

Second, our antenna usage optimization does not guarantee the number of streams received by each client is in proportion to the energy consumption of the client. That is, our solution seeks to cancel inter-cell interference by leveraging the spare antennas on clients. A client may inevitably use its multiple antennas to cancel the inter-cell interference from the interfering AP, without receiving more streams. Depending on the number of antennas on each AP and client, clients may have to

Table 5.2 : Number of antennas on the APs and clients in the three cases, for the study of the client starvation issue and energy-per-stream fairness issue in antenna usage optimization.

	AP1	AP2	Clients in Cell 1	Clients in Cell 2
Case A	2	2	1/2	1/3
Case B	2	3	4/4	3/3/3
Case C	4	4	4/4/4/4	4/4/4/4

dedicate different number of antennas to inter-cell interference cancellation. Since the power consumption of receive beamforming is proportional to the number of client antennas [51], clients are likely subject to an energy-per-stream fairness issue.

5.4.1 Client Starvation

Our antenna usage optimization may be subject to the client starvation issue when clients with more antennas are preferably selected over those with fewer antennas. To understand this issue, let us consider three cases where the number of antennas on the APs and clients are illustrated in Table 5.2.

For Case A, our algorithm enables three streams, by selecting the two clients in Cell 1 and the client with three antennas in Cell 2. Notably, this is the only set of clients that allows the maximum number of streams in the MU-MIMO network.

Consequently, our algorithm starves the client in Cell 2 with a single antenna. For 802.11ac, since the two APs can take turns to serve their clients, it does not starve any client in the two cells.

For Case B, our algorithm enables five streams, by requiring all clients in the two cells to be simultaneously served. Clearly, our algorithm starves no clients similar to 802.11ac, even though our algorithm enables two additional streams.

For Case C, our algorithm enables six streams, by choosing three clients from each cell to be simultaneously served. Important, since all clients have four antennas, our algorithm can choose any three of them in each cell. Therefore, even with our algorithm each client has an equal chance to be selected, and the two APs can choose different set of clients to serve during multiple MU-MIMO transmissions in a fair manner, without starving any one of them.

To summarize, we have the following conclusion regarding the client starvation issue in the antenna usage optimization: *the antenna usage optimization may starve a client if and only if the following two conditions regarding the clients in the same cell are met: (i) not all clients in the same cell can be included in the optimal client set that maximizes the number of streams in the MU-MIMO network, and (ii) the client has fewer antennas than any of the included client.* Clearly, Case B and Case C violate condition (*i*) and condition (*ii*) respectively so that neither of them is subject to the client starvation issue.

Preventing client starvation. To prevent client starvation, we propose the following adaptation to the antenna usage optimization: the optimization periodically enforces the inclusion of starved clients so that these clients will be served no fewer times than that with 802.11ac. We use Case A in Table 5.2 to explain how he adaptation works. In this case, 802.11ac would allow AP2 to serve the single-antenna client in Cell 2 in half of the MU-MIMO transmissions. To prevent starvation of this client, our adapted antenna usage optimization forcibly includes it in the selected client set for every other MU-MIMO transmission. Such transmission cannot deliver three streams due to the lack of client antennas, and therefore reduces the achievable rate improvement by our solution. However, it makes sure our solution does not serve any client less frequently compared to 802.11ac.

5.4.2 Energy-per-stream Fairness

Our antenna usage optimization can be subject to an *energy-per-stream fairness* issue. Battery-powered mobile clients have energy efficiency concerns, and receive beamforming leads to nearly proportionally increased power consumption to the wireless transceiver of the client [51]. In our optimization, clients may have to use multiple, and different number of antennas to receive a stream from its AP, given their interference cancellation responsibility. Consequently, to achieve the same number of streams, these clients may have to consume different amount of power, which we refer to as the energy-per-stream fairness issue. Note that the adaptation we apply in Chapter 5.4.1 ensures that no clients are served less frequently in our optimization than that in 802.11ac. However, this adaptation does not guarantee energy-per-stream fairness between clients: it does not necessarily serve a client with a frequency proportional to the number of streams the client receives.

Achieving energy-per-stream fairness. To achieve energy-per-stream fairness between clients, we propose the following adaptation to the antenna usage optimization: the optimization periodically pre-selects a set of clients with fewer number of antennas to appropriately balance the energy-per-stream of each client. We use Case B in Table 5.2 to explain how to perform such adaptation. In this case, our antenna usage optimization is able to enable five streams from the two cells, where each client needs to contribute all its spare antennas to cancel the inter-cell interference. As a result, to receive a single stream, clients in Cell 1 need to use four antennas while clients in Cell 2 only use three antennas. To achieve energy-per-stream fairness, after every three MU-MIMO transmissions to all clients, the adapted optimization lets AP1 serve the two clients in Cell 1 while AP2 remain silent. This way, for every four MU-MIMO transmissions, each client in either cell receives the same number of streams and thus our optimization achieves energy-per-stream fairness.

5.5 Channel Analysis for Beamforming Weight Optimization

In this chapter, we analyze the channel knowledge requirement for an AP or a client to optimize its transmit or receive beamforming weights. To calculate the beamforming weights that enable the optimal antenna usage, an AP or client must have certain channel knowledge based on which it cancels the intra-cell and inter-cell interference using the beamforming techniques presented in Chapter 2. In the following, we first study the two-cell example in Chapter 5.3.1 in order to simplify the analysis and obtain insightful findings, and then extend our analysis to a MU-MIMO network with arbitrary number of cells.

5.5.1 Illustrative Example

We reuse the example in Figure 5.1 to study the channel knowledge each AP and client requires to compute its optimal beamforming weights. We use $\mathbf{H}_{i\rightarrow j}$ ($\mathbf{h}_{i\rightarrow j}$) to denote the channel matrix (vector) from AP i to client j, and $\mathbf{w}_j$, $\mathbf{v}_j$ to denote the AP's transmit beamforming weight vector for client j, and the receive beamforming weight vector of client j, respectively.

Channel knowledge for AP1. AP1 uses its two antennas to send a stream to Client1 and cancel the inter-cell interference to Client4. To do this, $\mathbf{w}_1$ must be orthogonal to $\mathbf{h}_{1\to4}$, i.e., $\mathbf{w}_1=\mathbf{h}_{1\to4}^{\perp}$ where "$\perp$" represents the null space of a vector. As a result, AP1 only needs the knowledge of $\mathbf{h}_{1\to4}$.

Channel knowledge for AP2. AP2 performs zero-forcing beamforming to simultaneously send streams to Client3 and Client4 without cancelling the inter-cell interference. To do this, AP2 only needs the knowledge of $\mathbf{H}_{2\to3}\mathbf{v}_3$ and $\mathbf{h}_{2\to4}$. Note that $\mathbf{H}_{2\to3}\mathbf{v}_3$ is the virtual channel of Client3 (see Chapter 1): it combines the physical channel matrix from AP2 to Client3 ($\mathbf{H}_{2\to3}$) and the determined beamforming weight vector of Client3 ($\mathbf{v}_3$) as a single channel vector ($\mathbf{H}_{2\to3}\mathbf{v}_3$). Compared to a physical channel matrix, a virtual channel vector is much more efficient for a client to report since it needs fewer bits to be represented [78]. To cancel the intra-cell interference with zero-forcing beamforming, the knowledge of such virtual channels is enough for AP2 which sets $\mathbf{w}_4=(\mathbf{H}_{2\to3}\mathbf{v}_3)^{\perp}$ and $\mathbf{w}_3=\mathbf{h}_{2\to4}^{\perp}$.

Channel knowledge for Client1. Client1 uses its three antennas to receive its stream from AP1 and cancel the inter-cell interference from AP2. Since AP2 sends two streams to Client3 and Client4, Client1 needs to cancel both of them. To do this, Client1 simply cancels the signals sent from the two physical antennas at AP2. In other words, $\mathbf{v}_1$ is chosen as $\mathbf{v}_1=\mathbf{H}_{2\to1}^{\perp}$ where "$\perp$" here refers to the joint null space of all the rows of the matrix. Consequently, the required channel knowledge for Client1 is restricted to its own channels from AP2, $\mathbf{H}_{2\to1}$.

Channel knowledge for Client3. Unlike Client4 who has a single antenna and needs not cancel the inter-cell interference, Client3 uses its two antennas to receive its stream from AP2 and cancel the inter-cell interference from AP1. Therefore, $\mathbf{v}_3$ must be orthogonal to the signal vector from AP1, i.e., $\mathbf{v}_3$=$(\mathbf{w}_1\mathbf{H}_{1\to3})^\perp$, which suggests Client3 needs $\mathbf{w}_1$ to calculate $\mathbf{v}_3$. However, observe that $\mathbf{w}_1$=$\mathbf{h}_{1\to4}^\perp$ so that the required channel knowledge for Client3 actually becomes $\mathbf{h}_{1\to4}$ and its own channel from AP1, $\mathbf{H}_{1\to3}$.

5.5.2 Network of Two Cells

Motivated by the findings from the illustrate example, we next analyze the channel knowledge requirement of the beamforming weight optimization for a two-cell MU-MIMO network with arbitrary configuration. In particular, we prove three key theorems regarding such requirement summarized as follows: *to calculate the optimal beamforming weights based on the optimal antenna usage, an AP or a client only needs the channel knowledge owned by a particular set of clients in the network.* With this requirement, we can not only reduce the cell coordination effort, but also guarantee the optimality of the computed beamforming weights. We next elaborate the three theorems:

Theorem 5.1

To calculate the optimal beamforming weights, an AP only needs the channel knowledge owned by the clients it serves, and the clinets it interferes with holding the interference cancellation responsibility.

Theorem 5.2

To calculate the optimal beamforming weights, a client only needs the channel knowledge owned by certain clients in the same cell.

Theorem 5.3

For clients in the same cell, there exists a proper order of them with which each client can calculate the optimal beamforming weights solely based on the channel knowledge owned by clients ranked before it.

To prove these theorems, let us consider a two-cell network after performing antenna usage optimization. In the network, AP1 having N antennas serves N-K clients, and AP2 having M antennas serves M-J clients (see Chapter 5.3.2). Recall that the optimal use of an AP antenna is to either deliver a stream to a served client, or cancel the inter-cell interference to a client the AP interferes with. As a result, we can partition the antennas on AP1 into two sets: N-K antennas used to serve

N-K clients in Cell 1, and K antennas used to cancel the interference to K clients in Cell 2. We can apply similar antenna partitioning to AP2. Afterwards, we can also partition the clients in each cell according to their responsibility of cancelling the inter-cell interference. In Cell 1 and Cell 2, J and K clients rely on AP2 and AP1 to cancel the inter-cell interference, while the rest N-K-J and M-J-K clients use their own antennas to cancel the interference, respectively.

Proof of Theorem 5.1. We use AP1 for the proof of Theorem 5.1. To serve the N-K clients in Cell 1 and cancel the interference to the K clients in Cell 2, AP1 only needs to know the channels from itself to these clients. The channels from AP2 to theses clients are not needed since AP1 is not involved in the inter-cell interference from AP2. The channels from AP1 to the M-J-K clients in Cell 2 are also unnecessary, since these clients use their own antennas to cancel the interference.

Proof of Theorem 5.2 and Theorem 5.3. We use clients in Cell 1 for the proof of Theorem 5.2 and 5.3. First, the J clients do not need any channel knowledge owned by other clients to optimize its beamforming weights. This is because their antennas do not contribute to inter-cell interference cancellation. Instead, they can be used to improve the client SNR by employing MRC based on their own channels $\mathbf{H}_{1\to j}$. We call these J clients the *MRC clients*. Second, the N-K-J clients need certain channel knowledge owned by other clients to cancel the interference from AP2. To do this, they perform interference alignment toward the channels of the other J clients, so that

$$\text{span}(\mathbf{H}_{2\to 1}\mathbf{v}_1, \cdots, \mathbf{H}_{2\to J}\mathbf{v}_J) =$$

$$\text{span}(\mathbf{H}_{2\to J+1}\mathbf{v}_{J+1}, \cdots, \mathbf{H}_{2\to N-K}\mathbf{v}_{N-K}). \tag{5.10}$$

Through interference alignment, the virtual channels of the N-K-J clients, which we call the *IA clients*, are aligned to the channels of the MRC clients. When there are no MRC clients, the IA clients simply cancel the interference from the M physical antennas on AP2. When AP2 cancels the interference to the J MRC clients, the signal vector must be perpendicular to span$(\mathbf{H}_{2\to 1}\mathbf{v}_1, \cdots, \mathbf{H}_{2\to J}\mathbf{v}_J)$, which meanwhile does not creates interference to the N-K-J IA clients. Clearly, the IA clients only need the knowledge of the virtual channels, $\mathbf{H}_{2\to j}\mathbf{v}_j$, from the MRC clients, and the optimal client order is given by ranking the MRC clients before the IA clients. The relative order between the MRC clients or between the IA clients does not have an impact.

5.5.3 Network of More than Two Cells

The above three theorems hold true for a MU-MIMO network with more than two cells, which we briefly explain as follows. First, Theorem 5.1 is self-explanatory given its proof for the two-cell network. Second, for Theorem 5.2, observe that for a given interfering AP, the process of partitioning the clients into MRC clients and IA clients is still feasible. Then, a client only needs the channel knowledge from those that are identified as MRC clients while the client itself is identified as an IA client. Apparently such channel knowledge is restricted to the client's own cell. For Theorem 5.3, the best client order is decided by the number of antennas the client carries in an increasing manner. This is because an IA client always has more antennas than a MRC client does. Then, after sorting all clients in a cell based on their number of antennas, a client only needs the channel knowledge from clients that are ranked before it.

5.5.4 Practical Implications

The analysis on the channel knowledge requirement tells us that the beamforming weight optimization for an AP or a client can be potentially performed in a distributed way due to reduced cell coordination. Theorem 5.1 indicates that an AP does not need the channel knowledge owned by all clients, and does not need to share its channel knowledge with other APs. Theorem 5.2 suggests that a client does not need to acquire any channel knowledge from clients in other cells. Theorem 5.3 implies that even in a single cell, a client only needs the channel knowledge from a particular set of clients.

Chapter 6

The Design of MACCO

In this chapter, we discuss how to realize the client beamforming optimization presented in Chapter 4 in a single 802.11ac cell.

The successive beamforming optimization requires a client to have the channel knowledge of others. However, in 802.11ac networks, obtaining channel knowledge is difficult for clients since the channel sounding process is intended to provide the channel knowledge to the AP. As we argued in Chapter 2, to enable centralized beamforming weight optimization in 802.11ac networks, the 802.11ac protocol has to be radically modified to allow the AP to optimize the beamforming weight vectors for the clients, and explicitly deliver the optimization results to the clients. Such centralized optimization incurs considerable overhead to the 802.11ac channel sounding process and is therefore prohibitively expensive. The central question we seek to answer in this chapter is: *for a single 802.11ac cell, how do we practically achieve successive client beamforming optimization with negligible overhead?*

To answer this question, we present an 802.11ac-based protocol called MACCO. The key idea in MACCO is to leverage the channel sounding process in 802.11ac to let a client optimize its beamforming weight and improve the channel orthogonality in a distributed way. To realize this idea, MACCO features two designs: *local beamforming optimization* and *implicit channel acquisition*.

6.1 Local Beamforming Optimization

MACCO allows a client to locally optimize its beamforming weights and report the corresponding virtual channel to the AP. According to Chapter 1, with receive beamforming, a client can leverage the physical channel vectors between the AP and itself to create a virtual channel vector. In the channel sounding process of 802.11ac, the AP asks the clients to explicitly report the estimated channels from the AP to them (see Chapter 2.2). If the receive beamforming weight vectors of the clients are determined, the AP can perform transmit beamforming simply based on the virtual channels of the clients. MACCO leverages this to let a client report a virtual channel created by received beamforming instead of the physical channels. According to Chapter 1, the signal after the client's beamforming is a weighted combination of the signal received at each client antenna, whereas the weights constitute the beamforming weight vector. The resultant combined signal can be considered as being received at one single virtual antenna at the client. Therefore, the channel vector from the antennas on the AP to the virtual antenna on the client, i.e., the virtual channel, is a combination of the client's physical channel vectors and beamforming weight vector. By reporting the virtual channel, the client hides from the AP how to use its antennas for receive beamforming.

It is natural to ask: what are the benefits of such distributed beamforming optimization performed by the clients locally? Notably, an alternative design achieving the same objective is a centralized optimization where every client reports its physical channels to the AP, the AP optimizes the beamforming weight vectors for all the clients, and sends them back to the clients. However, such centralized optimization has two practical limitations.

First, the centralized optimization requires the clients to report more channel information to the AP compared to our distributed optimization , and the AP to deliver the optimized beamforming weight vectors to the clients, both of which consume additional channel air time as overhead. In contrast, MACCO clients only need to report their virtual channels and the AP does not need to deliver the optimization results at all. Observe that the physical channels consist of multiple vectors that correspond to the client antennas. The virtual channel, on the other hand, is a single vector irrespective of the number of client antennas. Therefore, reporting the physical channels would amplify the overhead from BF-R frames which is already substantial [57]. When the client uses N antennas for beamforming, the size of the

BF-R frame is N times larger. Moreover, the AP must send a frame to each client to deliver the optimized beamforming weight vector. Such process is also known as “feedforward” [56]. Similar to the feedback from clients, feedforward is expensive in 802.11ac networks since the frame must be sent at base rate (6 Mbps) for reliability. Adding a feedforward frame from the AP to each client can aggregately double the duration of the 802.11ac channel sounding process. It is important to note that the AP cannot “implicitly” deliver the beamforming weight vectors to the clients without using feedforward frames. A naive but incorrect design would let the AP encode the data frames in a way that allows each client to infer the optimal beamforming weight vector using maximum ratio combining. However, the client cannot correctly combine the signals for the data frame that is subject to intra-cell interference; to remove the interference, the client must know the optimal beamforming weight vector beforehand.

Second, the centralized optimization requires modification to the 802.11ac AP, i.e., the AP must optimize each client’s beamforming weight vector and feedforward the optimization results to the clients. MACCO, on the contrary, works with unmodified 802.11ac APs employing zero-forcing beamforming and 802.11ac clients. The compatibility with unmodified 802.11ac APs and clients is an important benefit of MACCO that allows rapid, incremental deployment of MACCO clients in current and future 802.11ac networks.

6.2 Implicit Channel Acquisition

MACCO allows a client to acquire other clients' optimized virtual channels through overhearing the BF-R frames of the latter. As we show in Chapter 4.1, poor channel orthogonality is the major factor that degrades the client SNR and achievable rate of MU-MIMO, which the client cannot improve without knowing the channels of other clients. Fortunately, during the channel sounding process of 802.11ac, a client can overhear the BF-R frames from other clients that reported their channels at an earlier time. Such BF-R frames contain the optimized virtual channel of the client. Since the wireless medium is shared by all clients, and the BF-R frame with a known format is not encrypted, a client can easily decode the BF-R frame from other clients and acquire the virtual channel in it. Such channel acquisition is considered "free" for the MACCO client to obtain, since the BF-R frame is mandated by 802.11ac and originally intended for the AP. With the virtual channel knowledge of other clients, a client can better optimize its beamforming weight vector aiming at a much improved channel orthogonality.

It is noticed that such overhearing in MACCO only offers partial virtual channel knowledge to each client, since the client must optimize its beamforming weight vector right before it reports the virtual channel. Nevertheless, such partial channel knowledge is sufficient to significantly improve the rate scalability of a MU-MIMO cell as we show in Chapter 6.3. We acknowledge that having full channel knowledge by each client may provide further improvement to the achievable rate of the cell. However, to the best of our knowledge, how to jointly optimize the beamforming weights of all clients with full channel knowledge remains an challenging and open problem, even for theoretical investigation.

6.3 Evaluation

In this chapter we evaluate MACCO with both experiments and simulations.

6.3.1 WARP Implementation

We implement MACCO on WARP [58], a flexible FPGA-based software-defined radio (SDR) platform. Our implementation is based on WARPLab [79], but incorporates extensive modifications. WARPLab is a framework that combines the advantages of FPGA and MATLAB to enable rapid physical layer prototyping that allows coordination of multiple boards. The extensible framework provides great flexibility to develop physical layer techniques, which is perfectly suitable for MU-MIMO designs. In WARPLab, one or more boards as APs or clients are connected to and controlled by a central computer running MATLAB, through gigabytes Ethernet cables. Baseband processing can be flexibly implemented in either hardware (FPGA on the WARP boards) or software (MATLAB on the central computer).

Given the WARPLab framework, we solve two challenges to implement MACCO.

Realtime vs. flexibility. The first challenge is to balance the realtime capability and flexibility of our implementation. On one hand, to make it fully real-time, we have to completely realize MACCO in hardware, which is prohibitively complex and inflexible. On the other hand, realizing all processing in software is simpler and more flexible, but incurs significant communication round-trip time (RTT) between the central computer and the WARP boards, since they need to exchange raw baseband samples.

To achieve a good tradeoff between realtime capability and implementation flexibility, we realize standard MU-MIMO processing including CSI estimation and beamforming in hardware, and the calculation of the transmit/receive beamforming weight vector and virtual channel vector in software. This way, the relatively complex matrix operations in MACCO can be easily accomplished in MATLAB, while hardware-friendly MU-MIMO processing is pushed onto FPGA. The transfer between the central computer and WARP boards is then reduced from raw baseband samples to the calculated beamforming weight vectors and virtual channel vectors.

Delay of over-the-air frames. The second challenge is the substantial delay of sending a frame over the air. This is again due to the substantial communication RTT between the central computer and the WARP boards. Even though we have largely reduced the volume of such transfer by realizing MU-MIMO processing on the WARP boards, the constant latency of forming the Ethernet packet in MATLAB, polling the Ethernet interface on the WARP board, and fetching the transferred samples into the WARP board buffer is still too long to satisfy the timing constraints of 802.11ac [1]. Our measurement shows that when a client receives the BF-P frame from the AP, it usually takes a few milliseconds for the client to respond with the BF-R frame. The aggregated delay from multiple BF-P and BF-R frames will make the channel sounding process much longer, rendering an potentially outdated wireless channel.

To tackle this, we emulate the channel reporting and overhearing in the sounding process of MACCO: when all client boards receive the NDP frame sent by the AP board over the air, they simply transfer the estimated physical channels to the central computer via Ethernet, and then the central computer emulates the BF-P and BF-R frames assuming proper channel knowledge for the AP and client boards. After the central computer calculates the transmit and receive beamforming weight vectors, it sends them back to the AP and client boards, and then the AP board starts the MU-MIMO transmission. We claim that such emulation does not lose authenticity due to two reasons. First, since the 802.11ac channel sounding is explicit, when a frame is successfully decoded, it correctly delivers the channel information. Second, 802.11ac considers the BF-P and BF-R frames as control frames and requires them to be sent with base rate (6 Mbps). Therefore, the likelihood of being not able to decode them by a client is fairly small. Emulating the BF-P and BF-R frames can greatly reduce the duration of the channel sounding process, allowing us to involve more (up to K=8) clients in the experiment.

Measurement Setup

We deploy a single MU-MIMO cell including one AP with eight antennas and up to eight clients. The deployment is within a lab room in a campus building shown in Figure 6.1. To conduct multiple groups of experiments to average out the performance, we change the locations of the AP and clients to create different wireless channels between them. We conduct the experiments at nighttime, on a clean Wi-Fi channel (# 14) to make sure the clients are not subject to external interference. We measure the client SNR and INR to calculate the achievable rate of each client and that of the cell.

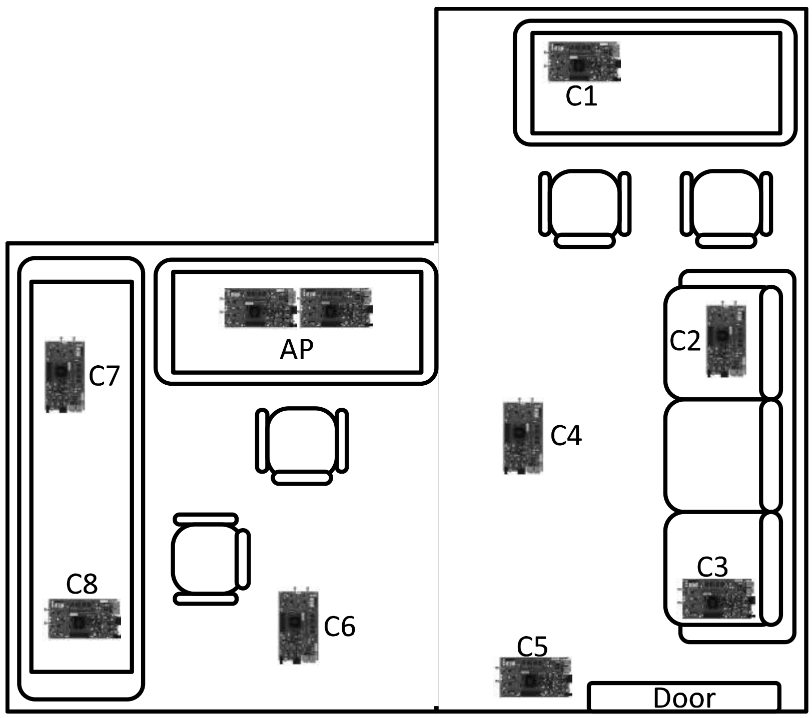


Figure 6.1 : Our experimental setup in a lab room for the evaluation of MACCO. An AP with eight antennas simultaneously serves up to eight MACCO clients each with four antennas.

6.3.2 Results

We next report our evaluation results. We evaluate the effectiveness of MACCO to improve the rate scalability of a MU-MIMO cell, and investigate the impact of various parameters such as the number of clients K, the number of AP antennas M, and the number of client antennas N.

We compare MACCO to the following alternative approaches: (i) *Omni* in which a client has a single antenna; (ii) *Antenna Selection (AntSel)* in which a client picks the antenna with the best physical channel; (iii) *Maximum ratio combining (MRC)* in which a client optimizes its beamforming weight vector to maximize its SNR based on its own physical channels; (iv) *Clustered MRC* in which the clients employ MRC but the AP only serves the best set of clients as a cluster at a time and schedules clusters in a TDMA fashion. There are two notes regarding the above alternative approaches for comparison. First, in our evaluation we approximate the achievable rate of clustered MRC, ignoring the overhead of clustering and scheduling the clients by the AP. In practice, serving more than one clusters of clients in a TDMA fashion requires extra control frames such as the NDP-A and NDP frames in the channel sounding process. Second, AntSel, MRC and Clustered MRC do not seek to leverage the channel knowledge of other clients to optimize the client beamforming, and thus cannot improve the channel orthogonality. However, the comparison between

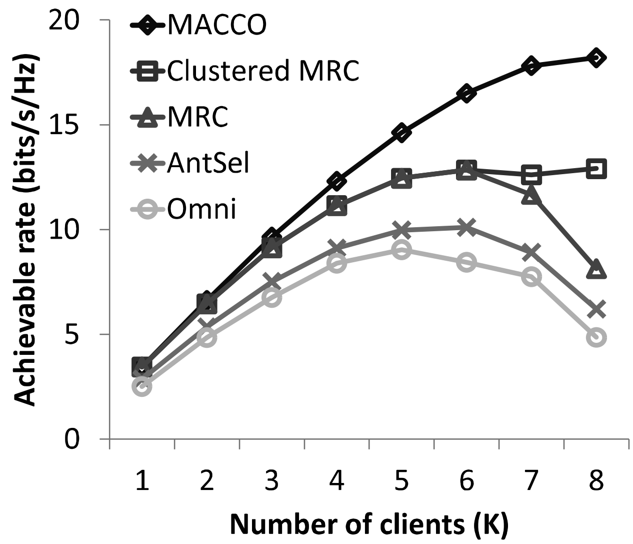


Figure 6.2 : Achievable rate for Omni, AntSel, MRC, Clustered MRC and MACCO as function of the number of clients, $\boldsymbol{K}$, with $\boldsymbol{M}$=8 antennas on the AP and low client SNR. MACCO achieves better rate scalability compared to others.

MACCO and these approaches is fair, since MACCO allows a client to acquire channel knowledge in an overhead-free manner, by overhearing the BF-R frames that are necessary in the 802.11ac channel sounding and destined to the AP.

Rate Scalability Improvement

To see how MACCO makes the achievable rate of a MU-MIMO cell more scalable to the number of clients, Figure 6.2, 6.3 and 6.4 show the achievable rate of a MU-MIMO cell where the AP employs M=8 antennas and serves K=$1, 2, \cdots, 8$ clients. We purposely choose three regimes with low, medium, and high average SNR at the clients, by adjusting the transmit power from the AP. The average SNR with respect

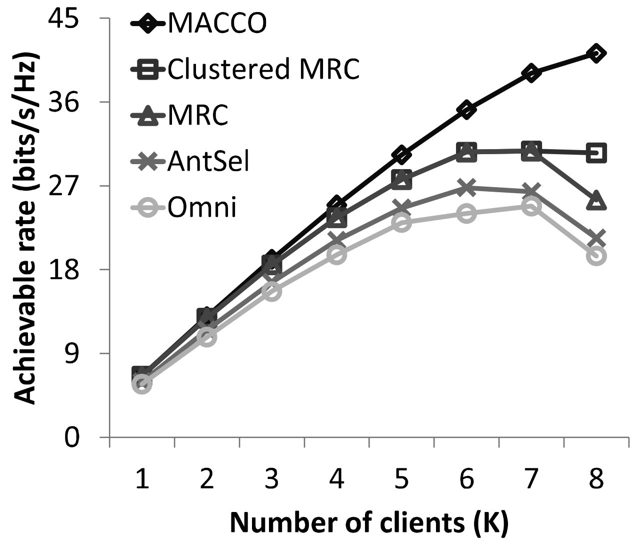


Figure 6.3 : Achievable rate for Omni, AntSel, MRC, Clustered MRC and MACCO as function of the number of clients, $\boldsymbol{K}$, with $\boldsymbol{M}$=8 antennas on the AP and medium client SNR. MACCO achieves better rate scalability compared to others.

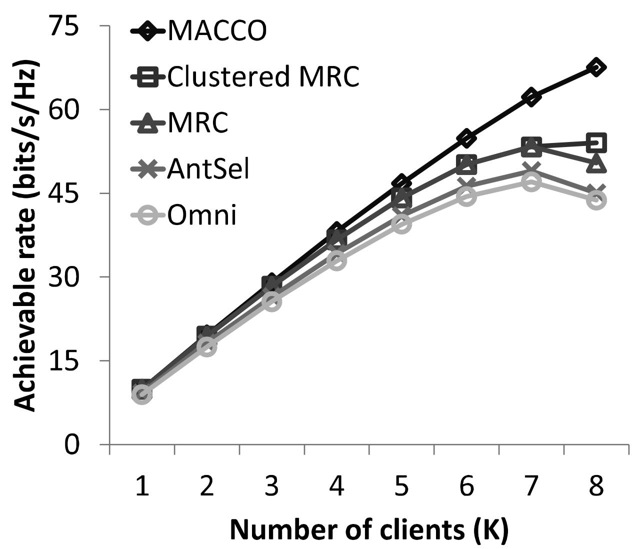


Figure 6.4 : Achievable rate for Omni, AntSel, MRC, Clustered MRC and MACCO as function of the number of clients, $\boldsymbol{K}$, with $\boldsymbol{M}$=8 antennas on the AP and high client SNR. MACCO achieves better rate scalability compared to others.

to a single client antenna is 1.47 dB, 11.02 dB, and 19.43 dB for the low, medium, and high SNR regimes, respectively. We obtain the following important observations from the figures.

- With Omni, AntSel and MRC, the achievable rate of the cell starts to drop as K approaches M. This is due to significantly reduced client SNR from ZFBF as we explained in Chapter 2.

- When the client employs AntSel instead of Omni, multiple antennas provide trivial rate improvement. This is because AntSel only offers antenna diversity gain, which is quite limited due to the physical proximity of the antennas on a single client.

- MRC achieves higher rate than AntSel does, since MRC allows the client to pick the best eigen channel instead of the best physical channel.

- Clustered MRC achieves similar rate as MRC does, but prevents the rate from dropping as K approaches M. This is because the AP stops adding more clients to serve as the rate starts to decrease due to client SNR reduction.

- MACCO outperforms alternative approaches by more effectively using the multiple client antennas based on the overheard channels. It works most effectively

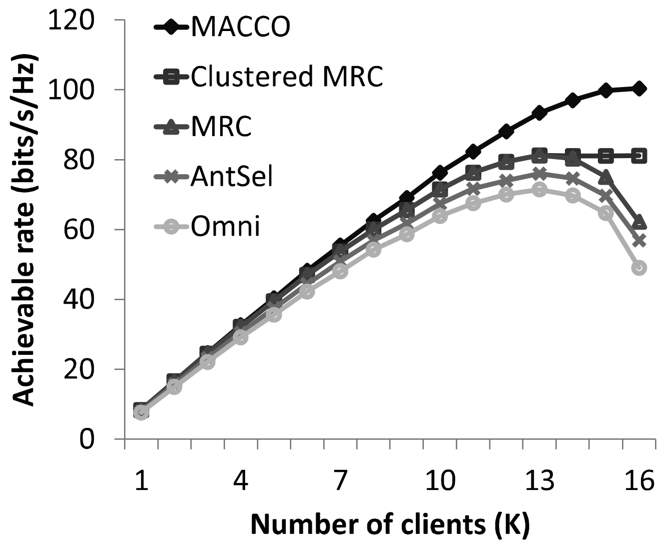


Figure 6.5 : (Simulated) rate achieved by MACCO and alternative approaches for large-scale MU-MIMO cells where $\boldsymbol{M}$=$\boldsymbol{K}$=16.

when K approaches M, making the achievable rate of the cell near-proportional to the number of clients.

- MACCO provides most rate improvement in the low SNR regime, where the rate near-linearly grows with SNR. This means MACCO is beneficial to clients that are on the edge of the cell.

Larger-scale MU-MIMO cells beyond 802.11ac. To investigate the effectiveness of MACCO beyond current 802.11ac that supports up to 8 AP antennas, we seek to evaluate it in MU-MIMO cells of even larger scale, e.g., with 16 and 32 AP antennas and clients. Unfortunately, we find it very difficult to perform experiments of such scale. While it is possible to build an AP with many antennas [6], it is hardly fea-

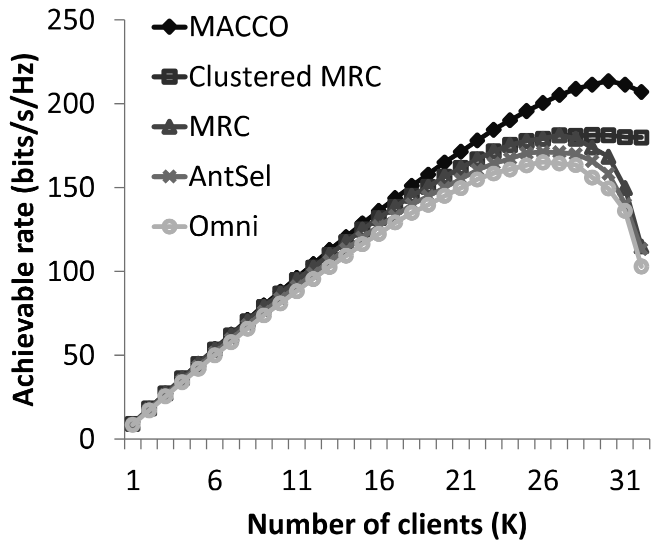


Figure 6.6 : (Simulated) rate achieved by MACCO and alternative approaches for large-scale MU-MIMO cells where $\boldsymbol{M}=\boldsymbol{K}=32$.

sible to properly deploy many multi-antenna clients since they need to be distributed and connected to the central computer using Ethernet cables. Therefore, we rely on simulation based on synthetic channels to evaluate the effectiveness of MACCO in large-scale MU-MIMO cells. We adopt the 802.11n/ac MIMO channel model based on the *Ricean* K factor [80] to generate MU-MIMO channel traces, assuming an i.i.d. MIMO channel from the AP to each client. To mimic a realistic indoor environment, in the model we choose the *Ricean* factor K as 10 which offers the MIMO channel both line-of-sight (LOS) and non-line-of-sight (NLOS) components.

Figure 6.5 and Figure 6.6 show the rate achieved by MACCO and alternative approaches under synthetic channels for the large-scale MU-MIMO cell. We first observe that MACCO is still able to considerably improve the rate scalability. However, the rate sub-linearity of MACCO becomes more obvious as K approaches M. The key reason is that as more clients are served, it becomes harder to improve the orthogonality between a greater number of channels. In fact, the rate scalability issue we tackle is intrinsically determined by the physical channel vectors of the clients. Even though MACCO improves the channel orthogonality, the resultant virtual channel can still suffer from certain SNR loss. Even with multiple client antennas, it is very hard to achieve proportionally increased rate of the cell especially for large K and M. Theoretically, MACCO is able to achieve perfect rate scalability only if each client is equipped with a very large number of antennas [5], which is apparently impractical. As a result, the key objective of MACCO is to maximize the benefit from the limited number of antennas on mobile clients to improve the achievable rate of the MU-MIMO cell, in an effective and practical way. Also note that when there are a large number of clients, the channel sounding process in 802.11ac becomes prohibitively expensive since its duration is proportional to the number of clients. We therefore suspect that large-scale MU-MIMO cells will adopt alternative channel estimation schemes that let the AP acquire channels implicitly [6]. MU-MIMO cells with implicit channel estimation require a radically different approach to solve the rate scalability problem, which is out of the scope of this thesis.

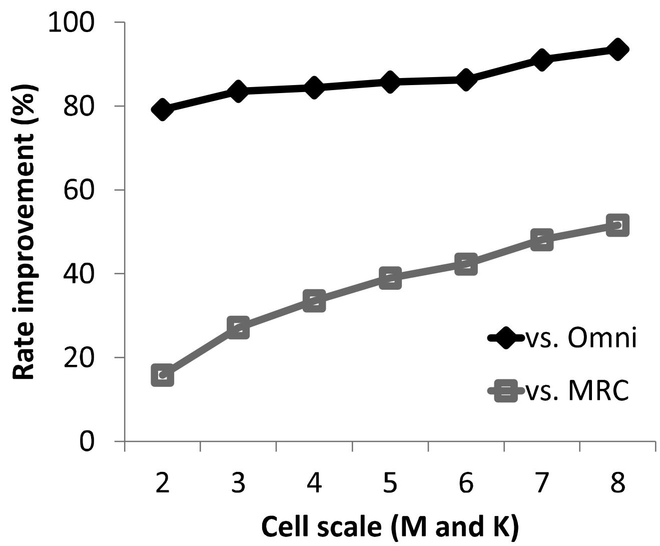


Figure 6.7 : Achievable rate improvement from MACCO with different cell scales ($\boldsymbol{M}$ and $\boldsymbol{K}$).

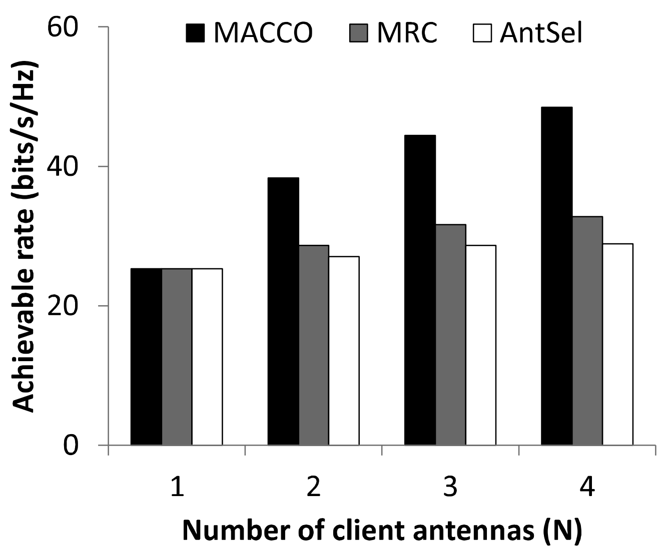


Figure 6.8 : Rate achieved by AntSel, MRC, and MACCO with different number of client antennas ($\boldsymbol{N}$).

The Impact of Antenna Numbers

We next study the impact of two important parameters on the effectiveness of MACCO: the number of AP antennas, and the number of client antennas.

Number of AP antennas. We first study the capacity gain from MACCO compared to Omni and MRC with different number of antennas on the AP (M), as shown in Figure 6.7. To investigate the full potential of MACCO, we assume a congested cell where K=M. It is seen that as M and K increases, the MU-MIMO cell enjoys more rate improvement from MACCO. This is because the rate scalability issue becomes more problematic as the AP is able to serve more clients. Moreover, the achievable rate improvement by MACCO compared to MRC is more obvious, due to the fact that MRC becomes less effective with more clients, without improving the channel orthogonality between them.

Number of client antennas. We then study how the number of antennas on the clients, N, affects the rate improvement from MACCO. We use M=K=8 as an example case, and show the results in Figure 6.8. One can see that with more client antennas, MACCO, MRC and AntSel can all achieve higher rate of the cell. However, the improvement from MRC and AntSel are very limited; on the contrary, MACCO is more capable of increasing the rate from more client antennas. This is because MACCO can more effectively leverage the precious degrees of freedom provided by the client antennas, by using them to improve the channel orthogonality.

Client Order does Not Matter

In this chapter we show that the client reporting order does not affect the performance of MACCO. When the MACCO clients successively send their BF-R frames, they acquire different amount of channel knowledge through overhearing. It is natural to hypothesize that the rate improvement from MACCO is dependent on the client reporting order. To see this, we perform simulation based on synthetic channels used Chapter 6.3.2, with up to M=K=16. However, our simulation results show that such client order has a negligible impact on the achievable rate of the cell (3.4% and 8% rate drop for the average and worst case) even when the clients have very different SNR (up to 30 dB). We argue that this is because the client with more channel knowledge does not necessarily enjoy more rate improvement. While this may sound counter-intuitive, our key insight is that MACCO increases the cell rate mainly by improving the channel orthogonality. Better channel orthogonality can usually lead to an improved rate for all involved clients while the client with the most channel knowledge does not necessarily enjoy the most improvement. In other words, the effectiveness of MACCO is toward the entire MU-MIMO cell instead of any individual client.

6.4 Rate and Energy Efficiency Tradeoff

The local beamforming optimization in MACCO allows the client to explore an interesting tradeoff between its achievable rate and energy efficiency. That is, the rate improvement from MACCO depends on two factors: (*i*) the knowledge of other clients' virtual channels, and (*ii*) the number of antennas for receive beamforming. First, knowing more channels allows the client to optimize its own virtual channel to produce better channel orthogonality, since the client can more precisely predict which direction the AP will project its data signals onto. Note, it does not contradict with our conclusion in Chapter 6.3.2: while MACCO does not provide higher gain to clients that report their virtual channels at a later time, for a particular client the number of overheard virtual channels does affect the channel orthogonality improvement. Second, more antennas provide the client more flexibility to adjust its virtual channel. They meanwhile offer an increased SNR gain from maximum ratio combining. However, both factors incur extra energy consumption. To overhear the channel knowledge from other clients, a client must keep its radio transceiver on to capture and decode the BF-R frame. To use more antennas, the client has to turn on more RF chains of the transceiver. Therefore, in MACCO a client has the opportunity to adaptively acquire channel knowledge from peer clients and use a subset of antennas based on its rate requirement and energy budget. Such achievable rate and energy tradeoff has been explored by the authors in [51,81,82] for clients in single-user MIMO networks, by adapting the number of antennas only. We believe that by allowing the client to adaptively overhear the channel knowledge, MACCO makes such rate and energy efficiency tradeoff even more interesting.

Chapter 7

The Design of CoaCa

In this chapter, we focus on the rate scalability problem in multiple MU-MIMO cells. We discuss how to realize the two-step optimization presented in Chapter 5 in an 802.11ac-based MU-MIMO network with multiple cells. In particular, we present CoaCa, a protocol that seamlessly integrates both the antenna usage optimization (Chapter 5.3) and the beamforming weight optimization (Chapter 5.5) into 802.11ac. CoaCa includes two key designs to achieve coordinated inter-cell interference cancellation, namely *interleaved channel sounding* and *channel reporting and overhearing*. With these two designs, each AP and client can locally perform the two-step optimization in a distributed but coordinated way.

7.1 Interleaved Channel Sounding

To provide the APs and clients with necessary information to optimize their antenna usage, CoaCa features *interleaved channel sounding*, in which the key idea is to let the APs in all cells send their NDP-A and NDP frame sequentially one after another, before each AP polls its served clients. The NDP-A frame can contain the information

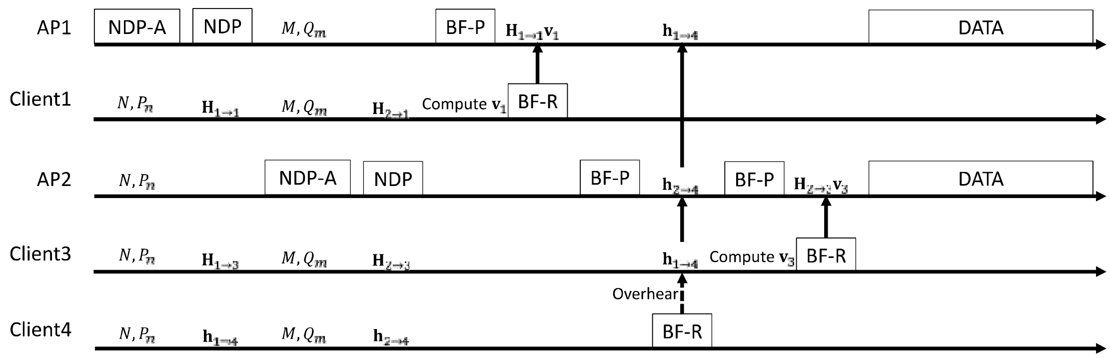


Figure 7.1 : Timeline of CoaCa where two APs concurrently serve their clients. To provide each AP and client the necessary information to optimize its antenna usage, an AP starts polling its clients with the optimal order only after all the APs have transmitted their NDP-A and NDP frames. To optimize its beamforming weights based on the optimal antenna usage, a client overhears the reported virtual channels from other clients in the same cell that report before it.

about the number of antennas on the AP and that on the clients the AP plans to serve. With interleaved channel sounding, an AP can broadcast such information to the entire MU-MIMO network, and all APs and clients can use the information to optimize their antenna usage with the same algorithm provided in Chapter 5.3 in a coordinated way. Moreover, the NDP frames allow the IA clients to estimate their physical channel matrix from the interfering APs, which is necessary to optimize their beamforming weights together with the overheard channel vectors from the MRC clients (see Chapter 7.2).

We illustrate the timeline of interleaved channel sounding in Figure 7.1 using the example in Figure 5.1. In the channel sounding process of 802.11ac, after an AP, say AP1, sends the NDP-A and NDP frame, the first served client, say Client1, will immediately respond with the BF-R frame after SIFS time. In the interleaved channel sounding of CoaCa unlike 802.11ac, no client in Cell 1 is allowed to immediately respond; instead, AP2 sends its NDP-A and NDP frame SIFS time after AP1 sends its NDP frame. Only after both AP1 and AP2 send their NDP-A and NDP frame, each of them sequentially polls their clients with a BF-P frame and their clients respond with their BF-R frames in the same order.

Viability. The interleaved channel sounding ensures correct behaviour of the involved APs and their served clients. To make sure the channel sounding from multiple APs are interleaved, an AP must be able to send its NDP-A and BF-P frame at the proper time with guaranteed medium access. In addition, the APs must sound their channels in a pre-determined order without collision. For example, according to Figure 7.1, AP2 must (*i*) send its NDP-A frame immediately after AP1 sends its NDP frame, and (*ii*) poll its clients only after Client1 sends its BF-R frame. CoaCa uses two techniques to ensure this coordinated behaviour, which we discuss based on Figure 7.1. First, CoaCa adopts CHAIN, a technique proposed in [83]. The key idea in CHAIN is that AP2 piggybacks its NDP-A or BF-P frame SIFS time after the ongoing frame from AP1 or Client1 finishes; this gives AP2 prioritized medium access since other 802.11 nodes must wait for at least DIFS time to contend for the medium. To determine their relative channel sounding order, AP1 and AP2 only need to coordinate once to initiate the transmissions in CHAIN. Second, to avoid collision between AP2 and Client1 who might also send its BF-R frame after SIFS time, CoaCa refrains Client1 from immediately responding. CoaCa achieves this by having AP1 specify a "fake" client with an invalid MAC address as the first responding client in the NDP-A frame. This way, all clients in Cell 1 yield to AP2.

Interoperability. CoaCa APs and clients can interoperate with unmodified 802.11ac clients. The key reason is that in the proposed interleaved channel sounding a CoaCa client still passively responds to the BF-P frame from its AP, similar to an unmodified 802.11ac client. In addition, CoaCa concatenates the number of antennas information (2 bytes) to the NDP-A frame and modifies its duration field, to ensure the frame can be decoded even by an unmodified 802.11ac client.

Overhead. The interleaved channel sounding introduces negligible overhead compared to 802.11ac. First, observe that the interleaved channel sounding requires each AP to send an extra BF-P frame to poll the first served client. Such overhead is not only negligible but also justified since (*i*) the BF-P frame is much shorter than the NDP-A and BF-R frame that constitute the major portion of channel sounding, and (*ii*) the extra BF-P frame eliminates the necessity for the first client to initiate the transmissions in CHAIN, which would otherwise require the client to overhear clients in other cells and negate the reduction of cell coordination in CoaCa. Second, the

BF-R frame may have to contain the estimated channels from not only the associated AP but also the interfering APs. Such extra channel information seemingly increases the size of the BF-R frame in a proportional manner. However, due to the use of virtual channels (see Chapter 7.2), such extra channel information can be more than compensated by the proportionally reduced information in each estimated channel, once the client is equipped with multiple antennas.

7.2 Channel Reporting and Overhearing

To provide the APs and clients with required channel knowledge to optimize their beamforming weights, similarly to the technique proposed in 6, in CoaCa *a client reports the necessary virtual channels in the BF-R frame and overhears other clients' BF-R frames for their virtual channels*, as illustrated in Figure 7.1. First, reporting the virtual channel $\mathbf{H}_{i\to j}\mathbf{v}_j$ instead of the physical channel $\mathbf{H}_{i\to j}$ can proportionally reduce the size of the BF-R frame which is known to incur substantial overhead [57, 67] to the channel sounding process of 802.11ac. This is because the virtual channel is a vector while the physical channel is a matrix that needs many more bits to encode when the client has multiple antennas. Reducing the size of the BF-R frame is especially beneficial for the MRC clients who must report the channels to not only its associated AP but also the interfering APs holding the interference cancellation responsibility (see Theorem 5.1). Such extra channel information as additional overhead can be more than compensated by the reduced size of the BF-R frame when the client has multiple antennas. Second, since the BF-R frame explicitly contains the virtual channels, a client that overhears the BF-R frame can easily acquire such knowledge by decoding the frame. The channel knowledge is guaranteed to be accurate once the BF-R frame is successfully decoded.

Decodability. An AP or a client is able to decode the overheard BF-R frames with high probability, given the following two observations. First, Theorem 5.1 and Theorem 5.2 indicate that an AP only needs to overhear the BF-R frames from the MRC clients it interferes with, and a client only needs to overhear the BF-R frames from the MRC clients in the same cell. This significantly reduces the likelihood that an AP or a client is too distant from the client it seeks to overhear. Second, the BF-R frame is considered a control frame and commonly sent at base rate (6 Mbps) in order to improve reliability [1]. This in turn extends its transmission range and reduces the possibility of frame decoding failure.

Sufficiency. Reporting and overhearing the virtual channels can provide the APs and clients just enough channel knowledge. First, according to Chapter 5.5.2, the virtual channel is sufficient for the APs to perform interference cancellation, and for the IA clients to perform interference alignment. Second, the AP can poll its clients with the optimal order to make sure a client has acquired enough virtual channels before it determines the optimal beamforming weights and reports its own virtual channel. This is because Theorem 5.3 indicates that if the clients in each cell send their BF-R frames following the optimal order, each client only needs the channel knowledge owned by clients ranked before it.

7.3 Cell Clustering

In this chapter, we address the scalability issue of CoaCa toward the number of cells in the MU-MIMO network. In particular, the following reasons make it hard to apply CoaCa to a large-scale network with more than a few cells. First and most importantly, the number of required antennas on the APs and clients to appreciate the multiplexing gain improvement from CoaCa significantly increases with more cells. Given that each AP and each client usually cannot afford more than eight and more than four antennas respectively, CoaCa cannot provide obvious achievable rate improvement when the network includes more than three cells. Second, as revealed in Chapter 5.3.3, the complexity of the recursive algorithm to identify the optimal antenna usage exponentially increases with the number of cells. Third, the duration of the interleaved channel sounding scales up proportionally to the number of cells, which may lead to having outdated estimated channels. Finally, to receive the NDP-

A frames and enable CHAIN, an AP must be in the overhearing range of all other APs, which becomes much more challenging with more cells.

We rely on a *cell clustering* technique to tackle the scalability issue of CoaCa. That is, we group all cells in the MU-MIMO network into clusters such that *within each cluster there are up to three cells.* Then, we apply CoaCa to each cluster, and employ standard CSMA/CA for the medium access between clusters. We are motivated by two reasons to allow up to three cells in each cluster. First, cell clustering with such scale does not reduce the effectiveness of CoaCa. As mentioned earlier, when an AP has no more than eight antennas and a client has no more than four antennas, to maximize the number of streams CoaCa usually does not need to allow more than three APs to concurrently transmit to their clients. Therefore, even if the MU-MIMO network included more than three cells, they would not lead to higher multiplexing gain. Notably, CoaCa seeks to most effectively leverage the available antennas from each AP and client in the network, by letting them coordinately cancel the inter-cell interference in a distributed way. Yet, it is incapable of going beyond such number of antenna constraint. To further increase the number of streams in the network, one must employ centralized solutions such as Network-MIMO [8–10] to convert interference into signals. Second, cell clustering with up to three APs can greatly simplify the operation of CHAIN in situations where only a subset of APs in a cluster intend to serve their clients. That is, in CHAIN, to ensure correct medium access the AP without packets to its clients may have to send an intermediate frame that triggers the NDP-A frame from the AP following it. With up to three APs in a cluster, this can be replaced by a much simpler design where the second and third AP use SIFS (16 μs) and PIFS (25 μs) in CHAIN to guarantee their relative priority, without having to send the triggering frame by the second AP.

Determining the included cells in a cluster can be accomplished by a cluster formation algorithm, which is outside of the scope of this work. We believe existing approaches based on assigning a master AP for managing each cluster, e.g., the one proposed by [76], can be adopted by CoaCa with appropriate adaptation.

7.4 Experimental Evaluation

Finally we experimentally evaluate the performance of CoaCa in real-world indoor environments.

7.4.1 WARP-based Implementation

Similar to MACCO, we implement CoaCa on WARP [58]. We again choose WARPLab [79] for our implementation and evaluation of CoaCa but extensively modify it to improve its real-time capability.

Since the baseband processing occurs in MATLAB, we cannot implement the 802.11ac MAC satisfying the timing constraint. Therefore, our evaluation of CoaCa focuses on its PHY, i.e., the improvement of the achievable rate of the network from the increased multiplexing gain achieved by antenna usage optimization and beamforming weight optimization. Similar to the implementation of MACCO, we move the process of client polling, virtual channel reporting and overhearing in the interleaved channel sounding into emulation. That is, after each AP sequentially sends its NDP-A and NDP frame that allows the clients to estimate their channels, we emulate the BF-P and BF-R frames on the central computer. To do this, we assume each AP and client possesses the same knowledge of the reported and overheard virtual channels, as if they were obtained from actual BF-R frames over the air. Note that such emulation is close-to-reality since the BF-R frame delivers the channel knowledge in an explicit and reliable way. Then, based on the channel knowledge, the APs and clients calculate their optimal beamforming weights, with which they coordinately cancel the inter-cell interference, and transmit and receive the data frames. We calculate the achievable rate of the MU-MIMO network based on the measured signal-to-noise-and-interference ratio (SINR) at each served client. While the actual throughput gain by MU-MIMO is known to be reduced by the channel sounding overhead [57, 67], we do not incorporate such overhead into our evaluation since it equally exists in 802.11ac according to Chapter 7.1.

7.4.2 Experimental Setup

In our experiments we construct a MU-MIMO network with two cells. We use each WARP board to serve as either an 802.11ac AP or an 802.11ac client so that the number of AP and client antennas is up to four. To evaluate the capability of CoaCa to eliminate inter-cell interference, we deploy the APs and clients within a single interference domain so that the capacity of each served client is interference instead of noise limited. Our experiments are conducted in a three-floor campus building that is representative of indoor environments. To make sure the interleaved channel sounding does not exceed the channel coherence time, we run the experiments at nighttime when we observe the wireless environment is static. To limit the interference within our MU-MIMO network, we select a clean Wi-Fi channel (# 14) that does not have any ongoing traffic.

7.4.3 Accuracy of Interference Cancellation and Interference Alignment

Since CoaCa relies on interference cancellation and interference alignment to eliminate inter-cell interference, we first evaluate their accuracy on WARP under realistic indoor wireless channels. Note that interference cancellation and alignment can be inaccurate due to (i) errors in the estimated and overheard channels used to compute the beamforming weights, and (*ii*) channel noise and hardware imperfection that distort the received signals. Toward this, we use three micro-benchmarks with two or three nodes to study how much inter-cell interference between an AP and a client can be eliminated. To evaluate interference cancellation by the AP, we assume an AP having two to four antennas cancels the interference to a client via transmit beamforming; to evaluate interference cancellation by the client, we assume a client having two to four antennas cancels the interference from an interfering AP with receive beamforming; to evaluate interference alignment by the client, we assume an AP cancels the interference to one client via transmit beamforming, and another client aligns its channel to that of the first client via receive beamforming. In each micro-benchmark we assume the beamforming AP or client owns the necessary channel knowledge, which is obtained right before the node transmits or receives. For all the experiments, we measure the original interference power and residual interference power after cancellation or alignment to compute the interference reduction. Note that we have properly scaled the transmit power from the AP to make sure the interference power is always above the noise floor of the client even after cancellation. This is to avoid underestimation of the effectiveness of interference cancellation and alignment when the client becomes noise limited.

Figure 7.2, 7.3 and 7.4 show the interference reduction (dB) achieved by inter-

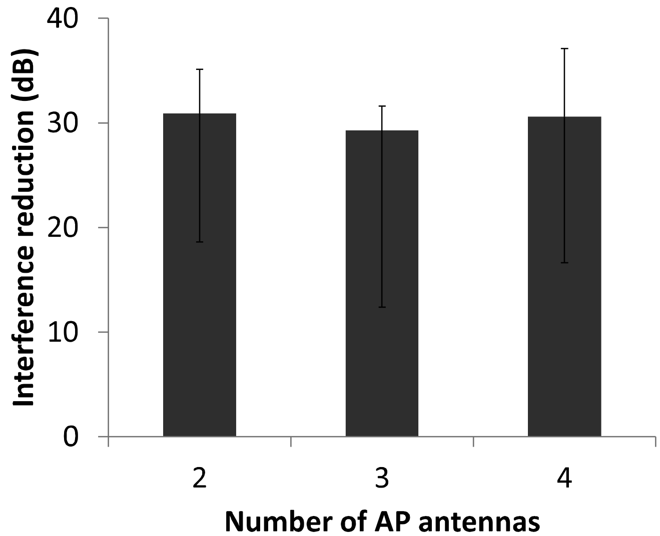


Figure 7.2 : Interference reduction from interference cancellation by the AP. The number of AP antennas does not have an obvious impact on the cancellation accuracy.

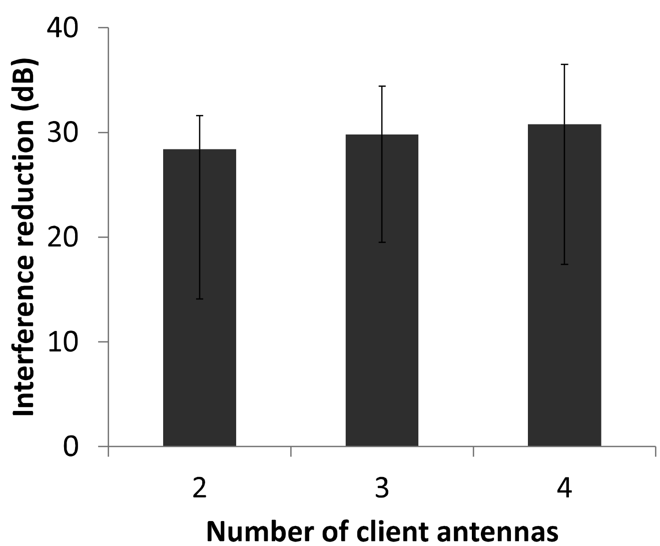


Figure 7.3 : Interference reduction from interference cancellation by the client. The number of client antennas does not have an obvious impact on the cancellation accuracy.

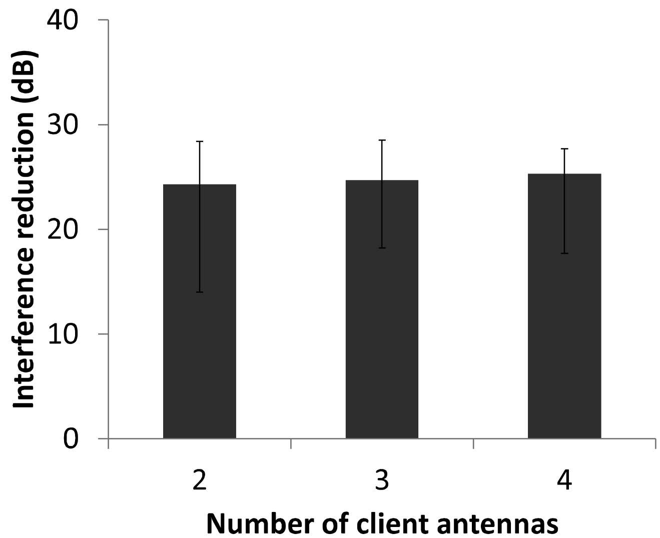
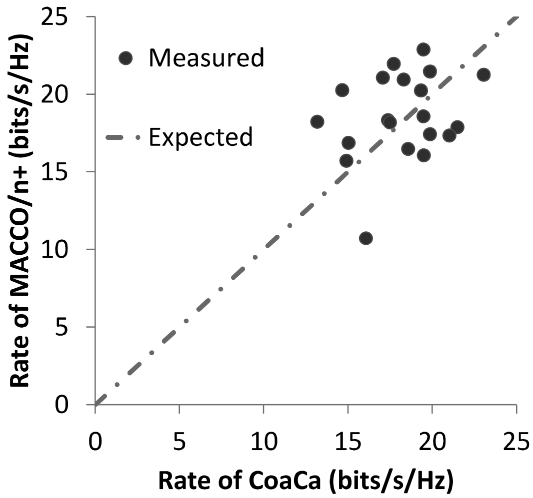


Figure 7.4 : Interference reduction from interference alignment by the client. The number of client antennas does not have an obvious impact on the alignment accuracy.

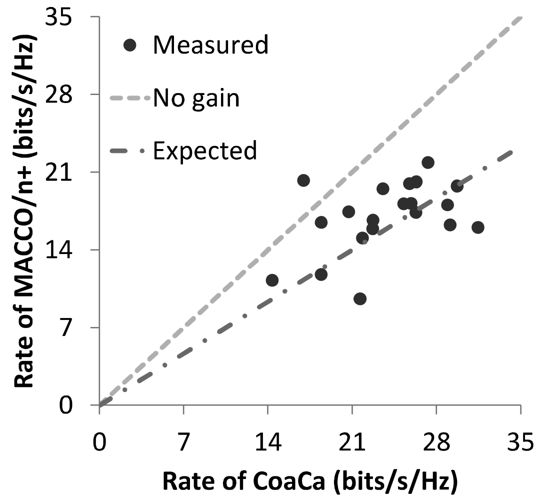
ference cancellation and interference alignment performed by the AP or client, with different number of antennas. We highlight the following findings from the figures. First, on average interference cancellation and alignment are able to reduce the interference by 30 dB and 25 dB, respectively. This is often sufficient to reduce the interference to below the noise floor in 802.11ac networks. Second, interference cancellation by the AP via transmit beamforming, and by the client via receive beamforming have similar accuracy. This is because with explicit channel estimation the AP and client own the same channel knowledge, and therefore the accuracy of interference cancellation is not subject to channel calibration errors that only exist in systems with implicit channel estimation such as [6]. Third, the number of antennas for interference cancellation and alignment does not have a clear impact on the accuracy. While somewhat counter-intuitive, we explain this finding with the observation that additional antennas cannot reduce channel estimation error or hardware nonlinearity. Last, the accuracy of interference alignment is lower than that of interference cancellation. This is because interference alignment suffers more from channel estimation errors by leveraging the channel knowledge from more than one clients. The authors in [65] reported consistent findings regarding the effectiveness of interference cancellation and alignment.

7.4.4 Achievable Rate Improvement

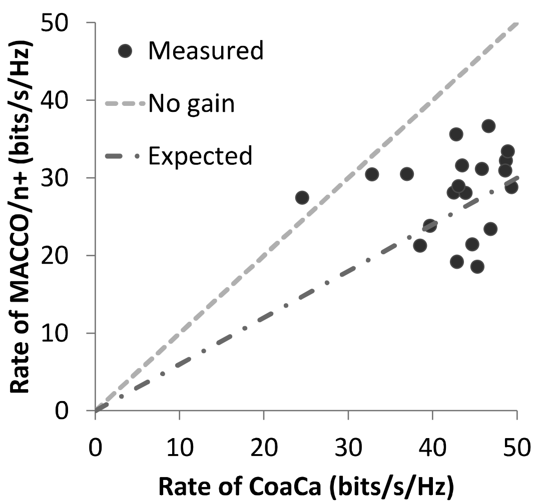
We next evaluate the effectiveness of CoaCa on improving the rate of a MU-MIMO network by achieving a higher multiplexing gain. We compare CoaCa with two existing schemes: (*i*) MACCO which does not cancel inter-cell interference by allowing a single AP to transmit and using the client antennas to improve the channel orthogonality and rate scalability; (*ii*) n+ [65] which allows an AP in one cell to serve its clients first and the AP in the other cell to opportunistically transmit at the same time. Note, given that each cell is congested, n+ cannot increase the multiplexing gain and may only provide a diversity gain as explored in [77]. Therefore, for a given channel condition we compare CoaCa with either MACCO or n+, whichever achieves



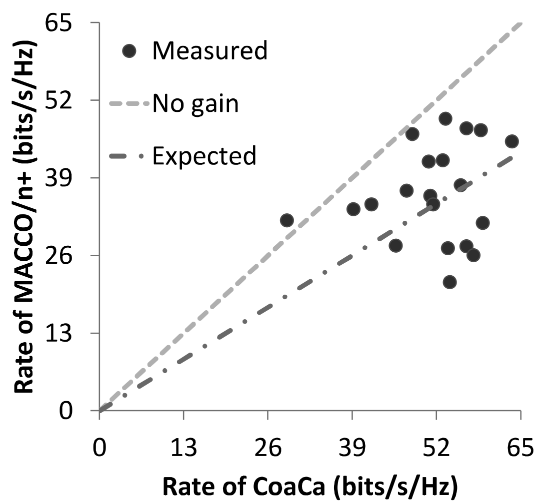
(a) Case 1



(b) Case 2



(c) Case 3



(d) Case 4

Figure 7.5 : Rate achieved by MACCO/n+ and CoaCa for four cases with different numbers of antennas on APs and clients. CoaCa outperforms MACCO/n+ when it delivers more streams. However the rate improvement from CoaCa is lower than the expected multiplexing gain increase due to imperfect channel orthogonality.

higher rate. For simplicity we use "MACCO/n+" to denote the better scheme among the two of them. We select four representative cases with different numbers of antennas on the APs and clients, summarized in Table 7.2. For each case, we run multiple instances of the experiment by deploying the APs and clients in different locations. It is important to mention that CoaCa improves the achievable rate by aiming to deliver more streams in the MU-MIMO network, which is known to be a suboptimal approach. Therefore, the actual rate improvement of CoaCa is dependent on the channel condition, i.e., the orthogonality between the channels of multiple served clients. Conducting the experiment under various channel conditions allows us to observe not only the average but also the wort-case performance of CoaCa. To randomize the channel conditions, we arbitrate the AP and client locations in each experiment instance, since previous work observed that in an indoor environment with rich multipath the channel condition does not have a clear dependence on the node locations [66, 78]. This way, our experiments cover channel conditions that are either beneficial or adverse to CoaCa.

Figure 7.6 shows one instance of our experiments for case 4 with two APs and eight clients. While the clients are in the physical proximity of their intended and interfering APs, we have properly scaled the transmit power of the APs to make sure the received signal to noise ratio (SNR), the interference to noise ratio (INR), and signal to interference ratio (SIR) are representative of those in realistic 802.11 networks. That is, the median SNR, INR, and SIR in our experiments are 18.6 dB, 16.2 dB, and 3.2 dB. To confirm they are representative, we have measured the SNR, INR and SIR of real-world APs in a campus building at about eighty different locations. The median SNR, INR, and SIR in these real-world indoor locations are 21.7 dB, 18.3 dB, and 3.7 dB, and Figure 7.7 shows two examples of the measured SNR and INR (dB) in the first and third floors of the building respectively. While the SNR and INR in our experiments are lower than their counterparts in real-world scenarios, the SIR that dominantly determines the interference condition is similar.

Table 7.1 shows even more real-world SNR, INR and SIR measurements. These measurements were conducted in different interference-limited scenarios such as shopping malls and apartment complexes. It is observed that in the two shopping mall scenarios, the median measured SIRs are close to that in our experiments. However, in a townhouse and in an apartment complex, the median measured SIRs are higher and lower than that in our experiments. This is because in a townhouse the interfering APs from neighbouring houses are relatively fewer and more distant, while in an apartment complex a client can easily see tens of interfering APs that are relatively closer. We acknowledge that our experiments cannot represent all real-world scenarios. Yet one can expect the performance of CoaCa since its effectiveness scales with

Table 7.1 : Median measured SNR, INR, and SIRin multiple interference-limited real-world scenarios.

Scenario	SNR (dB)	INR (dB)	SIR (dB)
Shopping mall 1	24.4	22.1	3.4
Shopping mall 2	27.1	21.4	4.7
Townhouse	30.4	18.6	7.9
Apartment complex	35.8	33.1	1.8

the dominance of the inter-cell interference.

Figure 7.5 shows our results. In each plot, one data point corresponds to the measured rate for CoaCa (X axis) and MACCO/n+ (Y axis) under a single experiment instance. The expected multiplexing gain increase from CoaCa, defined as the increase of the number of streams in the two MU-MIMO cells, is also plotted for comparison. We report the following important findings from Figure 7.5. First, when CoaCa delivers more streams it considerably improves the achievable rate, i.e., on average by 40%, 52% and 41% for Case 2, 3 and 4, respectively. When CoaCa cannot deliver more streams due to insufficient antennas on the APs or clients, e.g., in Case 1, on average it achieves similar rate to MACCO/n+. Second, the rate improvement from CoaCa is lower than the expected multiplexing gain increase (50%, 67%, and

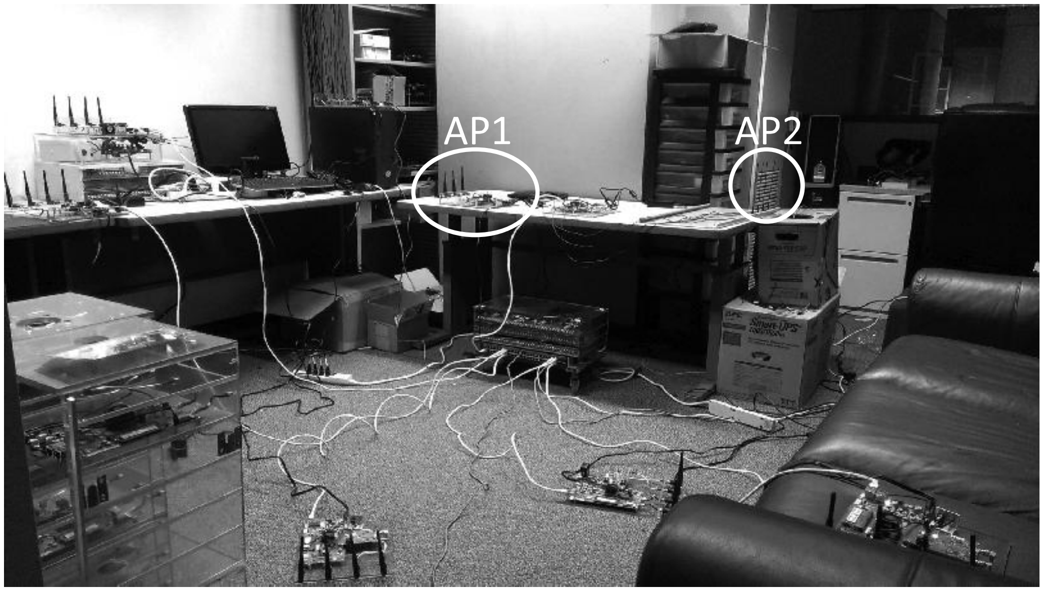


Figure 7.6 : One example experiment for case 4 with two APs and eight clients. The clients are close to not only the associated but also the interfering AP. However the transmit power of the APs is properly scaled down to mimic realistic Wi-Fi signal and interference strength.

Table 7.2 : Number of antennas on the APs and clients for Case 1-4.

	AP1	AP2	Cell 1 clients	Cell 2 clients
Case 1	2	2	2/2	2/2
Case 2	2	2	1/2	1/3
Case 3	2	3	4/4	3/3/4
Case 4	4	4	1/2/4/4	1/2/4/4

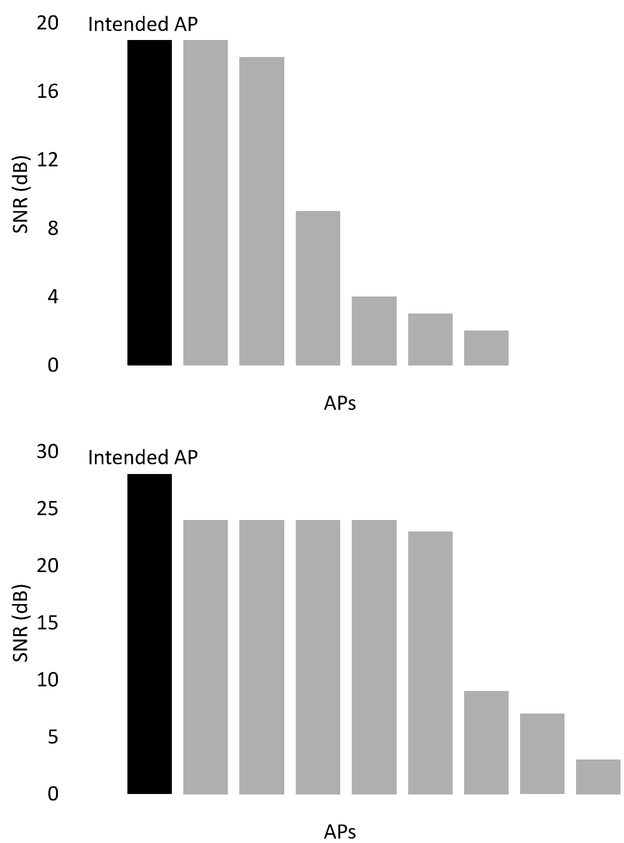


Figure 7.7 : Measured SNR/INR of realistic APs in two example indoor locations in a campus building: (top) first floor and (bottom) third floor.

50% for case 2, 3, and 4, respectively). This is because CoaCa is theoretically sub-optimal and cannot proportionally increase the rate due to its two-step optimization. The per-client rate can be reduced when the antennas are used to cancel the inter-cell interference instead of enhancing the client SNR. Finally, even in Case 2, 3 and 4 where CoaCa allows more streams, there is a small probability that CoaCa does not increase rate. This happens when the antenna usage optimization yields few spare AP and client antennas to enhance the SNR, and the channel orthogonality between clients are far from being orthogonal, such that the rate of each cell suffers from a serious scalability problem.

Average- and worst-case rate improvement of CoaCa. The worst-case performance of CoaCa confirms that due to the use of two-step optimization as a key heuristic CoaCa does not necessarily maximize the achievable rate. While CoaCa offers considerable rate improvement on average, it is important for us to understand how often such worst cases of CoaCa are encountered in practice. To see this, we rely on simulation to examine a large number of cases with various orthogonality between the client channels. Since channel orthogonality is the critical factor that determines the rate gain of CoaCa, we leverage real-world indoor MU-MIMO channels to construct a model of the channel orthogonality between clients. Then, using this model we generate a large number of synthetic MU-MIMO channels and simulate

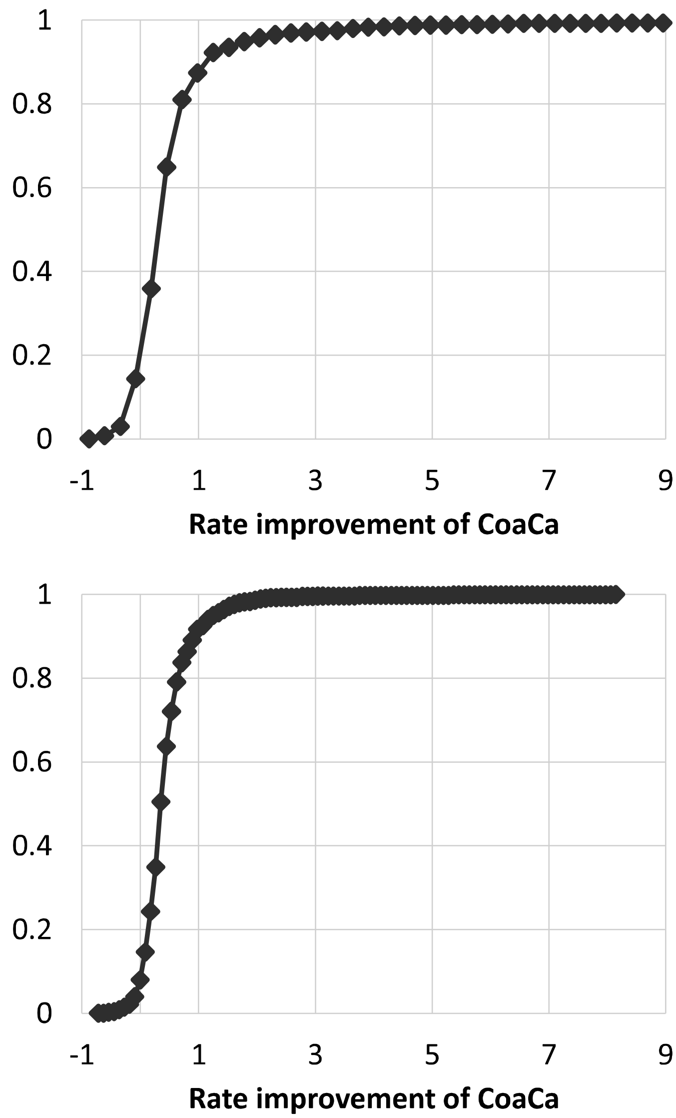


Figure 7.8 : Probability distribution of the rate improvement by CoaCa, under the client SNR of 0-20 dB (top) and 20-40 dB (bottom) respectively.

the performance of CoaCa. Figure 7.8 shows the CDF of the performance gain from CoaCa compared to single-cell solutions, with two different ranges of client SNR: 0-20 dB and 20-40 dB. It can be observed that when the client SNR is between 20-40 dB, CoaCa provides a rate gain in about 98% of the cases; when the client SNR decreases to 0-20 dB, CoaCa is effective in about 80% of the cases. This is because when the client SNR is higher, enabling a greater number of streams more likely increases the rate, since the rate of the client is a logarithmic function of the client SNR.

Dependency on the client SNR. We next evaluate how the client SNR impacts the achievable rate gain from CoaCa. We need to mention that when the client capacity is interference-limited, the client often has a moderate SNR (about 15-25 dB according to our measurements). Therefore we emulate a low SNR regime for each case above, by further reducing the transmit power on the two APs by approximately 15 dB. We demonstrate the results in Figure 7.9 where the average achievable rate for n+, MACCO, and CoaCa are compared for both high (measured) SNR and low (emulated) SNR regimes. Clearly, when the client SNR is reduced, CoaCa achieves lower rate gain: for Case 2-4 the improvement decreases to 28%, 35%, and 30%, respectively. This is because with lower SNR, the logarithmic relationship between rate and SNR becomes closer to linear, and cancelling the inter-cell interference leads to higher rate reduction with a diminished SNR.

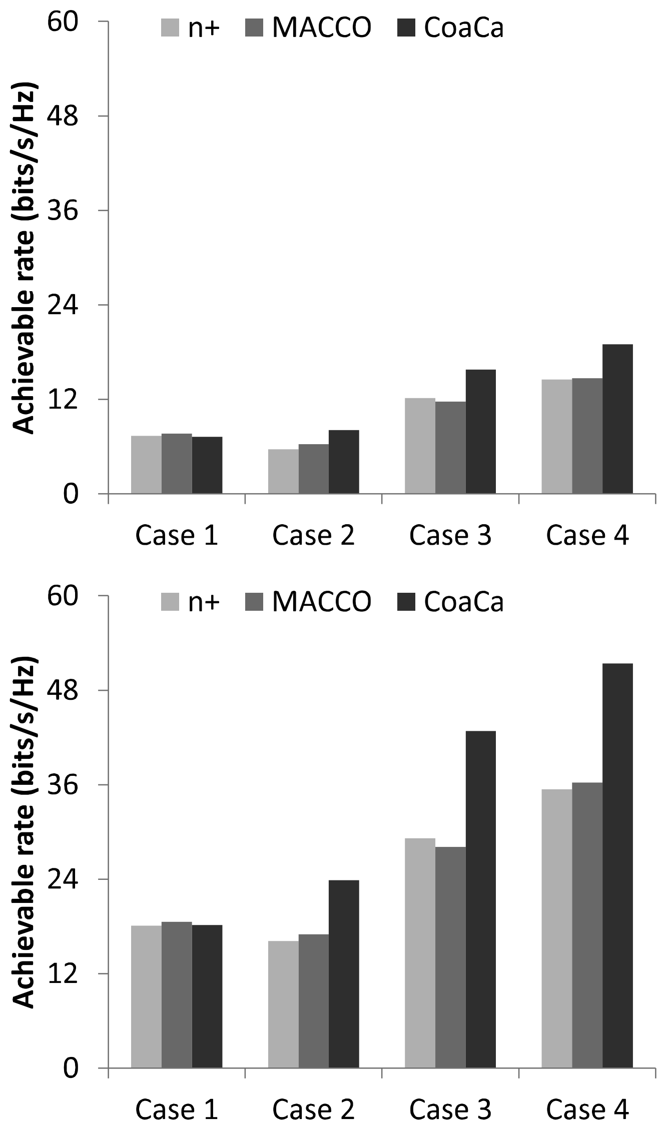


Figure 7.9 : Rate achieved by 802.11n+, MACCO, and CoaCa with high client SNR (top) and low client SNR (bottom). CoaCa achieves higher rate gain with higher client SNR, due to the logarithmic relationship between rate and SNR.

7.4.5 Tradeoff between Achievable Rate and Fairness

The two adaptations of our antenna usage optimization we discussed in Chapter 5.4 enable a tradeoff between fairness and achievable rate in CoaCa: to prevent client starvation or achieving energy-per-stream fairness between clients, CoaCa has to allow a smaller number of streams in the MU-MIMO network, which usually reduces the achievable rate. When such adaptations are applied to the antenna usage optimization in CoaCa, the multiplexing gain and achievable rate improvement of CoaCa needs to be revisited: one must apply a scalar λ to the improvement to reflect the rate loss for achieving fairness. λ can be calculated as follows:

$$\lambda = \frac{g_{ac}T_{ac} + g_c T_c}{g_c T_{ac} + g_c T_c}$$

where g_{ac} and g_c are the average number of streams increase achieved by CoaCa with and without adaptation, T_{ac} and T_c the average number of MU-MIMO transmissions for CoaCa with and without adaptation, respectively. Clearly, when either adaptation is applied to CoaCa for more transmissions, the achievable rate improvement of CoaCa reduces.

To confirm this we next apply the adaptation to CoaCa to prevent client starvation and repeat our experiments for Case 2 (same as Case A in Chapter 5.4). In this case, to make sure no client is served less frequently in CoaCa compared to that in 802.11ac, we apply the enforcement of including the single-client in Cell 2 in CoaCa in half of the MU-MIMO transmissions. As a result, we have $g_{ac} = 0, g_c = 0.5$ and $T_{ac} = T_c$ so that $\lambda = 0.5$. Our evaluation results show that compared to the results in Figure 7.8, the average achievable rate improvement of CoaCa decreases from 40% to 18% which is consistent to our calculation. In other words, in half of the time CoaCa enables the same number of streams as 802.11ac and does not provide a rate improvement.

Chapter 8

Conclusion

In this thesis we tackle a fundamental problem of MU-MIMO networks: the achievable rate of the network does not scale to the number of served clients. When the MU-MIMO network consists of a single cell, poor channel orthogonality leads to a trivially improved or even reduced rate of the cell as the AP simultaneously serves more clients. When the MU-MIMO network includes multiple cells, inter-cell interference prevents the APs from serving their clients at the same time and thus limits the number of streams and achievable rate of the network. Toward solving such rate scalability problem of MU-MIMO networks, we make the following research contributions in this work.

First, we unveil a novel but effective way of tackling the rate scalability problem which is radically different from existing approaches. That is, unlike existing approaches based on Massive-MIMO or Network-MIMO that aim to enhance the MU-MIMO AP, our approach focuses on modern clients that often have multiple antennas and leverages these client antennas for receive beamforming. Beamforming allows the client to create a virtual channel that either is more orthogonal to the channels of other clients in the same cell, or cancels the interference from APs in other cells. While approaches based on Massive-MIMO or Network-MIMO are prohibitively expensive to deploy in practice, our approach applies to arbitrary MU-MIMO network that is either distributed (e.g., home WLAN) or centralized (enterprise WLAN), since it does not require any modification to existing network infrastructure.

Second, we provide two theoretical solutions regarding the optimization of client beamforming. Importantly, these solutions do not rely on the underlying protocol and are therefore applicable to any MU-MIMO network with beamforming clients. For a MU-MIMO network with a single cell, our solution optimizes client beamforming in a successive way such that the SNR and rate of each client is maximized based on the beamforming of all previous clients. For a MU-MIMO network with multiple interfering cells, our solution optimizes client beamforming in a coordinated way in order to cannel inter-cell interference and maximize the number of streams from all cells. While our solutions are theoretically suboptimal and do not necessarily achieve the capacity of the MU-MIMO network, they outperform all existing solutions and significantly reduce the channel knowledge requirement for client beamforming optimization to achieve better practicality.

Third, we devise two 802.11ac-based protocols, namely MACCO and CoaCa, that realize the two solutions respectively with negligible overhead. MACCO works in a single 802.11ac cell and seeks to make the rate of the cell more scalable to the number of clients; CoaCa works in multiple interfering cells and aims to maximize the number of streams from all cells by cancelling the inter-cell interference. Both MACCO and CoaCa leverage the channel sounding process of 802.11ac to let the beamforming clients acquire necessary channel knowledge by simply overhearing the channel reports from other clients. Overhearing is an efficient and reliable way for clients to acquire channel knowledge in 802.11ac networks without introducing any additional frame exchange as overhead. It is meanwhile compatible with unmodified 802.11ac APs and clients. While in this thesis the two protocols are based on 802.11ac, the theoretical solutions we propose are valid for other MU-MIMO networks as well.

Finally, we provide WARP-based implementations of MACCO and CoaCa and experimental evaluation of their performance in real-world indoor environments. Our evaluation results clearly show the effectiveness of MACCO and CoaCa to improve the rate scalability of MU-MIMO networks with one or multiple cells respectively. While due to the lack of suitable platforms we cannot demonstrate the feasibility of integrating MACCO and CoaCa into commercialized 802.11ac devices, we believe our proposals offer important theoretical and practical insights to the future amendments of the 802.11 protocol family as well as other wireless protocols that feature MU-MIMO.

Bibliography

[1] O. Bejarano, E. W. Knightly, and M. Park, "IEEE 802.11ac: From Channelization to Multi-User MIMO," *IEEE Communication Magazine*, 2013.

[2] Long Term Evolution (LTE). `http://http://en.wikipedia.org/wiki/LTE_\%28telecommunication\%29`.

[3] D. Tse and P. Viswanath.

[4] B. Hochwald and S. Vishwanath, "Space-Time Multiple Access: Linear Growth in the Sum Rate," in *Proc. Annual Allerton Conference On Communication, Control and Computing*, 2002.

[5] J. Hoydis, S. Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?," *arXiv:1107.1709v2 [cs.IT]*, 2011.

[6] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical Many-antenna Base Stations," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2012.

[7] Q. Yang, X. Li, H. Yao, J. Fang, K. Tan, W. Hu, J. Zhang, and Y. Zhang,

"BigStation: Enabling Scalable Real-time Signal Processing in Large MU-MIMO Systems," in *Proc. ACM SIGCOMM*, 2013.

[8] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming Intercell Interference in Indoor Wireless Systems," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers (ACSSC)*, 2007.

[9] H. Rahul, S. Kumar, and D. Katabi, "JMB: Scaling Wireless Capacity with User Demands," in *Proc. ACM SIGCOMM*, 2012.

[10] H. V. Balan, R. Rogalin, A. Michaloliakos, and K. Psounis, "Achieving High Data Rates in a Distributed MIMO System," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2012.

[11] A. R. Behzad, Z. M. Shi, S. B. Anand, L. Lin, K. A. Carter, M. S. Kappes, T.-H. Lin, T. Nguyen, D. Yuan, S. Wu, Y. C. Wong, V. Fong, and A. Rofougaran, "A 5-GHz Direct-conversion CMOS Transceiver Utilizing Automatic Frequency Control for the IEEE 802.11a Wireless LAN Standard," *IEEE Journal of Solid-State Circuits (JSSC)*, 2003.

[12] I. Vassiliou, K. Vavelidis, T. Georgantas, S. Plevridis, N. Haralabidis, G. Kamoulakos, C. Kapnistis, S. Kavadias, Y. Kokolakis, P. Merakos, J. C. Rudell, A. Yamanaka, S. Bouras, and I. Bouras, "A Single-chip Digitally Cali-

brated 5.15-5.825-GHz 0.18-μm CMOS Transceiver for 802.11a Wireless LAN," *IEEE Journal of Solid-State Circuits (JSSC)*, 2003.

[13] W. Kluge, L. Dathe, R. Jaehne, S. Ehrenreich, and D. Eggert, "A 2.4GHz CMOS Transceiver for 802.11b Wireless LANs," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2003.

[14] R. Ahola, A. Aktas, J. Wilson, K. R. Rao, F. Jonsson, I. Hyyrylainen, A. Brolin, T. Hakala, A. Friman, T. Makiniemi, J. Hanze, M. Sanden, D. Wallner, Y. Guo, T. Lagerstam, L. Noguer, T. Knuuttila, P. Olofsson, and M. Ismail, "A Single Chip CMOS Transceiver for 802.11 a/b/g WLANs," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2004.

[15] K. Vavelidis, I. Vassiliou, T. Georgantas, A. Yamanaka, S. Kavadias, G. Kamoulakos, C. Kapnistis, Y. Kokolakis, A. Kyranas, P. Merakos, I. Bouras, S. Bouras, S. Plevridis, and N. Haralabidis, "A Dual-band 5.15-5.35-GHz, 2.4-2.5-GHz 0.18-μm CMOS Transceiver for 802.11a/b/g Wireless LAN," *IEEE Journal of Solid-State Circuits (JSSC)*, 2004.

[16] T. B. Cho, D. Kang, C.-H. Heng, and B.-S. Song, "A 2.4-GHz Dual-mode 0.18-μm CMOS Transceiver for Bluetooth and 802.11b," *IEEE Journal of Solid-State Circuits (JSSC)*, 2004.

[17] L. Perraud, M. Recouly, C. Pinatel, N. Sornin, J.-L. Bonnot, F. Benoist, N. Massei, and O. Gibrat, "A Direct-conversion CMOS Transceiver for the 802.11a/b/g WLAN Standard Utilizing a Cartesian Feedback Transmitter," *IEEE Journal of Solid-State Circuits (JSSC)*, 2004.

[18] M. Zargari, M. Terrovitis, S. H.-M. Jen, B. J. Kaczynski, M. Lee, M. P. Mack, S. S. Mehta, S. Mendis, K. Onodera, H. Samavati, W. W. Si, K. Singh, A. Tabatabaei, D. Weber, D. K. Su, and B. A. Wooley, "A Single-chip Dual-band Tri-mode CMOS Transceiver for IEEE 802.11a/b/g Wireless LAN," *IEEE Journal of Solid-State Circuits (JSSC)*, 2004.

[19] Z. Xu, S. Jiang, Y. Wu, H.-Y. Jian, G. Chu, K. Ku, P. Wang, N. Tran, Q. Gu, M.-Z. Lai, C. Chien, A. F. Chang, and R. D. Chow, "A Compact Dual-band Direct-conversion CMOS Transceiver for 802.11a/b/g WLAN," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2005.

[20] P. Zhang, L. Der, D. Guo, I. Sever, T. Bourdi, C. Lam, A. Zolfaghari, J. Chen, D. Gambetta, B. Cheng, S. Gowder, S. Hart, L. Huynh, T. Nguyen, and B. Razavi, "A Single-chip Dual-band Direct-conversion IEEE 802.11a/b/g WLAN Transceiver in 0.18-μm CMOS," *IEEE Journal of Solid-State Circuits (JSSC)*, 2005.

[21] D. G. Rahn, M. S. Cavin, F. F. Dai, N. H. W. Fong, R. Griffith, J. Macedo, A. D. Moore, J. W. M. Rogers, and M. Toner, "A Fully Integrated Multiband MIMO WLAN Transceiver RFIC," *IEEE Journal of Solid-State Circuits (JSSC)*, 2005.

[22] E. Song, Y. Koo, Y.-J. Jung, D.-H. Lee, S. Chu, and S.-I. Chae, "A 0.25-μm CMOS Quad-band GSM RF Transceiver Using An Efficient LO Frequency Plan," *IEEE Journal of Solid-State Circuits (JSSC)*, 2005.

[23] Y.-H. Hsieh, W.-Y. Hu, S.-M. Lin, C.-L. Chen, W.-K. Li, S.-J. Chen, and D. J. Chen, "An Auto-I/Q Calibrated CMOS Transceiver for 802.11g," *IEEE Journal of Solid-State Circuits (JSSC)*, 2005.

[24] L. Nathawad, D. Weber, S. Abdollahi, P. Chen, S. Enam, B. Kaczynski, A. Kheirkhahi, M. Lee, S. Limotyrakis, K. Onodera, K. Vleugels, M. Zargari, and B. Wooley, "An IEEE 802.11a/b/g SoC for Embedded WLAN Applications," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2006.

[25] K. Muhammad, Y.-C. Ho, T. L. Mayhugh, C.-M. Hung, T. Jung, I. Elahi, C. Lin, I. Deng, C. Fernando, J. Wallberg, S. Vemulapalli, S. Larson, T. Murphy, D. Leipold, P. Cruise, J. Jaehnig, M.-C. Lee, R. B. Staszewski, R. Staszewski, and K. Maggio, "The First Fully Integrated Quad-Band GSM/GPRS Receiver in a 90-nm Digital CMOS Process," *IEEE Journal of Solid-State Circuits (JSSC)*,

2006.

[26] Y. Palaskas, A. Ravi, S. Pellerano, B. R. Carlton, M. A. Elmala, R. Bishop, G. Banerjee, R. B. Nicholls, S. K. Ling, N. Dinur, S. S. Taylor, and K. Soumyanath, "A 5-GHz 108-Mb/s 2 × 2 MIMO Transceiver RFIC With Fully Integrated 20.5-dBm P_{1dB} Power Amplifiers in 90-nm CMOS," *IEEE Journal of Solid-State Circuits (JSSC)*, 2006.

[27] A. Behzad, K. A. Carter, H.-M. Chien, S. Wu, M.-A. Pan, C. P. Lee, Q. Li, J. C. Leete, S. Au, M. S. Kappes, Z. Zhou, D. Ojo, L. Zhang, A. Zolfaghari, J. Castanada, H. Darabi, B. Yeung, A. Rofougaran, M. Rofougaran, J. Trachewsky, T. Moorti, R. Gaikwad, A. Bagchi, J. S. Hammerschmidt, J. Pattin, J. J. Rael, and B. Marholev, "A Fully Integrated MIMO Multiband Direct Conversion CMOS Transceiver for WLAN Applications (802.11n)," *IEEE Journal of Solid-State Circuits (JSSC)*, 2007.

[28] T.-M. Chen, Y.-M. Chiu, C.-C. Wang, K.-U. Chan, Y.-H. Lin, M.-C. Huang, C.-H. Lu, W.-S. Wang, C.-S. Hu, C.-C. Lee, J.-Z. Huang, B.-I. Chang, S.-C. Yen, and Y.-Y. Lin, "A Low-Power Fullband 802.11a/b/g WLAN Transceiver With On-Chip PA," *IEEE Journal of Solid-State Circuits (JSSC)*, 2007.

[29] Z. Li, W. Ni, J. Ma, M. Li, D. Ma, D. Zhao, J. Mehta, D. Hartman, X. Wang,

K. K. O, and K. Che, "A Dual-Band CMOS Transceiver for 3G TD-SCDMA," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2007.

[30] M. Hammes, C. Kranz, J. Kissing, D. Seippel, P.-H. Bonnaud, and E. Pelos, "A GSM Baseband Radio in 0.13 μm CMOS with Fully Integrated Power-Management," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2007.

[31] S. Kousai, D. Miyashita, J. Wadatsumi, A. Maki, T. Sekiguchi, R. Ito, and M. Hamada, "A 1.2V 0.2-to-6.3GHz Transceiver with Less Than -29.5dB EVM@-3dBm and a Choke/Coil-Less Pre-Power Amplifier," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2008.

[32] L. Nathawad, M. Zargari, H. Samavati, S. Mehta, A. Kheirkhahi, P. Chen, K. Gong, B. Vakili-Amini, J. Hwang, M. Chen, M. Terrovitis, B. Kaczynski, S. Limotyrakis, M. Mack, H. Gan, M. Lee, S. Abdollahi-Alibeik, B. Baytekin, K. Onodera, S. Mendis, A. Chang, S. Jen, D. Su, and B. Wooley, "A Dual-Band CMOS MIMO Radio SoC for IEEE 802.11n Wireless LAN," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2008.

[33] O. Degani, M. Ruberto, E. Cohen, Y. Eilat, B. Jann, F. Cossoy, N. Telzhensky, T. Maimon, G. Normatov, R. Banin, O. Ashkenazi, A. Ben Bassat, S. Zaguri, G. Hara, M. Zajac, E. Shaviv, S. Wail, A. Fridman, R. Lin, and S. Gross, "A 1×2

MIMO Multi-Band CMOS Transceiver with an Integrated Front-End in 90nm CMOS for 802.11a/g/n WLAN Applications," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2008.

[34] M. Locher, J. Kuenen, A. Daanen, H. Visser, B. H. Essink, P.-P. Vervoort, R. Kopmeiners, W. Alkema, W. Redman-White, R. Balmford, and R. El Waffaoui, "A Versatile, Low Power, High Performance BiCMOS MIMO/Diversity Direct Conversion Transceiver IC for WiBro/WiMAX (802.16e)," *IEEE Journal of Solid-State Circuits (JSSC)*, 2008.

[35] T. Sowlati, B. Agarwal, J. Cho, T. Obkircher, M. El Said, J. Vasa, B. Ramachandran, M. Kahrizi, E. Dagher, W.-H. Chen, M. Vadkerti, G. Taskov, U. Seckin, H. Firouzkouhi, B. Saeidi, H. Akyol, Y. Choi, A. Mahjoob, S. D'Souza, C.-Y. Hsieh, D. Guss, D. Shum, D. Badillo, I. Ron, D. Ching, F. Shi, Y. He, J. Komaili, A. Loke, R. Pullela, E. Pehlivanoglu, H. Zarei, S. Tadjpour, D. Agahi, D. Rozenblit, W. Domino, G. Williams, N. Damavandi, S. Wloczysiak, S. Rajendra, A. Paff, and T. Valencia, "Single-chip Multiband WCDMA/HSDPA/HSUPA/EGPRS Transceiver with Diversity Receiver and 3G DigRF Interface without SAW Filters in Transmitter/3G Receiver Paths," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2009.

[36] L. Lin, N. Wongkomet, D. Yu, C.-H. Lin, M. He, B. Nissim, S. Lyuee, P. Yu, T. Sepke, S. Shekarchian, L. Tee, P. Muller, J. Tam, and T. Cho, "A Fully Integrated 2×2 MIMO Dual-band Dual-mode Direct-conversion CMOS Transceiver for WiMAX/WLAN Applications," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2009.

[37] M. Ingels, V. Giannini, J. Borremans, G. Mandal, B. Debaillie, P. Van Wesemael, T. Sano, T. Yamamoto, D. Hauspie, J. Van Driessche, and J. Craninckx, "A 5mm^2 40nm LP CMOS 0.1-to-3GHz Multistandard Transceiver," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2010.

[38] J. Deguchi, D. Miyashita, Y. Ogasawara, G. Takemura, M. Iwanaga, K. Sami, R. Ito, J. Wadatsumi, Y. Tsuda, S. Oda, S. Kawaguchi, N. Itoh, and M. Hamada, "A Fully Integrated 2×1 Dual-Band Direct-Conversion Mobile WiMAX Transceiver With Dual-Mode Fractional Divider and Noise-Shaping Transimpedance Amplifier in 65 nm CMOS," *IEEE Journal of Solid-State Circuits (JSSC)*, 2010.

[39] Q. Huang, J. Rogin, X. Chen, D. Tschopp, T. Burger, T. Christen, D. Papadopoulos, I. Kouchev, C. Martelli, and T. Dellsperger, "A Tri-band SAW-less WCDMA/HSPA RF CMOS Transceiver with On-chip DC-DC Converter Con-

nectable to Battery," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2010.

[40] S. Abdollahi-Alibeik, D. Weber, H. Dogan, W. W. Si, B. Baytekin, A. Komijani, R. Chang, B. Vakili-Amini, . Lee, H. Gan, Y. Rajavi, H. Samavati, B. Kaczynski, S.-M. Lee, S. Limotyrakis, H. Park, P. Chen, P. Park, M. S. Chen, A. Chang, Y. Oh, J. J. Yang, E. C. Lin, L. Nathawad, K. Onodera, M. Terrovitis, S. Mendis, K. Shi, S. Mehta, M. Zargari, and D. Su, "A 65nm Dual-band 3-stream 802.11n MIMO WLAN SoC," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2011.

[41] M. Nilsson, S. Mattisson, N. Klemmer, M. Anderson, T. Arnborg, P. Caputa, S. Ek, L. Fan, H. Fredriksson, F. Garrigues, H. Geis, H. Hagberg, J. Hedestig, H. Huang, Y. Kagan, N. Karlsson, H. Kinzel, T. Mattsson, T. Mills, F. Mu, A. Mrtensson, L. Nicklasson, F. Oredsson, U. Ozdemir, F. Park, T. Pettersson, T. Phlsson, M. Plsson, S. Ramon, M. Sandgren, P. Sandrup, A. Stenman, R. Strandberg, L. Sundstrom, F. Tillman, T. Tired, S. Uppathil, J. Walukas, E. Westesson, X. Zhang, and P. Andreani, "A 9-band WCDMA/EDGE Transceiver Supporting HSPA Evolution," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2011.

[42] A. Cicalini, S. Aniruddhan, R. Apte, F. Bossu, O. Choksi, D. Filipovic, K. Godbole, T.-P. Hung, C. Komninakis, D. Maldonado, C. Narathong, B. Nejati, D. O'Shea, X. Quan, R. Rangarajan, J. Sankaranarayanan, A. See, R. Sridhara, B. Sun, W. Su, K. Van Zalinge, G. Zhang, and K. Sahota, "A 65nm CMOS SoC with Embedded HSDPA/EDGE Transceiver, Digital Baseband and Multimedia Processor," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2011.

[43] O. Oliaei, M. Kirschenmann, D. Newman, K. Hausmann, H. Xie, P. Rakers, M. Rahman, M. Gomez, C. Yu, B. Gilsdorf, and K. Sakamoto, "A Multiband Multimode Transmitter without Driver Amplifier," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2012.

[44] P. Madoglio, A. Ravi, H. Xu, K. Chandrashekar, M. Verhelst, S. Pellerano, L. Cuellar, M. Aguirre, M. Sajadieh, O. Degani, H. Lakdawala, and Y. Palaskas, "A 20dBm 2.4GHz Digital Outphasing Transmitter for WLAN Application in 32nm CMOS," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2012.

[45] R. Kumar, T. Krishnaswamy, G. Rajendran, D. Sahu, A. Sivadas, M. Nandigam, S. Ganeshan, S. Datla, A. Kudari, H. Bhasin, M. Agrawal, S. Narayan, Y. Dharwekar, R. Garg, V. Edayath, T. Suseela, V. Jayaram, S. Ram, V. Murugan, A. Kumar, S. Mukherjee, N. Dixit, E. Nussbaum, J. Dror, N. Ginzburg,

A. EvenChen, A. Maruani, S. Sankaran, V. Srinivasan, and V. Rentala, "A Fully Integrated 2×2 b/g and 1×2 a-Band MIMO WLAN SoC in 45nm CMOS for Multi-Radio IC," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2013.

[46] L. Ye, J. Chen, L. Kong, P. Cathelin, E. Alon, and A. Niknejad, "A Digitally Modulated 2.4GHz WLAN Transmitter with Integrated Phase Path and Dynamic Load Modulation in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2013.

[47] P. Rossi, N. Codega, D. Gerna, A. Liscidini, D. Ottini, Y. He, A. Pirola, E. Sacchi, G. Uehara, C. Yang, and R. Castello, "An LTE Transmitter Using A Class-A/B Power Mixer," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2013.

[48] J.-W. Lai, C.-H. Wang, K. Kao, A. Lin, Y.-H. Cho, L. Cho, M.-H. Hung, X.-Y. Shih, C.-M. Lin, S.-H. Yan, Y.-H. Chung, P. C. P. Liang, G.-K. Dehng, H.-S. Li, G. Chien, and R. B. Staszewski, "A 0.27mm^2 13.5dBm 2.4GHz All-Digital Polar Transmitter Using 34%-Efficiency Class-D DPA in 40nm CMOS," in *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2013.

[49] IEEE, "International Solid-State Circuit Conference (ISSCC)." `http://isscc.org/`.

[50] IEEE, "Journal of Solid-State Circuits (JSSC)." `http://sscs.ieee.org/`

`ieee-journal-of-solid-state-circuits-jssc.html`.

[51] H. Yu, L. Zhong, A. Sabharwal, and D. Kao, “Beamforming on Mobile Devices: A First Study,” in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2011.

[52] W. Shen, Y. Tung, K. Lee, K. C. Lin, S. Gollakota, D. Katabi, and M. Chen, “Rate Adaptation for 802.11 Multiuser MIMO Networks,” in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2012.

[53] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels,” *IEEE Trans. Signal Processing*, 2004.

[54] M. Ku and D. Kim, “Tx-Rx Beamforming with Multiuser MIMO Channels in Multiple-cell Systems,” in *Proc. Int. Conf. Advanced Communication Technology*, 2008.

[55] M. Codreanu, A. Tolli, M. Juntti, and M. Latva-aho, “Joint Design of Tx-Rx Beamformers in MIMO Downlink Channel,” *IEEE Trans. Signal Processing*, 2007.

[56] C. Chae, D. Mazzarese, and T. I. R. W. Heath, “Coordinated Beamforming

for the Multiuser MIMO Broadcast Channel With Limited Feedforward," *IEEE Trans. Signal Processing*, 2008.

[57] X. Xie, X. Zhang, and K. Sundaresan, "Adaptive Feedback Compression for MIMO Networks," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2013.

[58] Rice University, "Wireless Open Access Research Platform (WARP)." `http://warp.rice.edu/trac`.

[59] K. Tan, H. Liu, J. Fang, W. Wang, J. Zhang, M. Chen, and G. M. Voelker, "SAM: Enabling Practical Spatial Multiple Access in Wireless LAN," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2009.

[60] M. Costa, "Writing on Dirty Paper," *IEEE Trans. Information Theory*, 1983.

[61] H. Weingarten, Y. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *IEEE Trans. Information Theory*, 2006.

[62] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication Over MIMO X Channels: Interference Alignment, Decomposition, and Performance Analysis," *IEEE Trans. on Information Theory*, 2008,.

[63] V. R. Cadambe and S. A. Jafar, "Interference Alignment and Degrees of Freedom of the K-User Interference Channel," *IEEE Trans. on Information Theory*, 2008.

[64] S. Gollakota, S. D. Perli, and D. Katabi, "Interference Alignment and Cancellation," in *Proc. ACM SIGCOMM*, 2009.

[65] K. C. Lin, S. Gollakota, and D. Katabi, "Random Access Heterogeneous MIMO networks," in *Proc. ACM SIGCOMM*, 2011.

[66] E. Aryafar, N. Anand, T. Salonidis, and E. W. Knightly, "Design and Experimental Evaluation of Multi-User Beamforming in Wireless LANs," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2010.

[67] O. Bejarano, E. Magistretti, O. Gurewitz, and E. W. Knightly, "MUTE: Sounding Inhibition for MU-MIMO WLANs," in *Proc. IEEE Int. Conf. Sensing, Communication and Networking (SECON)*, 2014.

[68] N. Anand, R. E. Guerra, and E. W. Knightly, "The Case for UHF-Band MU-MIMO," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2014.

[69] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diago-

nalization," *IEEE Trans. Signal Processing*, 2006.

[70] N. Jindal, "Antenna Combining for the MIMO Downlink Channel," *IEEE Trans. Wireless Communications*, 2008.

[71] T. Bansal, B. Chen, K. Srinivasan, and P. Sinha, "RobinHood: Sharing the Happiness in a Wireless Jungle," in *Proc. ACM Int. Workshop Mobile Computing Systems and Applications (HotMobile)*, 2014.

[72] S. Kumar, D. Cifuentes, S. Gollakota, and D. Katabi, "Bringing Cross-Layer MIMO to Today's Wireless LANs," in *Proc. ACM SIGCOMM*, 2013.

[73] H. Dahrouj and W. Yu, "Coordinated Beamforming for the Multicell Multi-antenna Wireless System," *IEEE Trans. Wireless Communications*, 2010.

[74] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated Multipoint: Concepts, Performance, and Field Trial Results," *IEEE Communications Magazine*, 2011.

[75] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated Multipoint Transmission/Reception Techniques for LTE-advanced [Coordinated and Distributed MIMO]," *IEEE Trans. Wireless Communications*, 2010.

[76] X. Zhang, K. Sundaresan, M. A. Khojastepour, S. Rangarajan, and K. G. Shin, "NEMOx: Scalable Network MIMO for Wireless Networks," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2013.

[77] B. Chen, K. C. Lin, and H. Wei, "Harnessing Receive Diversity in Distributed Multi-User MIMO Networks," in *Proc. ACM SIGCOMM (Poster)*, 2013.

[78] H. Yu, L. Zhong, and A. Sabharwal, "Achieving Better Channel Orthogonality for Improved User Scaling of Multi-user MIMO." *Technical Report 2014-06-28*, Rice University, 2014.

[79] Rice University, "WARPLab." `http://warp.rice.edu/trac/wiki/WARPLab`.

[80] V. Erceg, L. Schumacher, P. Kyritsi, A. Molisch, and D. S. Baum, "TGn Channel Models." IEEE Doc. 802.11-03/940r2, 2004.

[81] C. Li, C. Peng, S. Lu, and X. Wang, "Energy-based Rate Adaptation for 802.11n," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, 2012.

[82] H. Yu, L. Zhong, and A. Sabharwal, "Power Management of MIMO Network Interfaces on Mobile Systems," *IEEE Trans. VLSI*, 2012.

[83] Z. Zeng, Y. Gao, K. Tan, and P. R. Kumar, "CHAIN: Introducing Minimum Con-

trolled Coordination into Random Access MAC," in *Proc. IEEE Conf. Computer Communications (INFOCOM)*, 2011.