

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

A Scalable and Cost Effective Architecture for High Gain Beamforming Antennas

Permalink

<https://escholarship.org/uc/item/6pv725qz>

Author

BAKR, Omar Mohammed

Publication Date

2010

Peer reviewed|Thesis/dissertation

**A Scalable and Cost Effective Architecture for High Gain Beamforming
Antennas**

by

Omar Mohammed Bakr

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ali M. Niknejad, Chair
Professor Eric A. Brewer
Professor Paul K. Wright

Fall 2010

**A Scalable and Cost Effective Architecture for High Gain Beamforming
Antennas**

Copyright 2010
by
Omar Mohammed Bakr

Abstract

A Scalable and Cost Effective Architecture for High Gain Beamforming Antennas

by

Omar Mohammed Bakr

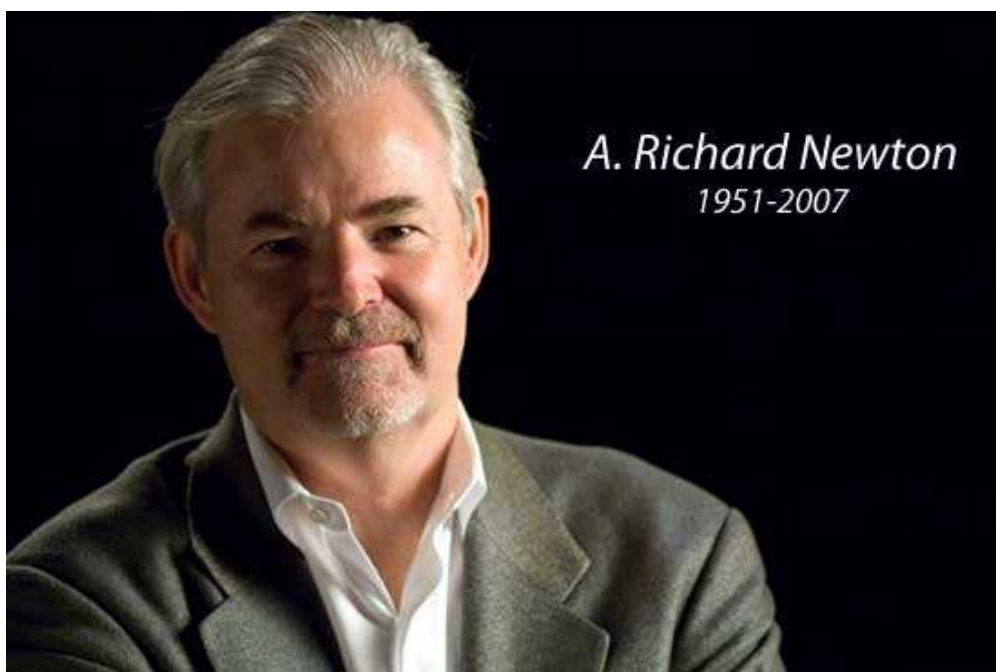
Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Ali M. Niknejad, Chair

Many state-of-the-art wireless systems, such as long distance networks (point-to-point, point-to-multipoint, and mesh) and high bandwidth networks using mm-wave frequencies, require high gain antennas to overcome adverse channel conditions. These networks could be greatly aided by adaptive beamforming antenna arrays, which can significantly simplify the installation and maintenance costs (e.g., by enabling automatic beam alignment), and improve the capacity of these networks. Such networks typically require gains ranging from 20-30dBi with wide scanning range in both dimensions. To achieve this, arrays with hundreds or even thousands of antennas are required, which cannot be done with existing techniques that do not scale very well beyond 10-20 antennas.

In this dissertation, we examine and address the main challenges presented by large arrays, starting from electromagnetic/antenna and radio circuit design and proceeding to the signal processing and algorithms domain. We propose 3-dimensional antenna array structures that realize large gains and scan angles at a much reduced size and form factor compared with conventional planar antennas. At the circuit level, we propose a hybrid RF/digital beamforming radio architecture that takes advantage of low cost silicon integration to reduce the overall component count and power consumption levels of the system without limiting the capacity. We consider different techniques for implementing compact beamformers reliably at high radio frequencies, and present signal processing techniques based on adaptive filtering methods for optimizing those beamformers. The performance implications of low precision analog beamformers and implementation errors are also analyzed and quantified, and computationally efficient vector quantization techniques that take advantage of the size and scale of the arrays to compensate for low precision are proposed. We validate our approach with mathematical proofs and computer simulations.



To my mentor and friend.

Contents

List of Figures	v
List of Tables	ix
1 Introduction	1
1.1 The wireless revolution	1
1.2 Rural connectivity and the digital divide	1
1.3 Challenges for growing and expanding the reach of wireless networks	2
1.4 Wireless networks today	4
1.5 Adaptive beamforming systems	5
1.6 High gain beamforming systems	9
1.7 Research goals and contributions	9
1.8 Organization of the Dissertation	10
2 Background	11
2.1 Antenna basics	11
2.1.1 Antenna radiation patterns	11
2.1.2 Directivity	12
2.1.3 Antenna efficiency	16
2.1.4 Antenna bandwidth	17
2.1.5 Polarization	17
2.1.6 Free space propagation	18
2.2 Static versus steerable antennas	20
2.2.1 Other types of antenna arrays	25
2.2.2 Scanning versus adaptive arrays	25
2.3 The wireless channel	26
2.3.1 Channel fading	27
2.3.2 Channel diversity	32
2.3.3 Frequency selective channels	33
2.3.4 Spatial multiplexing	34
3 Radio architectures	36
3.1 Comparison of Architecture	36
3.1.1 Hybrid arrays	39
3.2 RF beamformer design	41

3.2.1	A multi-input/multi-output RF beamforming module	41
3.2.2	Signal combining/splitting	46
3.2.3	RF phase shifting and complex multiplication techniques	47
3.2.4	Vector combining complex multiplier implementation	55
3.3	Conclusion	59
4	Antenna designs	62
4.1	Array factor and radiation pattern calculations	63
4.1.1	Impact of antenna spacing	63
4.1.2	Impact of array geometry	64
4.1.3	Planar arrays	66
4.1.4	Side lobes	72
4.1.5	Antenna coupling	73
4.2	3-dimensional arrays	75
4.2.1	Lobe cancellation	77
5	Adaptive beamforming and channel estimation	82
5.1	Adaptive filter techniques	83
5.1.1	Gradient descent	85
5.1.2	Other adaptive filtering techniques	87
5.2	Channel Estimation via Adaptive Filtering	87
5.2.1	Performance evaluation	89
5.3	Null-steering	93
5.4	Adaptive filtering with interference cancellation	99
5.5	Frequency synchronization	103
5.6	Multi-beam (multi-receiver) arrays	105
5.7	Transmit beamforming	107
5.8	Conclusion	109
6	Beamforming performance with weight errors	110
6.1	Beamforming	111
6.1.1	Analysis of Imperfect Phase Shifts	112
6.1.2	Simulation Results	113
6.1.3	Worst case performance analysis	116
6.2	Beam-nulling	121
6.2.1	Analysis of Errors in Weights	122
6.2.2	Simulation Results	125
6.3	Conclusion	125
7	Error mitigation with vector quantization	132
7.1	Quantization problem statement	133
7.2	Scalar quantization of beamforming weights	134
7.2.1	Spatial matched filter	135
7.2.2	Zero-Forcing beamformer	137
7.3	Vector quantization of antenna weights	156

7.3.1	An approximate lower bound on achievable SIR	156
7.3.2	Constructive algorithms for improving interference suppression	175
7.3.3	Improving the beamforming gain using vector quantization	201
7.4	Conclusion	205
8	Conclusion and future work	209
	Bibliography	211

List of Figures

1.1	Long distance wireless networks for rural connectivity	3
1.2	Mobile data traffic	3
1.3	Antenna pattern options	6
1.4	Network sectorization and cell splitting	7
1.5	Adaptive beamforming antennas	8
2.1	3D antenna radiation pattern plots	13
2.2	2D antenna radiation pattern plots	14
2.3	Basic antennas patterns	15
2.4	Free space path loss	18
2.5	Phase response of a uniform linear array	21
2.6	Linear array pattern	23
2.7	Switched antenna array	26
2.8	2x2 MIMO scenario	34
3.1	Radio architectures for beamforming	37
3.2	Hybrid beamforming architectures	42
3.3	Multi-input/multi-output RF frontend	43
3.4	Hybrid RF beamforming chip on a PCB	43
3.5	Transmit/Receive building blocks	45
3.6	Multi-input/Multi-output RF complex multiplier block	45
3.7	Power combining structures	48
3.8	Wilkinson power combiner measurements (Part 1)	49
3.9	Wilkinson power combiner measurements	50
3.10	Reflective-type phase shifters	51
3.11	Switched transmission line phase shifter	53
3.12	Vector combining based phase rotator	54
3.13	Polyphase filter networks	57
3.14	Wideband 4-stage polyphase network	58
3.15	Gain control (digital implementation)	59
3.16	Programable gain block (unit cell architecture)	60
4.1	Impact of number of antennas/antenna spacing on the array pattern	65
4.2	Broadside/endfire configurations of a linear array	66
4.3	Impact of array geometry on beamwidth and pattern shape	67

4.4	Broadside versus endfire azimuth beamwidth	68
4.5	Beam pattern of a linear array for different steering vectors	69
4.6	Array response versus frequency	71
4.7	Azimuth endfire pattern of rectangular arrays	72
4.8	Side lobe level versus array geometry	73
4.9	Dipole radiation pattern (polar plot)	74
4.10	3-dimensional rectangular array of monopoles	78
4.11	3-dimensional cylindrical array of monopoles	79
4.12	Patch arrays	80
4.13	3D array layout with $\lambda/4$ shifting	80
4.14	Lobe cancellation using $\lambda/4$ spacing	81
5.1	Block diagram for a generic adaptive filter	86
5.2	System identification via adaptive filtering	88
5.3	Channel estimation through adaptive filtering (receive beamforming)	90
5.4	Adaptive filter beamforming performance without noise	92
5.5	Adaptive filter beamforming performance with noise (NLMS)	94
5.6	Adaptive filter beamforming performance with noise (sign-LMS)	95
5.7	Adaptive filter beamforming performance with channel variation	96
5.8	Interference cancellation scenario	97
5.9	Beam-nulling using the LMS algorithm	100
5.10	Beam-nulling using the LMS algorithm with time-varying channels	101
5.11	Channel estimation through adaptive filtering with known interferers	103
5.12	Iterative beam-nulling using the LMS algorithm	104
5.13	DMMSE vs LMS in the presence of frequency offsets	106
5.14	Computation of digital beamforming weights	108
5.15	Channel estimation through adaptive filtering (transmit beamforming)	109
6.1	LOS beamforming scenario	111
6.2	Beamforming gain as a function of phase errors	114
6.3	Simulated beamforming gain for different phase errors distributions	115
6.4	Array patterns with quantization (Part 1)	117
6.5	Array patterns with quantization (Part 2)	118
6.6	Array patterns with quantization (Part 3)	119
6.7	Spatial domain wireless channel representation	123
6.8	Mean square error angle σ_θ^2 versus weight errors	126
6.9	Average interferer power in the presence of errors	127
6.10	Average desired power (beamforming gain) in the presence of errors	128
6.11	Cosine angle approximation for the beamforming gain	129
6.12	Cosine angle approximation for the beamforming gain	130
7.1	SIR under QMF (Rayleigh channel, Part 1)	138
7.2	SIR under QMF (Rayleigh channel, Part 2)	139
7.3	Desired signal under QMF (Rayleigh channel, Part 1)	140
7.4	Desired signal under QMF (Rayleigh channel, Part 2)	141

7.5	Interferer power under QMF (Rayleigh channel, Part 1)	142
7.6	Interferer power under QMF (Rayleigh channel, Part 2)	143
7.7	SIR under QMF (LOS channel/uniform linear array, Part 1)	144
7.8	SIR under QMF (LOS channel/uniform linear array, Part 2)	145
7.9	Desired signal under QMF (LOS channel/uniform linear array, Part 1)	146
7.10	Desired signal under QMF (LOS channel/uniform linear array, Part 2)	147
7.11	Interferer power under QMF (LOS channel/uniform linear array, Part 1)	148
7.12	Interferer power under QMF (LOS channel/uniform linear array, Part 2)	149
7.13	SIR under QMF (LOS channel/uniform square array, Part 1)	150
7.14	SIR under QMF (LOS channel/uniform square array, Part 2)	151
7.15	Desired signal under QMF (LOS channel/uniform square array, Part 1)	152
7.16	Desired signal under QMF (LOS channel/uniform square array, Part 2)	153
7.17	Interferer power under QMF (LOS channel/uniform square array, Part 1)	154
7.18	Interferer power under QMF (LOS channel/uniform square array, Part 2)	155
7.19	SIR under QZF (Rayleigh channel, Part 1)	157
7.20	SIR under QZF (Rayleigh channel, Part 2)	158
7.21	Desired signal under QZF (Rayleigh channel, Part 1)	159
7.22	Desired signal under QZF (Rayleigh channel, Part 2)	160
7.23	Interferer power under QZF (Rayleigh channel, Part 1)	161
7.24	Interferer power under QZF (Rayleigh channel, Part 2)	162
7.25	SIR under QZF (LOS channel/uniform linear array, Part 1)	163
7.26	SIR under QZF (LOS channel/uniform linear array, Part 2)	164
7.27	Desired signal under QZF (LOS channel/uniform linear array, Part 1)	165
7.28	Desired signal under QZF (LOS channel/uniform linear array, Part 2)	166
7.29	Interferer power under QZF (LOS channel/uniform linear array, Part 1)	167
7.30	Interferer power under QZF (LOS channel/uniform linear array, Part 2)	168
7.31	SIR under QZF (LOS channel/uniform square array, Part 1)	169
7.32	SIR under QZF (LOS channel/uniform square array, Part 2)	170
7.33	Desired signal under QZF (LOS channel/uniform square array, Part 1)	171
7.34	Desired signal under QZF (LOS channel/uniform square array, Part 2)	172
7.35	Interferer power under QZF (LOS channel/square linear array, Part 1)	173
7.36	Interferer power under QZF (LOS channel/square linear array, Part 2)	174
7.37	Average SIR under the GCD algorithm (Rayleigh, Part 1)	177
7.38	Average SIR under the GCD algorithm (Rayleigh, Part 2)	178
7.39	Average desired signal under the GCD algorithm (Rayleigh, Part 1)	179
7.40	Average desired signal under the GCD algorithm (Rayleigh, Part 2)	180
7.41	Average interferer power under the GCD algorithm (Rayleigh, Part 1)	181
7.42	Average interferer power under the GCD algorithm (Rayleigh, Part 2)	182
7.43	Average SIR under the iGCD algorithm (Rayleigh, Part 1)	184
7.44	Average SIR under the iGCD algorithm (Rayleigh, Part 2)	185
7.45	Average desired signal under the iGCD algorithm (Rayleigh, Part 1)	186
7.46	Average desired signal under the iGCD algorithm (Rayleigh, Part 2)	187
7.47	Average interferer power under the iGCD algorithm (Rayleigh, Part 1)	188
7.48	Average interferer power under the iGCD algorithm (Rayleigh, Part 2)	189
7.49	Average SIR under the LSB-GCD algorithm (Rayleigh, Part 1)	191

7.50	Average SIR under the LSB-GCD algorithm (Rayleigh, Part 2)	192
7.51	Average desired signal under the LSB-GCD algorithm (Rayleigh, Part 1) . .	193
7.52	Average desired signal under the LSB-GCD algorithm (Rayleigh, Part 2) . .	194
7.53	Average interferer power under the LSB-GCD algorithm (Rayleigh, Part 1) .	195
7.54	Average interferer power under the LSB-GCD algorithm (Rayleigh, Part 2) .	196
7.55	Comparison of vector quantization algorithms	197
7.56	iGCD under different channel models	198
7.57	GCD performance with different initializations	199
7.58	SIR under exhaustive search	200
7.59	SIR under GCD for switched antenna arrays	202
7.60	Interference power under GCD for switched antenna arrays	203
7.61	Desired signal power under GCD algorithm for switched antenna arrays . . .	204
7.62	Beamforming gain (GCD versus scalar quantization)	206
7.63	Beamforming gain (GCD versus exhaustive search)	207
7.64	Beamforming gain under the GCD algorithm with different initializations . .	208

List of Tables

4.1	Simulated azimuth beamwidths of broadside and endfire linear arrays	69
4.2	Simulated array gain versus the spacing between antennas	75

Acknowledgments

“Dear Rich,

It has been over four years since we last spoke. I still remember that conversation as if it had happened yesterday. There was a lot more I wanted to say, but did not realize that this conversation was going to be our last. So I will try to do my best to express my thoughts in a few words.

The last seven years are by far the most special. Meeting you and working with you was THE highlight of it all. I learned a lot from you in a very short period of time. Most importantly, you taught me how to think big, how science and engineering can be a great instrument for changing the world to the better, and how I can do it all while having the time of my life. Thank you for giving me this opportunity. Thank you for your patience and confidence in me. It helped a lot every time I faced a new challenge knowing that you believed I can take it on.

I cannot express enough my feelings of gratitude. However, there is another reason why I am writing this to you. I have no doubt that you remember (you never forget) that when we last talked, I asked you “Rich, what can I do?” Your answer was “You can finish your dissertation!”. That was your last request. Well here it is. It took a little longer than you would have liked, but we finally made it, and I say “we” because your advice and guidance were key in making it this far.

Although Rich I feel relieved that this is over, the real journey has only begun. It’s a journey to give back, a journey to make a difference in the lives of others like you made a difference in mine, a journey that I had never thought I would embark on without you, but life is a full of surprises. I have seen you accomplish a lot in so little time. I would have been more than content with a fraction of that, but now I know that you expect nothing less.

Rest in peace my friend, and Go Bears!”

Well, it was a great journey, probably the most exciting and stimulating of my life. I am very fortunate that I did not have to travel alone. By the grace of God, I was always blessed with great companions throughout the journey, and would like to express my gratitude to each and every one of them.

I cannot imagine making it this far in life without the unconditional love and support I received from my parents. Every morning I wake up, I feel very confident knowing I could always count on their prayers and knowing that they are always behind me. They always put my happiness and well being before everything. From the moment I opened my eyes, their lives revolved around mine (that may have changed a little bit after their first grandchild was born). Most importantly, they gave me a great brother and two great sisters. Together, we make a great team.

I was very fortunate to have the Blue & Gold blood run in my family before coming to Berkeley. When you have relatives like Khalid Alireza and his son Raaid, it’s only a matter of time before you start drinking the Berkeley Cool-Aid, and I am so glad I did. Lucky for me, they’re not the kind of people who would convince you to do something and leave you on your own. They were there for me from the beginning till the end. For that I will be ever grateful. I also won’t forget the memorable summers I spent with my uncles Hisham and Yousuf Alireza.

UC Berkeley was a great research and learning environment. Being in that environment for so many years gave me the opportunity to interact and work with many great minds, both faculty and students and staff. It would be very difficult for me to single out every person who has had a positive impact on my experience in all these years (that would be a dissertation topic of its own). Therefore, I apologize to those whom I forgot to mention.

I would like to first thank my dissertation advisors, Professors Ali Niknejad and Eric Brewer and Paul Wright, for their strong support especially following Rich Newton's passing. The moral support they provided during that period meant a lot more to me than anything else.

With my research spanning multiple areas, I was very fortunate to be a member of several multidisciplinary research groups. Most importantly, the Technology Infrastructure for Emerging Regions (TIER) group founded by Professor Eric Brewer, and the Berkeley Wireless Research Center (BWRC) started by Professors Bob Brodersen and Jan Rabaey. Without exception, every member in these groups was always generous with their assistance whenever I needed them. The discussions I had with folks like RJ Honicky, Rabin Patra, Sergiu Nedeveschi, Sonesh Surana, Michael Rosenblum, and Melissa Ho (TIER), and Sayf Alalusi, Ben Wild, Amin Arbabian, Ehsan Adabi, and Vinayak Nagpal from the BWRC helped enrich my PhD years. Folks like Wei-hung Chen, Debo Chowdhury, Michael Mark and David Sobel were happy and patient to share their research ideas and technical expertise.

RJ Honicky deserves a special mention since we worked closely together in the earlier years of my PhD. He was very helpful in many class projects and with my qualifying exam preparation. I will never forget our trip to Ghana, which was made possible by the efforts of Dr. Kristi Raube and Professor Anrew Issacs (the entertainment during that trip was courtesy of Samir Mehta and Ronnie Chatterji).

One of my lucky coincidences in the past several years was meeting Kevin Jones. Prior to meeting Kevin, I was looking for someone to collaborate with me on the project by building the array hardware. I could not have found anyone better than Kevin, whose hard work and dedication helped me focus on other aspects in the project and finish my dissertation in time. In the last couple of years, I received extra support from him and Rabin Patra and Sergiu Nedeveschi in building and testing the equipment we built.

Another luxury students enjoy at UC Berkeley is the ability to walk into the office of any professor during their office hours. I took full advantage of this and paid several professors in the department multiple visits. These visits helped shape and refine some of the research ideas I developed during these years. If I remember correctly, the professors that I bugged most were Kannan Ramchandran, David Tse, and Ken Gustafson. I thank them for their patience and the insights they provided.

I also recently co-authored several papers with Mark Johnson, Raghu Mudumbai, and Professor Upamanyu Madhow of Santa Barbara. It was a great pleasure working with them. I would also like to thank Raghu for reviewing my dissertation and providing feedback.

I cannot forget the support I received from members of the staff at UC Berkeley, especially Tom Boot, Brian Richards, Kevin Zimmerman (BWRC), Ruth Gjerde, La Shana Porlaris (EECS), and Bill Oman (College of Engineering). I have never seen a staff member so dedicated to a lab like Tom Boot was to the BWRC.

Living near Silicon Valley for all these years, I could not resist the urge of starting a company. So a little over two years ago, along with Kevin Jones, Rabin Patra, and Sergiu

Nedevschi, I started Tarana Wireless, to take the ideas we developed at Berkeley and venture into the real world. We were later joined by Dale Branlund, who brought a vast amount of industrial knowledge and experience to the team. It has been a great adventure and learning experience working with these folks. However, starting a company while writing a dissertation was no piece of cake (take it from me) , but I was very fortunate to be working with these guys who did most of the heavy lifting and took a huge burden off my shoulder, which allowed me to focus on finishing my research. For that, I am ever grateful.

Throughout my life, I have been to many places, and in most cases, I was not only blessed with new friends, but with new family as well. Berkeley was no exception.

My old friend from the MIT days, and now my brother Lik Mui was there for me the moment I stepped into SFO Airport. He waited patiently for 5 hours as I got through Immigration and Customs. He and his parents would never let go until they made sure I was fully settled and my apartment was fully furnished. Their home was my home, and they treated me as one of their own. You can never meet a better person or a better family. I was so honored that he chose me to be one of the groomsmen at his wedding.

Judy Olson welcomed a complete stranger to her home and treated him like her own son. He was always welcome to drop in any time he desired, and she and her husband Mamade would cook for him and take care of him.

One of the many beautiful outcomes of my friendship with Rich was meeting Steve Beck and Candice Eggers. These two are among the kindest individuals I have met. Steve is a very unique individual that approaches science with an artistic mind, and art with a scientific mind. Over the past four years, I have enjoyed I lot of stimulating discussions with Steve. We would discuss a variety of topics, including art, science, religion and world affairs. I wish him all the best with his beautiful NOOR project, and of course I will not forget the special BBQ chicken dinners I had with him and his beautiful wife Candice.

Throughout my stay in Berkeley, I did not develop a friendship that is closer than the one I developed with Mohamed Muqtar (or the mayor of Berkeley as some like to call him). He took care of me when I was still a stranger in Berkeley. Almost half the people I know in Berkeley, I got to know through him. He got me interested in Cal athletics, and thanks to him I now know half of the athletic department. I have always appreciated his support and candid advice.

Of course I cannot forget my old friends Mamdouh Salama, Bob Randolph, and Tom Greene, who continue to keep an eye on me from very far away.

Also, during most of my stay at Berkeley, I almost made my home at Raza's Kitchen and EZ Deli, where I dined almost every day, and where the staff always gave me a special treatment.

Finally, I would like to acknowledge King Abdullah University for Science and Technology (KAUST) and the Saudi Ministry of Higher Education for funding my research in the last 7 years.

Chapter 1

Introduction

1.1 The wireless revolution

In the past decades, we have seen a constant penetration of wireless technology in everyday life. Today, wireless is replacing wires as the dominant medium for communication, and is rapidly changing the way people interact and do business. Earlier this decade, the number of landline phone subscriptions started to decline rapidly. Instead, people are relying more heavily on cell phones and the Internet for voice calls. With the rapid proliferation of wireless local area network (WLAN) technology (e.g. 802.11/WiFi), people can access the Internet anywhere at home and in the office without necessarily needing to be close to a wired connection. In fact, even the connectivity between computers and devices (e.g. printers, scanners, mice, keyboards, digital cameras) in close proximity is increasingly becoming wireless through technologies like WiFi, Bluetooth and ultrawideband (UWB). These networks are sometimes referred to as wireless personal area networks or WPANs. More recently, with the introduction of the third and fourth generation cellular networks that have been designed and optimized for fast data delivery, many people have started replacing their wired Internet services (e.g. cable and DSL) with wireless [48]. Therefore, it is becoming increasingly common for people to maintain full Internet connectivity most of the time, continuously moving from one network to another to maintain coverage¹.

In addition to traditional web surfing, the main applications that are driving innovations (e.g. latency, bandwidth and coverage) in wireless technology today include gaming, voice of IP or VoIP, and video. These applications usually have very tight latency requirements or high bandwidth requirements or both.

1.2 Rural connectivity and the digital divide

Despite the proliferation of wireless technologies in the last several decades, its impact has been mostly limited to urban and suburban areas especially in developed regions. Rural

¹For example, at home or in the office, the Internet connection will usually be through a WLAN, then probably switch to the cellular network when the device moves out of the range of the home network. In rare situations when the cellular coverage is spotty, the device might switch to a satellite connection, and thus maintaining connectivity all the time.

areas on the other hand are still left behind in the Internet revolution and digital divide is widening [35]. Solving the rural problem is essential for the geopolitical future of any country because if it persists, pressure will continue to mount on urban areas. The main challenge in rural areas is a combination of low population densities and limited purchasing power which often make wired infrastructure solutions economically unsustainable. Most these areas are usually scattered and separated by large empty spaces, and any wired infrastructure will have to be maintained over these empty areas. Wireless on the other hand, can skip over these empty areas and only serve the area(s) of interest, which dramatically changes the economics of these areas. This has been recently demonstrated with long distance point-to-point and point-to-multipoint links that connect rural areas with nearby urban centers [44, 39, 55]. Figure 1.1 shows a fixed wireless point-to-point and point-to-multipoint network connecting rural villages to urban centers with long distance links.

This dissertation presents a study of one of the key components in enabling wireless systems to meet future capacity demands and expanding their reach beyond the traditional urban/suburban domain, namely the design of the antenna subsystem. We propose a solution based on adaptive beamforming technology. We present an architecture that combines both RF and digital techniques to achieve high gain and capacity at much lower the cost and complexity compared with existing techniques. In this architecture we consider the challenges in building the antennas and radio circuits as well as implementing the adaptive signal processing algorithms.

In the rest of the chapter, we discuss the main challenges for deploying present and future wireless networks. We also discuss the limitations of existing techniques used in today's networks which rely on static antennas. We then propose adaptive beamforming systems as a potential solution for addressing those challenges. Finally, we conclude the chapter with main goals and contributions of this research and the organization of the rest of the dissertation.

1.3 Challenges for growing and expanding the reach of wireless networks

In spite of recent advances in wireless technology, today's wireless networks face many challenges in both urban and rural areas. In urban areas, the growing demand for data traffic is putting a lot of strain on existing infrastructure. This growth is spurred by the increased use of small data hungry devices such as smart phones and netbooks. Both the number of devices as well as the demands from each individual device are growing. Figure 1.2 shows the expected data traffic increasing by almost a factor of 40 in the next 5 years. With spectrum being a finite and scarce resource, the options for coping with this growth in data demand are limited to increasing the network density through better spatial reuse or moving to less crowded frequency bands with usually poor propagation characteristics (e.g. mm-wave frequencies). Therefore, a technology that enables more efficient spatial reuse and overcomes poor propagation at high frequencies would takes us a long way in solving the capacity problem.

In rural areas, the main challenges for deploying wireless networks are the distance,

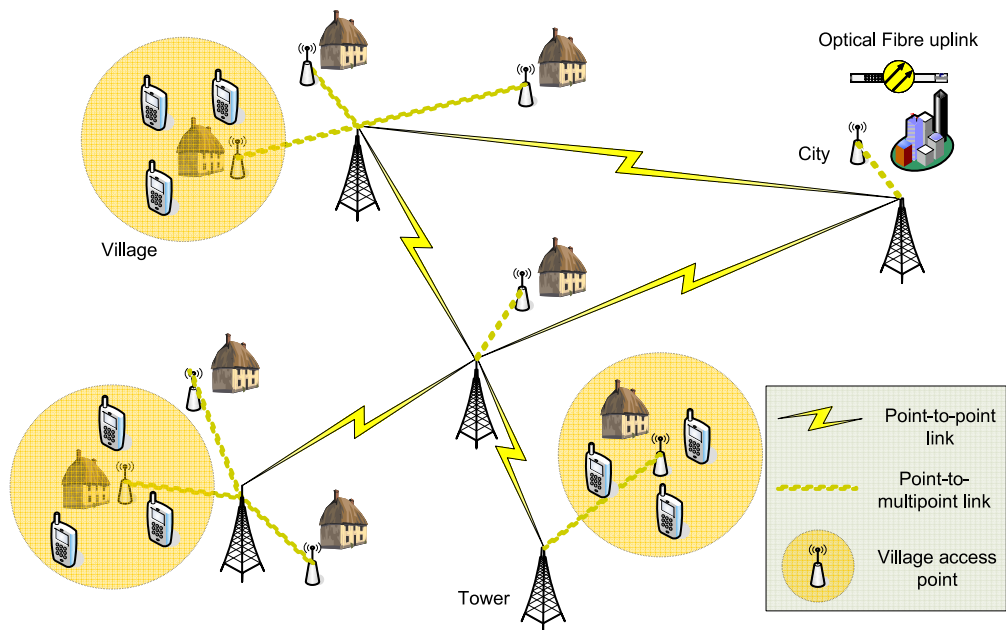


Figure 1.1: Long distance wireless point-to-point and point-to-multipoint links for connecting rural villages to urban centers.

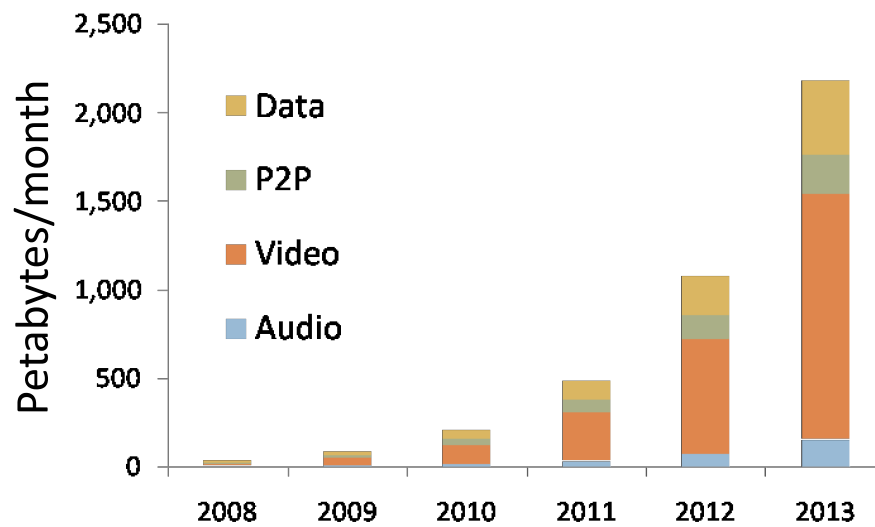


Figure 1.2: Projected mobile traffic. Source: CISCO VNI Project [2].

shortage of skilled labor, and shortage of reliable energy sources and frequent power outages. The long distance makes it very hard to achieve a line of sight (LOS) without big towers. Even when LOS is available (e.g. a building or a hill), very high gain antennas are required to cover the distance, which can be as high as 50-100 miles. These high gain antennas have very narrow beams that are difficult to align at both ends of the link at long distances [44]². To make things worse, the alignment has to take place near the top of the tower. Therefore, these tower have to be very stable and climbable, which makes them significantly more expensive. In fact, the towers usually represent the bulk of the capital cost of these networks [35]. Also, the antenna alignment is not a one-time process, but needs to be repeated when the antennas get misaligned, which can happen for a variety of reasons including wind, temperature and air density variations. The shortage of skilled labor in these areas makes it harder to maintain the network without frequent truck-rolls. These truck-rolls are slow and increase the downtime and make up the bulk of the operating cost. The lack of reliable energy sources and access to the power grid means that unless the network equipment can operate off of portable energy sources (e.g. solar cells) and backup batteries, the network will experience a lot of downtime. Finally, the current architecture of these networks makes it difficult to expand the network by adding new and redundant links to increase the reliability and reduce the downtime. Therefore, to address these problems, we need a technology that is power efficient and can self-align and thus eliminate the need for heavy and expensive towers and significantly reduce the required maintenance.

1.4 Wireless networks today

Traditionally, most wireless systems, both fixed and mobile, have been deployed using either omni-directional antennas or fixed-beam directional antennas. Such architecture presents several challenges to achieving high capacity and good coverage at low cost. Omni-directional antennas transmit and receive equally in every direction, which results in very good coverage. However, this coverage comes at the expense of the range and capacity. Since the antenna radiates equally in all directions, a lot of the energy gets wasted in directions that do not contain any desired receivers. This will significantly degrade the range of the system. Furthermore, since the energy propagates equally in all directions, it might cause unnecessary interference to neighboring networks and thus reduce the achievable capacity of the overall system. The same thing happens on the receive side where the antenna will have a difficult time distinguishing desired signals from interferers since it receives equally from all directions.

Directional antennas on the other hand, have very narrow and focused beams. Therefore, they have a much better range and are more energy efficient the omni-directional antennas. The improved energy efficiency comes at the expense of a larger aperture and narrow coverage area, thus, rendering these antennas unsuitable for mobile and dynamic systems. Furthermore, even for fixed systems, directional antennas require very accurate alignment.

There have been several approaches to address the capacity/coverage problem. Cellular networks have taken a two step approach. In the first step, each cell, which used to be

²Studies have shown that antennas typically need to be aligned within 30% of their half-power-beamwidth (HPBW) in order to be effectively used [62].

covered by a single base station with omni-directional antennas is sectorized (typically 3 sectors are used as shown in Figure 1.4a)³. This process effectively scales the capacity of the system by the number of sectors since each sector now handles only a small fraction of the users handled by the original base station. However, this increase in capacity is nowhere near what is required to meet future data demands⁴. In the second step towards achieving the target capacity, cells are split into smaller cells as shown in Figure 1.4b. Each of the smaller cells (sometimes referred to as micro or pico cells) have the same capacity as the larger cells (also known as macro cells). The resulting capacity increase is proportional to the increase in the total number of cells (base stations). However, this migration is still very expensive. Although, micro and pico are much cheaper than macro cells since they have smaller footprint and much lower transmit powers, the overhead of adding new cells is very high, even higher than the cells themselves⁵. Therefore, a solution that increases capacity with the minimal number of cells can significantly reduce cost⁶.

Signal propagation in dense urban environments also presents a challenge to antennas with static patterns, both directional and omni-directional. In these environments, both shadowing and fading are difficult to overcome and reduce the quality of the signal.

1.5 Adaptive beamforming systems

It is clear from the discussion above that the solution to the capacity and range challenges of today and tomorrow's networks would have to combine the key features of omni-directional (wide coverage) and directional systems (high range, energy and spectral efficiency through spatial reuse). A solution that satisfies these requirements can be implemented with adaptive or dynamic beamforming (beam-steering). Beamforming refers to the technique that, unlike static antennas, can dynamically shape the beam pattern and focus the beam at wide range of directions. Shaping and reshaping the beam can be accomplished without any moving parts using an array of antennas. With this capability, beamforming can achieve the required coverage area since the beam can be steered in any direction. At the same time, it can extend the range of the system since the beam is focused at a very narrow angle at any given time (Figure 1.3 compares the beam pattern of a beam-steering system with directional and omni-directional antennas). Using these features, beamforming technology can address the rural and urban deployments as follows. For rural networks, since the beams can be electrically steered in any direction without any moving parts, no antennas need to be manually aligned and no towers need to be climbed. Therefore, both the capital and operating costs are significantly reduced. Also, since the beamforming antennas can cover wide angles, the network can now easily support point-to-multipoint and mesh links and achieve better redun-

³Sectorization is the process of dividing a cell into multiple sectors with narrower coverage. Each sector is effectively a separate base station and uses a directional antenna. Collectively, the sectors achieve full coverage.

⁴In practice, different neighboring sectors still interfere with one another, which degrades the capacity. Also, a sectorized cell has a much higher footprint than a non-sectorized cell, which increases the overall deployment costs.

⁵These costs include site rental and backhaul costs.

⁶Another disadvantage of a static cell layout is that the user density is usually dynamic, and thus, the load will in general not be balanced.

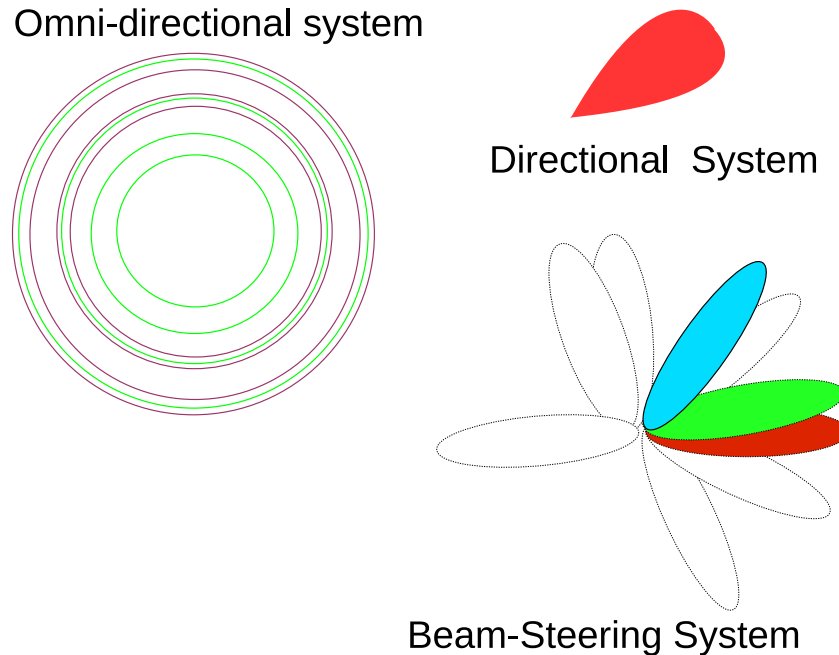
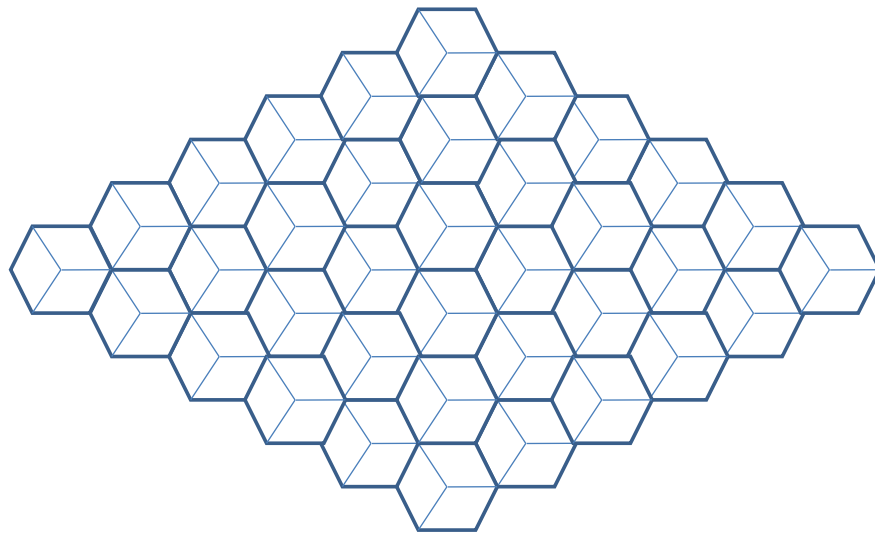


Figure 1.3: Beam patterns for the three different antenna types: omni-directional, directional, and beamforming (beam-steering).

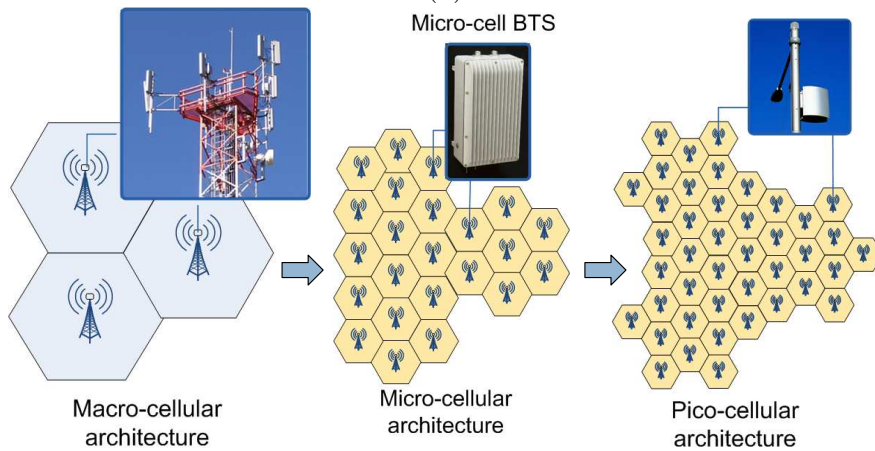
dancy. For urban networks, beamforming can increase the capacity in different ways. First, nodes equipped with beamforming capability can communicate by pointing beams towards one another instead of simply broadcasting. That means that interference with nodes in the network is reduced, and another pair of nodes in the network can also be communicating at the same time without causing interference, and thus, the capacity of the overall network increases by increasing the number of pairs that can communicate simultaneously in a given geographic area⁷. Second, in addition to pointing the beam, adaptive beamforming systems have the capability to shape the radiation pattern such that radiation is minimized in the direction of interferers (also known as null-steering and beam-nulling) as shown in Figure 1.5. While studies have shown that simple beam pointing can significantly increase the system capacity [30], directing beam nulls or zeros in the direction(s) of interferers can increase the capacity even further. The ability to null-steer enables beamforming systems equipped with multiple baseband radios to form multiple simultaneous beams (each beam carrying an independent data stream) to multiple users as long as each beam has nulls in the direction of other beams as demonstrated in Figure 1.5. This technique for increasing capacity is called spatial multiplexing⁸. Using this approach, a cellular base station can increase its capacity by the number of simultaneous beams. That means a single base station can achieve the capacity of multiple base stations, and thus reducing the network cost and overhead

⁷This technique is sometimes referred to as spatial reuse.

⁸The multiple simultaneous beams can also be directed at the same radio provided that it is also equipped with a multi-baseband beamforming system. This technique is called multi-input/multi-output or MIMO and usually requires an environment with a lot of reflections and scattering [59].



(a)



(b)

Figure 1.4: (a) Improving system capacity with sectoring (3 sectors per cell)(b) Improving system capacity by increasing network density using smaller cells.

by reducing the total number of base stations. Compared with a sectorized base station, a multi-beam beamforming system can do a better job at reducing the cross-talk between different beams and thus ensuring more reliable performance. Furthermore, a beamforming system has a much smaller footprint than a sectorized system since all the beams share the same antenna frontend.

Another advantage of beamforming is in applications and systems that operate in high frequency bands (e.g. mm-wave). Since most of the frequency spectrum that is suitable for dense urban cellular communication (e.g. below 5GHz) is mostly licensed. The only opportunity to expand the data rate in the frequency domain is to leverage the unused frequency bands near the mm-wave range (e.g. 60GHz and above). The main advantage of these frequency bands is the high bandwidth availability. However, the propagation characteristics of these bands is very poor even for short distances⁹. To overcome these conditions, it is necessary to use highly directive antennas. Fortunately, high antenna gain can be achieved at a considerably smaller antenna size due to the high carrier frequency. That means these directional antennas can be used with mobile units as well. However, a fixed narrow beam system is not suitable for mobile applications as mentioned above. That makes beamforming the only viable solution for these applications.

Beamforming antennas also perform much better than static beam antennas in presence of heavy shadowing and fading. When the line of sight is obstructed, the beamforming system can explore alternative paths through reflections and focus the radiation in those directions. Also, in the presence of fading, since the beamforming system has a wide variety of beam patterns to choose from, the probability of being in deep fade is reduced significantly¹⁰.

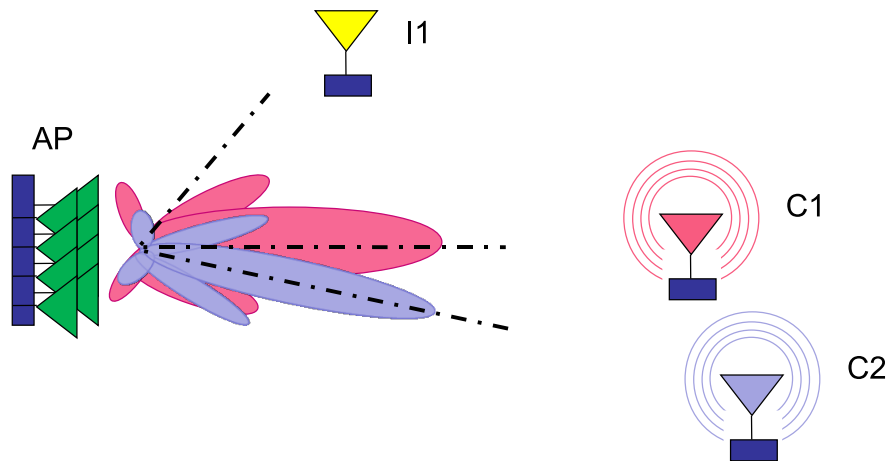


Figure 1.5: An adaptive antenna system with multiple basebands forming multiple simultaneous (spatially orthogonal) beams and canceling interference from other sources by steering nulls in their directions.

⁹Several factors contribute to poor propagation in addition to the free space path loss. These include oxygen absorption and low antenna efficiency [4]. These will be discussed in more detail in Chapter 2.

¹⁰This property is also known as antenna or spatial diversity [59].

1.6 High gain beamforming systems

As pointed out in Section 1.3, the data demand in dense urban areas is expected to increase by a factor of 40 over the next 5 years (Figure 1.2). In order to keep up with this demand, the spatial reuse or capacity density of the networks will have to increase by similar factor, which means that the size and complexity of beamforming systems must increase by a similar amount. Similarly, for rural networks, beamforming systems have to provide the same gain as the directional antennas they replace (usually of the order of 20-30dBi). Beamforming systems delivering these orders of magnitude of gain are very expensive to build and have traditionally been reserved for high budget military and defense applications, and unfortunately, beamforming systems that are used commercially today in WiFi and cellular networks do not scale very well to meet future demands. Developing techniques for building inexpensive high gain beamforming systems is the focus of this dissertation.

1.7 Research goals and contributions

The goal of this dissertation is to investigate the challenges for building beamforming systems with high gain (20-30dB) and wide scanning range ($180 - 360^\circ$). Such a study requires a close examination of the system at different levels starting from the antenna and circuit design levels, and proceeding all the way to the signal processing level. There are several key observations. First, design choices made at one level usually have a lot implications on other levels as well. For example, the choice of adaptive signal processing algorithms is highly dependent on the design of the beamforming circuit. Therefore, these design choices are not separable, but must be jointly made. Second, the optimal design needs to combine both analog and digital techniques. The scope of this research covers the interdependence between the different components of the system (e.g. antenna, circuit, signal processing) as well as the optimal partitioning of functionality between these components. In particular, the research contributions include:

- 1) A hybrid RF/digital radio architecture for beamforming, including the design of a multi-input/multi-output beamforming module in RF, to combine the scalability and power efficiency of RF beamforming with flexibility and speed of digital beamforming.
- 2) 3-dimensional antenna array structures with wide scanning range to reduce the size of the antenna and improve beam-shaping.
- 3) Signal processing techniques based on adaptive filtering for dynamically estimating and optimizing both RF and digital beamforming weights. The proposed framework provides resiliency against noise, interference, synchronization errors and channel variation.
- 4) A comprehensive analysis and simulation of the impact of errors in the beamforming weights on the performance of adaptive array systems.
- 5) Vector quantization techniques for improving interference suppression capabilities of large arrays in the presence of quantization errors.

1.8 Organization of the Dissertation

The rest of the dissertation is organized as follows. Chapter 2 presents an overview of beamforming and adaptive antenna technology and reviews some wireless communication concepts related to the topic. A survey of some of the prior work in this area is also presented. Chapter 3 discusses the implications of the radio architecture on various aspects related to system cost and performance including power consumption and gain and capacity. A hybrid architecture that combines RF and digital techniques is proposed that can meet the system gain and capacity requirements with the lowest cost and power requirements. The implementation of the RF frontend of the system that contains the beamforming functionality is also discussed in detail. Chapter 4 discusses the antenna design constraints, and presents design techniques for building compact antenna structures that meet the gain and coverage requirements of the system. In Chapter 5, we address the signal processing component of the system that makes it adaptive, and present computationally efficient techniques for optimizing the RF and digital beamforming weights of the system that are resilient to noise, interference, and channel variations. Chapter 6 discusses the errors in a beamforming system and analyzes their impact on the overall performance. Chapter 7 presents techniques for mitigating quantization errors. Finally, the dissertation concludes with a summary of the key results and a discussion of future research in Chapter 8.

Chapter 2

Background

The concept of electronic beam-steering or beamforming has been used since World War II for a variety of applications that extend beyond communications, both military and civilian. These applications include radar, radio astronomy, oil exploration, and medical imaging. This chapter provides an overview of beamforming techniques and covers some basic concepts related to beamforming. It starts with a brief overview of antenna technology and theory, then introduces the concept of electrical beam-steering and describes some of the common implementation techniques. The chapter concludes with a discussion on modeling the wireless channel. Although most of the concepts described in this chapter are very general, there is a little more emphasis on terrestrial wireless communication since it is the focus of the dissertation.

2.1 Antenna basics

The antenna is the component of the radio system that converts a guided wave on a transmission line to an electromagnetic wave propagating in an unbounded medium (usually air or free space) and vice versa in the receive case¹. An antenna is usually characterized by its radiation patterns, directivity, radiation efficiency, bandwidth, and polarization. These properties depend on the size and shape and the material of the antenna, as well as its surrounding. Since antennas are passive elements, they exhibit the same behavior in both transmit and receive modes (also known as reciprocity). In this section, we provide an overview of each these basic antenna properties².

2.1.1 Antenna radiation patterns

The radiation pattern of an antenna describes the relative field intensity (or power density) at different points on a sphere in the far-field. The radius of the sphere is irrelevant

¹A more comprehensive discussion and analysis of antennas and antenna related topics can be found in [12] and [54] and [60]

²In this chapter and the rest of the dissertation, we shall focus on the far-field properties of the antenna. Far-field refers to points in space at a distance from the antenna that is much larger than the dimensions of the antenna.

since we are only concerned with the relative values (usually relative to the point (direction) of maximum intensity). The radiation pattern is usually described in terms of the *normalized radiation intensity* $F(\theta, \phi)$, which is defined as the ratio of power density in the direction (θ, ϕ) to the maximum power density³. Therefore, $F(\theta, \phi)$ is dimensionless and has a maximum value 1 or 0dB [60]. The radiation pattern is usually plotted in polar (2D) or spherical (3D) coordinates in both linear and dB scales. In polar plots, either θ or ϕ is constant. Rectangular plots are also used, but they are less common. Figures 2.1/2.2 show examples of these plots, both polar and rectangular.

There are many types of radiation patterns, three of them are of practical interest. The first is the isotropic beam pattern. An isotropic pattern has equal power density at all points on the sphere (i.e. the 3D plot is a perfect sphere and the polar plots are perfect circles). Therefore, the antenna pattern is unbiased. Although an antenna with isotropic radiation is not realizable in practice⁴, it is used as the standard or yard-stick against which all other antennas are measured as described in Section 2.1.2. The second basic pattern of interest is the omni-directional pattern, which is the closest that can be achieved in practice to an isotropic pattern. An omni-directional pattern radiates equally in a plane and is directional in the planes perpendicular to it. For example, a half-wave dipole (shown in Figure 4.9) radiates equally in the plane that perpendicular to its axis. The 3D radiation pattern of the dipole is donut shaped (Figure 2.1)⁵. Finally, there is directional pattern which is usually biased towards certain direction(s) in both the azimuth and elevation planes. Plots of the three basic patterns are shown in Figure 2.3.

2.1.2 Directivity

Directivity D of an antenna is defined as the inverse of the average normalized radiation intensity F_{av} [12]:

$$D = \frac{1}{F_{av}}, \quad F_{av} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi F(\theta, \phi) \sin \theta d\theta d\phi \quad (2.1)$$

The average normalized radiation intensity is the integral of $F(\theta, \phi)$ over a unit sphere. The directivity is a measure of how “focused” the beam is. Antennas with low directivity will spread the radiated energy in many directions (a wide angle), whereas antennas with high directivities focus the beam in very narrow angles. By definition, the directivity cannot be less than 1. The isotropic antenna has the minimum directivity $D_{iso} = 1 = 0\text{dBi}$. The directivity of any other antenna is specified relative to the isotropic antenna in the dimensionless unit dBi⁶. The directivity of an antenna is proportional to its effective area

³ θ is the angle with respect to the positive z-axis ($0 < \theta < \pi$), and ϕ is the angle on the xy plane with respect to the positive x-axis (counter clockwise). Sometimes, ϕ is referred to as the azimuth (horizontal) angle, and θ as the elevation angle. Together, they specify a unique direction in the 3-dimensional space.

⁴The only antenna that can achieve an isotropic pattern is a point source, which is not possible in practice since a real antenna must have physical dimensions. In fact, the smallest realizable antennas like the short dipole or the half-wave dipole are biased towards some directions compared to others [60].

⁵In practice, the radiation pattern of a short dipole will not be perfectly omni-directional since the feed network breaks some of the symmetry.

⁶The directivity is assumed to be in the direction of maximum intensity. However, to compute the directivity in an arbitrary direction (θ_d, ϕ_d) , we simply multiply D by $F(\theta_d, \phi_d)$.

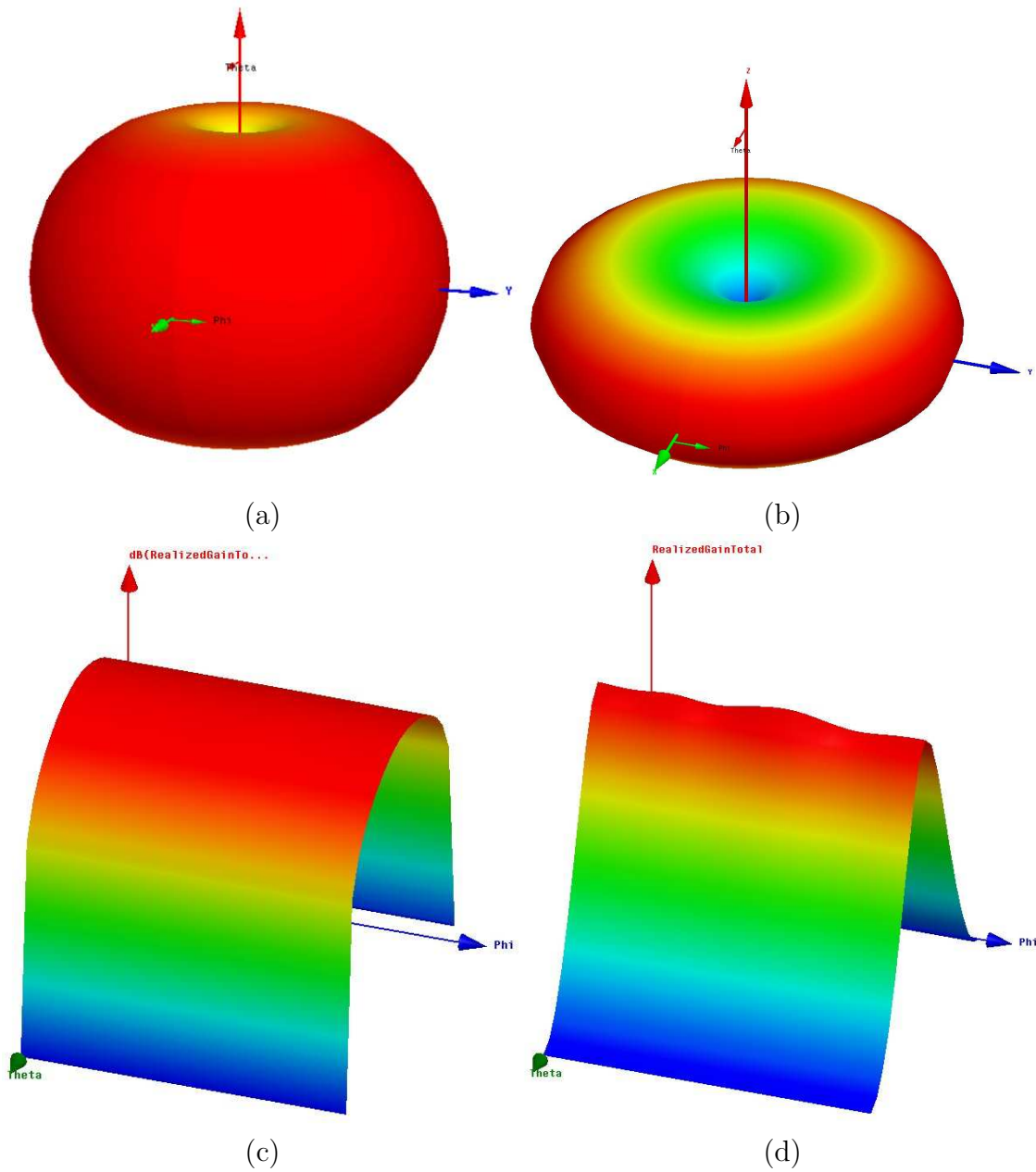


Figure 2.1: Different representations of an antenna pattern in three dimensions: (a) 3D plot in spherical coordinates (dB scale). (b) 3D plot in spherical coordinates (linear scale). (c) 3D plot in rectangular coordinates (dB scale). (d) 3D plot in rectangular coordinates (linear scale).

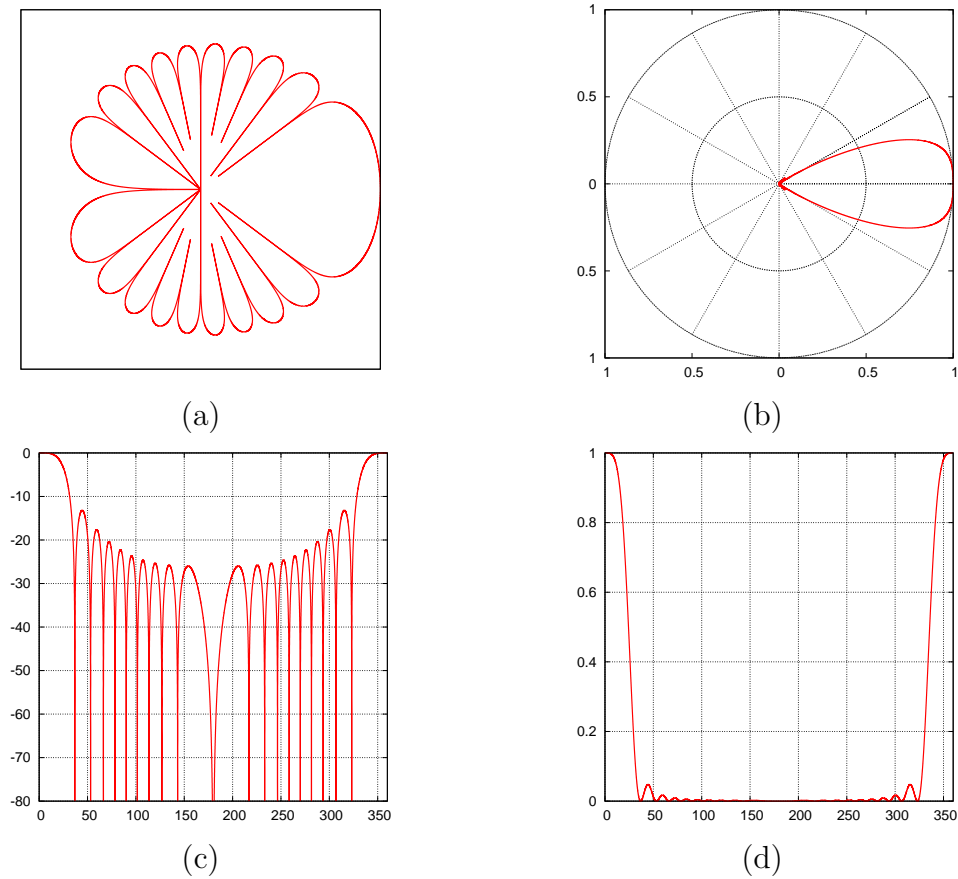


Figure 2.2: Different representations of an antenna pattern in two dimensions: (a) 2D plot in polar coordinates (dB scale) with either θ or ϕ fixed. (b) 2D plot in polar coordinates (linear scale) with either θ or ϕ fixed. (c) 2D plot in Cartesian coordinates (dB scale) with either θ or ϕ fixed. (d) 2D plot in Cartesian coordinates (linear scale) with either θ or ϕ fixed.

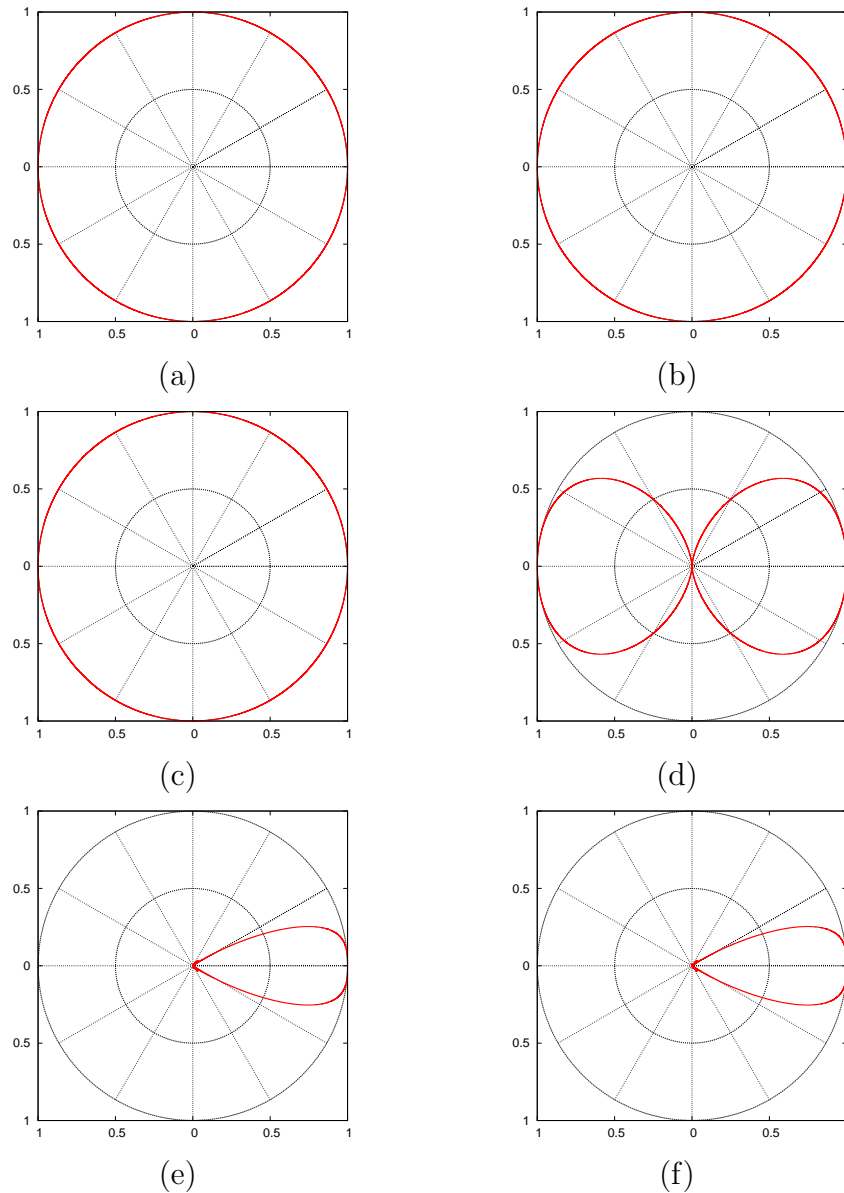


Figure 2.3: Three basic patterns (isotropic, omni-directional, directional) in polar coordinates (linear scale). (a) Isotropic pattern (azimuth). (b) Isotropic pattern (elevation). (c) Omni-directional pattern (azimuth). (d) Omni-directional pattern (elevation). (e) Directional pattern (azimuth). (f) Directional pattern (elevation). Note that the patterns of the omni-directional and directional antennas depend on the antenna orientation. Here, we assume that the omni-directional antenna is oriented such that the radiation is uniform in the xy -plane, and the directional antenna is oriented such that the beam is focused along the x -axis.

[54]. In fact, we can write the relationship between the directivity D and the effective area A_e as [60]:

$$A_e = \frac{\lambda^2 D}{4\pi} \quad (2.2)$$

where λ is the carrier wavelength. Therefore, for a fixed area, the directivity increases with frequency.

Beamwidth

Another measure of the directivity is the beamwidth measured near the point(s) on the sphere of peak radiation intensity. There is no fixed definition of the beamwidth. However, the most popular is the half power beamwidth (HPBW), which is the measure of the angle over which the radiation intensity of the antenna is at least one-half the value of its maximum value. The directivity is inversely proportional to the beamwidth:

$$D \propto \frac{1}{HPBW} \quad (2.3)$$

The constant of proportionality depends on many factors including the size and number of lobes in the pattern. Also, depending on the context, the HPBW may sometimes refer to the solid angle that covers the contiguous area on the unit circle around the peak in which the intensity is at least one-half the value of its maximum value. Solid HPBW is proportional to the product of the azimuth and elevation HPBWs [60]. Other measures of the beamwidth include the null to null beamwidth (B_{nn}) which is a measure of the smallest angle between the peak and the first null (or between two nulls), and the 10dB beamwidth, which measures the angle from the peak up to the point where the intensity drops by more than 10dB. Equation 2.3 exemplifies the tradeoff of high directivity antennas. Whereas, the energy efficiency improves with directivity, the coverage area get smaller as well.

2.1.3 Antenna efficiency

In all practical cases, antennas are not perfect lossless radiators. In fact, not all the power accepted by antenna is actually radiated. The ratio of the radiated power P_{rad} to the power accepted or transmitted P_t is radiation efficiency ξ of the antenna. From a transmission line's perspective, an antenna is merely an impedance. The real part of this impedance can be divided into two components: a *radiation resistance* R_{rad} and a *loss resistance* R_{loss} . ξ can be expressed in terms of R_{rad} and R_{loss} [60]:

$$\xi = \frac{R_{rad}}{R_{rad} + R_{loss}} \quad (2.4)$$

The gain G of the antenna is defined as product of the directivity D and the efficiency ξ . Like the directivity, the gain of an antenna is specified in dBi, which is the ratio of the gain of the antenna to the gain of a lossless isotropic antenna (in dB). Note that the gain of the antenna can be less than 1 even though the directivity is always greater than 1 if the efficiency of the antenna is bad. The efficiency usually drops at very high frequencies (e.g. mm-wave) where the material loss is high.

2.1.4 Antenna bandwidth

From a transmission line's prospective, an antenna is just an impedance. If not well matched, a significant fraction of the transmit/receive power will not be transferred to/from the antenna. The bandwidth of the antenna is defined as the size of the frequency band over which the square magnitude of the reflection coefficient $|\Gamma|^2$ is smaller than a given threshold⁷. Therefore, in order to ensure maximum power radiation, we need antennas with high radiation efficiency and are well matched⁸.

We note, however, that a lot of the antenna properties like the impedance and radiation pattern can change depending on the medium and the surroundings (especially nearby metal objects). An antenna that is measured in isolation will usually exhibit slightly different behavior when used with an entire system.

2.1.5 Polarization

An electromagnetic (EM) wave cannot be fully described by frequency, phase, and amplitude since these are scalar quantities. To complete the picture, we also include the polarization of the wave, which fully captures its vector nature. In general, radiated EM waves propagate in spheres. As these spheres become larger, they can be approximated by plane waves over small observation regions. If we fix one of these planes in space, then the figure that is traced by the electric field vector on this plane with time define the polarization of the wave [54]. In general, this figure will be an ellipse, which means that the polarization can be fully specified by the orientation of the ellipse and the ratio of the major axis to the minor axis (that ratio is sometimes referred to as the *axial ratio* or *AR*). There are, however, some special cases that often occur in practice. When $|AR| = \infty$, the ellipse becomes a line, and results in linear polarization. On the other hand, when $|AR| = 1$, the ellipse becomes a circle, and results in either right-hand circular polarization (RHCP) or left-hand circular polarization (LHCP) depending on the direction of the rotation.

The polarization of an antenna is defined as the polarization of the wave it radiates⁹. For example, dipole and monopole antennas produce linearly polarized waves¹⁰. Helix antennas produce circularly polarized waves (the direction of rotation follows the windings of the helix). Patch antennas are interesting since their polarization can be controlled by changing the feed network, so they can be designed to produce any type of polarization [12].

The polarization of an antenna is important in the context of wireless communication for two reasons. First, antennas only accept (receive) waves (or components of waves) that have the same polarization. For example, if an antenna is linearly polarized with a polarization vector along the x-axis (horizontally polarized), then it will only receive the component(s) of incoming waves that are polarized along the x-axis. If the incoming wave is linearly polarized

⁷The threshold is application dependent, but for many applications, an impedance match of -10dB or better is considered sufficient.

⁸For some applications, a narrow band impedance match is desirable to minimize out of band radiation.

⁹The definition is a bit ambiguous since the radiated waves may have different polarizations in different directions. However, the polarization of the antenna usually refers to the polarization along the main beam [54].

¹⁰If the dipole is oriented vertically, then the waves are vertically polarized, and if it is oriented horizontally, then the waves will be horizontally polarized, and so on.

with the polarization vector pointing at a direction that is 45° from the x-axis, then only half the power will be received. Similarly, if the polarization vector is 90° from the x-axis (vertically polarized), then the entire wave will not be received. Using this property, two wireless systems can operate simultaneously on the same frequency band if they each use antennas with orthogonal polarization vectors, which means that in theory we can double the capacity by doubling the spectral efficiency by using different polarizations. Therefore, antenna polarization adds an additional degree of freedom to the system. A vertically polarized antenna is orthogonal to a horizontally polarized antenna, and RHCP antenna is orthogonal to LHCP antenna. Second, in many practical scenarios, the polarization of the wave will impact its propagation through the atmosphere and the reflection coefficients of different materials. The exact relationship will also be frequency dependent.

In addition to the properties discussed in this section, antenna design is also affected by other factors such as cost and packaging. Some antennas are easier to manufacture than others. For example, dipoles and microstrip antennas are easier to print on circuit boards than helix or horn antennas.

2.1.6 Free space propagation

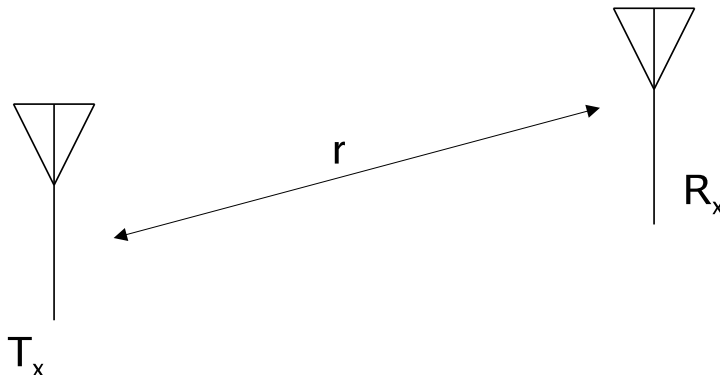


Figure 2.4: A transmit antenna T_x and a receive antenna R_x separated by a distance r meters in free space.

Consider a simple example of a transmit antenna T_x and a receive antenna R_x that are separated by a distance r meters in free space¹¹ as shown in Figure 2.4. If we assume that the main beams of both antennas are along the line connecting both antennas, then P_{R_x} , the power received by R_x , is related to P_{T_x} , the power radiated by T_x , by the Friis equation [60]:

$$\frac{P_{R_x}}{P_{T_x}} = \frac{G_{T_x} G_{R_x} \lambda^2}{(4\pi r)^2} \quad (2.5)$$

The ratio $L = P_{R_x}/P_{T_x}$ is called the path loss (the free space path loss in this case). By definition, $L \leq 1$. G_{T_x} and G_{R_x} are the antenna gains of T_x and R_x respectively, and λ is the carrier wavelength.

¹¹We assume that r is much larger than the dimensions of both antennas.

The inverse relationship between the path loss L and the distance r in Equation 2.5 can be intuitively derived as follows. Since the radiated waves propagate in spheres, at a distance r , the transmit power will be spread over an area that is proportional to a sphere of radius r . Since the receive antenna has a fixed capture area, the fraction of power captured by the receive antenna is inversely proportional to the area of the radius r sphere or $1/r^2$. Equation 2.5 can also be written in terms of antenna directivity:

$$L = \frac{P_{R_x}}{P_{T_x}} = \frac{\xi_{T_x} \xi_{R_x} D_{T_x} D_{R_x} \lambda^2}{(4\pi r)^2} \quad (2.6)$$

The relationship in Equations 2.5 and 2.6 seem to suggest that the free space path loss gets worse as the carrier frequency increases. However, in both equations, we assumed that the directivities of both antennas are fixed. If instead, we fix the effective area of each antenna, and rewrite Equation 2.6 in terms of the effective area:

$$L = \frac{P_{R_x}}{P_{T_x}} = \frac{\xi_{T_x} \xi_{R_x} A_{T_x} A_{R_x}}{(\lambda r)^2} \quad (2.7)$$

then the path loss seems to improve as we increase the frequency. There is no contradiction since we now assume a fixed area instead of a fixed directivity. When the effective area is constant, then we can get more directivity by increasing the frequency¹².

In Equation 2.5, we assumed that the main beams of both T_x and R_x are along the line connecting the two antennas, which is not always the case in practice. In order to account for the misalignment, we multiply the radiation intensities of both T_x and R_x at their respective pointing angles:

$$L = \frac{G_{T_x} G_{R_x} F_{T_x}(\theta_{T_x}, \phi_{T_x}) F_{R_x}(\theta_{R_x}, \phi_{R_x}) \lambda^2}{(4\pi r)^2} \quad (2.8)$$

where $(\theta_{T_x}, \phi_{T_x})$ represents the direction of the receive antenna with respect to the transmit antenna, and $(\theta_{R_x}, \phi_{R_x})$ represents the direction of the transmit antenna with respect to the receive antenna.

Note that the figure of interest is the signal to noise ratio at the receiver or SNR. The signal component of SNR (the numerator) is computed from the Friis equation. The noise component (the denominator) is usually dominated by the thermal noise, which is proportional to the absolute temperature T (usually measured in Kelvin) and the signal bandwidth B (usually measured in Hertz). The expression for the noise power (variance) N_0 at the receiver [47]¹³:

$$N_0 \sim kTB \quad (2.9)$$

¹²In practice, the propagation medium will not be free space, and the loss from medium is usually higher at high frequencies although the relationship is neither linear nor strictly increasing [45]. For example, some bands like 60GHz are very sensitive to oxygen absorption. Thus, the medium loss at 60GHz is going to be larger than 80-100GHz even though the frequency is lower. The loss is also affected by other parameters like polarization.

¹³Noise is a random process that is modeled as a complex white stationary Gaussian process[47]. The constant of proportionality in Equation 2.9 depends on many factors, but most importantly on the receiver noise figure [28]. The constant k in Equation 2.9 is the Boltzmann constant ($k = 1.3806503 \times 10^{-23} m^2 \cdot kg \cdot s^{-2} \cdot K^{-1}$).

$$\Rightarrow \text{SNR} \sim \frac{P_{R_x}}{N_0}$$

The capacity C of a link refers to maximum theoretical data rate that can be achieved over that link, and is bounded by the Shannon limit which is a function of the SNR [47]:

$$C = B \log(1 + \text{SNR}) \quad (2.10)$$

The units for C is bits/sec. C increases linearly with the bandwidth and logarithmically with the SNR¹⁴. The quantity C/B is known as the spectral efficiency and has units of bits/sec/Hz.

Equation 2.10 holds for both wired and wireless channels alike as long as the noise is white, additive and Gaussian. Channels that have this property are called additive white Gaussian noise or AWGN channels. For wireless channels however, there is another factor that needs to be accounted for, which is interference from other sources. In high interference environments, the quantity of interest is the signal to interference plus noise ratio (SINR). Even though the statistics of interference are usually different from noise, they are often assumed to be white and are treated like noise for simplicity¹⁵.

2.2 Static versus steerable antennas

The antennas we have discussed so far in this chapter are passive antennas, whose radiation characteristics are determined solely by their mechanical structure. Examples include helix antennas, reflector antennas, and waveguide antennas. The only way to redirect the beam pattern is by physically moving or rotating the antenna. However, there exists another class of active antennas whose properties depend on their electrical characteristics (e.g. impedance, signal excitations) which can be changed to produce different properties without altering the mechanical structure. This class of antennas is sometimes referred to as electrically steerable antennas or beamforming antennas or adaptive antennas and is the focus of this dissertation.

Electrical beam-steering can be accomplished by means of an antenna array. The simplest example is a uniform N -element linear array shown in Figure 2.5. The array is uniform in the sense that all the antennas have the same radiation pattern (typically omni-directional), and the inter-element spacing is the same throughout the array. Suppose for simplicity that the array elements lie on the x -axis, and the spacing between two consecutive antennas in meters is d . Therefore, the total length of the array is Nd . Furthermore, assume that there is an incoming signal $s(t)$ from an angle θ off the x -axis and a distance r from the first antenna in the array or (Ant₀) as shown in Figure 2.5. We assume for simplicity that $s(t)$ is a very narrow band or single tone signal at frequency f_0 corresponding to wavelength λ_0 (i.e. $s(t) = e^{j2\pi f_0 t} = e^{j2\pi v_s t / \lambda_0}$)¹⁶. We also assume that the source of this signal is in the far field of the antenna array (i.e. $r \gg Nd$), so that the received signal at each antenna in

¹⁴Note that the SNR itself is a function of the signal bandwidth.

¹⁵For wireless channels, there is also another statistical factor that impacts the capacity. This factor is a result of the multipath nature of the channel, and is called fading. Fading is discussed in more detail in Section 2.3.

¹⁶ v_s is the velocity of light in the medium.

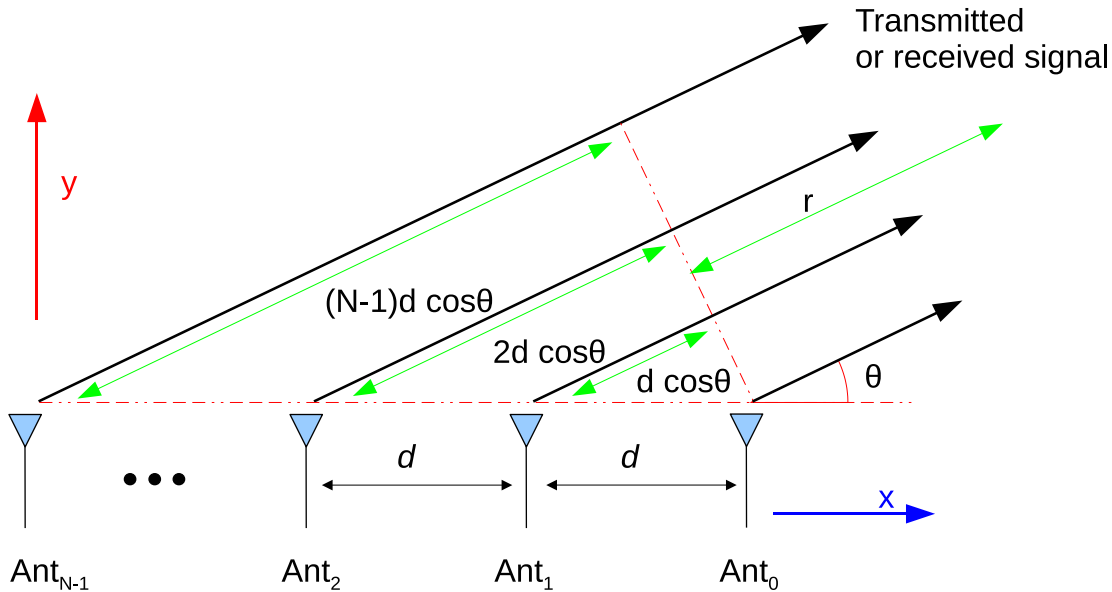


Figure 2.5: N-element linear antenna array.

the array will have roughly the same amplitude, which we normalize to 1. The same signal arrives at different elements of the array with different delays. Those delays can easily be computed from the geometry of the array and the angle of arrival and the speed of light in the medium. Figure 2.5 shows those delays for a uniform linear array as a function of the angle of arrival θ . At a given frequency, these delays translate into phase shifts. If we rotate the phase of the incoming signal at each antenna Ant_i by $i\vartheta \quad \forall_{0 \leq i < N}$, then we can derive the expression of the gain of the array (the aggregated signal from all antennas) as a function of the angle of arrival θ and the progressive phase shift ϑ as follows:

$$\text{relative delay at Antenna } i: \tau_i = \frac{id \cos \theta}{v_s} = i\tau \quad \forall_{0 \leq i < N}$$

$$\text{relative phase shift at Antenna } i: \varphi_i = 2\pi f_0 \tau_i = 2\pi i (\cos \theta) \frac{d}{\lambda_0} = i\varphi \quad \forall_{0 \leq i < N}$$

$$\text{E-field at Antenna } i: E_i(t) = e^{j2\pi f_0(t - \tau_i) + i\vartheta} = e^{j2\pi f_0 t} e^{ji(\vartheta - \varphi)}$$

where $\tau = d \cos(\theta)/v_s$ and $\varphi = 2\pi f_0 \tau$. We also normalized the delays and the phases such that $\tau_0 = \varphi_0 = 0$. From these expressions, we can calculate the E-field at the output of the array:

$$\begin{aligned} E_{array}(t) &= \sum_{i=0}^{N-1} E_i(t) = e^{j2\pi f_0 t} \sum_{i=0}^{N-1} e^{ji(\vartheta - \varphi)} \\ &\quad \Delta \triangleq \vartheta - \varphi \\ \Rightarrow E_{array}(t) &= e^{j2\pi f_0 t} \sum_{i=0}^{N-1} e^{ji\Delta} = e^{j2\pi f_0 t} \frac{\sin(N\frac{\Delta}{2})}{\sin(\frac{\Delta}{2})} \end{aligned}$$

$$\Rightarrow |E_{array}(t)|^2 = \left| \frac{\sin(N\frac{\Delta}{2})}{\sin(\frac{\Delta}{2})} \right|^2$$

The term $|E_{array}(t)|^2$ denotes the increase in the power of the received signal at the output of the array relative to power of the received signal at an individual antenna element. However, we also have to account for the noise the output array as well. Since we are adding noise from N independent sources¹⁷, the noise power at the output of the array scales by a factor of N relative to the noise at a single antenna. Therefore, the boost in SNR at the output of the array relative to a single antenna:

$$\frac{\text{SNR}_{out}}{\text{SNR}_{in}} = \frac{\text{SNR}_{array}}{\text{SNR}_{antenna}} = \frac{1}{N} \left| \frac{\sin(N\frac{\Delta}{2})}{\sin(\frac{\Delta}{2})} \right|^2 \quad (2.11)$$

The ratio of the SNR at the output of the array to SNR at the input of the array (output of a single antenna) is called the array gain or array factor (AF), which describes how the power of the radiation pattern varies based on the angle of arrival θ (or angle of departure in the transmit case). The same derivation also applies to a transmit array¹⁸. We can see from Equation 2.11, the AF is maximized when $\Delta = 0$, and the maximum value $AF_{max} = N$. That means that the maximum array gain is proportional to the number of antenna elements and occurs when the signals from each antenna element are all added in-phase¹⁹. Figure 2.6 shows the normalized array factor of a 20-element uniform linear array with $\lambda/2$ inter-element spacing, with an angle of observation $\theta = \pi/6 = 30^\circ$ from the axis of the array, and $\Delta = 0$.

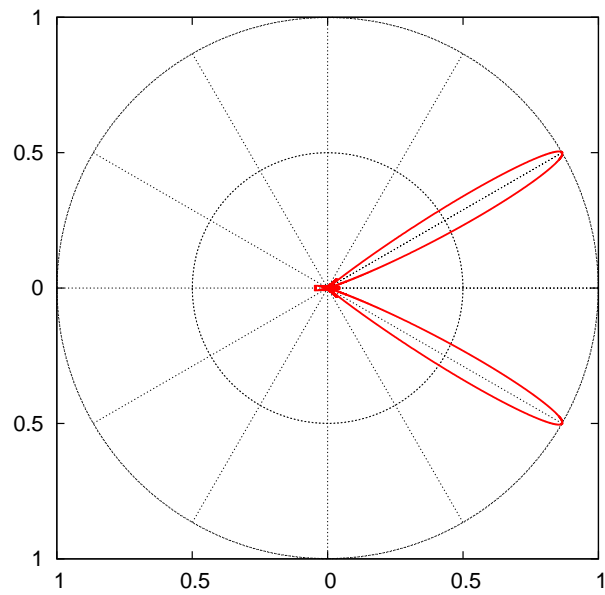
Several aspects of the array factor plot, shown in Figure 2.6b, are of interest in a wireless communication system. First and the most obvious is the direction where the beam peaks, usually referred to as the main lobe. The direction of the main lobe is steered towards the desired target by controlling the phase excitation ϑ . The width of the main lobe is inversely proportional to the array gain and can be decreased by increasing the effective area of array (by adding more elements without changing the spacing). Second, the other mini-peaks or local maximas are called sidelobes. While they may not impact the beam towards the desired signal, they can contribute to the overall interference in the system. The level of these side lobes can be lowered by controlling the relative amplitudes of the signals at the antennas. While the phase is used to steer the direction, the amplitude adds more leverage in shaping the pattern. In Figure 2.6b, the amplitude is uniform, which results in highest size lobe levels [46]. This is equivalent to multiplying the antenna weights with a rectangular window. However, by ‘‘tapering’’ the window (e.g. using a triangular window), the side lobe levels can be reduced significantly²⁰. Many techniques used for time-domain filter design apply in

¹⁷we assume that the noise is independent across both time and space.

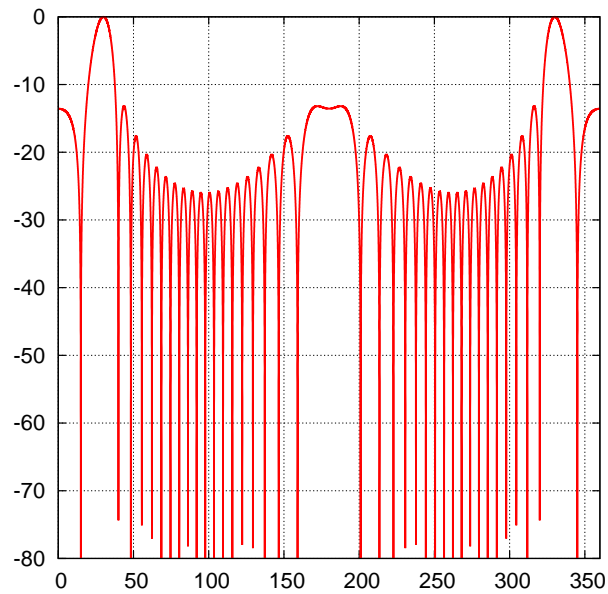
¹⁸In the receive case, the factor $1/N$ in AF is a result of adding N independent noise sources. In the transmit case, the $1/N$ factor comes from the fact that the transmit power from each antenna is scaled down by N such that the total power transmitted from all antennas is normalized to 1, which is equal to a single antenna case.

¹⁹Since $\Delta = 0$ is only true for a single frequency, the maximum value of the array factor $AF_{max} = N$ is only attained at that frequency. In order to get a wideband response, using a true programmable time delay is necessary.

²⁰Tapering has the effect of reducing the effective area of the array, since using weights that less than the



(a)



(b)

Figure 2.6: Beam pattern of a 20-element linear array with $\theta = \pi/6 = 30^\circ$ and $\Delta = 0$ (a) 2D linear polar plot. (b) 2D rectangular plot. For both plots the elevation angle is fixed at $\pi/2$ (i.e. xy-plane).

this context as well [42]²¹. A more comprehensive discussion on window design, tapering, and side lobe suppression can be found in [12, 46]. Third, the location of minimas or zeros, also called beam nulls, also impact the interference in the system. Beam nulls are usually steered towards big jammers. Like the side lobes, null-steering requires controlling both the relative phase and amplitude of the signal at each antenna. Null-steering is discussed in more detail in Chapters 5, 6, and 7. Finally, notice that the signal also peaks in a direction other than the desired direction $\theta = 30^\circ$. This other peak occurs at $\theta' = -30^\circ$, and is called a grating lobe. The presence of a grating lobe means that signals coming from the direction of the grating lobe have the same relative phase responses at the antennas as the desired signal, and are thus indistinguishable. As a result, these grating lobes, which result from the symmetry in the array geometry, are not desirable²². The grating lobes can be reduced by designing the array geometry to break the symmetry. This will be discussed in Chapter 4.

If the antenna elements of the array were all isotropic, then the radiation pattern will be exactly the array factor. However, the individual antenna elements are not going to be isotropic since such antennas are not realizable in practice. In this case, the overall radiation pattern is the product of the array factor with the pattern of the individual antenna (also known as the element factor or EF) [54]:

$$F(\theta, \phi) = AF(\theta, \phi)EF(\theta, \phi) \quad (2.12)$$

where $F(\theta, \phi)$ is normalized radiation pattern of the array. In practice, mutual coupling between neighboring antennas and edge effects make it difficult to have a uniform pattern for all the antenna elements even if those elements were identical. In this case, Equation 2.12 will not yield accurate results. However, for large arrays, the edge effects and pattern distortions can be smoothed, in which case, the overall radiation pattern can be approximated by the product of the array factor with the “average” element factor [54]:

$$F(\theta, \phi) = AF(\theta, \phi)EF_{av}(\theta, \phi) \quad (2.13)$$

Although we used a simple example of a uniform linear array to illustrate the basic properties of phased array antennas, the same concepts and techniques can be used to derive the properties and radiation patterns of arrays with arbitrary geometries. Some of these geometries will be explored in more depth in Chapter 4.

Another advantage of phased array antennas is spatial power combining. On the transmit side, in addition to the increased directivity, the power transmitted by the individual power amplifiers attached to each antenna are also combined in space. Thus, the signal power at the receiver is enhanced by a factor of N^2 (a factor of N comes from the array gain and

peak weight means that some of the antennas are only partially on. That means that the price of lower side lobe levels will be a wider main lobe and reduced gain [46]. The shape and level of the side lobes can also be changed by changing the geometry of the array.

²¹The set of complex weights that multiply the signals at each antenna can be viewed as a filter in the spatial domain (similar to the time domain). The corresponding beam pattern can be viewed as the angular domain representation of the filter (similar to the frequency domain). The relationship between the spatial domain and angular domain is very similar to the relationship between time domain and the frequency domain.

²²The number of grating lobes also increases when we increase the spacing between the antenna elements, especially when the spacing exceeds λ .

another factor of N comes from the spatial power combining). This also means that a single large power amplifier can be replaced by an array of smaller amplifiers that can potentially be implemented more efficiently using a low cost silicon technology like CMOS.

The implementation of signal amplitude change and phase rotation as well as signal combining and splitting is covered in more depth in Chapter 3.

2.2.1 Other types of antenna arrays

Phased array antennas are by far the most popular method for producing a electrically steerable beam. However, there are other ways of utilizing antenna arrays to steer the beam without changing the phase of the signal. These methods include switched array antennas and switched parasitic antennas. Here is a brief overview of each of these methods.

- Switched array antennas

This implementation of beam-steering uses an array of directional antennas (e.g. parabolic dishes) [12]. These antennas can be arranged in a circle for example, with each of them pointing in a different angle or covering a different sector (the main lobes of the different antennas usually have some overlap to ensure continuity of coverage as shown in Figure 2.7). The beam-steering is implemented by turning on the antenna or group of antennas pointing in the desired direction and turning off the rest of the antennas. The main advantage of switched arrays is simplicity. However, they do not provide the same level/granularity of control over the shape of the beam pattern as do phased arrays. Furthermore, whereas all antennas in a phased array simultaneously work together to form a beam, only one or a few of the antennas in a switched array effectively capture a signal from a given direction. Therefore, in order to provide the same level of directivity, each antenna in the switched array will have to be as large as the entire phased array. Also, since only a few antennas are used at any given time, the array does not take advantage of spatial power combining to increase the transmit power.

- Switched parasitic antennas

In a switched parasitic antenna, only one of the antennas in the array is excited by a signal; the rest of the antennas around that driven element are parasitic elements [57, 32]. These parasitic elements act like scatterers. The radiation pattern is shaped by changing the electrical characteristics of these scatterers.

Most of the discussion in this dissertation is focused on adaptive antennas based on phased arrays for the clear advantages they provide. However, many of the ideas and techniques that are discussed apply to other beam-steering implementations as well.

2.2.2 Scanning versus adaptive arrays

Beamforming systems can be classified based on their beam shaping capability. For example, one class of beamforming systems is scanning arrays. As the name suggests, scanning arrays can scan the direction of the beam across a wide range of angles, but do not have the ability to create more sophisticated beam patterns with different lobe sizes. However, this capability is sufficient for many applications including radar and line of sight communications.

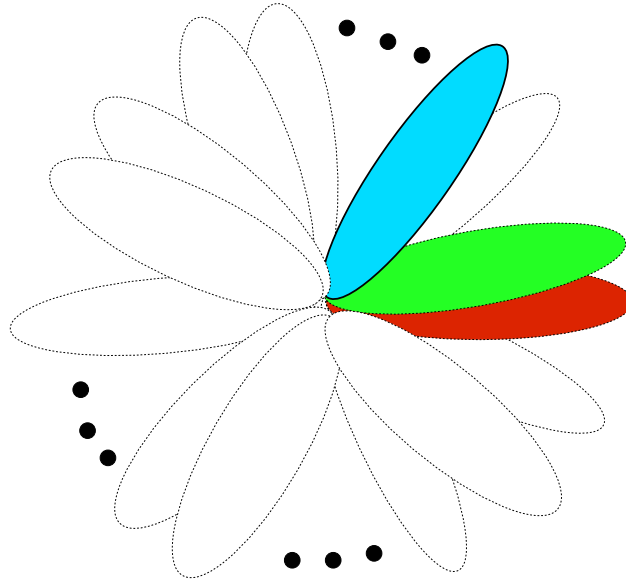


Figure 2.7: A switched array of directional antennas with overlapping beams arranged in a circle.

Popular techniques for implementing scanning arrays include the Butler and Blass matrices [15, 14]. Another popular class of beamforming systems is adaptive arrays. Adaptive antenna arrays can dynamically compute the beam pattern that optimizes certain aspect of the desired signal. The beam may not necessarily peak in any particular direction. In fact, in multipath scattering environments (described in Section 2.3), the signal comes from many directions, and the optimum beam pattern gathers the energy from all these directions. Also, the optimum beam pattern may not result in the maximum strength. Instead, it tries to maximize the another aspect of the signal, usually the signal to interference and noise ratio or SINR. To maximize the SINR, the adaptive array tries to maximize the desired signal while placing nulls in the direction(s) or interferers. Also, as the name suggests, the process is dynamic and adapts to changes in the environment and does not require prior knowledge of the channel statistics. Adaptive array systems add a lot of advantages to wireless systems in urban and suburban environments compared with other types of antenna arrays, and thus, are the focus of this dissertation.

2.3 The wireless channel

Communication over a wireless channel is inherently different from traditional wired communication due to the limitless nature of the medium. Because waves are free to travel in any direction, they can bounce, reflect, and scatter off of any object. The scattering, bouncing, and reflections affect the characteristics (amplitude, phase, polarization) of the waves in very arbitrary ways. Furthermore, the movement of these objects make the channel dynamic in nature. These effects can be detrimental to a wireless system if not well understood and accounted for. In this section, we present a brief overview of the basic properties of wireless

channels and their implications for system design, especially adaptive array systems²³.

2.3.1 Channel fading

One of the main consequences of waves bouncing and scattering off of objects is that the signal usually arrives at the destination from multiple paths. These paths usually have different lengths (delays), arrive from different directions (angles), and have different strengths. The statistics of these quantities depend on the environment (e.g. the number and location of scatterers). In the absence of multipath, the variation in the wireless channel occurs mainly due to the change in distance between the transmitter and the receiver as these radios move around. The signal “fades” as the distance between the two radios grows. However, the time scale required for a significant change in the signal amplitude is usually large, and this change will be largely frequency independent²⁴. On the other hand, when there are multiple paths, the signal can still fade even when the transmitter and receiver are close to each other. In this case, fading occurs when the signals from different paths are out of phase at the receiver, which creates a destructive interference. This type of fading has several characteristics that distinguish it from the traditional “slow” fading²⁵:

- While the delay from each path is only a function of the length of that path and is independent of frequency, the phase of the signal is frequency dependent. Therefore, when the signals from different paths are added at the receiver, the magnitude of the sum will also be frequency dependent. Therefore, the fading will be a function of frequency.
- Since any change in the path length that is of the order of the carrier wavelength can rapidly change the phase, any change in the position of the radios or the scatterers that is of the order of the carrier wavelength can lead to rapid variations in the amplitude of the received signal. This is why this type fading is referred to as “small-scale fading” [59].

Therefore, small-scale fading is a function of both frequency and space²⁶. For the rest of the dissertation, we shall use the term fading to refer to small-scale fading. A multipath environment is usually referred to as a fading environment. Using ray tracing, we can represent a wireless channel between two radios in a multipath fading environment in the time domain with a linear time varying system[59]:

$$h(\tau, t) = \sum_i a_i(t)\delta(\tau - \tau_i(t)) \quad (2.14)$$

where $a_i(t)$ is the amplitude response of path i as a function of time, and $\tau_i(t)$ is the delay of path i (also a function of time). Equation 2.14 describes the wireless channel from the

²³The material presented in this section is covered in more detail in [59].

²⁴With the exception of the fact that atmospheric attenuation is frequency dependent [45].

²⁵In addition to distance, slow or large-scale fading also occurs due to shadowing by large objects like buildings and hills in non-line of sight environments [59].

²⁶Fading can also be a function of time when the some of the radios and/or reflectors are mobile. However, the temporal variation is a consequence of the spatial variation.

transmit antenna to the receive antenna or the wireless channel in RF (passband) domain. However, since most of the processing happens in the baseband, a transfer function that describes the channel from the transmit baseband to the receive baseband would provide a lot more insight into the channel characteristics. Recall that the passband signal is an “up-converted” version of the baseband signal (i.e. the passband is obtained by multiplying the baseband signal by a complex exponential $e^{j2\pi f_0 t}$) [47]. Therefore, in order to convert back to the baseband, we multiply by the complex conjugate ($e^{-j2\pi f_0 t}$), where f_0 is the carrier frequency:

$$h_b(\tau, t) = \sum_i a_i(t) e^{j2\pi f_0 \tau_i(t)} \delta(\tau - \tau_i(t)) = \sum_i a_{b_i}(t) \delta(\tau - \tau_i(t)) \quad (2.15)$$

Therefore, the carrier frequency gets canceled. What remains is a phase term $e^{j2\pi f_0 \tau_i(t)}$ that results from the delay of path i . Equation 2.15 is called the baseband equivalent model [59], and $h_b(\tau, t)$ is the channel response at the baseband ($a_{b_i}(t)$ is the complex response of each path at the baseband). Notice that when a path length changes by a quarter wavelength, the phase response rotates by $\pi/2$. At 1GHz, a quarter wavelength is a little bit less than 8cm. Therefore, the phase response of each path can change rapidly for mobile nodes (on the order of a few milliseconds).

Most of the wireless systems of interest are digital and have finite resolution limited by the signal bandwidth B_s (symbol interval $T_s \approx 1/B_s$). The relationship between the digital (discrete-time) and analog (continuous-time) signal is:

$$x_b(t) = \sum_n x[n] p(t - nT_s) \quad (2.16)$$

where $p(t)$ is a pulse signal that is band-limited to B_s (e.g. a square pulse or a sinc pulse). When $x[n]$ is replaced by the unit sample $\delta[n]$, the discrete-time channel response is obtained by convolving $h_b(\tau, t)$ with $p(t)$ and sampling the output at multiples of T_s :

$$h_b(\tau, t) * p(t) = \sum_i a_{b_i} p(\tau - \tau_i(t))$$

If we sample at $t = mT_s$, then the l^{th} tap (delay) of the discrete-time channel response is ($l = m - n$ is the discrete-time equivalent of τ):

$$h_l[m] = \sum_i a_{b_i}(mT_s) p(lT_s - \tau(mT_s)) \quad (2.17)$$

Based on Equation 2.17, if $p(t) = \Pi(t/T_s)$ (i.e. a square pulse with width T_s), then the first channel tap is the sum of paths that arrive in the interval $[0, T_s]$, and the second tap is the sum of paths that arrive in the interval $[T_s, 2T_s]$, and so on (i.e. the l^{th} tap is the sum of paths that arrive in the interval $[(l-1)T_s, lT_s]$).

There are several key parameters (quantities) that are used to capture the basic characteristics of a fading environment²⁷. These quantities are statistical in nature, and are usually specified by their root mean square (rms) values. These key parameters are [59]:

²⁷These parameters come in pairs. With each pair describing a certain aspect of the environment in complementary domains (e.g. time/frequency and spatial/angular domains).

- Delay spread: the delay spread T_d is the difference in the time of arrival between the shortest path and the longest path [59]²⁸:

$$T_d \triangleq \max_{i,j} |\tau_i(t) - \tau_j(t)| \quad (2.18)$$

Two factors contribute to high delay spreads: long distance and rich scattering. If the environment has a lot of the scattering but short distances (e.g. an indoor environment), then the delay spread will be low (e.g. of the order of tens of nanoseconds). Similarly, if the transmitter and the receiver are far apart, but without any scattering (e.g. line of sight), then the delay spread will also be very low (almost zero). On the other hand, urban and suburban outdoor environments are both long distance and rich scattering. In these environments, the delay spread can be of the order of microseconds (sometimes tens of microseconds).

- Coherence bandwidth: the coherence bandwidth W_c is the frequency domain complement to the delay spread and is defined as the size of the frequency band over which the channel statistics remain relatively consistent. That means that when part of the band is in deep fade, then with high probability the entire band will also be in deep fade. Two frequencies that are more than a coherence bandwidth apart are assumed to have independent statistics. The coherence bandwidth is the inverse of the delay spread:

$$W_c \sim \frac{1}{T_d} \quad (2.19)$$

When the signal bandwidth B_s is smaller than the coherence bandwidth, the channel response will be almost flat over the entire signal bandwidth, a situation usually referred to as flat fading [59]. In the time domain, the delay spread will be smaller than the symbol time T_s , and thus, most significant paths will fall into a single tap (the Fourier transform of a single tap channel is constant over all frequencies.). On the other hand, when $B_s \gg W_c$, the fading will be frequency-selective. In the time domain, $T_s \ll T_d$, and the channel response will be represented by multiple taps. In our analysis, we focus on flat fading without loss of generality, since a frequency selective channel can be transformed into a set of parallel narrow sub-channels, each of them is flat fading and can be processed independently (see Section 2.3.3).

- Coherence distance: the coherence distance r_c is defined as the minimum distance between a constructive interference (channel peak) and destructive interference (channel minimum) [59]. In other words, r_c is the minimum distance between two points in space where the channel statistics are independent. The coherence distance not only depends on the environment, but also on the carrier wavelength λ_0 . This dependence is a result of the phase (and its rate of change) of each path being a function of the wavelength. Also, depending on the environment, the coherence distance is different in different directions. The coherence distance is independent of the delay spread. However, it depends on another quantity “the angular spread”, which is described next.

²⁸Paths with very low energies are generally neglected.

- Angular spread: The angular spread θ_d is defined as the maximum angle between two paths with significant energy²⁹. Unlike the delay spread, the angular spread does not depend on the relative lengths of the different paths, but instead on their angles of arrival. The angular spread depends on the location of the scatterers/reflectors relative to the radio. If the radio is located near the scatterers, then the angular spread will be large, and it will be low when the radio is far away from the scatterers. For example, in an indoor environment, the angular spread is usually high. In an urban outdoor environment (a cellular network for example), the angular spread will be high at the mobile stations (located near the ground), and low for the base stations (usually at a higher elevation). The angular spread is inversely proportional to the coherence distance:

$$r_c \sim \frac{1}{\theta_d} \quad (2.20)$$

These parameters, together with mobility related parameters, discussed in Section 2.3.1, capture the key features of the propagation environment that help the designers of wireless systems improve the overall system design.

Channel variation

The channel variation at fast time scales is mainly due to mobility of the transmitter and/or the receiver as well as, but to a lesser extent, the mobility of the reflectors³⁰. The combined effect of the velocity of the radios and the multipath nature of the environment on the wireless channel can be captured with two complementary parameters:

- Coherence time: the coherence time T_c of a time varying channel refers to the minimum time required for the statistics of the channel to change. Effectively, the statistics of channel change when the wireless node moves a coherence distance away from the original location. That means that the coherence time depends on both the coherence distance and the velocity of the node(s)³¹:

$$T_c \sim \frac{v_t}{r_c} \quad (2.21)$$

where v_t is the velocity of the node along the direction of the coherence distance. The coherence time is low in high mobility environments. This situation is usually referred to as fast fading. On the other hand, fixed wireless networks are slow fading environments. T_c is usually of the order of a few milliseconds for fast fading channels.

- Doppler spread: the mobility of a wireless node slightly changes the frequency of the signal by an amount that is proportional to the ratio of velocity v_t of the node to the speed of the light c :

$$f' = f(1 \pm v_t/c)$$

²⁹The angular spread must also specify the direction or the plane of the angle.

³⁰Changes in atmospheric conditions (e.g. temperature, air density, rain) also lead channel variations, but these happen at a much slower scale.

³¹The coherence time depends on the velocity of both the transmitter and the receiver as well as the reflectors.

where the sign depends on the direction in which the node is moving. The change in frequency ($\Delta f = fv_t/c$) is called a Doppler shift. Since the velocity is going to be different along different paths (velocity is a vector quantity that is also a function of direction), the Doppler shift will be different for different paths. Let $f'(i)$ be the frequency (Doppler shift) of path i . The the Doppler spread D_s is defined as the maximum in difference in Doppler shift between any pair of paths:

$$D_s \triangleq \max_{i,j} |f'(i) - f'(j)| \quad (2.22)$$

The Doppler spread is also inversely proportional to the coherence time:

$$D_s \sim \frac{1}{T_c} \quad (2.23)$$

Channel models

We can classify the wireless channel based on the propagation environment into three broad categories:

- Line of sight

The line of sight or LOS is the simplest model and assumes a single direct path between the transmitter and the receiver. In terrestrial communications, there will be usually at least one reflection (e.g. from the ground). Therefore, a perfect LOS channel is unattainable in practice. None the less, LOS is a good approximation to many real channels. In an LOS channel, fading can be approximated with the Friis equation with an additional exponential factor that captures the loss from the medium³²:

$$L(r, f) = \frac{G_{T_x} G_{R_x} \lambda^2 e^{-\alpha(f)r}}{(4\pi r)^2} \quad (2.24)$$

where f is the carrier frequency. The LOS channel is sometimes referred to as the AWGN channel. The delay, angular, and Doppler spreads are zero for an LOS channel.

- Rayleigh fading channel

The Rayleigh channel model is on the opposite extreme of the LOS model. Whereas the LOS model is completely deterministic, the Rayleigh model is completely probabilistic. The Rayleigh model is based on the assumption that each channel tap is sum of the responses from a large number of independent paths with different lengths [59]. Since the lengths of the paths are much larger than the carrier wavelength, it is reasonable to assume that the phase response of each path is uniformly distributed between 0 and 2π and is statistically independent across paths. Since there is a large number of paths, the sum of their complex responses can be approximated with a complex

³²The loss exponent is usually dependent on both the medium and the frequency. When the medium is air, the exponential factor in Equation 2.24 is negligible at low frequencies. It becomes significant at high frequencies. In fact, at frequencies around 60GHz for example, oxygen absorption is a major factor, and the exponential term dominates in long distances [45].

Gaussian random variable (using the Central Limit Theorem) [59]. Therefore, the discrete complex baseband channel taps are modeled as independent complex Gaussian random variables. The square magnitude (power) of each of these taps is exponentially distributed [59].

The key difference between the Rayleigh fading model and the AWGN model is that in a Rayleigh model, the channel SNR also becomes statistical. The main consequence of that the bit error probability (BER) improves as $1/\text{SNR}$ instead of $e^{-\text{SNR}}$ as is the case in AWGN channels [59]. This is one of the major hurdles faced by wireless systems.

The Rayleigh channel model is usually adopted for its simplicity. In Chapters 5, 6, and 7, we rely on the Rayleigh channel assumption to simplify the analysis.

- Rician fading channel

In practice, the channel is neither completely deterministic nor completely random. The Rician channel model is used to model channels that lie somewhere between these two extremes (i.e. channels that have a deterministic component and a random component). The Rician model is effectively a linear combination of the LOS model and the Rayleigh. The relative contributions of the deterministic (LOS) and random (Rayleigh) components to the overall channel model is determined by a parameter called the K -factor. A larger K -factor means the channel is more deterministic, and a smaller K -factor means the channel is more random. The K -factor itself is a random variable. The distribution of the magnitude of this random variable is Rician [59].

2.3.2 Channel diversity

The wireless channel response varies in both frequency and space. As we mentioned earlier in this section, the channels become almost statistically independent at locations that are more than a coherence distance apart, and frequencies that are more than a coherence bandwidth apart. There are different ways for exploiting this property of the channel in order to improve the channel reliability (i.e. achieve a BER better than $1/\text{SNR}$):

- By placing multiple receivers (i.e. antennas) or transmitters at multiple locations that are more than a coherence distance apart, the probability of the channel in all the locations being in deep fade is reduced considerably. This technique is called spatial diversity [59]. When implemented at the receiver, it's called receive diversity, and transmit diversity when implemented at the transmitter. There are several ways of combining the signals from different antennas to obtain diversity. The simplest way is to choose the strongest channel, and discard all others. This is called diversity combining. The second method, called maximum ratio combining or MRC³³, is little more sophisticated, and involves combining all the channels in phase in order to obtain the power gain in addition to the diversity gain.

Incidentally, spatial diversity is one of the biggest advantages of using antenna arrays. We can also consider this diversity method in the angular domain. The idea is that

³³In the signal processing literature, MRC is usually referred to as the matched filter.

if the array contains two or more antennas that are more than a coherence distance apart, in the angular domain, this means that the array has an angular resolution that is finer than the angular spread of the channel, which means that the array has the ability to resolve some paths and significantly attenuate others in order to reduce the probability of deep fade.

- By communicating over a bandwidth that is larger than the coherence bandwidth, we can reduce the probability of the entire band being in deep fade. This method is called frequency diversity [59]. The time domain interpretation of this phenomenon is as follows. When the channel bandwidth is larger than the coherence bandwidth, the symbol time will be smaller than the delay spread. Therefore, the channel will be composed of multiple taps. Since the statistics of these taps are independent, the probability of all of them simultaneously being in deep fade is also reduced.

The total diversity in the system will be the product of all these diversity methods. That means if the receiver has N_{R_x} antennas that are more than a coherence distance apart from each other, and the transmitter has N_{T_x} that are at least a coherence distance apart from each other³⁴, and the channel has P effective taps in the time domain, then overall diversity rank of the channel $L = N_{R_x} N_{T_x} P$. Diversity improves the BER from $\sim 1/\text{SNR}$ to $\sim 1/\text{SNR}^L$. In the limit as $L \rightarrow \infty$, the channel approaches an AWGN channel.

There are other methods for achieving diversity. Time diversity is used in mobile and cellular networks. The idea is that the statistics of the channels change after a coherence time. So diversity can be achieved by coding and interleaving (spreading) codewords over multiple coherence times [59]. Time diversity is a product of spatial diversity (radios moving to different points in space). Another method is polarization diversity. With polarization diversity, two antennas with opposite polarizations can have independent channel statistics even when they are less than a coherence distance apart. The reason this is the case is that wave propagation and reflection are affected by polarization. However, the diversity rank that can be achieved from polarization is limited since there can only be two orthogonal polarization states at any given time.

2.3.3 Frequency selective channels

Frequency selective channels improve the system performance by adding more diversity. However, one of the challenges for communicating at the maximum rate over a frequency selective channel is that since the channel has multiple taps in the time domain, the current symbol will be interfered with by the delayed versions of previous symbols. This situation is referred to as inter-symbol interference or ISI [59]. There are several approaches for dealing with ISI. We describe two of these approaches that are relevant to this dissertation:

- Linear equalization: a linear equalizer is a time domain filter that is chosen such that when convolved with the multitap channel response, the output will be approximately single tap. This method is called equalization because it tries to make the channel response equal at all frequencies across the signal bandwidth. The order of the filter

³⁴The coherence distance at the transmitter and the receiver need not be the same. In fact, they are usually different because the angular spreads are different.

has to be larger than the channel response in order to get satisfactory results. While linear equalization is not optimal, it provides a good balance between performance and complexity.

- Orthogonal Frequency Division Multiplexing (OFDM): this technique transforms a single wide band channel into a group of narrow band channels, called subcarriers. The data are encoded on the subcarriers in the frequency domain, and then converted to the time domain (using the Inverse Discrete Fourier Transform (IDFT)) to be transmitted. Since each subcarrier is a narrow band signal, the channel response will be flat for single subcarrier, but different for different the subcarriers. On the receive side, the receiver converts the incoming signal back into the frequency domain (using the Discrete Fourier Transform) to extract the data off of the subcarriers [59].

Other methods include spread spectrum and optimal maximum likelihood detection using the Viterbi algorithm [59].

For beamforming arrays, a frequency selective or ISI means that the single complex weight (amplitude + phase) used at each antenna is no longer optimal. The optimal beamformer will have to be a complex multitap filter at each antenna. The process of computing the optimal time domain weights for the filters at each antenna is called space/time adaptive processing or STAP.

2.3.4 Spatial multiplexing

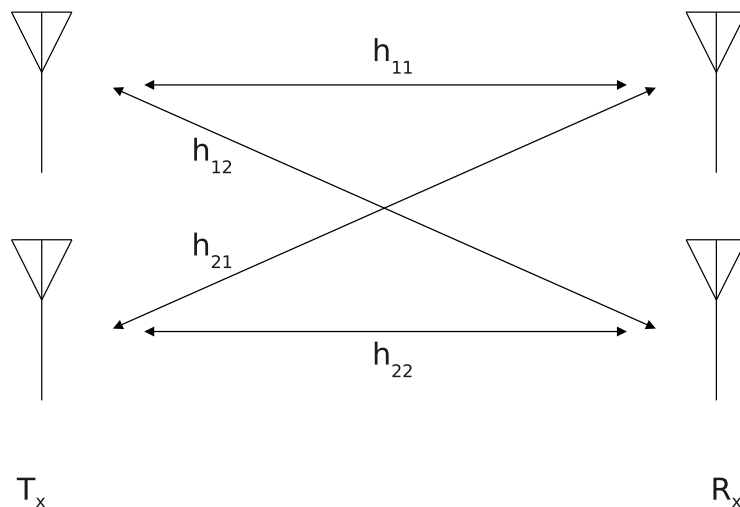


Figure 2.8: A two antenna transmitter antenna T_x communicating with a two antenna receiver in a Rayleigh fading channel.

Consider the scenario shown in Figure 2.8, where a two antenna transmitter is communicating with a two antenna receiver in a Rayleigh channel. Also, assume that the antennas on both the transmitter and the receiver are spaced more than a coherence distance apart

so that the complex coefficients of the channel matrix $\mathbf{H} = \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix}$ are statistically independent (we also assume that the channels are flat fading without loss of generality). If the transmitter transmits data stream $x_1[n]$ from the first antenna and $x_2[n]$ from the second antenna, then the received signals $y_1[n]$ and $y_2[n]$ at the first and second antennas of the receiver respectively are:

$$\begin{bmatrix} y_1[n] \\ y_2[n] \end{bmatrix} = \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix} \begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} + \begin{bmatrix} v_1[n] \\ v_2[n] \end{bmatrix}$$

Or in more compact vector form:

$$\mathbf{y}[n] = \mathbf{H}\mathbf{x}[n] + \mathbf{v}[n]$$

where $v_i[n]$ is an additive white Gaussian noise signal at antenna i at the receiver. If the receiver multiplies the input signal $\mathbf{y}[n]$ by \mathbf{H}^{-1} :

$$\hat{\mathbf{y}}[n] = \mathbf{H}^{-1}\mathbf{y}[n] = \mathbf{x}[n] + \mathbf{H}^{-1}\mathbf{v}[n]$$

then both data streams $x_1[n]$ and $x_2[n]$ can be recovered at the receiver. This technique is called multiple-input/multiple output (MIMO) or spatial multiplexing³⁵, and it is a consequence of the Rayleigh fading channel [59]. A prerequisite for spatial multiplexing is that the spatial channel matrix \mathbf{H} be invertible. In general, the transmitter will have N_{Tx} antennas and the receiver will have N_{Rx} antennas. In this case, the channel matrix \mathbf{H} will be $N_{Tx} \times N_{Rx}$. The maximum number of independent data streams that can be communicated simultaneously over the channel is equal to the rank of \mathbf{H} . When the matrix is not square, then the individual streams can be recovered by doing a singular value decomposition on \mathbf{H} . The number of “significant” non-zero singular values of \mathbf{H} will be approximately the effective rank of \mathbf{H} . If $N'_{Tx} < N_{Tx}$ is the maximum number of antennas on the transmitter that are more than a coherence distance apart from each other, and $N'_{Rx} < N_{Rx}$ is the maximum number of antennas on the receiver that are more than a coherence distance apart from each other, then the effective rank of the channel matrix \mathbf{H} is approximately $\min(N'_{Tx}, N'_{Rx})$.

MIMO and spatial multiplexing are key advantages provided by adaptive arrays for increasing system capacity with better spatial reuse. Another easier and more intuitive way to visualize this problem is through the angular domain. From the angular domain prospective, we can think of the transmitter as encoding different data streams on different paths. If the beamwidth of the receiver is narrower than the angle separating the paths at the receiver (angular spread), then the receiver array can resolve each path independently. The narrower the beams at both the transmitter and the receiver relative to the angular spreads, the more independent streams can be transmitted. In general, there is a one-to-one correspondence between the spatial and angular domains.

³⁵The antennas do not necessarily have to be attached to the same node.

Chapter 3

Radio architectures

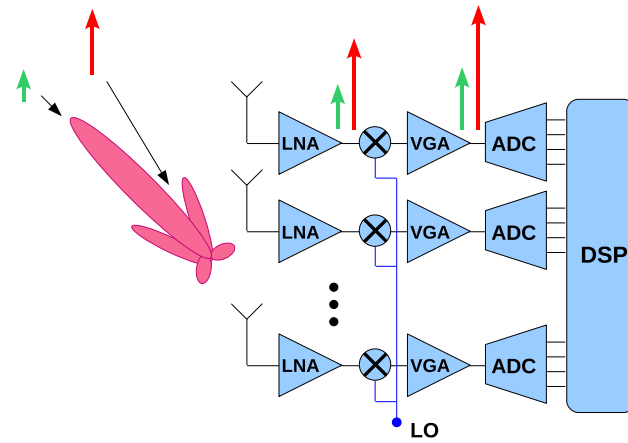
The practical implementation of phased-array systems has been greatly aided by recent advances in silicon technology (especially CMOS), which have made possible the integration of a large number of radios on a single die. Integrating the components of a multi-antenna radio on a single chip can significantly reduce the overall cost and power consumption of a phased array. However, there are several possible architectures, which each present a different combination of benefits and obstacles. The fundamental difference lies in how the signals from the various antennas are combined. In this chapter, we present an overview of three beamforming architectures (digital, LO, RF), and discuss their advantages and drawbacks. Although we focus on receive side of the system, the same ideas apply to transmit as well. We propose a hybrid RF/digital architecture that combines the benefits of both the RF and digital beamforming architectures in order to meet the overall system capacity and gain requirements at a much lower cost and power requirement than each of the RF and digital architectures used individually. The chapter is organized as follows. In Section 3.1, we provide an overview of the three basic architectures, and discuss the pros and cons of each, and describe our proposed RF/digital hybrid architecture. In Section 3.2, we discuss the RF beamformer implementation.

3.1 Comparison of Architecture

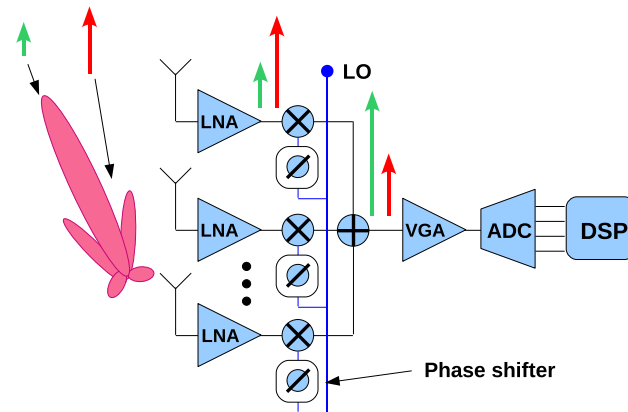
An antenna array must perform two basic operations to dynamically adapt its beam pattern: change the phase and amplitude of a signal (i.e. multiplying by a complex number or providing a programmable complex gain), and add multiple signals from different paths¹. This core functionality can be implemented at any stage in the radio transceiver chain (e.g. RF, IF, local oscillator (LO), analog baseband, digital baseband), as shown in Figure 3.1. The different array architectures provide different trade offs among cost, complexity, power consumption, and performance. Here, we analyze three of the most popular architectures.

The work presented in this chapter was done in collaboration with Ehsan Adabi and Ali Niknejad [9].

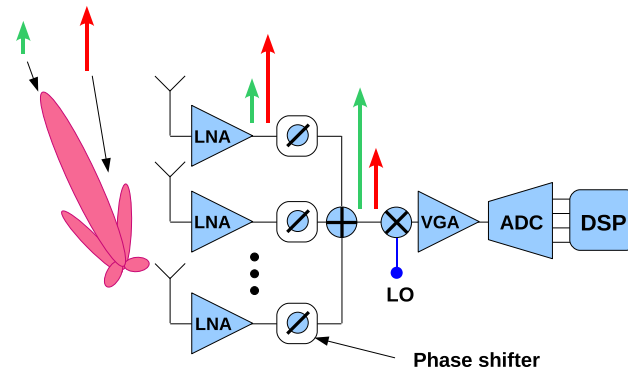
¹In the context of transmit beamforming, signal combining is replaced by signal splitting.



(a)



(b)



(c)

Figure 3.1: Radio architectures for beamforming: (a) Digital baseband beamforming (b) LO beamforming (c) RF path beamforming. While the figures only show the receiver chains, the same concepts also apply to transmit beamformers as well.

- a) Digital baseband beamforming (Figure 3.1a): In this architecture, the core functionality is implemented in the digital signal processing domain, after the ADC. Performing the beamforming functions in the digital domain provides the best performance, since the system can exploit high-speed DSPs and implement complex, high-resolution algorithms in digital domain. It also makes the system highly flexible, as it is relatively easy to subsequently reprogram the beamforming algorithm. The reconfigurability of this architecture makes it easier to implement wideband, multitap true time delay filters, and thus enables support for multiple data streams in both the spatial and frequency domains. A technique known as space-time adaptive processing (STAP).

However, there is a price to be paid for this flexibility. First, a digital beamformer requires a complete transceiver chain per antenna (degree of freedom). This will result in a significant increase in cost and power consumption, which will in turn limit the scalability (in the number of antennas) of this architecture. The complexity of the digital baseband processor will also increase in terms of processing power and number of input/output (I/O) ports. For example, a 20-antenna digital beamformer operating on a 40MHz signal (sampled at twice Nyquist rate) needs to process 1.6Gsample/sec. If each sample is quantized to 8 bits by the ADC, then this translates to 12.8Gbps to be processed and at least 20 I/O lines that are 640Mbps each! Second, each of these transceivers will have to be tightly synchronized in both time and frequency. Synchronization errors especially in frequency can significantly degrade the performance. This tight synchronization is harder to achieve at higher frequencies (e.g. mm-wave bands). At those frequencies, even if all local oscillators (LOs) are to be locked to the same reference clock, the phase noise and jitter can still be significant. Finally, since the spatial filtering and interference nulling does not occur until the signal reaches the digital baseband, all the blocks at the receiver prior to the DSP will have to deal with the interferers and jammers. Therefore, they should have sufficient linearity and dynamic range to cope with the large blocker levels. This will increase the complexity and power requirements of these blocks. Furthermore, the linearity of the transceiver is not only affected by in band jammers, but also out of band jammers. These, however, can be dealt with using time domain filters with sharp cutoff either at RF or IF and analog baseband. These filters, however, will require a lot of area and sometimes expensive material (especially if implemented at RF) and thus increase the overall cost since they are replicated for each transceiver.

Multi-input/Multi-output systems, which have become popular in the latest IEEE 802.11 and 802.16 standards, are examples of popular systems that are based on digital beamforming. However, most practical systems can only support a small number of antennas or degrees of freedom (typically 4-8).

- b) LO frequency beamforming (Figure 3.1b): In this architecture, phase shifting is performed in the LO and clock distribution network, while signal combining is performed in IF or analog baseband (depending on the LO frequency). This has the advantage of reducing the number of ADCs and baseband circuitry, and relaxing the dynamic range requirement on them due to spatial filtering after the signal combining. However, the system needs a highly symmetric distribution network, to provide the exact desired

phase shift, and it also requires significant power consumption in the buffer stages that are needed to provide a strong LO signal at each of the mixers. In addition, the LO architecture, like the digital baseband architecture, requires that different signal paths be frequency synchronized. Errors in synchronization can lead to significant performance degradation. A further limitation is that in general LO phase shifting provides no control over signal amplitudes². Examples of this architecture can be found in [25, 38]

- c) RF path beamforming (Figure 3.1c): In this architecture, both complex multiplication and signal addition are performed directly in the RF path. The advantages of this architecture are that it provides the lowest component count and power consumption, which yields the lowest cost and best scalability. It also avoids the synchronization problem of LO and digital beamforming. However, one drawback is that the design of RF phase shifting components is much more challenging than digital beamforming algorithms. In addition, the performance is lower than the other architectures, since practical phase shifters have non-ideal characteristics (e.g., loss, noise, nonlinearity), which are now directly in the RF path. These problems become even more pronounced as the frequency increases. Despite the challenges of implementing the beamforming functions in the RF domain, this architecture is the best candidate for meeting our scalability requirements. The implementation of this architecture using integrated circuit (IC) and printed circuit board (PCB) technology will be discussed in more detail in Section 3.2.

We also note that while the beamforming is performed in the RF domain under this architecture, the signal processing and parameter optimization can be and is more efficiently performed in the digital domain. However, the same signal processing techniques used under the digital beamforming architecture cannot be reused in this context since there is a fundamental difference: the RF beamformer presents the signal processing unit with only a single input, which is the aggregate signal. The implications of this limitation are discussed in more detail in Chapter 5.

3.1.1 Hybrid arrays

We showed in this section that different architectures have their advantages and limitations. While it may not be possible to meet the system requirements of a particular application (e.g. gain, capacity, cost, power consumption) with one of these architectures, a combination or hybrid architecture is more likely to achieve those desired features since it will combine the desired features of multiple architectures. For example, a well designed hybrid RF/digital can achieve high gain and scalability at low cost and power (RF beamforming), and at the same time achieve high capacity (digital beamforming). Although exact digital/RF breakdown is application dependent, here are some general guidelines:

- 1) The total capacity of the system is proportional to product of the signal bandwidth and the number of simultaneous spatial streams. The number of simultaneous spatial data

²This can be remedied by using programmable gain blocks in IF or analog baseband prior to combining the signals at the receiver, which imposes additional cost and complexity.

streams depends on the number of digital I/Os or transceivers³. The bandwidth of the individual signal also places some limitations of maximum size of an RF subarray. The size of the RF subarray has to be such that the maximum difference in time of travel or delay (in any direction between any two elements of the subarray) is small relative to the symbol duration (inverse of bandwidth). If this condition does not hold, then not only will the phase shift approximation breakdown, but it may also be difficult to implement programmable delay blocks at RF that can support these levels of delay.

- 2) The required rate of adaptation also determines the number of digital I/Os. Since the RF beamformer only has access to the aggregate signal, the RF beamforming weights are generally slower to adapt than the digital weights. Chapter 5 contains a more detailed discussion on this topic.
- 3) The total number of antennas in the array depend on the required gain and steering range. If the steering (scanning) range is narrow, then a small number of high gain directional antennas can be used. However, when the required scanning angle is wide, then a large number of small antennas need to be used. The number of antennas can be increased by either increasing the number of RF subarrays or the size of each subarray. The latter option is more cost effective, since it adds more antennas in the RF domain without increasing the digital processing and I/Os.

Although the main system requirement that drives the need for more antenna gain is distance or range, gain is also strongly related to both the capacity and power requirements of the system since higher gains result in narrower beams that reduce transmit power requirements and improve the isolation between different spatial streams and are better at canceling external interference. Furthermore, the higher gain and improved isolation increase the reliability of the channel which reduce the probability of error and enable higher signal constellations that result in better spectral efficiency.

- 4) The total power consumed is usually dominated by two factors: the RF power amplifiers and the digital baseband processing unit.

Examples of such a hybrid or hierarchical architecture are shown in Figure 3.2. In Figure 3.2a, the total area is divided into reticles of subarrays. Within each subarray, beamforming is achieved by RF phase shifting. Then the subarrays are connected and synchronized via the lower clock frequency, and programmed digitally. The beams of each subarray can be unique or can be combined with neighboring arrays to create a stronger beam. Multiple target tracking at different directions and different frequency bands will be controlled by the high-speed DSPs that program the subarrays. While this hybrid architecture enables multiple simultaneous beams, each beam only uses one subarray. Figure 3.2b shows an alternative multibeam hybrid architecture that enables each beam to use the entire array. In this architecture, the analog (RF) beamforming architecture is extended to support multiple beams without significantly increasing the complexity. In order to support K beams, we need K sets of complex weights. Therefore, the total number of phase shifters or complex

³In theory, the number of digital I/Os has to be at least the number of spatial streams, but is usually a lot more in practice.

multipliers will increase by a factor of K . The original $N \times 1$ beamformer (where N is the total number of antennas) is transformed into an $N \times K$ beamformer by replicating the complex multipliers and signal adders. If $N \gg K$, then we are not significantly increasing the complexity, because most of the RF components (e.g. PAs, LNAs) are still shared. Under this architecture, each stream uses the full array.

3.2 RF beamformer design

In Section 3.1, we evaluated multiple architectures for beamforming transceivers. Of these architectures, the one based on RF beamforming is of interest both as a standalone and as a frontend in a Hybrid RF/digital architecture due to its scalability in both power and area. The main challenge of the RF beamformer is implementing the core beamforming functionality (signal combining/splitting and phase/amplitude correction) in the RF domain. In this section, we consider the implementation of these blocks in more detail. Before we get into the circuit details, we first present a high level design of a multi-input/multi-output beamforming module in RF, and discuss the main requirements that this module needs to satisfy in order to meet the overall system specs. We then present different implementation options for the basic beamforming building blocks and evaluate them in light of those requirements. We note, however, that many of these requirements cannot be perfectly satisfied, and the extent to which they are met is application and frequency dependent⁴.

3.2.1 A multi-input/multi-output RF beamforming module

Like any conventional antenna, the size of an adaptive array system is also proportional to the gain and the wavelength. Therefore, silicon integration is not an option for the antennas except at extremely high frequencies ($\geq 100GHz$). However, the antennas can still be printed or mounted on a printed circuit board (PCB) at very low cost. In this case, a large portion of the cost will be the number of components mounted on the board. This can be reduced by integrating RF blocks from multiple channels (antennas) into a single IC as shown in Figure 3.3. Figure 3.3 shows a high level block that takes several inputs at RF (from multiple antennas) and outputs multiple linear combinations of these signal (also at RF). The linear combinations are programmed through a digital control channel. Figure 3.4 shows a chip that supports 16 RF channels mounted on a PCB and feeding monopole antennas on a ground plane with microstrip lines.

For phased array systems that require antenna gains of 20dBi or more, the number of antennas can easily be of the order of hundreds. It is almost impossible to support such a large number in a single silicon chip. However, even it were possible, it is still not a good idea to put that many RF channels on a single chip. First, the traces between the antennas and the chip will get too long and loss will increase. The advantage of using a small chip is that it

⁴In this chapter we use the term low frequency to refer to frequency bands with wavelengths that are much larger than the size of the module itself. For example, the wavelength at 5GHz is about 6cm, whereas, the area of the entire module may not exceed a few square millimeters. On the other hand, the wavelength at 60GHz is only a few millimeters and is comparable with size of the chip. However, there really is not an exact cutoff frequency separating the two groups.

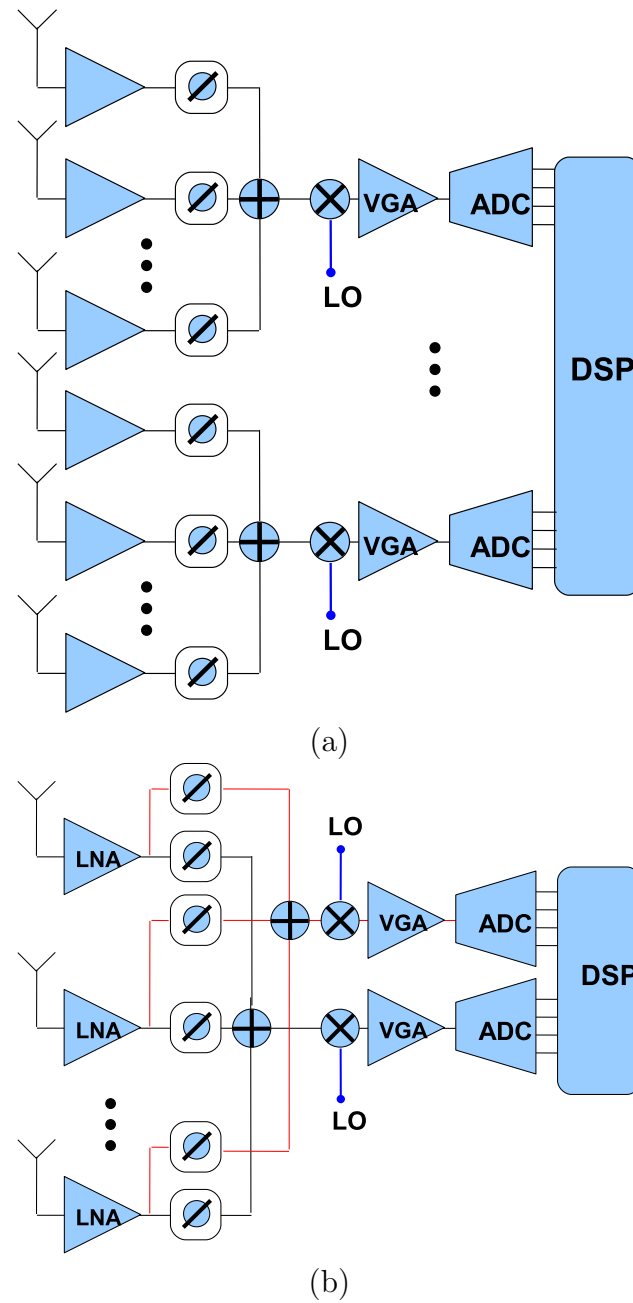


Figure 3.2: Two different architectures for a hybrid (hierarchical) RF/digital beamforming: (a) Large array divided into smaller subarrays connected digitally. Each subarray implemented in the RF domain. (b) A single array with multiple parallel RF beamformers. The outputs of the RF beamformers are connected and processed digitally.

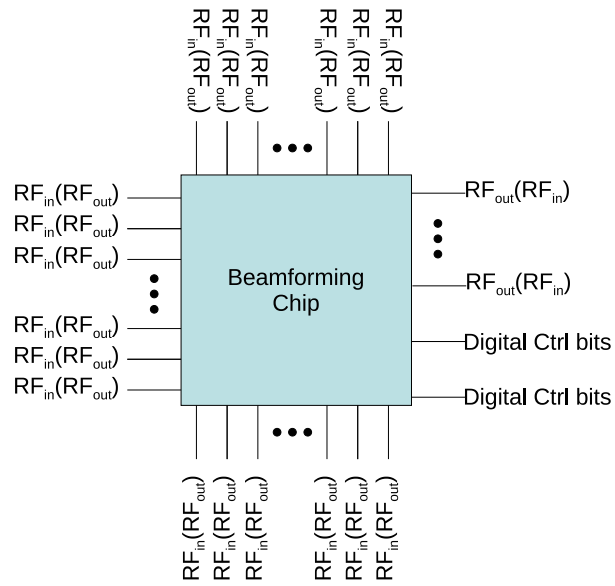


Figure 3.3: Multi-input/multi-output RF frontend.

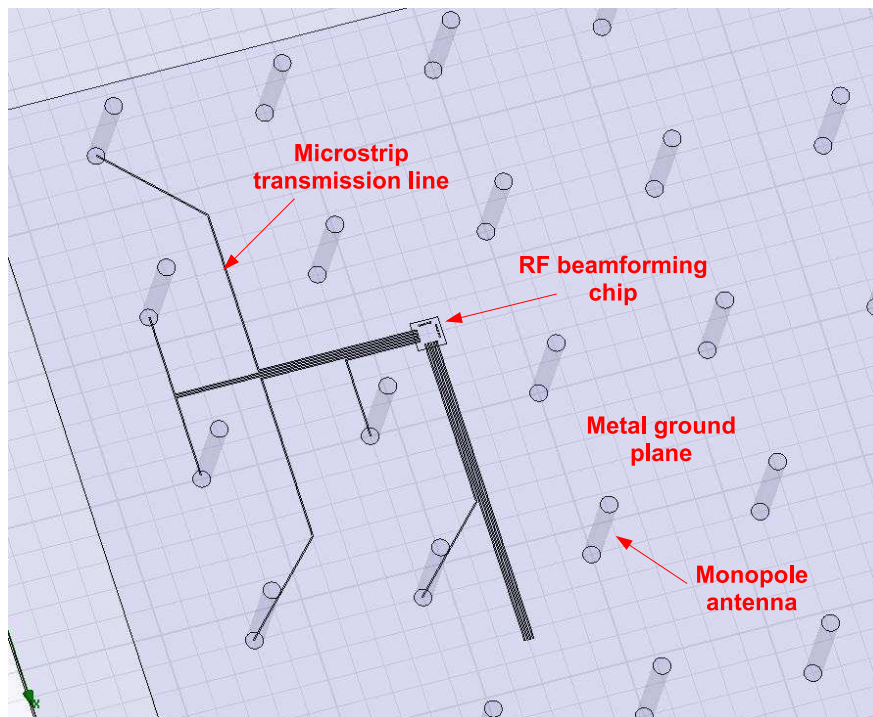


Figure 3.4: RF beamforming chip with 16 RF I/Os connected to monopole antennas with microstrip lines on a PCB.

is going to be close to the antennas as shown in Figure 3.4. Second, the cost of the IC drops in high volumes. If an IC supports 200 channels, it is unlikely to see wide adoption since it is very application specific (only applications that require around 200 hundred antennas). However, the advantage of the RF_{in}/RF_{out} chip architecture is that it can be cascaded. We can take advantage of this feature to support antenna arrays of virtually any size. For example, suppose we have an IC that supports 16 antennas and outputs 4 different linear combinations, and we want to build a 64 antenna array. Then we can use 4 of these ICs to connect to 64 antennas, and the output of these ICs is fed to a 5th IC. The four outputs of the 5th IC are then fed to 4 transceivers that convert them into digital basedband to be processed by a DSP unit (the reverse is true for transmit).

The multichannel RF IC would integrate the conventional RF blocks (e.g. PAs, LNAs, VGAs, switches) as well as the new beamforming related blocks (e.g. complex multipliers and adders). These components are shown in Figure 3.5a for the receiver and Figure 3.5b for the transmitter. While both figures show a 4x4 complex vector multiplier block, the numbers of inputs $M = 4$ and outputs $N = 4$ do not have to be restricted (with $M \geq N$)⁵. Figure 3.6 shows these $M \times N$ complex modulators as black boxes that take the M input (complex) signals at RF s_{i1}, \dots, s_{iM} and produce N programmable linear combinations of these inputs:

$$s_{ok} = \sum_{j=1}^M c_{jk} s_{ij} \quad \forall 1 \leq k \leq N$$

where c_{jk} are programmable complex coefficients (gains). The equations can also be written more compactly in matrix form: $\mathbf{S}_o = \mathbf{C}\mathbf{S}_i$.

The design of these blocks impacts system level performance such as interference rejection. Therefore, there are certain requirements for these blocks in order to meet the system level specs. These requirements are:

- 1) Independent phase and amplitude control must be provided at each signal path. This is necessary to give the beamformer tight control over the beam nulls and side lobes to suppress interference.
- 2) Coupling (cross talk) between different paths must be minimized. That means that amplitude and phase control on one path should not affect signals on other paths. Coupling impacts the accuracy of beam nulls.
- 3) The resolution of the phase and amplitude control must be as high as possible, and should be robust to process and temperature variations.
- 4) Since all these blocks are in the signal path, the distortion of these blocks must be minimized.
- 5) The mismatch between transmit and receive paths must be minimized in order to maintain channel reciprocity.
- 6) Power consumption and area requirements should be as little as possible.

⁵The numbers of inputs and outputs are usually chosen as powers of 2 for simplicity.

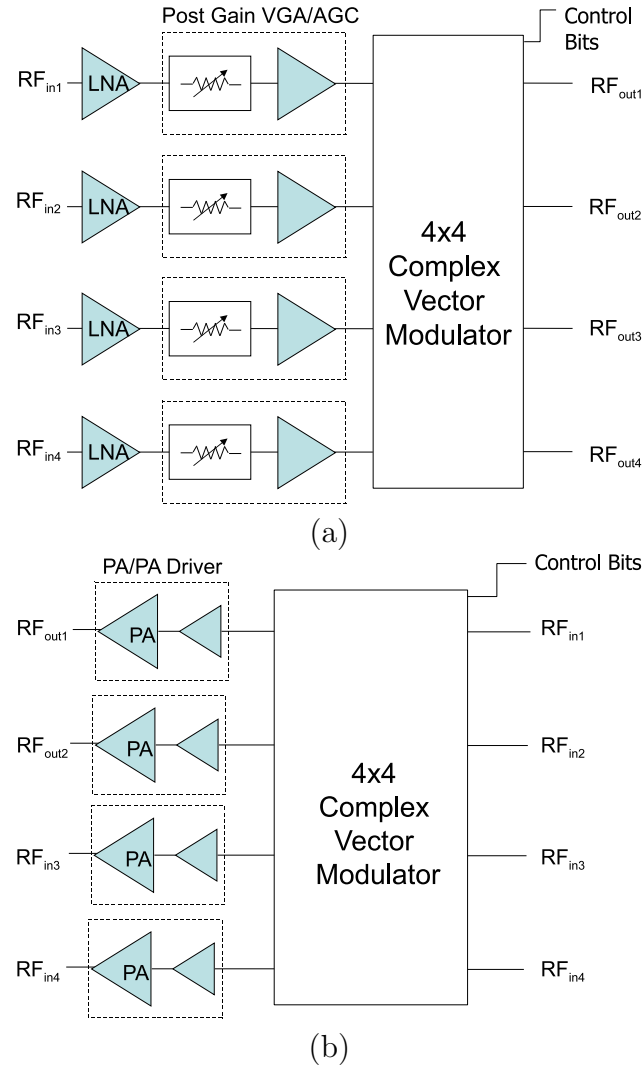


Figure 3.5: Building blocks for RF beamforming frondends: (a) Receiver. (b) Transmitter.

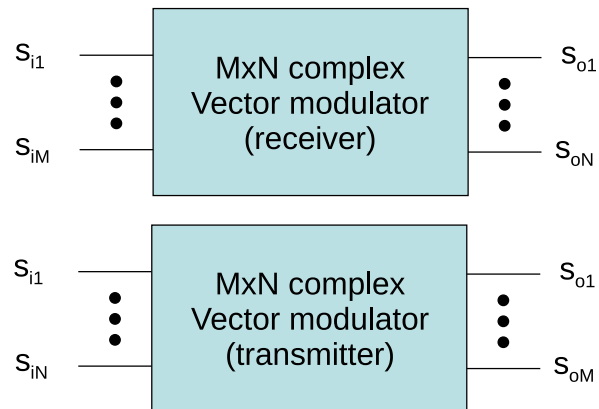


Figure 3.6: Multi-input/Multi-output RF complex multiplier block.

If we can meet these requirements with a passive implementation, then it is always more desirable to an active implementation for several reasons. First, a passive implementation draws a lot less DC power, and is usually a lot more linear. Second and more importantly, passive networks tend to be symmetric (i.e. input/output are interchangeable). That means that the same components can be shared among both the transmit and receive chains. This not only reduces the area requirements, but also reduces the possibility of mismatch between the transmit and receive beamformers. However, as will see, it is not always possible to meet all these requirements with passive circuits.

The rest of the chapter covers the implementation of the basic components of the complex multiplier block in RF: signal combining/splitting and phase amplitude control.

3.2.2 Signal combining/splitting

There are several circuit techniques that can be used for signal combining and splitting at RF. Different choices are appropriate for different frequency bands. For low frequencies signal combining/splitting is more efficiently done in the current domain [25]. When this is not possible, power domain combining and splitting is the only option. Popular circuit techniques for power combining use transformers and transmission lines:

- 1) Transformer-based power combiners: a transformer-based power combiner is shown in Figure 3.7a. It has two desirable characteristics: low loss and compact structure, which makes it appropriate for integrated circuit implementations and a wide range for frequency bands. The main disadvantage of the transformer is that it is narrow band. Examples of this type of combiner can be found in [20, 24, 43].
- 2) Wilkinson power combiners: the Wilkinson is a popular microwave structure for building multi-way power combiners and dividers. A two-way/equal-split Wilkinson divider is shown in Figure 3.7b. The Wilkinson uses quarter-wave transmission line impedance transformers, and provides three desired properties:
 - Simultaneous impedance match at all ports.
 - High isolation between combined ports.
 - Lossless when all output ports are matched.

Figure 3.11 shows a 2-way/equal-split Wilkinson divider (designed for 2.4GHz) in microstrip printed on an FR4 board. The measurements from this divider are shown in Figures 3.8/3.9. Figure 3.8a shows an insertion loss of about .5dB resulting from non-ideal transmission lines. Figures 3.9ab show a low input return loss from all ports. Figure 3.8b, shows a high isolation between the divided ports near the frequency of interest. The the isolation and input return loss are wideband when all the ports are matched.

The main disadvantage of the Wilkinson is its size. Since it is proportional to the wavelength, it is more appropriate for high frequencies (e.g. mm-wave or higher), especially for integrated solutions. Furthermore, the insertion loss of the Wilkinson is usually higher than the transformer in practice. Also, it is difficult to implement the

Wilkinson network differentially. A more comprehensive analysis of the Wilkinson and other microwave-based power combiners can be found in [45]

Other power combining/splitting networks that combine both transmission lines and transformer have also been proposed [40].

3.2.3 RF phase shifting and complex multiplication techniques

Several methods have been used to implement a programmable complex gain at radio frequencies. Here are some examples of these techniques:

- 1) Extended-resonance/path-sharing phase shifters⁶: the underlying concept behind this technique is a ladder network that evenly distributes the signal across multiple paths while at the same time shifting the phase by progressive steps, and thus combining the signal splitting and phase shifting functionality into a single network [23]⁷. The progressive phase-shift is usually accomplished by adding a varactor at each stage of the ladder to control the shunt impedance. The voltage across the varactor controls the resonance and phase response (the same varactors and the same voltage are used in all stages of the ladder). In addition to combining the signal splitting and phase shifting functionality into a single network, the other advantage of this architecture is that the phases across the elements of the array are controlled by a single parameter (voltage). The network is also passive, which means it can be used for both transmit and receive in a time-division duplexing (TDD) system. However, the price of this simplicity is that the progressive phase distribution is only appropriate for uniform linear arrays. For arrays of arbitrary geometry, the phase-shifts across the different elements will in general not be progressing. Even for linear arrays in a rich scattering environment, the channel phase responses are unlikely to be progressive. Furthermore, since the array requires a single control parameter, this parameter is required to be high resolution. The required resolution increases logarithmically with the number of antennas in the array⁸. Also, a variable control block is required for amplitude tapering if side-lobe suppression is desired⁹.
- 2) Reflective-type phase shifters: a reflective-type phase shifter consists of a hybrid 90° coupler, which can be implemented with lumped circuit elements as shown in Figure

⁶The path-sharing structure can be used with other types of phase-shifters as well. However, extended-resonance has a more compact structure

⁷The concept of a progressive phase shift is discussed in Section 2.2 in the context of linear arrays.

⁸In Chapters 6 and 7, we analyze the impact of phase and amplitude errors resulting from quantization and other error sources on the performance of the array. However, in those chapters, we assume that the beamforming weights are independently chosen for each antenna, and thus the errors can also be assumed to be independent. This assumption does not hold for this phase-shifting technique since the errors propagate. A more comprehensive analysis of this technique and similar techniques can be found in [41].

⁹A path-sharing based phase shifting approach is more appropriate in an LO based beamforming architecture. The signal distribution network progressively splits the signal at every junction. At each split, an amplifier is used to compensate for the loss. Unlike, the RF signal, the LO signal is a lot less sensitive to the gain mismatch at each split especially since it does not require very high linearity.

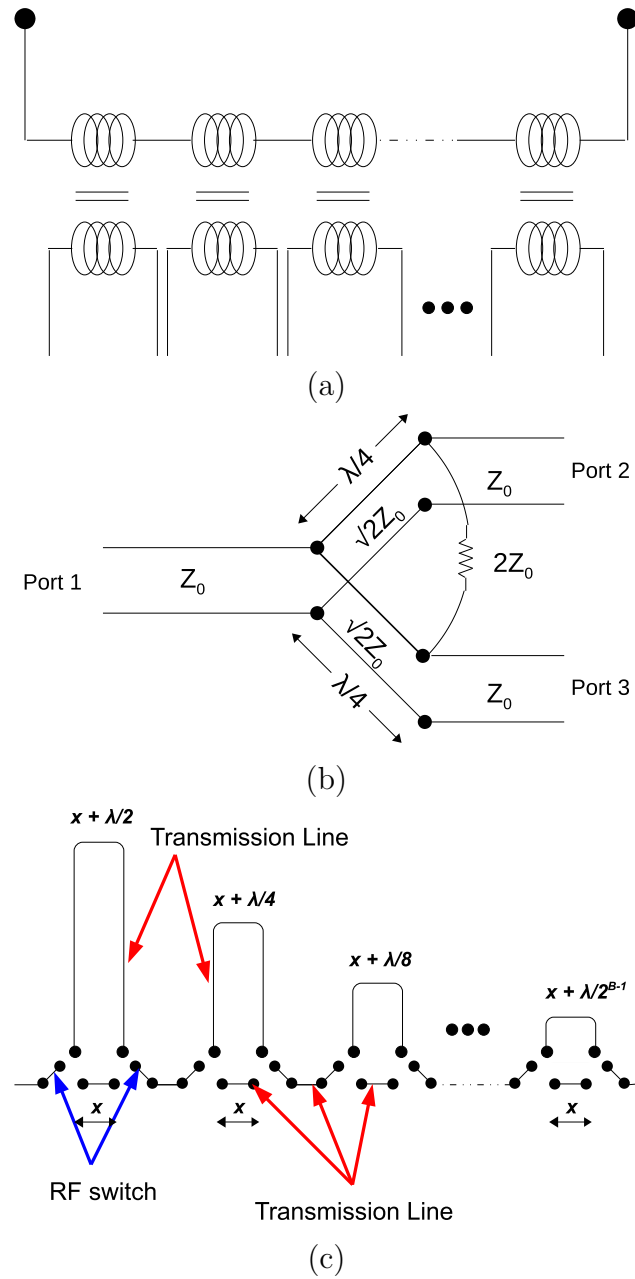
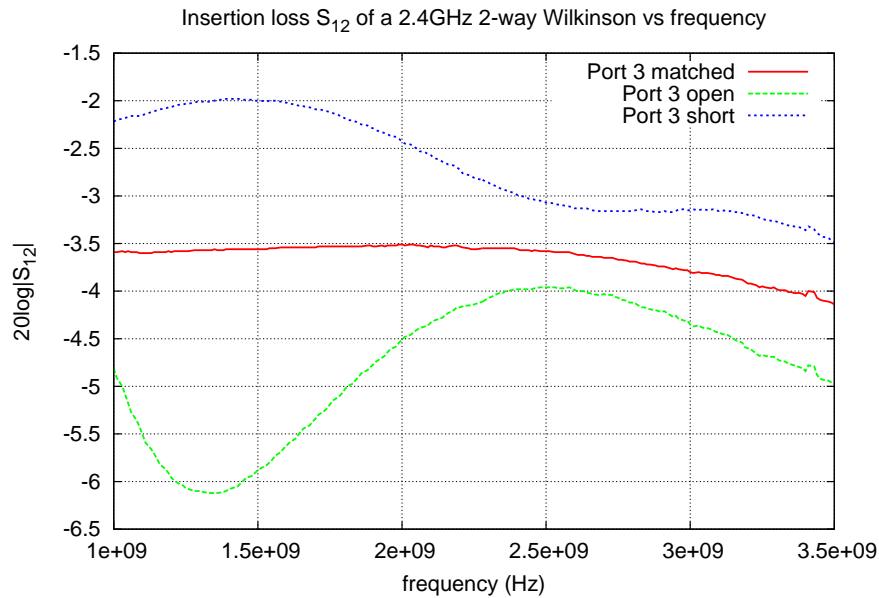
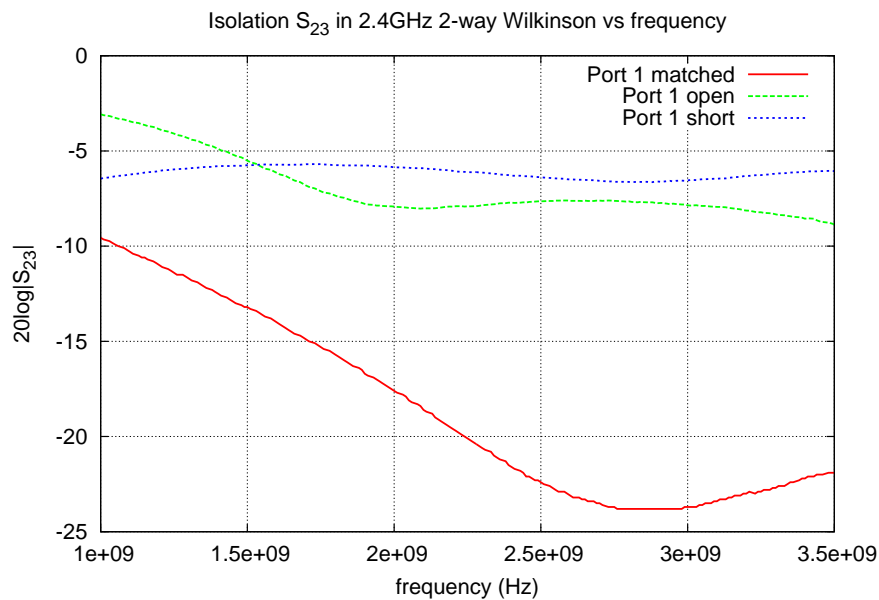


Figure 3.7: Power combining and delay structures: (a) Transformer-based power combiner. (b) Wilkinson power combiner. (c) Discrete programmable delay line implemented with switched transmission lines.



(a)



(b)

Figure 3.8: Measured performance of a 2-way Wilkinson power combiner/splitter (printed on an FR4 PCB and designed for 2.4GHz): (a) Insertion loss ($S_{12} = S_{21}$) in dB (the ideal loss is 3dB resulting from a 2-way power division). (b) Isolation ($S_{23} = S_{32}$) in dB between the split ports (the ideal isolation is $-\infty$). In both graphs, the third port was terminated with 50Ω .

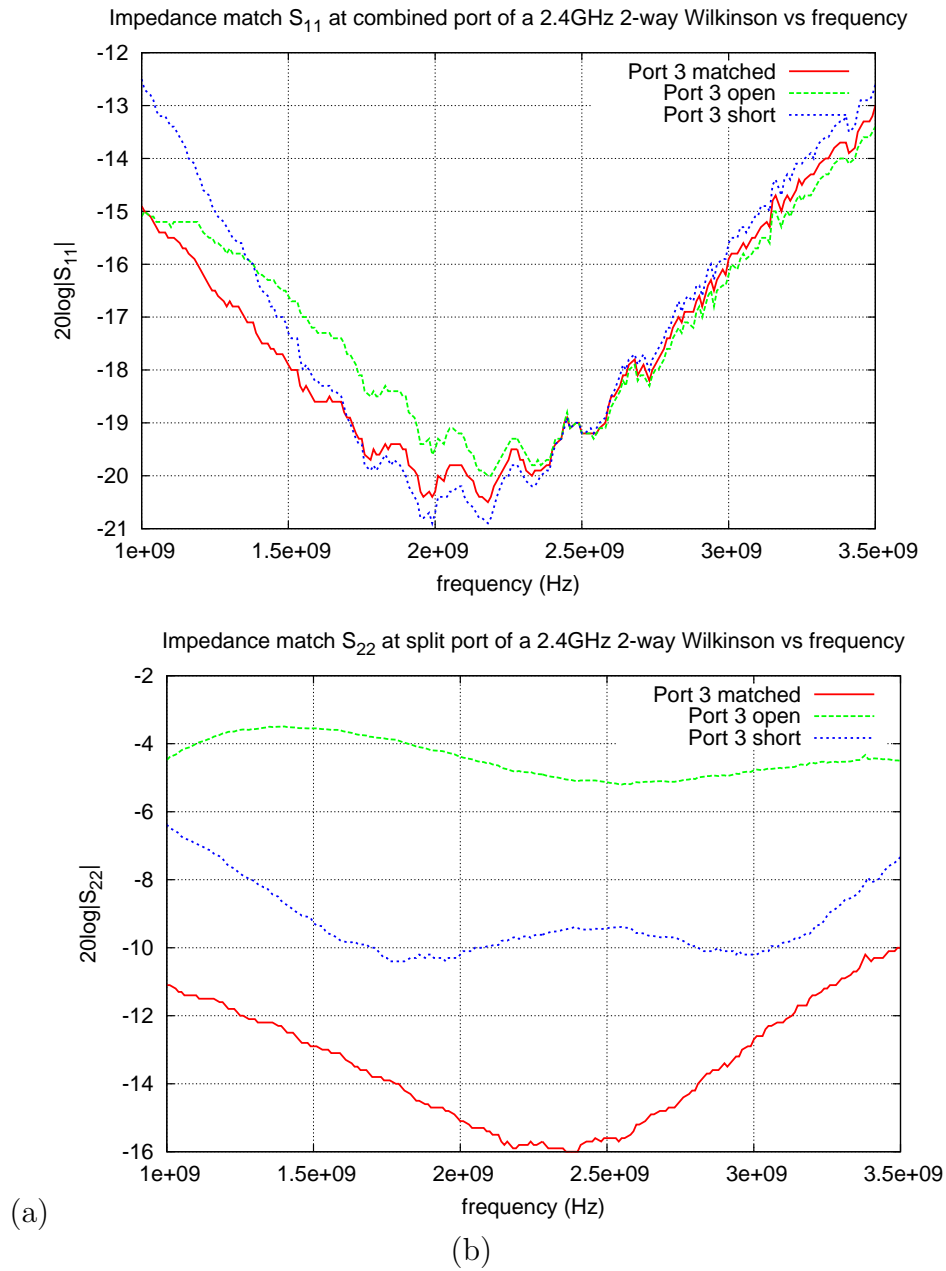


Figure 3.9: Measured performance of a 2-way Wilkinson power combiner/splitter (printed on an FR4 PCB and designed for 2.4GHz): (a) Input return loss S_{11} (impedance match) at Port₁ (combined port). (b) Input return loss $S_{22} = S_{33}$ (impedance match) at Port₂/Port₃ (split port).

3.10 or with transmission lines, and two reflective loads¹⁰. The reflective loads are usually resonant circuits consisting of inductors and varactors. The amount of phase shift at the output depends on the reflection coefficient $\rho = (Z_L - Z_0)/(Z_L + Z_0)$. The phase response is controlled by changing the load impedance. By controlling the voltage across the varactor, the load impedance can be changed from capacitive to inductive. The range of the phase-shift depends on the range of the varactor. Since the load impedance is purely imaginary, the theoretical magnitude of the reflection is always $|\rho| = 1$. Therefore, the amplitude response remains constant, regardless of the phase response. Also, since the phase shifter is composed of only passive components its both highly linear and symmetric¹¹.

Integrated circuit implementations of this phase-shifter have been demonstrated at ISM bands [63]. However, even a lumped implementation occupies a significant area (usually dominated by inductors), which is a disadvantage compared to a vector combining phase rotator described below. Also, a variable gain block after the phase shifter is required for tighter control over the shape beam pattern.

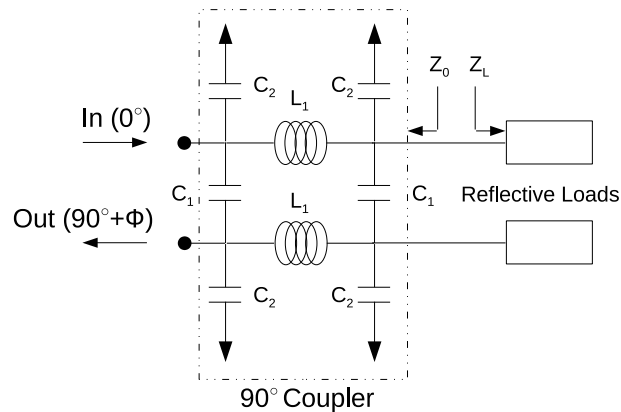


Figure 3.10: Reflective-type phase shifter using a hybrid 90° coupler implemented with lumped circuit elements.

- 3) Programmable delay lines: in wideband applications, the phase-shift approximation no longer holds and a true time delay is required. A simple and straight forward way to implement a programmable delay is using switched transmission lines as shown in Figure 3.7. Under this structure, the length of the signal path is controlled through the RF switches. If the transmission lines are low loss, then the gain variation across the different delay settings is very small. Since the transmission lines are passive, they can be used for both transmit and receive¹². A discrete 2-bit switched transmission line

¹⁰The lumped circuit implementation is more compact than the transmission line implementation, and thus more appropriate for lower frequency bands. The values of the lumped circuit elements depend on the center frequency f_c and the characteristic impedance Z_0 [63].

¹¹The varactor is usually an active component and thus exhibits some non-linearity.

¹²This is true even if the system uses frequency division duplexing or FDD as long as the wireless channel is a line of sight, since the broadband delay covers both transmit and receive frequency bands.

delay implementation on an FR4 PCB for 2.4GHz is shown in Figure 3.11. The main disadvantage of this technique is that it is very difficult to integrate except at very high frequencies (e.g. mm-wave frequencies and above). Even for a discrete implementation, every additional control bit requires a pair of RF switches in addition to a transmission line section. This increases both the cost and the footprint.

A synthesized transmission line is a more compact implementation of a broadband programmable delay and is more suitable for integration [3]. Unlike switched transmission lines that use distributed components, synthesized transmission lines can be implemented with lumped circuit elements that are a lot easier to integrate at low frequencies. We note, however, that when the array gets very large, the relative delays between array elements on opposite sides of the array are usually much larger than the maximum delays provided by the programmable delay lines whether they are implemented with switched or synthesized transmission lines¹³. In this case, the only solution is to divide the array into smaller subarrays such that the maximum delay within each subarray is small relative to the symbol duration. Larger delays are better handled in the digital domain. Also, this class of controlled delays or phase-shifters provide no amplitude control. Amplitude control will require adding a variable gain block (Section 3.2.4).

- 4) Vector combining: a fundamentally different technique from the ones listed above in that it enables simultaneous (in-direct) control of both the signal phase and amplitude by independently controlling the amplitudes of the real and imaginary components of the signal. The basic idea behind this approach is that if we can split the signal into two orthogonal components that are 90° apart in phase (e.g. sine and cosine), then we can generate any amplitude/phase combination by controlling the amplitudes of these two components. The underlying mathematical equations relating the different quantities:

$$A \cos(\omega t) + B \sin(\omega t) \equiv C \cos(\omega t + \varphi)$$

$$A = C \cos(\varphi), B = C \sin(\varphi), C = \sqrt{A^2 + B^2}, \varphi = \arctan \frac{B}{A}$$

The equations show how the amplitude C and phase φ are generated as a function of the real Cartesian coefficients A and B . Figure 3.12 shows the general structure of a vector combining based complex multiplier. The figure highlights two basic functions required by this architecture, which will be discussed in detail in Section 3.2.4: I/Q splitting to obtain two orthogonal components (e.g. real and imaginary) of the RF signal, and amplitude control achieved by a variable (programmable) gain block at each path in addition to the digital/analog control circuits for programming the gain. Vector combining has several advantages that make it the most appropriate choice for an integrated solution for a wide range of frequency bands. First, it provides reasonably accurate control over the phase and amplitude of the signal. Second, it can be implemented with active circuit elements (e.g. transistors) that use very little area compared to transmission lines or lumped circuit elements like inductors and transformers. This feature is very useful for a multichannel implementation that requires

¹³These delays can span multiple symbol intervals and thus result in intersymbol interference (ISI).

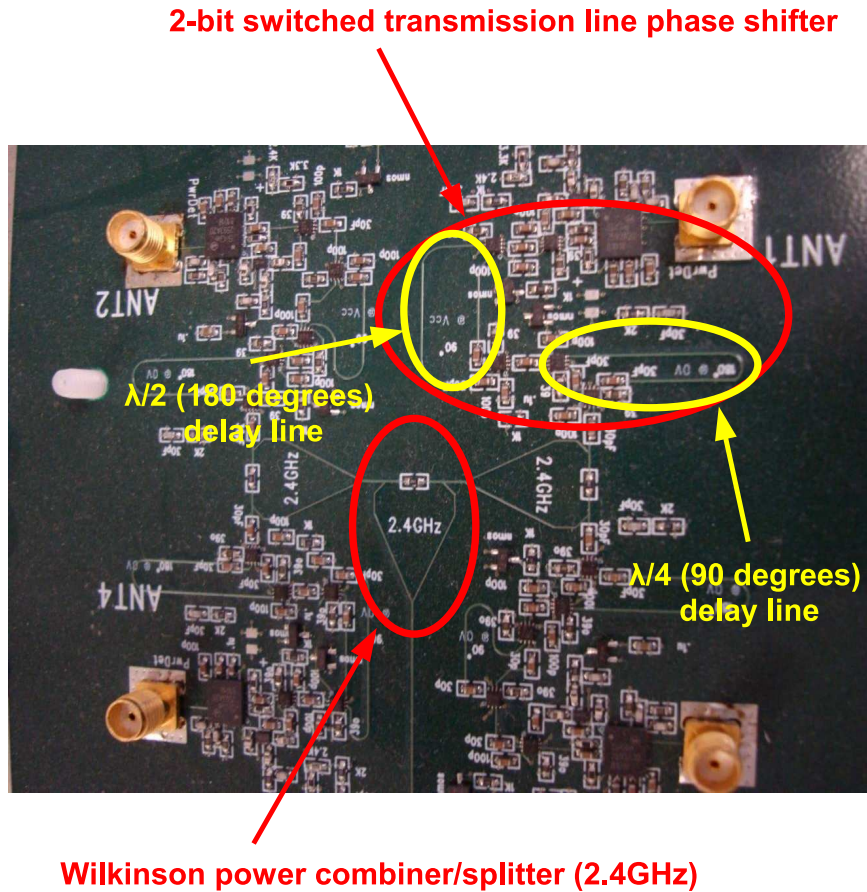


Figure 3.11: A 2-bit switched transmission line phase shifter composed of $\lambda/2$ and $\lambda/4$ transmission line sections and RF switches for controlling the delay path.

a large number of complex multipliers (Figure 3.2b). Finally, an active/differential implementation of the vector combiner can significantly reduce coupling and improve reverse isolation and enable the signals to be combined and split in the current domain.

Integrated circuit implementations in CMOS of this scheme have been demonstrated at both lower bands and mm-wave frequencies [43, 5]. While an active implementation is good for reducing area and improving isolation, the price is paid in higher power consumption, lower linearity, and reduced transmit/receive symmetry (one set of programmable complex gains is required for the transmitter and another for the receiver)¹⁴.

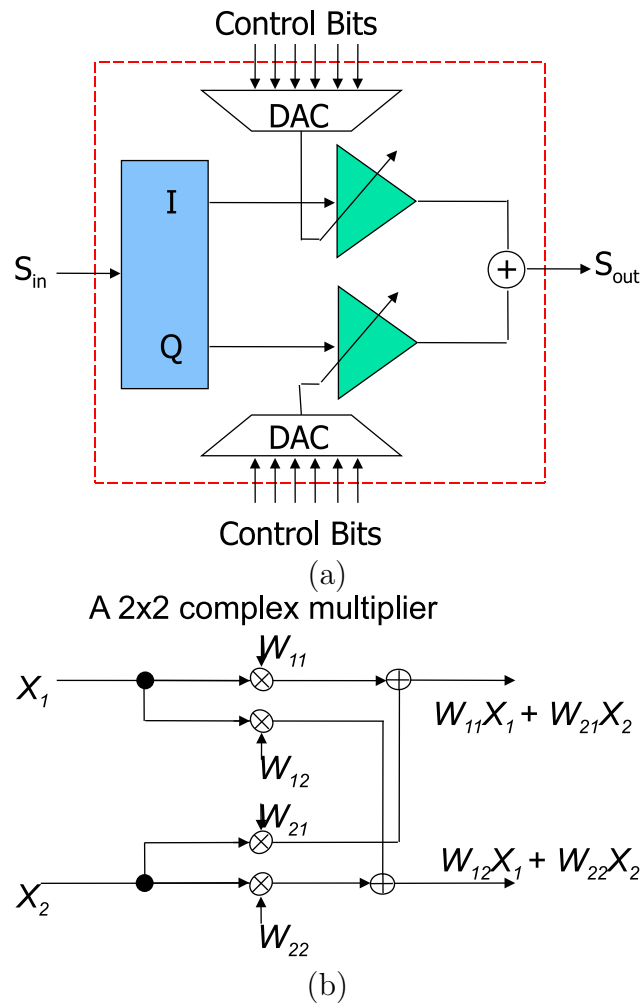


Figure 3.12: (a) A digitally controlled complex multiplier (programmable complex gain) in the RF domain. (b) A 2×2 complex multiplier matrix in the RF domain.

¹⁴The MOS transistors can be used in triode mode, in which case they behave like passive resistors. This improves linearity, symmetry and power consumption. However, isolation will be a challenge.

3.2.4 Vector combining complex multiplier implementation

In this section, we discuss the two main components of a vector combining architecture, namely the I/Q signal generation and the variable gain block.

I/Q signal generation

A Cartesian vector combining based phase rotator requires the signal be split into two orthogonal components that are 90 degrees apart in phase, usually referred to as the in-phase (I) and quadrature (Q) (or real and imaginary) components of the signal¹⁵. There are several ways of obtaining the I/Q components in the RF domain:

- 1) Polyphase filters: a polyphase filter is a popular technique for splitting an RF signal into two orthogonal components that are 90 degrees apart in phase. Originally designed for image reject mixers, a polyphase filter is a differential passive network of resistors and capacitors¹⁶ shown in Figure 3.13. The outputs at the I and Q ports are 90 degrees apart at all frequencies¹⁷. However, the bigger challenge is to get the output amplitudes matched since the I path has low-pass characteristic, while the Q path is high-pass. Therefore, the amplitude match occurs at a single frequency $f_c = (2\pi RC)^{-1}$. There are several ways to achieve a wideband amplitude match. First, we can cascade multiple stages of the polyphase filter. If resistor/capacitor values $R_1/C_1, \dots, R_K/C_K$ for K stages are such that their geometric mean is the desired center frequency ($f_c = \prod_{i=1}^K (2\pi R_i C_i)^{-1/K}$), then a wideband response can be achieved. In general, each stage provides an approximately constant gain (within .2dB) within 10% of f_c . The amplitude response of a 4-stage polyphase filter is shown in Figure 3.14. Second, if the system needs to support a narrow band signal over a wide range of frequencies, then the gain mismatch can be compensated for using a variable gain block (described in the next section) or by using variable components in the polyphase filter itself. For example, if the resistors (or the capacitors) in Figure 3.13 are replaced by variable resistors (or capacitors)¹⁸, then the center frequency can be varied by adjusting the value of these resistors.

The RC polyphase network has several other advantages:

- It's very compact since it does not contain any inductors, transformers, or distributed elements. Furthermore, the same polyphase network can be shared among multiple channels¹⁹.

¹⁵Orthogonality is not a necessary condition. As long as the components are linearly independent, then any phase between 0 and 2π can be synthesized in theory by choosing the appropriate coefficients. However, the more orthogonal the components, the more accurate the results are in practice.

¹⁶A single ended implementation also exists.

¹⁷While the actual phase for both I and Q channels varies over frequency, the phase difference is always 90 degrees.

¹⁸This can be implemented with transistors in triode mode.

¹⁹The order of the I/Q block and the variable gain block in Figure 3.12 can be interchanged without affecting the output. Thus, the variable gain block from different paths can be combined before the polyphase filter

- Since the polyphase network is passive, it has a very high linearity and burns no DC power and is bidirectional. That means that the same network can be shared between the transmitter and the receiver, reducing the transmit/receive mismatch.
- Both the inputs and outputs are in the RF domain. Therefore, it can be used in a cascadable architecture to build larger arrays.

The polyphase network has some disadvantages as well. First, it is very lossy (every stage results in roughly a 6dB loss). Second, it requires a very careful layout to insure symmetry and minimize the mismatch resulting from parasitics and component variation due to temperature and process.

A more comprehensive treatment of polyphase networks can be found in [28, 13].

- 2) Mixers: another way to produce the I/Q components of the signal is at the output of the mixer stage. The advantage of this method is that I/Q components are produced at the LO stage. Although this I/Q generation technique still requires a polyphase filter, a lot the drawbacks of the polyphase such as loss and amplitude mismatch are avoided since they only occur at the LO path²⁰. Therefore, there is little or no amplitude mismatch between the I and Q channels at the output of the mixer (except for possible mismatch in the mixer circuit itself as a result of process variation). This scheme, however, has several disadvantages. First, since signal combining occurs post mixing, the mixers must have high linearity²¹. Second, the output (combined) signal is no longer in RF domain. That means that the blocks cannot be cascaded to assemble a larger array²². Finally, this approach is only appropriate in a heterodyne radio architecture [43].

Variable gain blocks

The design of variable gain blocks²³ for beamformers and phase-rotating slightly differs from conventional VGA design used in automatic gain control circuits (AGCs) in several ways:

- 1) The phase response of the variable gain block has to be independent of the gain setting especially at the frequencies of interest.
- 2) Both the input and output impedances of the variable gain block must also be independent of the gain setting. A change in the input and output impedances usually impacts the isolation between different channels and results in phase variation. A simple and straight forward way to isolate the input and output impedances is by adding buffer

²⁰Since the LO is very narrow band, a single polyphase stage is sufficient. Any amplitude mismatch can easily be corrected with amplifiers without linearity being an issue since the LO is a “single tone” signal.

²¹A passive mixer with very high linearity can be used. However, passive mixers are very lossy and thus negating some of the advantages of this approach.

²²An intermediate block or chipset can be designed to take the output signals from each outer RF chip at an intermediate frequency (IF) and add them in the IF domain. However, using this technique requires that the LOs feeding the mixers in each of these RF chips to be tightly synchronized.

²³We use the term variable gain block to refer to both amplifiers and attenuators. Attenuators may be more desirable in some applications due to their high linearity and low power consumption [22].

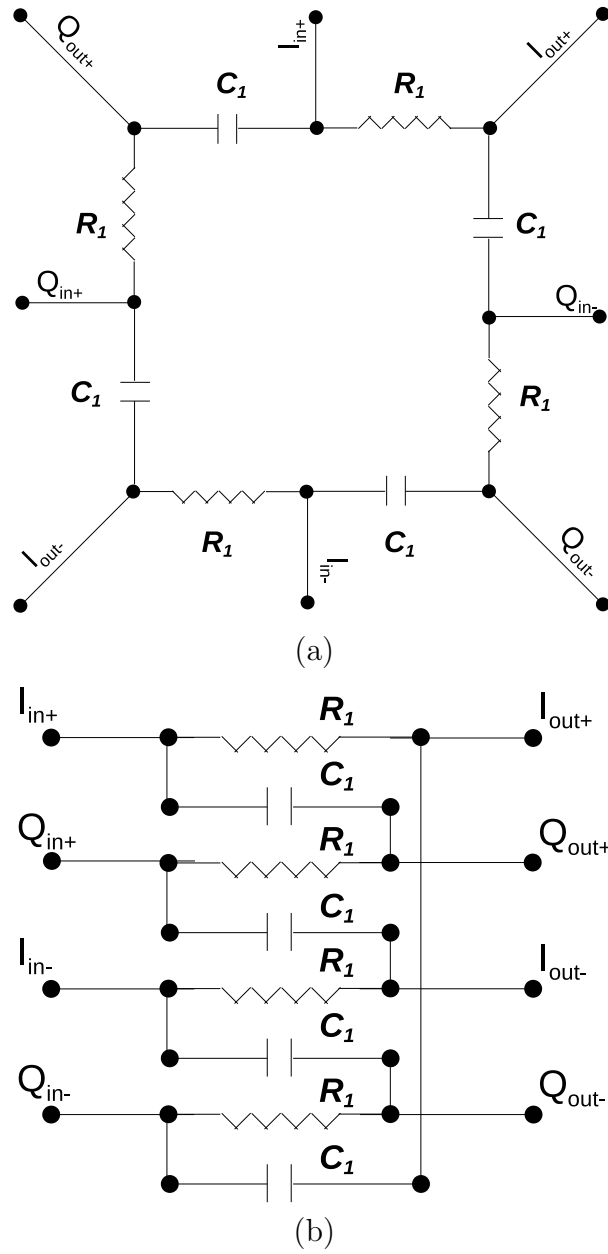


Figure 3.13: Two equivalent ways of drawing a polyphase filter.

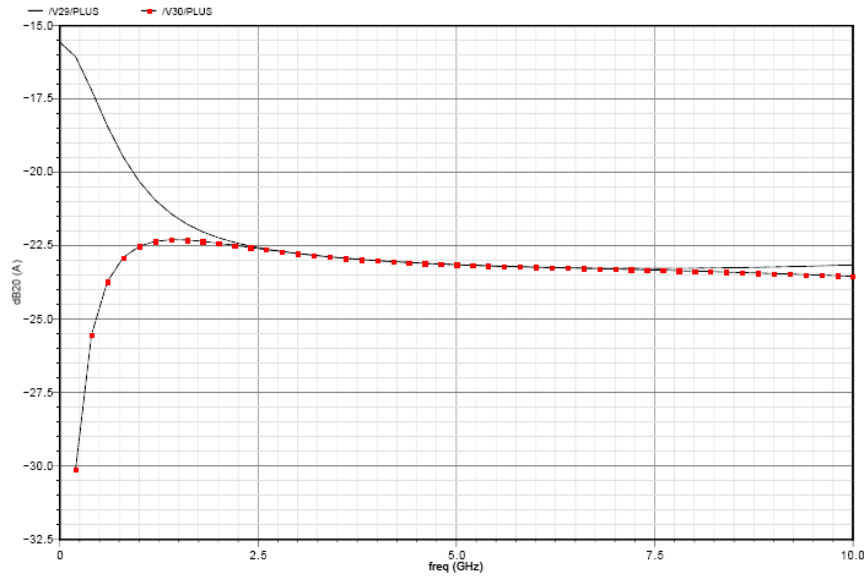


Figure 3.14: A wideband 4-stage polyphase network with very good amplitude match.

stages at the input and the output. However, reducing the number of buffer stages is usually more desirable since they affect both the power consumption and the linearity.

- 3) The gain steps must be uniform and precise.
- 4) Linearity of the variable gain block is critical since most in-band and out of band jammers will not be filtered except after the variable gain stage. Furthermore, better linearity will also improve the accuracy of the gain steps since they will be less dependent on the amplitude of the input signal.
- 5) The design of the variable gain block should be robust to temperature and process variations. Unlike a single antenna system, mismatch between the different channels can degrade the performance. Therefore, these these variations need to be minimized or dynamically calibrated. In general, variation and mismatch can be reduced with a good layout, and can be calibrated with a combination of analog and digital techniques.

The linearity and precision are usually improved by using amplifiers with discrete (as opposed to continuous) control steps. For example, many variable gain amplifiers use transistors with a variable (continuous) bias voltage. In this case, the linearity will be very sensitive to the gain setting, and the gain steps are going to be more sensitive to temperature and process variations. A discrete implementation replaces a single variable-bias transistor with bank of switch/fixed-bias transistors as shown in Figure 3.15. Each of the transistors in the bank is hard switched (i.e either set to a fixed bias voltage V_{bias} or turned off completely). The gain is controlled by changing the effective size of the virtual transistor represented by the bank (by turning on the right number of transistors). The gain steps will depend on the relative sizes of these small transistors. In order to achieve uniform gain steps and reduce the variation, this concept can be further extended by designing a single unit-gain cell and

replicating it as shown in Figure 3.16(a). In this figure, each gain block contains twice the number of unit-gain cells as the preceding block. Each block is differential and is controlled by a single bit, which controls the sign of the output. The overall gain is controlled by controlling the signs of the sub-blocks. There have been several reported implementations that take advantage of this concept [61, 34]. The unit cell in these architectures, shown in Figure 3.16(b), is composed of two differential pairs that are identical except for a sign flip. Only one of these pairs is on at any given time, and the circuit is controlled by differential logic. When control bit $B = 1$, then the current source M_1 is turned on (i.e. set to the correct bias voltage), and M_2 is turned off. In this case, the differential pair M_3 and M_4 is turned on, and M_5 and M_6 are off, and vice versa when $B = 0$. This architecture has several advantages. First, the bias current remains constant and independent of the gain setting. This helps reduce the variation of the circuit properties as a function of control signal. Second, the input and output impedances of the variable gain block are also independent of the gain configuration. That means better isolation between channels and little or no phase variation. This also allows the signal splitting and combining to occur in a single stage in the current domain without requiring additional input and output buffers. Since the architecture uses a unit-gain cell, mismatch is also reduced.

In many respects, the design and layout of variable gain blocks for phase shifting are very similar to the design of digital to analog converters (DACs).

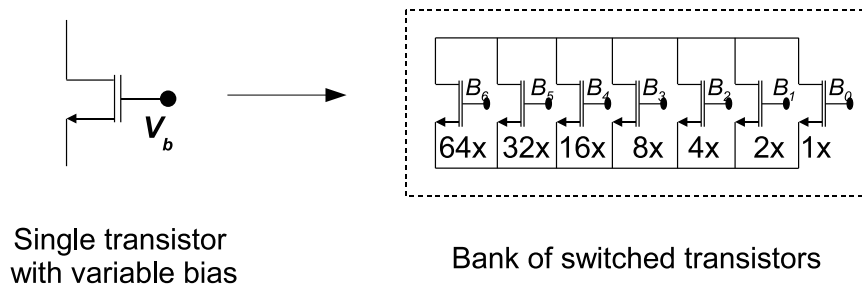


Figure 3.15: Replacing a single variable-bias transistor with a bank of digitally controlled transistors. Each of the transistors in the bank are either On or Off (hard switched). By turning different transistors On and Off, the effective size of the transistor is varied.

3.3 Conclusion

In this chapter, we considered different beamforming architectures and examined their impact on the cost, complexity and power requirements of an adaptive array system. We proposed a hybrid (hierarchical) RF/digital architecture that combines the scalability and power efficiency of RF beamforming with the speed and flexibility of digital beamforming. The relative sizes of the RF and digital components of the system are determined by the capacity and gain requirements. We also proposed a multi-input/multi-output RF IC architecture that integrates multiple channels on a single chip and can be cascaded to support large array with more channels. We examined different techniques and circuits for implementing

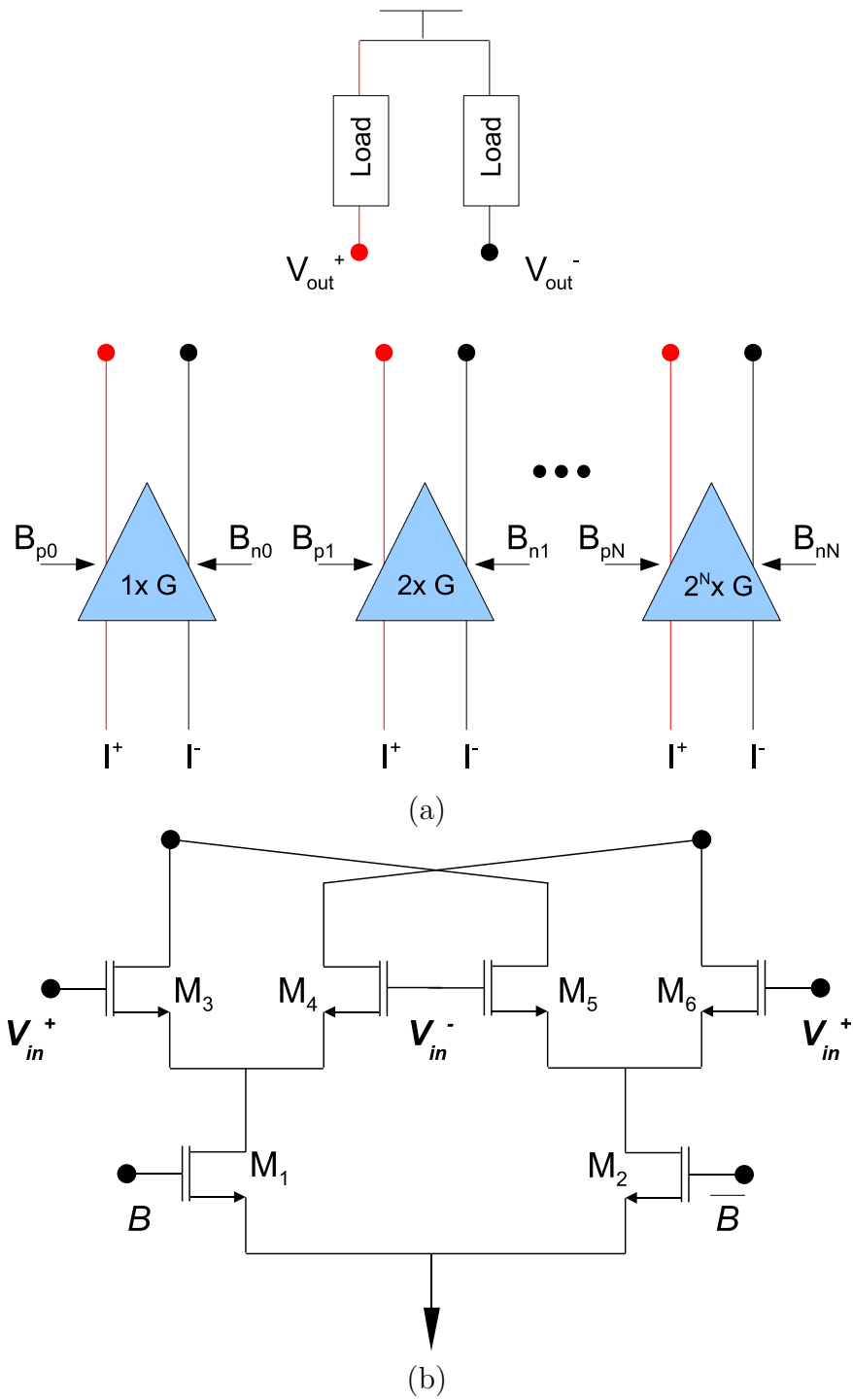


Figure 3.16: (a) A programmable gain block using a unit-gain cell architecture. (b) Unit cell topology.

phase shifting and signal combining/splitting. We considered a programmable complex gain block implementation using a vector combining technique with digitally-controlled variable gain blocks. We showed that this can be a promising architecture for meeting the desired requirements of low footprint and accurate phase and amplitude control. Using this technique, signal combining and splitting can be implemented efficiently in the current domain in a single stage. The adaptive signal processing (beamforming) techniques for estimating and optimizing both the RF and digital beamforming weights are presented in Chapter 5.

Chapter 4

Antenna designs

Antenna design is very critical for beamforming systems. Key performance metrics of a beamforming system like range, coverage, and capacity depend very strongly on the antenna properties discussed in Chapter 2. For example, the range can be extended with higher antenna array directivity, and the coverage (scan angle) can be extended by using arrays with elements that have wide beams. Capacity is influenced by the antenna in two different ways. First, narrower beams reduce interference and result in better spatial reuse. Second, if the array has polarization diversity (i.e. contains antennas with orthogonal polarization states), then this property can also be used to scale capacity even further (by transmitting different data streams on different polarizations) [21]¹. Ideally, we would like these quantities to be as large as possible. However, the price usually paid is a bigger aperture with a large number of elements. The size and form factor are major contributors to the overall cost of the system. In Chapter 3, we discussed techniques for designing the radio circuit component of the beamforming system that can reduce the cost and power consumption of the main circuit blocks by taking advantage of low cost silicon integration. This chapter focuses on antenna design, with the goal of developing new antenna array design techniques that meet the system requirements (gain, capacity, coverage) with compact form factors that are easy to build and manufacture at low costs.

The chapter consists of two parts. In the first part, we examine the impact of different design choices (e.g. the structure of the individual antennas and their arrangement within the array) on the properties and performance of the array (e.g. gain and radiation patterns). In the second part, we introduce new 3-dimensional array structures and geometries that achieve the desired results in practice with smaller and more compact form factors compared

The work presented in this chapter was done in collaboration with Ehsan Adabi, Kevin Jones, Jundong Park and Ali Niknejad [9].

¹Using polarization for diversity or multiplexing is most beneficial in environments that have little or no spatial diversity (i.e. environments with little scattering like pure line of sight). In environments with rich scattering, there is usually enough spatial diversity to enable independent streams to be transmitted on different antennas with the same polarization. In fact, having different polarizations in a rich scattering environment may not add a significant advantage since the scattering and reflections usually transform the polarization randomly.

with conventional planar antennas. We also present examples that are easy to build and manufacture using low cost printed circuit board (PCB) technology. In general, reducing the size of the antenna is more challenging than reducing the size of the radio circuit because, unlike the radio subsystem, the size of the antenna is limited by the laws of electromagnetics and is largely unaffected by technology.

4.1 Array factor and radiation pattern calculations

As discussed in Section 2.1.2, the gain (or directivity) of an antenna is inversely proportional to the beamwidth measured in solid angle units². This gain is also proportional to the effective area of the antenna. This rule applied to antenna arrays as well.

In Chapter 2, we considered a simple example of a uniform linear array. In this chapter, we generalize the analysis to arrays of arbitrary geometries. To simplify the analysis, we factor the overall radiation pattern into two components: the array factor and the element factor. The array factor (AF) is the pattern of the array when all the antennas are replaced with isotropic antennas³. The element factor (EF) is the radiation pattern of a single antenna element in the array⁴.

The overall theoretical gain and beam pattern of an array of isotropic antennas depends on three factors: the number of antennas, the spacing between the antennas, and the arrangement of the antennas. In this section, we examine the impact of each of these factors on the properties of the beam pattern such as beamwidth, side lobes and grating lobes. Although, these different factors are interrelated and it is difficult to quantify each individually, we can still provide some general guidelines on how to take advantage of them in designing the array.

4.1.1 Impact of antenna spacing

Let N be the number of antennas that are arranged in a linear array (along the y-axis) with equal spacing d between the antennas as shown in Figure 4.2. Let's also assume that the observation angle is along the x-axis (i.e. the array is configured to maximize reception from this angle). When the spacing between the antennas is fixed, then the overall gain of the array will be proportional to the number of antennas N (the width of the main lobe will be inversely proportional to N). On the other hand, if the spacing between the antennas d is increased while number of antennas is fixed, then the gain increases with d up to a point, and remains flat beyond that point. The break point is approximately $d \approx \lambda/2$ (λ is carrier wavelength) [12]. The beamwidth of the main lobe will always be inversely proportional to

²This is often approximated with the 3dB beamwidth (Ω_{3dB}) or 10dB beamwidth (Ω_{10dB}) or the null-to-null beamwidth (Ω_{NN}). The unit used for solid angles is often called steradians [60].

³Recall from Section 2.1.1 that isotropic antennas have beam patterns that are shaped like spheres (i.e. radiate equally in all directions (azimuth and elevation)) and are represented by dimensionless point sources. An isotropic antenna is an abstract concept and is not realizable in practice.

⁴We assume for simplicity that the antenna elements of the array are identical. However, this does not guarantee that the beam patterns will be identical. As discussed in Section 2.1.4, the properties of the antenna not only depend on the structure of the antenna itself, but also on the surroundings, especially those that are nearby. In Section 4.1.5, we show some simulation results that confirm this observation.

d . However, when $d > \lambda/2$, grating lobes start to appear. The grating lobes keep the overall directivity constant by compensating for the narrowing of the main lobe. These results are demonstrated in Figure 4.1. In Figure 4.1a, we simulated three uniform linear arrays with different number of antennas ($N = 4, 8, 16$ respectively) and fixed spacing ($d = \lambda/2$) along the y-axis with an angle of observation along the x-axis ($\phi = 0$). The figure shows that the width of the main lobe being reduced by a factor of 2 for every factor of 2 increase in the number of antennas. In Figure 4.1b, we simulated three uniform linear arrays with different inter-element spacing ($d = \lambda/2, \lambda, 2\lambda$) and a fixed number of antennas along the y-axis with an angle of observation along the x-axis ($\phi = 0$). The figure shows that the width of the main lobe being reduced by a factor of 2 for every factor of 2 increase in d . However, the number of peaks (grating lobes) doubles as well. We also included in Figure 4.1b the beam pattern of an 8-element linear array with $\lambda/4$ spacing in order to illustrate that an increase in spacing can improve the directivity especially when spacing is small to begin with (i.e. $d < \lambda/2$). Figure 4.1b, shows that the 8-element/ $\lambda/4$ -spaced and 4-element/ $\lambda/2$ -spaced linear arrays have almost identical beam patterns.

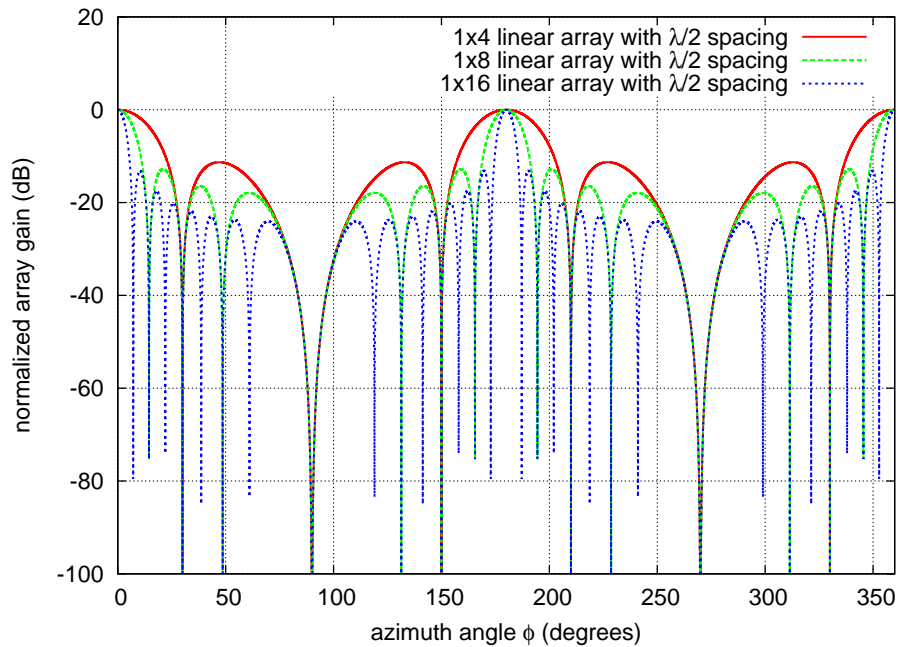
4.1.2 Impact of array geometry

A linear array is not a symmetric structure in the sense that it looks different from different angles of observations. For example, on the one hand, when the array is observed from an angle that is orthogonal to the axis of the array, the array appears as a line (i.e. the entire array is observed). On the other hand, when the array is observed from the same direction as the axis of the array, the array appears as a single point (i.e. only the first element is observed). The direction orthogonal to the axis of the array is known as the array broadside, and the direction along the axis of the array is known as the array endfire (see Figure 4.2). Therefore, it is natural to expect the radiation pattern to be different for different observation angles. For example, consider the two extreme cases (broadside/endfire) plotted in Figure 4.3. Figure 4.3a, shows the azimuth beamwidth much narrower in the broadside direction than the endfire direction, which may suggest that there is more gain in the broadside than the endfire. However, gain of the array cannot be based on the azimuth pattern alone; we have to factor in the elevation pattern as well⁵. Just by looking at the symmetry of the structure of a linear array, it is easy to see that it provides no elevation gain (i.e. no beam resolution) in the broadside direction⁶. Similarly, in the endfire direction, the elevation pattern will look identical to the azimuth pattern. Figure 4.3b shows the elevation patterns for both broadside and endfire. Since the overall solid beamwidth can be approximated with product of the azimuth beamwidth and the elevation beamwidth (i.e. $\Omega_{3dB} \approx \phi_{3dB}\theta_{3dB}$), then in order for the gain to be linear in the number of antennas N (in both cases), we must have the following:

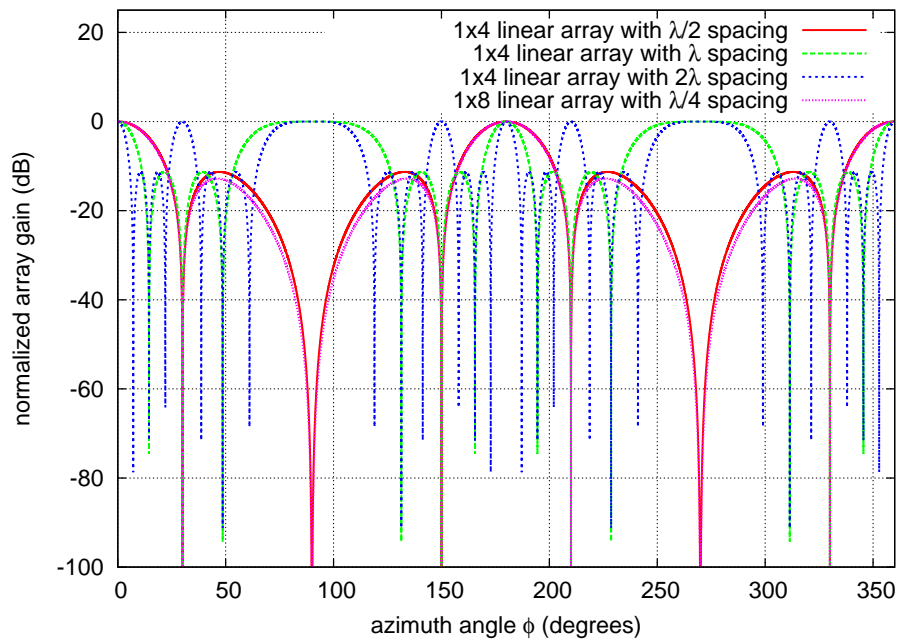
- $\phi_{3dB} \propto 1/N$ for broadside.
- $\phi_{3dB} \propto 1/\sqrt{N}$ for endfire.

⁵Recall that the gain is based on the solid angle beamwidth Ω .

⁶The array cannot distinguish signals arriving from different elevation angles in the broadside direction since they all arrive at the antenna elements with zero relative delay. In fact, the broadside pattern of a linear array is omni-directional (see Section 2.1.1).



(a)



(b)

Figure 4.1: Impact of number of antennas/antenna spacing on the array pattern. (a) The main lobe gets narrower as we increase the number of elements while keeping the spacing fixed. (b) Grating lobes are created when the antenna spacing is increased beyond $\lambda/2$. In both figures, we simulated a uniform linear array along the y-axis with an angle of arrival along the x-axis.

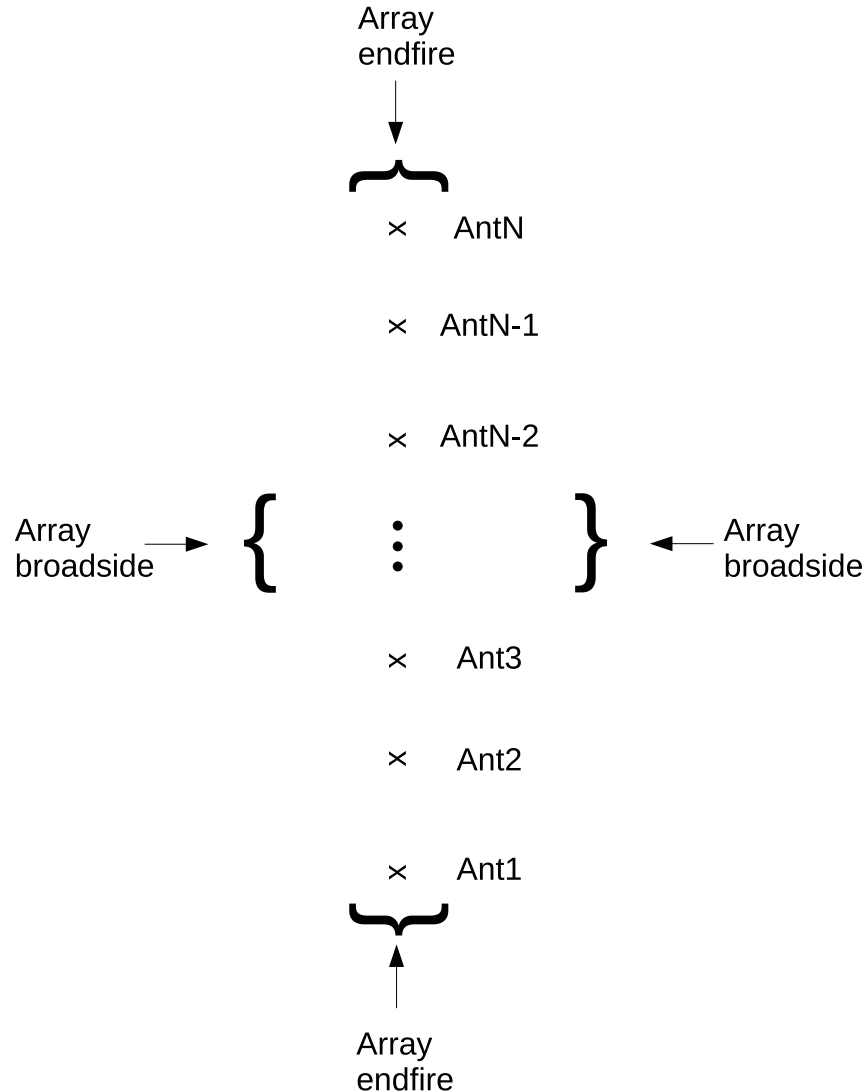
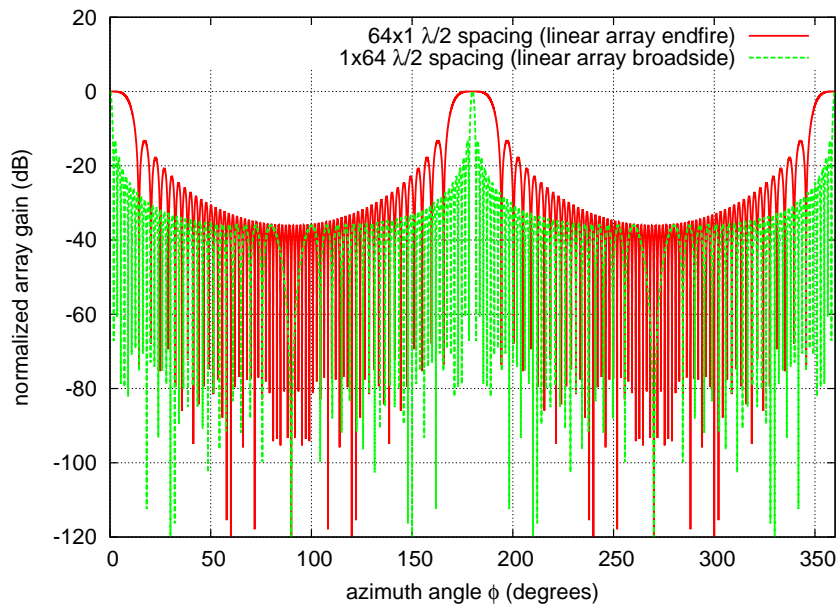


Figure 4.2: Broadside/endfire configurations of a linear array.

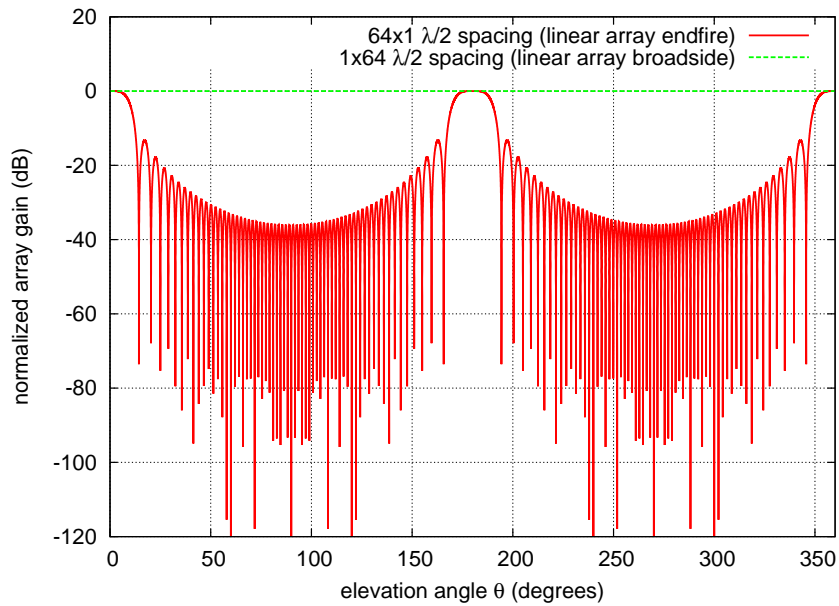
In order to verify this, we simulated the endfire and broadside azimuth patterns for $N = 10, 100, 1000$ in Figures 4.4a (endfire) and 4.4b (broadside) and Table 4.1.2. The results clearly show that $\phi_{3dB} \propto 1/N$ for broadside and $\phi_{3dB} \propto 1/\sqrt{N}$ for endfire. Both Figures 4.4ab show a 100-element endfire array and a 10-element broadside array having very similar beamwidths. In general, angles of arrival that fall in between these two extreme cases will behave somewhere in the middle. The azimuth beam pattern of a linear array for different steering vectors (observation angles) is shown in Figure 4.5.

4.1.3 Planar arrays

One of the main disadvantages of linear arrays is their inability to scan the beam in more than one dimension. For this reason, most practical applications that utilize large phased

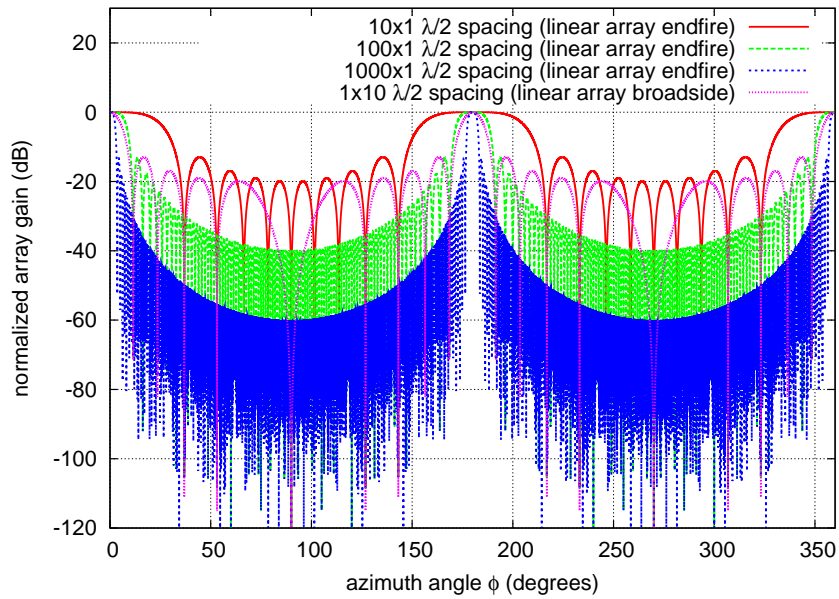


(a)

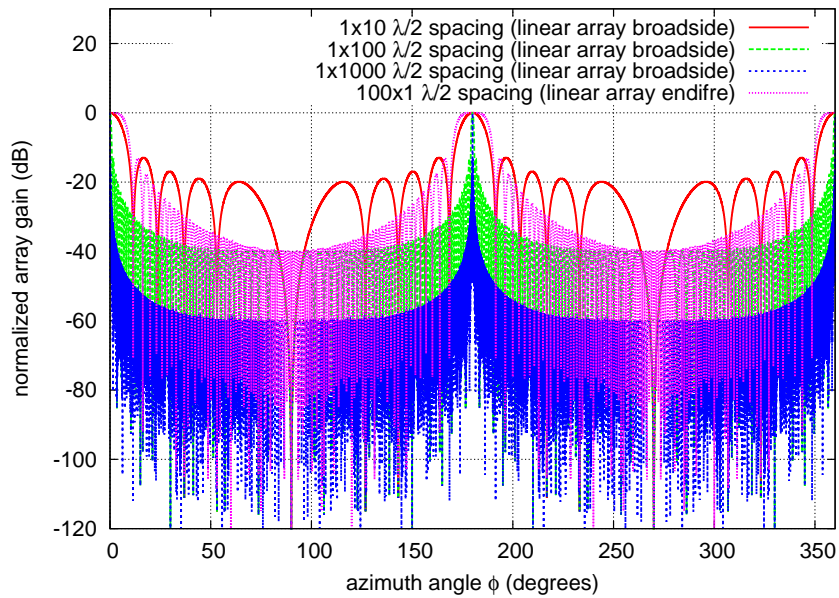


(b)

Figure 4.3: (c) Impact of the array geometry on the azimuth beamwidth (array steered towards $\phi = 0^\circ$ (azimuth) and $\theta = 90^\circ$ (elevation)). (b) Impact of the array geometry on the elevation beamwidth (array steered towards $\phi = 0^\circ$ (azimuth) and $\theta = 90^\circ$ (elevation)).



(a)



(b)

Figure 4.4: (a) Azimuth radiation pattern for linear endfire arrays with different number of antennas. (b) Azimuth radiation pattern for linear broadside arrays with different number of antennas. In both figures, the arrays are uniform with $\lambda/2$ spacing.

geometry	uniform broadside ($\lambda/2$)			uniform endfire ($\lambda/2$)		
	1×10	1×100	1×1000	10×1	100×1	1000×1
ϕ_{3dB}	10.02°	1°	0.1°	48.62°	15.24°	4.8°
ϕ_{10dB}	17.02°	1.7°	$.17^\circ$	63.2°	19.7°	6.24°
ϕ_{NN}	23.06°	2.3°	$.22^\circ$	73.8°	22.94°	7.24°

Table 4.1: Simulated azimuth beamwidths of uniform broadside and endfire linear arrays ($\lambda/2$ spacing).

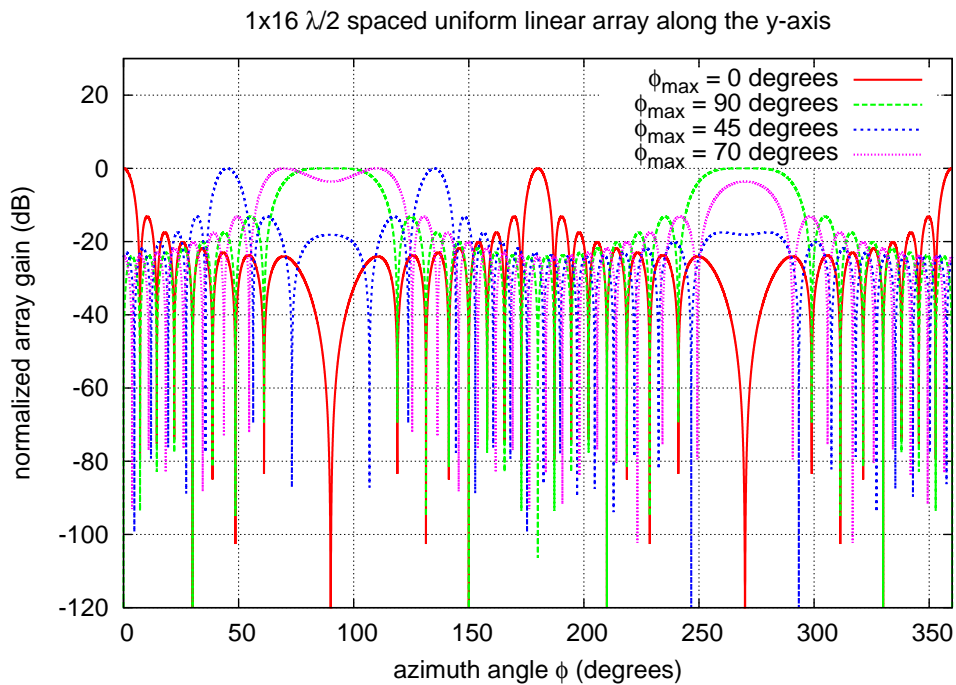


Figure 4.5: Beam pattern of a linear array for different steering vectors.

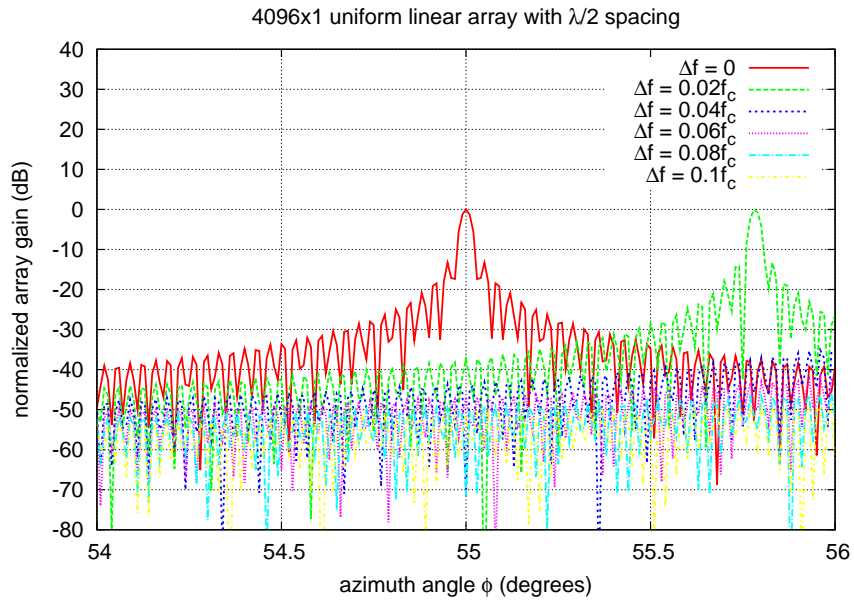
array antennas use arrays with planar geometries. In addition to the ability to scan in every direction in space, planar arrays have several other advantages⁷. They have more compact form factors since they pack antennas in two dimensions instead of just one. This allows more flexibility in choosing the geometry of the array in order to achieve certain beam properties. The compact form factor also improves the bandwidth of phased array systems. Recall that the incoming (or outgoing) signal arrives at the elements of the array with different delays. For narrow band (long symbol interval) signals, these delays can be approximated with phase shifts. However, when these delays become large, this approximation starts to break down. The delays become larger when the array becomes larger (especially between elements at opposite sides of the array). In planar arrays, the maximum amount time required to travel from one end of the array to another is going to be much smaller than a linear array of comparable gain. Figure 4.6 plots the array factor for different deviations from the center frequency for both a 4096-element linear array (Figure 4.6a) and a 4096-element square array (Figure 4.6b). The Figure shows that for the linear array, the gain drops well below 30dB off the peak when the frequency drifts by 2% from the center frequency. On the other hand, for the square array, the gain is still within 10dB of the peak when the frequency drifts by 2%.

Planar arrays do have a drawback, however. Unlike linear arrays, where the array factor is proportional to number of antennas for all observation angles, the array factor of a planar array is not uniform. In general, the gain of a planar array is larger in the broadside direction (direction orthogonal to the plane of the array) than in endfire directions (directions on the same plane as the array elements)⁸. In order to demonstrate this, we simulate the endfire azimuth pattern of an $M \times N$ rectangular array on the xy-plane (M is the dimension along the x-axis or the depth of the array, and N is the dimension along the y-axis) with an angle of arrival along the x-axis (i.e. $\phi = 0, \theta = \pi/2$). We fixed $N = 10$ and varied the depth of the array M from 1–400 to see the impact on the azimuth beam pattern. The plots are shown in Figure 4.7. The figure shows little improvement in the width of the main lobe as we increase the depth of the array from 1 to 400. In fact, the main lobe starts decreasing only when $M > 100$. The planar array is behaving like a linear endfire array of linear broadside arrays. Since the azimuth endfire beamwidth is proportional to $1/\sqrt{M}$ and broadside beamwidth is proportional to $1/N$, the broadside pattern dominates when $N > \sqrt{M}$, and the endfire pattern dominates when $M < \sqrt{M}$. Therefore, the azimuth endfire beamwidth ϕ_{3dB} of a rectangular array is proportional to $\min(1/N, 1/\sqrt{M})$ instead of $1/(N\sqrt{M})$ ⁹.

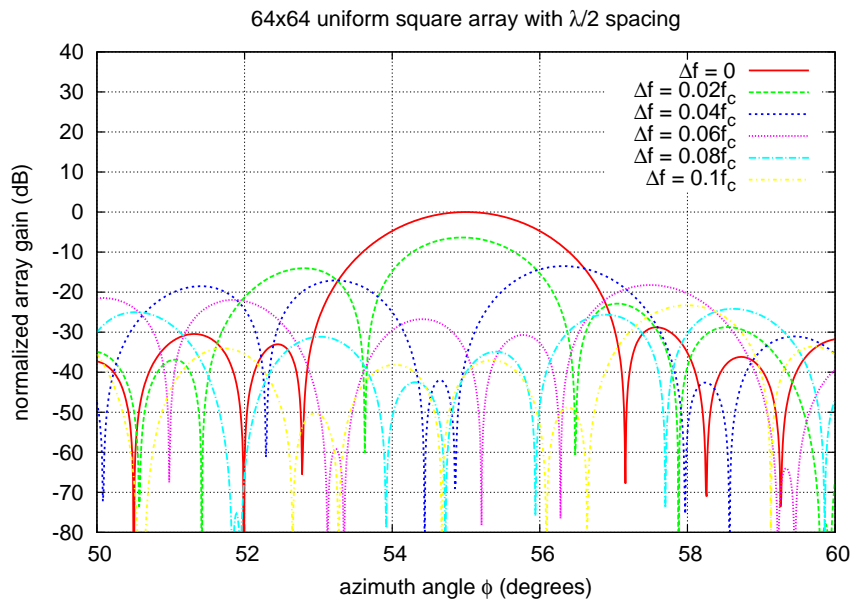
⁷In a real system, the scanning range will be limited by the antenna elements. However, in this section our focus is on the array factor.

⁸Recall that the gain (or directivity) of an antenna is proportional to its effective area. The effective area of a planar array is larger in the broadside direction than in the endfire direction.

⁹That means that total gain in the endfire direction will be proportional to $N\sqrt{M}$ (or $M\sqrt{N}$ depending on the direction of arrival). The extra factor of \sqrt{M} comes from the elevation gain. On the other hand, the overall broadside gain of a planar array is NM (N is the azimuth gain and M is the elevation gain) [12]. Therefore, the difference between the broadside and endfire gains is a factor of \sqrt{N} or \sqrt{M} . Although the plots in Figure 4.7 show that the side lobe levels decrease significantly as the depth of the array increases, the gain from side lobe reduction diminishes very quickly and cannot make up for the \sqrt{M} lost gain.



(a)



(b)

Figure 4.6: (a) Array factor of a 4096-element uniform linear array with $\lambda/2$ spacing for different deviations from the center frequency (f_c). (b) Array factor of a 4096-element uniform square array with $\lambda/2$ spacing for different deviations from the center frequency (f_c).

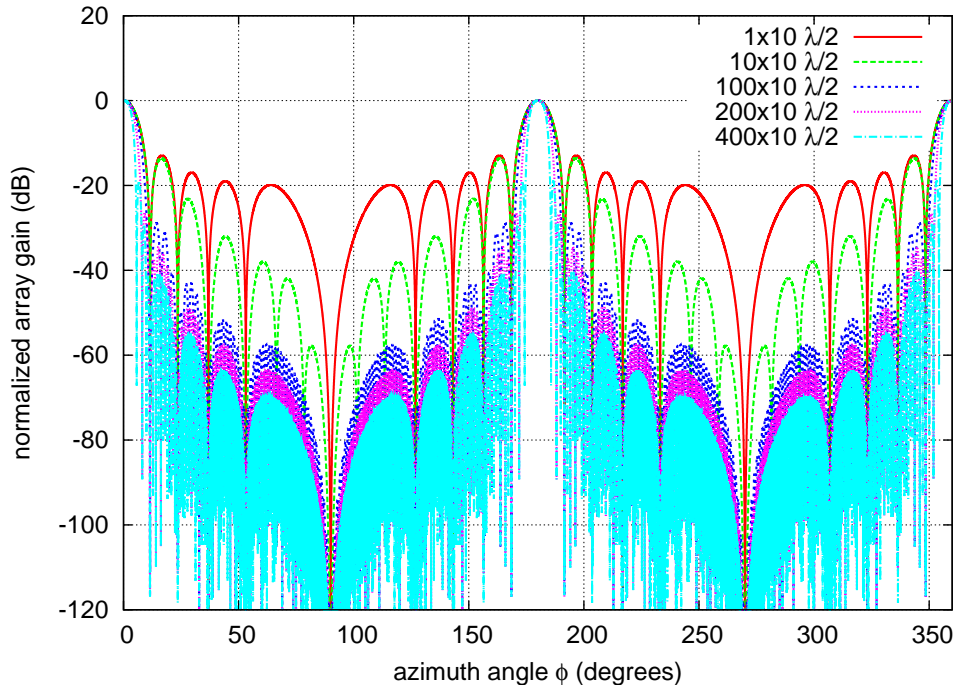


Figure 4.7: Azimuth endfire pattern of rectangular arrays. Different curves corresponds to different array depths M .

4.1.4 Side lobes

Any radiation outside the main lobe of the beam is usually undesirable because it creates unnecessary interference. In a well designed antenna, most of the radiation outside the main lobe comes from side lobes that are weaker than the main lobe¹⁰. However, for some applications, even the side lobe radiation levels are unacceptable. So far in our analysis, we have assumed that the antennas are excited with equal amplitudes. This will lead to the highest directivity because the entire array aperture is fully utilized [12]. Unfortunately, this pattern of excitation also leads to the higher side lobe levels. If we look back at pattern plots for linear arrays and planar arrays (e.g. Figures 4.3 and 4.4), we see that although the width of the first side lobe becomes narrower as we increase the size of the array, its amplitude relative to the main lobe remains constant (approximate -13dB below the main lobe) [12]. There are two ways to control the side lobes levels. The first way is by controlling the amplitude of the signal excitation. As we mentioned in Section 2.2, whereas controlling the relative phases of the signal excitations helps steer (direct) the beam, controlling the relative amplitudes of the signal excitations helps shape the beam. For a uniform linear array, choosing the relative amplitudes to reduce the side lobe levels is very similar to the well known problem of designing a digital/time domain finite impulse response (FIR) filter, and lot of the techniques for designing time domain filters apply to spatial domain filters

¹⁰As we have seen earlier, radiation can also come from grating lobes which are comparable to the main lobe. However, grating lobes are a lot easier to avoid than side lobes as will be shown in Section 4.2. Grating lobes can also be filtered out by the element pattern.

as well [42, 12]. The second way to reduce the side lobe levels, which is less obvious, is by properly designing the geometry of the array¹¹. To get an idea how the side lobe levels are affected by the array geometry, consider the pattern plots of a square array shown in Figure 4.8. The figure shows the pattern of a 10×10 square array antenna for two observation angle $\phi = 0, 45^\circ$. The figure clearly shows that when the beam is steered towards a 45° azimuth angle, then the side lobe levels are lower than when it is steered towards the x-axis. The reason is that the shape of the array looks different from a 45° angle. Section 4.2 discusses some techniques for designing the array geometry to minimize side lobes. Note that regardless of the technique used (e.g. controlling amplitude and/or geometry), the reduction in side lobe levels usually comes at the price of reduced directivity [12].

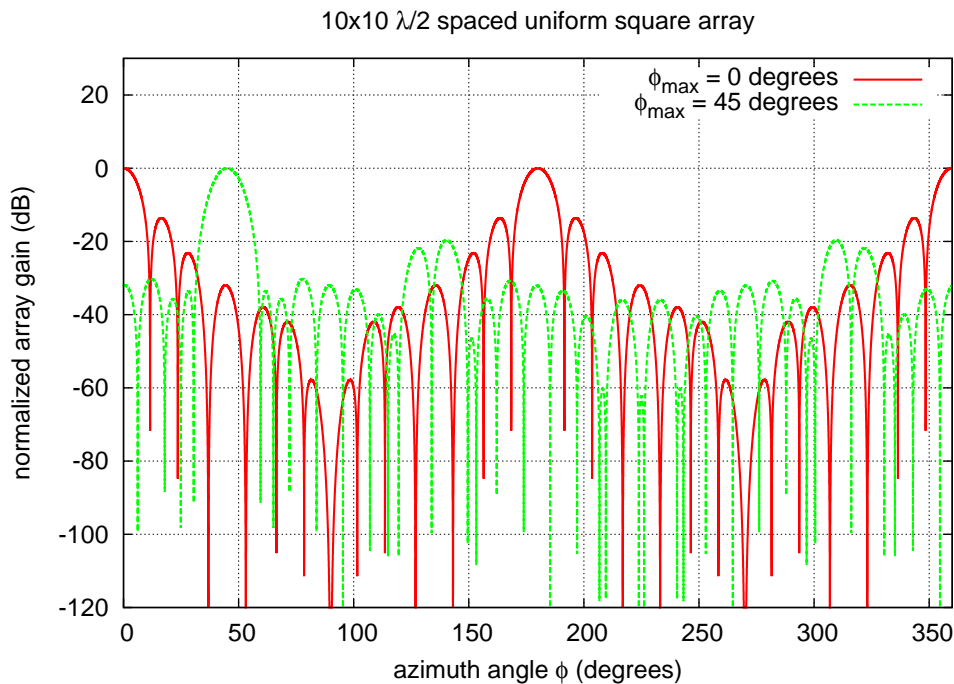


Figure 4.8: Side lobe level for different angles of arrival.

4.1.5 Antenna coupling

In the preceding sections, we focused our analysis on the array factor. In this section, we consider the impact of the array geometry on the element factor. The design and properties of the antenna element are critical because they determine the coverage and polarization state of the system. In Section 4.1.1, we showed that the minimum spacing d between antennas in the array required to achieve the maximum theoretical array gain is $d_{\min} = \lambda/2$. In practice, however, the required d_{\min} maybe larger than $\lambda/2$. Real antenna elements have physical dimensions and are not dimensionless point sources, and thus, the presence of neighboring

¹¹This mostly applies to planar arrays.

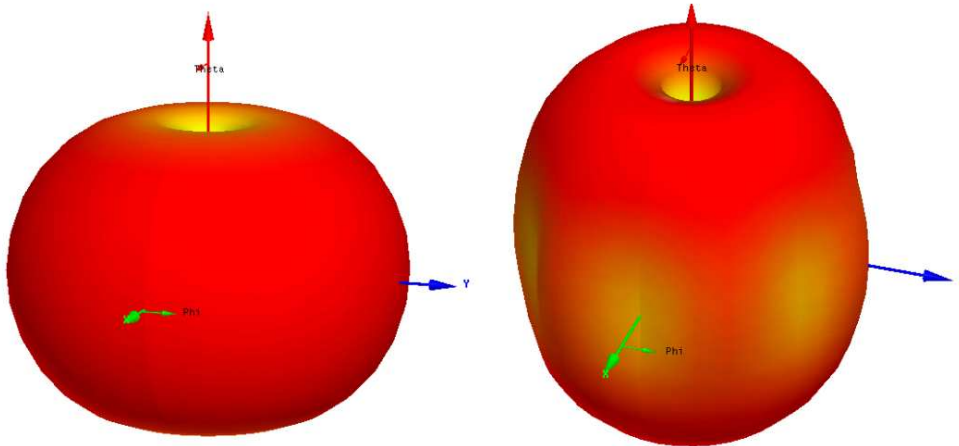


Figure 4.9: (a) 3-dimensional radiation pattern of an isolated dipole. (b) 3-dimensional radiation pattern of a dipole in a circular array with $\lambda/2$ spacing.

antennas creates coupling that distorts the radiation pattern (usually results in a net loss), and alters the impedance of the antenna (affects the bandwidth). Naturally, the coupling goes down as the spacing between the antenna elements increases.

Figure 4.9a shows the radiation pattern of an isolated dipole, and Figure 4.9b, shows the distortion in the pattern after placing the dipole in a circular array with $\lambda/2$ spacing. As a result, the realized beamforming gain is usually less than what is predicted by theory (the loss can be as large as 2-3dB). The loss will also depend on the angle of arrival since the distortion is not uniform. The distortion will also depend on the type and polarization of the antennas being used (e.g. patches, monopoles, dipoles), since different types will be affected differently by different array geometries [12]. In some cases, the distortion in the beam pattern might result in some antennas having increased gain in some directions relative to others. Since these directions are usually different for different antennas in the array (different antennas see different surroundings), these non-idealities in the beam patterns might average out in a large array and the total gain will be comparable to the case of a uniform array. However, in this case, we do not get the full benefit of spatial power combining since a significant portion of the array might not be radiating efficiently in a given direction.

In order to show the impact of antenna spacing on coupling and pattern distortion, we simulated the peak beamforming gain (accounts for both the element factor and the array factor) of different uniform square arrays (with different antenna spacings) of monopole antennas with finite ground plane in HFSS [1]¹². The results are shown in Table 4.2. When we increase the size of the array from 3×3 to 5×5 , the expected theoretical gain would be $10 \log(25/9) \approx 4.4 \text{ dB}$. However, the simulation results show that increasing the number of antennas from 3×3 to 5×5 by reducing the spacing from λ to $\lambda/2$ yields only 1.8dB additional gain. While increasing the number of antennas by reducing the spacing from 2λ to λ yields 4.3dB additional gain (almost the entire predicted gain). This means that

¹²The finite ground plane also contributes to the loss in symmetry of antenna patterns.

geometry	spacing	$\lambda/2$	λ	2λ
3×3		–	10.2dB	10.3dB
5×5		12dB	14.5dB	–

Table 4.2: Simulated array gain versus the spacing between antennas.

increasing the spacing from λ to 2λ does not improve the coupling situation by much, which means that in this case λ spacing provides a good trade off between gain and coupling¹³.

Although the increase in spacing necessary to avoid coupling might give rise to grating lobes, in practice, several of these lobes will be filtered (attenuated) by the beam pattern of individual antennas. For those that are not, we present an array topology in Section 4.2 for canceling or significantly attenuating them. We note, however, that even though $\lambda/2$ spacing may not yield the expected (theoretical) beamforming gain, there are other benefits to having more antennas such as increased diversity and transmit power (by having more radiating elements).

4.2 3-dimensional arrays

In Section 4.1, we looked at how different design choices (number of antennas, spacing, array geometry) influence the radiation pattern and gain of the array. In this section, we propose a new class of array structures that utilize all 3-dimensions, and analyze them in light of the constraints discussed in Section 4.1. As we mentioned earlier, the goal is to achieve as much directivity, coverage (in both azimuth and elevation), and diversity (in both space and polarization) as possible. In this context, using 3-dimensional arrays may seem like an obvious choice given that they use space a lot more efficiently. However, the advantages of using a 3-dimensional array structure are a little more subtle than that. Recall that the directivity of an antenna array (or any antenna) is proportional to the surface area. we showed in Figure 4.7 that increasing the depth on array does not to lead narrower beams. Also, 3-dimensional structures are generally more difficult to build and manufacture. This might lead us to conclude that using 3D array adds little or no value to the system. However, the advantages of using 3D arrays become clear after further examination:

- We showed in Section 4.1.3 that the array factor of a planar area is not uniform in all directions since the effective area is larger in some directions (broadside) than others. On the other hand, when all three dimensions are used (e.g. a cubic array), the array geometry and the gain become more uniform in all directions.
- Although increasing the depth of the array does not improve the width of the main lobe, it can reduce the side lobe levels significantly. Also, as we will demonstrate later in this section, the depth of the array can be exploited to cancel grating lobes that

¹³The optimal spacing will depend on the geometry and radiation patterns of the individual antennas.

arise from large spacing between elements of the array. This property can significantly improve the beam shaping capability of the array.

- Since more antennas can be packed in a three dimensional structure than a planar structure, the aggregate transmit power (spatial power combining) can be increased without increasing the transmit power of individual power amplifiers. Furthermore, having more antennas, while not necessarily contributing to the overall directivity of array, provides better redundancy and reliability. If one antennas fails, then there are several backups.
- A larger number of antennas also leads to better noise averaging. Although the directivity of an array is proportional to the number of antennas on the surface N^2 , where N is the number of antennas (length) in each dimension, there is a total of N^3 receiving antennas. The background noise picked up by those antennas will not be completely uncorrelated¹⁴, and thus, the signal to background noise ratio at the output of the array will not improve beyond N^2 . However, the thermal noise generated by the receivers will be independent across different elements, and thus, the component of the noise at the output of the array that is due to thermal noise can be reduced by N^3 relative to the desired signal. In most systems, the thermal noise dominates over the background noise¹⁵. Therefore, a 3-dimensional array receiver can improve the overall noise figure of the system¹⁶.
- Since we can pack a larger number of antennas in the same amount of space, and not every antenna contributes equally to the gain, we can design different antennas with different properties in order to achieve pattern and polarization diversity. Since it is difficult to achieve all desired properties (e.g. coverage, polarization) with a single type of antenna elements, using a mixture of different types is the only viable solution in practice.

Figures 4.10 and 4.11 show examples of 3-dimensional arrays. Figure 4.10 shows a cubic array composed of multiple stacked planar arrays (sheets). Each sheet is a square array of monopole antennas. Figure 4.11 shows a similar array with a cylindrical shape. In addition to having a more compact form factor than planar arrays and conventional static directional antennas, the arrays shown in Figures 4.10 and 4.11 have other advantages as well:

- In these array designs, our choice for an antenna element was a monopole antenna with a ground plane. We chose monopoles because they are easy to manufacture and assemble, and have an omni-direction pattern, which enables beam steering in

¹⁴Antennas pickup noise from the environment. The noise temperature of the background noise depends on the angle of arrival [45]. The correlation of the background noise picked up by different antennas in the array depend on many factors including directivity.

¹⁵The contribution of the thermal noise to the overall noise at the output of a receiver is captured by the noise figure (NF). The noise figure is defined as the ratio of the SNR at the input of the radio to the SNR at the output of the radio [28].

¹⁶Depending on the number of antennas, this property can be utilized in different ways. For example, if we relax the noise figure requirements of a receiver, we might be able to design the amplifiers with less power or better linearity. However, this is still an open problem that requires further investigation.

any direction in the azimuth plane¹⁷. Furthermore, since monopoles do not radiate vertically, the presence of a ground plane (a wavelength away) from above has little or no effect on their beam patterns.

- Each sheet can be realized with a printed circuit board (PCB), and the presence of a metal ground plane simplifies the process of mounting the digital and analog electronic components. The feedlines for the antennas can be implemented using microstrip lines on the PCB, which significantly reduces the cost of manufacturing and assembly. The loss from microstrip lines is not an issue since the amplifiers (e.g. PAs and LNAs) can be placed very close to the antennas.
- The different sheets (PCBs) can be connected with one another by mounting them on another PCB.
- Compared with array structure shown in Figure 4.12 (a sectorized array of patch antennas), which also provides the desired 360° coverage and can be configured to have polarization diversity, the 3D arrays in Figures 4.10 and 4.11 can be built to improve on the transmit power and noise figures of the system¹⁸.

4.2.1 Lobe cancellation

Increasing the antenna spacings beyond $\lambda/2$ has other undesirable effects beyond the increase in size. The most important one is the creation of multiple main lobes, also known as grating lobes, most of which will point towards undesired directions [12]. These grating lobes appear in the theoretical array factor of the antenna. In fact, even a linear array with $\lambda/2$ spacing exhibit grating (back) lobes in both the broadside and the endfire directions because of the symmetry, with the number of grating lobes increasing as we increase the spacing. One way to break the symmetry and eliminate or reduce the size of grating lobes in theory is to reduce the spacing between the antennas to $\lambda/4$. Figure 4.1 shows the theoretical array patterns of two uniform square arrays with the same area but different spacings: $\lambda/2$ and $\lambda/4$. In Figure 4.14a, with both array patterns configured to point towards 90° , we see that the $\lambda/2$ spacing has a back lobe pointing towards 270° that gets reduced by almost 25dB when the spacing is $\lambda/4$. In Figure 4.14b, we see that when both arrays are pointing towards 45° , the back lobes for both spacings are at a significantly lower power level than the main lobe. The reason for this is that looking through the diagonal of the array, the spacing between the antennas is no longer $\lambda/2$, but $\frac{\sqrt{2}}{2}\lambda$ instead, which automatically breaks the symmetry since $\frac{\sqrt{2}}{2}\lambda \approx \frac{3}{4}\lambda$. However, in both figures we notice the level of side lobes

¹⁷Ideally, we would like to be able steer the beam in any direction in space. Unfortunately, isotropic antennas are not realizable in practice. However, if complete coverage is required, it can be achieved with a mixture of antennas in the array such that their “aggregate” pattern covers the entire space. For example, we can put patch antennas on the top and bottom sheets in Figures 4.10, 4.11 in order to cover the vertical directions. Also, we can utilize other types of antenna that are easy to print on a PCB and provide a different polarization from the monopole. Another way of achieving different polarizations is by changing the orientation of some sheets relative to others.

¹⁸In the patch array, since only a few sectors are pointing at the desired direction at any given time, the transmit power from other sectors is not fully utilized.

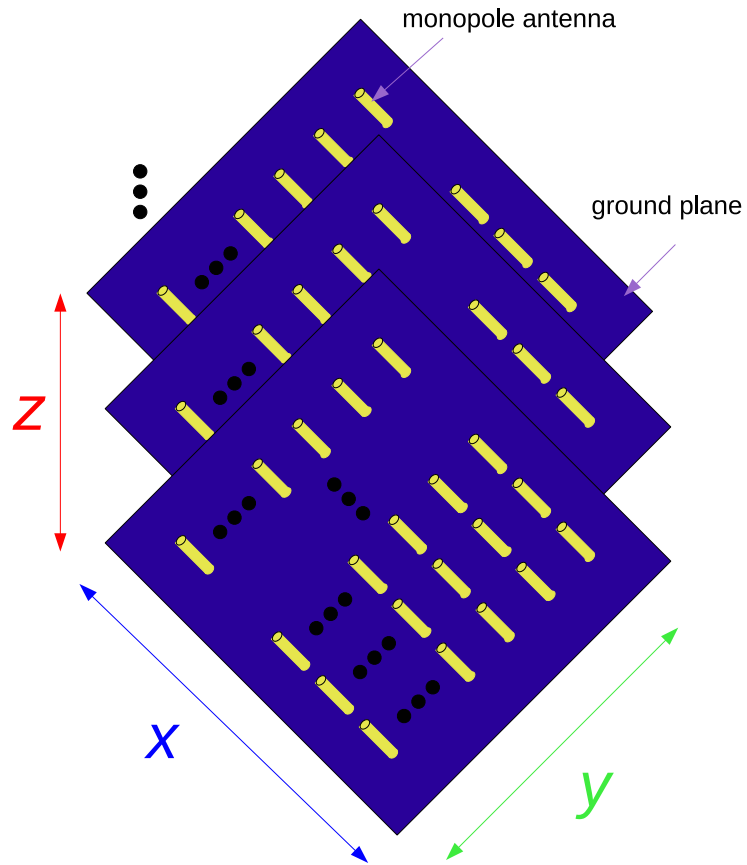


Figure 4.10: 3-dimensional rectangular array of monopoles.

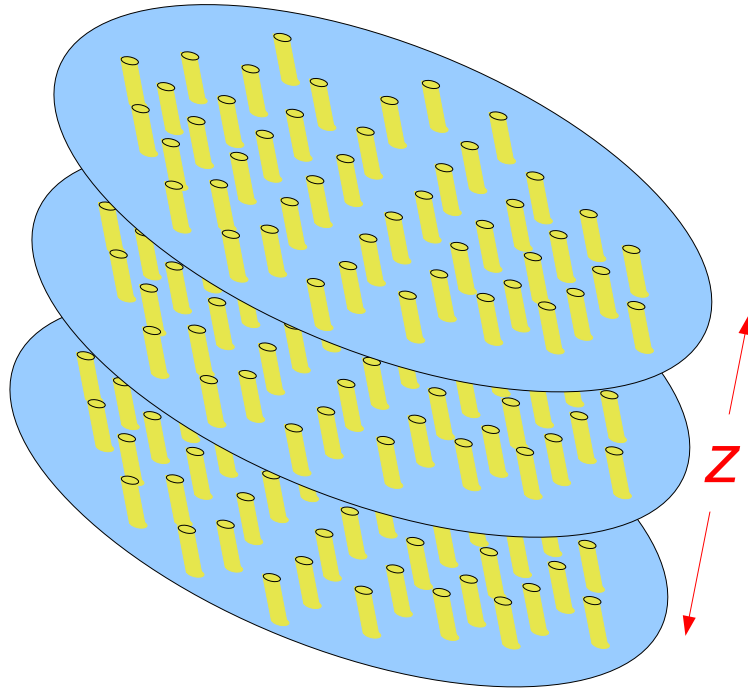


Figure 4.11: 3-dimensional cylindrical array of monopoles.

is generally lower for the $\lambda/4$ spacing¹⁹, and that adding more antennas by reducing the spacing to $\lambda/4$ does not affect the width of the desired main lobe, which is what we expect.

As pointed out in Section 4.1.5, in practice, $\lambda/4$ spacing can do more harm than good since it further increases coupling and pattern distortion between neighboring antennas without increasing the gain. In fact, we may end up with a net loss in the overall gain. However, if the array has a 3-dimensional structure (multiple sheets), we make a small modification to the relative placement of the sheets in order to emulate the lobe cancellation feature of a $\lambda/4$ array. By shifting the different sheets relative to one another by $\lambda/4$ in the direction of the x-axis and the y-axis (as shown in Figure 4.13), the antennas projected onto the xy plane appear like a single uniform planar array with $\lambda/4$ spacing. In this case, the back lobes and other grating lobes get canceled²⁰. We also don't have to worry about coupling, since spacing between antennas is still greater than $\lambda/2$ (as long as the different sheets are far enough from each other). Also, since the beam pointing towards 45° has a lower first side lobe than a beam pointing towards 90° , we can reduce the side lobes by not only shifting the sheets relative to one another, but also by changing their relative orientations as well or by using other planar structures for the sheets that have better side lobe properties (e.g. circular arrays shown in Figure 4.11).

¹⁹The fact that the first side lobe is lower for a 45° observation angle is consistent with the discussion in Section 4.1.4

²⁰To achieve full cancellation of grating lobes, the required number of sheets is such that the total number of antennas is larger than the number of antennas on a single plane with $\lambda/4$ spacing. Also, the sheets do not have to be stacked vertically in order to achieve the grating lobe cancellation.

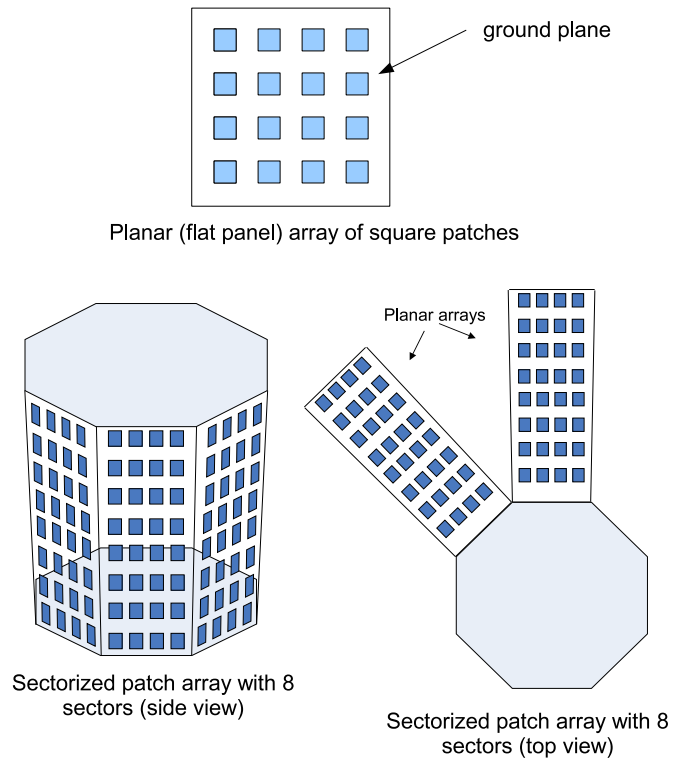


Figure 4.12: Planar and sectorized patch arrays (8 sectors).

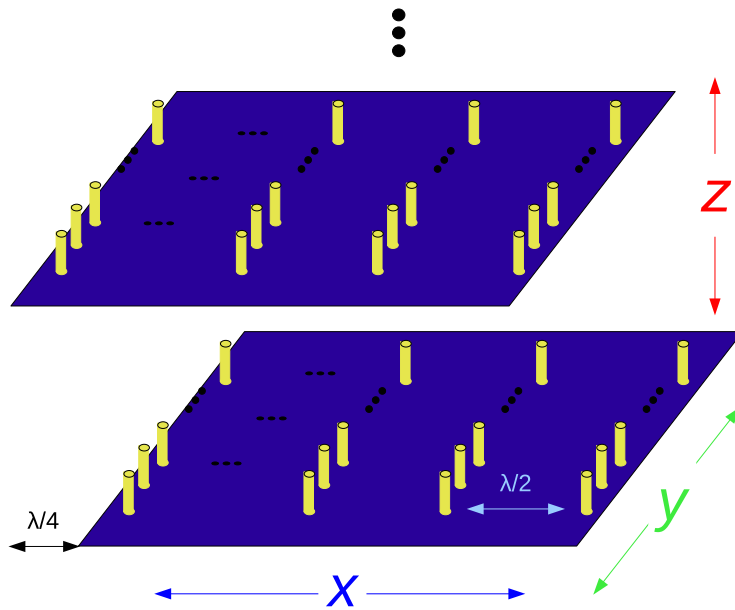
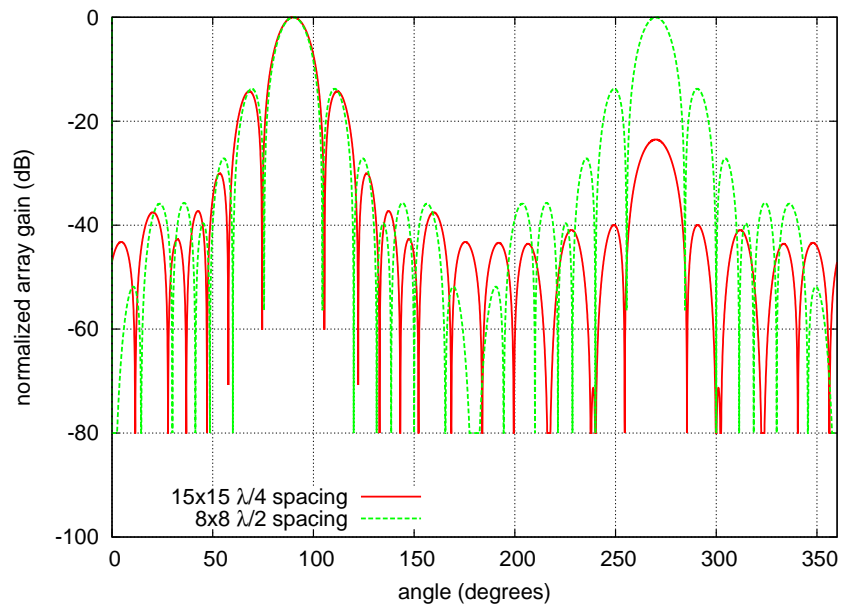
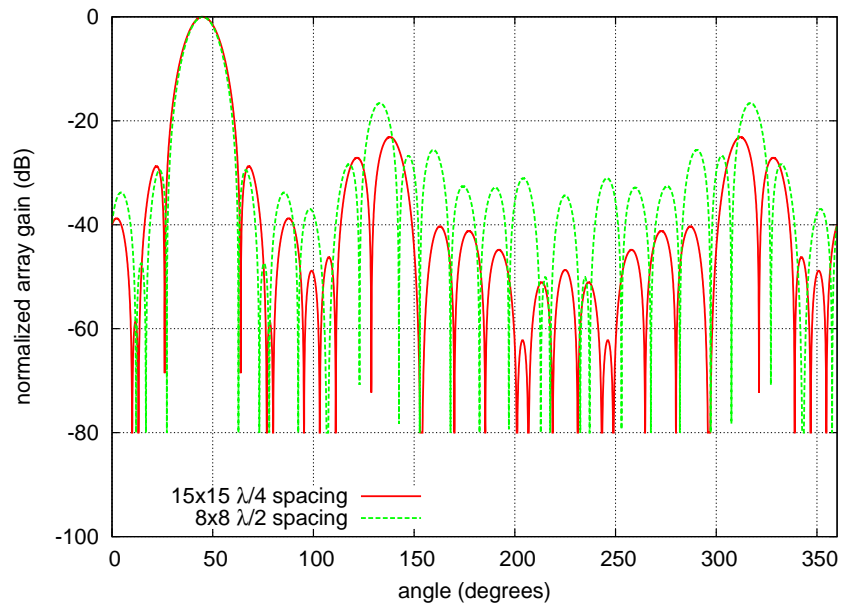


Figure 4.13: A multi-sheet/3-dimensional rectangular array of monopoles with $\lambda/4$ shifting.



(a)



(b)

Figure 4.14: Comparing the theoretical array patterns (in the azimuth (xy) plane) of two uniform square arrays with same area, but different spacings ($\lambda/2$ and $\lambda/4$). (a) Both arrays steered to towards 90° . (b) Both arrays steered to towards 45° .

Chapter 5

Adaptive beamforming and channel estimation

At a high level, a multi-antenna receiver maximizes the SNR by steering its main beam toward the desired signal and nulling out any interfering signals. If the receiver knows the channels responses of the various transmitters, then the problem becomes trivial. In practice, though, the channels from the various transmitters are unknown a priori, and will vary in time. Thus, the system must be able to adaptively estimate the channels and compute the optimum beamforming weights. Designing optimal methods¹ to accomplish this task, also known as adaptive array processing techniques, is heavily influenced by the radio architecture of the array and its hardware capabilities. Most of the previous work in this area has focused on the digital beamforming architecture, while the RF beamforming architecture has received very little attention.

The fundamental difference between the digital baseband and RF architectures that impacts beamforming is that the digital architecture offers access to the input signals from each individual antenna element in the array (prior to combining), whereas the RF architecture only offers access to the aggregated signal(s)². Thus, in a pure digital architecture, the signals from each antenna are digitized and stored. This capability enables the signals to be manipulated and combined in multiple ways. That means that the optimum beamforming weight vector(s) can be computed and updated while data is being received. Furthermore, multiple spatial data streams can be extracted simultaneously. On the other hand, the RF architecture enables only a single way to aggregate the signals at any given time. This means that the amount of time and training samples required to estimate the optimum beamforming weights is much larger than the digital architecture, which creates an additional overhead

The work presented in this chapter was done in collaboration with Mark Johnson, Raghuraman Mudumbai, Kannan Ramchandran and Ben Wild [11, 10, 8, 9].

¹The performance criteria will vary from system to system. These include speed, accuracy, and complexity. There is usually a trade off between these factors, and the optimal combination will depend on the system requirements.

²Multi-beam and hybrid RF/digital architectures discussed in Chapter 3 can have $M > 1$ different aggregated outputs. In general, however, $M \ll N$, where N is the number of antennas.

since it is difficult update the weights while data is received³. This makes most beamforming techniques developed for the digital architecture unsuitable for the RF beamforming architecture. Therefore, new beamforming techniques need to be designed that are tailored for the RF architecture, which is the focus of this chapter.

Most of the previous work on beamforming schemes that rely only the aggregate signal at the output of the array were based on greedy energy maximization [36, 37]. In [36], a scalable algorithm that utilized a single bit feedback was proposed. The algorithm proceeds by adding small random perturbations to the beamforming weights at each step. If the new weights result in a higher SNR at the output of the array, then the new weights are retained. Otherwise, they are discarded. It was shown that if the magnitude of the perturbations is small enough, then the algorithm is guaranteed to converge, and the number of steps required grows linearly with the number of antennas N . However, the main disadvantage of this class of algorithms is that they are slow to converge and are very sensitive to noise and interference⁴. Also, while the convergence time scales linearly with the number of antennas, the constant factor is usually too large for most practical systems.

In this chapter, we present a more flexible and robust framework for computing the complex beamforming weights based on adaptive adaptive filtering. The chapter is organized as follows. In Section 5.1, we present a brief overview of popular adaptive filter techniques. In Section 5.2, we show how to use these techniques to estimate the optimum beamforming weights. The problem of null-steering and the null-steering performance are discussed in Section 5.3. In Section 5.4, we consider the problem of beamforming in the presence of interference. In Section 5.5, we discuss the problem of frequency synchronization and fast fading channels and show how the adaptive filter framework can be modified to handle fast changing channels and large frequency offsets. We discuss multichannel and hybrid RF/digital architectures in Section 5.6. Finally in Section 5.7, we conclude with a brief discussion on transmit beamforming.

5.1 Adaptive filter techniques

Statistical channel estimation and filter design and optimization occur in a wide variety of applications. Although different applications have different requirements, a lot of these problems share a common structure and can be written as:

$$\mathbf{w}^H \mathbf{x}[n] = d[n]$$

$$\mathbf{x}[n] = [x[n], x[n-1], \dots, x[n-K+1]]^\top$$

where $\mathbf{x}[n]$ is a complex column vector (sequence) of order K of data at time n , $d[n]$ is a complex scalar denoting the desired (training) signal at time n , and \mathbf{w} denotes a discrete time complex order K linear filter⁵. The interpretation of the data and the desired signal are

³The size of this overhead depends on the rate of channel variation as well as the system requirements.

⁴It is difficult to reliably detect energy levels even in moderate SNR regimes.

⁵In this chapter, we denote scalars in lower case, vectors in bold lower case, and matrices in bold upper case.

application dependent⁶. However, the goal is to find the linear combination⁷ of data samples that best approximates the desired signal at time n . In practice, the “best” or optimum solution is defined in terms of the minimum mean square error (MMSE) criterion:

$$e[n] = d[n] - \mathbf{w}^H \mathbf{x}[n]$$

$$\mathbf{w}_{MMSE} = \underset{\mathbf{w}}{\operatorname{argmin}} E[\|e[n]\|^2]$$

This solution is obtained by solving the Normal Equations (sometime referred to as the Wiener-Hopf equations) [50]:

$$\mathbf{w}_{MMSE} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{dx} \quad (5.1)$$

$$\mathbf{R}_{xx} = E[\mathbf{x}[n]\mathbf{x}[n]^H], \quad \mathbf{r}_{dx} = E[d[n]^H \mathbf{x}[n]]$$

\mathbf{R}_{xx} is the $K \times K$ autocorrelation matrix of $x[n]$, and \mathbf{r}_{dx} is $K \times 1$ cross-correlation vector of the desired signal $d[n]$ and $\mathbf{x}[n]$ (the data vector)⁸. The solution in Equation 5.1 is also referred to as the Wiener or least squares solution. This solution is derived from the orthogonality principle, which states that the power of the error signal is minimized when the error signal is orthogonal (uncorrelated) with the data⁹:

$$E[e[n]x[m]] = E[(d[n] - \mathbf{w}_{MMSE}^H \mathbf{x}[n])x[m]] = 0 \quad \forall_{n,m}$$

$$\Rightarrow E[d[n]x[m]] - E[\mathbf{w}_{MMSE}^H \mathbf{x}[n]x[m]] = 0 \Rightarrow r_{dx}[n-m] = \mathbf{w}_{MMSE}^H \mathbf{r}_{xx}[n-m]$$

$$\Rightarrow r_{dx}[k] = \mathbf{w}_{MMSE}^H \mathbf{r}_{xx}[k]$$

where \mathbf{r}_{xx} is the k^{th} column of \mathbf{R}_{xx} . To solve for the K coefficients of \mathbf{w}_{MMSE} we need K equations (i.e. $k = 0, \dots, K-1$), which is equivalent to Equation 5.1. In general, the second order statistics \mathbf{R}_{xx} and \mathbf{r}_{dx} are not known a priori, and are usually estimated from sample averages ($\overline{\mathbf{R}}_{xx}$ and $\overline{\mathbf{r}}_{dx}$ are used instead of \mathbf{R}_{xx} and \mathbf{r}_{dx}). In practice, however, the normal equations are not usually solved by directly inverting the \mathbf{R}_{xx} (or $\overline{\mathbf{R}}_{xx}$) for several reasons:

- 1) The data sequence $x[n]$ is not available all at once, but instead is received over time. Therefore, waiting for the entire sequence to be available may introduce unnecessary delay.
- 2) The data set $\{x[n]\}$ maybe too large that computing and inverting R_{xx} maybe computationally unfeasible.
- 3) The statistics of the signal $x[n]$ are usually not known a priori and frequently vary with time, and thus using the entire data set to compute the optimum weights vector (filter) may lead to inaccurate results.

⁶In communication systems, the desired signal is usually a known pilot signal defined by the system a priori.

⁷In this chapter, we focus on finite impulse response filters (FIR). We will also assume that all signals have zero means (i.e. $E[x[n]] = E[d[n]] = 0$).

⁸We assume that these second order statistics are stationary. Otherwise, we need to include the time index (e.g. $\mathbf{R}_{xx}[n], \mathbf{r}_{dx}[n]$).

⁹The results are simplified by assuming that $x[n]$ and $d[n]$ are wide sense stationary and jointly wide sense stationary signal, which means that their first and second order statistics are independent of time [47].

This motivated the design of a class of filters, known as adaptive filters, that addresses these issues. Adaptive filters have the following properties:

- 1) The computation of the beamforming weights proceeds on a sample by sample basis. That is the filter taps are incrementally updated when a new sample (or a group of samples) is received, as opposed waiting for the entire sequence. This is sometimes referred to as online versus batch mode computation. This property addresses the latency problem.
- 2) The computational complexity of updating the filter taps when new samples are received is far less than direct matrix inversion. This addresses the complexity problem.
- 3) The algorithms do not assume any prior knowledge of the signal statistics.
- 4) More recent samples have stronger influence on the values of the filter taps than older samples. This, together with Properties (1) and (3), gives the algorithm the ability to track changes in the signal statistics, and is what makes those filters truly adaptive.

Figure 5.1 shows the basic structure and building blocks that most adaptive filters have in common. Adaptive filters are closed loop systems. The error signal (or a function of it) in the feedback loop is used to update the filter taps. Different adaptive filters have different update equations, which in turn depend on the error criterion and other factors.

5.1.1 Gradient descent

One of the most popular classes of adaptive filters are based on the concept of Gradient descent. Gradient descent defines how the filter taps get updated in order to minimize some error criterion ξ . The filter taps update proceeds as follows:

$$\mathbf{g}_{n+1} = \mathbf{g}_n - \mu \nabla_{\mathbf{g}_n} \xi_n \quad (5.2)$$

where \mathbf{g}_n is the vector of filter taps at time (iteration) n , $\nabla_{\mathbf{g}_n} \xi_n$ is the gradient of the error criterion ξ_n with respect to the filter taps \mathbf{g}_n , and μ is the step size (a scalar that determines the rate of adaptation and convergence). Since the gradient defines the direction of the maximum rate of change, the error is minimized by moving in the opposite direction of the gradient (and hence the name gradient descent). One of the most popular error criteria is the mean square error (MSE). The generic minimum mean square error gradient descent problem is shown in Figure 5.1. The goal is to select the filter taps such that the output of the filter, when the input is the random sequence $\{w[n]\}$, is as close as possible to the desired sequence $\{d[n]\}$ in the mean square sense. The MSE is defined as follows:

$$\begin{aligned} \xi_n &= E[|e[n]|^2] \quad \text{where} \quad e[n] = d[n] - \mathbf{g}_n^T \mathbf{w}_n \\ &\Rightarrow \nabla_{\mathbf{g}_n} \xi_n = \nabla_{\mathbf{g}_n} E[e[n]e^*[n]] \\ &= \nabla_{\mathbf{g}_n} E[(d[n] - \mathbf{g}_n^T \mathbf{w}_n)(d^*[n] - \mathbf{g}_n^H \mathbf{w}_n^*)] = -E[e[n]\mathbf{w}_n^*] \end{aligned} \quad (5.3)$$

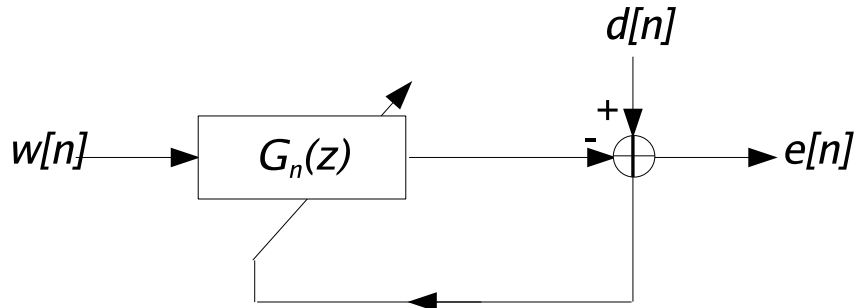


Figure 5.1: The standard MMSE adaptive filtering problem.

$$\Rightarrow \mathbf{g}_{n+1} = \mathbf{g}_n - \mu \nabla_{\mathbf{g}_n} \xi_n = \mathbf{g}_n + \mu E[e[n] \mathbf{w}_n^*] \quad (5.4)$$

The difficulty in applying gradient descent directly lies in the fact that computing the expectation in Equation 5.4 requires prior knowledge of the first and second order statistics of the signals, which defeats the purpose of using a “blind” algorithm. Therefore, in most implementations, expected values are replaced by sample averages. By far the simplest and most popular implementation is the LMS, which replaces the expected values of random variables with their instantaneous values¹⁰, and works well in practice. The LMS update equations are as follows:

desired sequence: $\{d[n]\}$, input sequence: $\{w[n]\}$

l -tap FIR filter: $\mathbf{g}_n = [g_n[0], g_n[1], \dots, g_n[l-1]]^T$

$\mathbf{w}_n = [w[n], w[n+1], \dots, w[n+l-1]]^T$, error sequence: $e[n] = d[n] - \mathbf{g}_n^T \mathbf{w}_n$

Filter update: $\mathbf{g}_{n+1} \leftarrow \mathbf{g}_n + \mu e[n] \mathbf{w}_n^*$ (5.5)

The following condition is sufficient to guarantee that the filter taps converge (in the mean square sense):

$$\mu \leq \frac{1}{\text{tr}\{\mathbf{R}_{ww}[n]\}}$$

where $\mathbf{R}_{ww}[\mathbf{n}] = E[\mathbf{w}_n \mathbf{w}_n^H]$ is the autocorrelation matrix (of order l) of the random sequence $\{w[n]\}$ (which can potentially be time varying). In general, a smaller value of μ leads to better convergence, while a larger value leads to faster adaptation and better tracking, but can make the algorithm less stable¹¹. The convergence time τ of the LMS algorithm is roughly:

$$\tau \sim \frac{1}{\mu} \frac{\lambda_{max}}{\lambda_{min}} \quad (5.6)$$

¹⁰This is why it is sometimes referred to as the “stochastic” gradient descent algorithm.

¹¹The value of μ doesn’t have to be fixed, and can be varied as a function of time (μ_n) to achieve a better compromise between convergence and adaptation.

where λ_{max} and λ_{min} are the largest and smallest eigenvalue of \mathbf{R}_{ww} respectively. The adaptive filter can be initialized to any random value (usually the filter taps are set to 0) without affecting the asymptotic convergence time of the algorithm.

The LMS algorithm works very well in practice and is very widely used in communication systems for channel estimation and equalization. One of the reasons the LMS is very popular is its low computational complexity. The number of operations required to update the filter taps is proportional to the number of taps (i.e. $O(l)$). There are several variants of the LMS algorithm that optimize some aspect of the performance like stability (e.g. Normalized LMS) or computational efficiency (e.g. sign-LMS)¹².

5.1.2 Other adaptive filtering techniques

There are other adaptive filters that are widely used that are not based on gradient descent. The most popular example is recursive least squares (RLS). RLS usually leads to much faster and more accurate convergence than LMS, and does not depend on the eigenvalue spread as is the case with LMS. However, this improvement in speed comes at the price of extra computational complexity (the number of operations required for each update step is quadratic in the number of filter taps ($O(l^2)$)).

Even though both LMS and RLS do not require any prior knowledge of the signal statistics, they both require a reference (training) signal (e.g. $\{d[n]\}$) that is known a priori¹³. There is another class of adaptive algorithms that do not rely on a known reference signal, but instead adapt based on some property of the data signal (e.g. constant modulus). Such algorithms are usually much slower to converge, and are not always guaranteed to converge to the correct result.

A more comprehensive treatment of adaptive filters can be found in [50, 27, 26].

5.2 Channel Estimation via Adaptive Filtering

System identification is one of the classical applications or formulations of the adaptive filter problem. The general system identification problem is illustrated in Figure 5.2. The algorithm generates a random (usually white) sequence $w[n]$ that is fed into the system with unknown impulse response \mathbf{h} or transfer function $H(z)$ ($H(z)$ is the z -transform of \mathbf{h})¹⁴. as well as to an adaptive filter g_n (transfer function $G_n(z)$)¹⁵. that is used to approximate the

¹²In the standard LMS algorithm, the size of the update to the vector \mathbf{g}_n in Equation 5.5 will be directly proportional to the magnitude of the data (regression) vector $\mathbf{x}^*[n]$. This can have some undesired effects on the performance of LMS. Normalized LMS avoids this problem by dividing the step size μ by $|\mathbf{x}[n]|^2$ at each step. Sign-LMS or sign-error LMS on the other hand, simplifies the required computation in the update step by quantizing the real and imaginary parts of the error signal $e[n]$ to a single bit (the sign bit). However, the saving in computation comes at the expense of slower convergence [50].

¹³The reference signal can sometimes be predicted from the data itself. For example, in a communication system, after a signal is decoded, the output of the decoder can be used to train the adaptive filter. Such a technique is called decision-based or decision-directed adaptive filtering. However, even in such cases, the LMS or RLS algorithm still needs to be initialized (boot-strapped) using a reference (pilot) signal that is known a priori. This pilot needs to be injected at regular intervals in order to avoid instability.

¹⁴We assume that the unknown system is linear.

¹⁵Sometimes we denote a filter with a subscript n (e.g. g_n) when the filter taps are time varying.

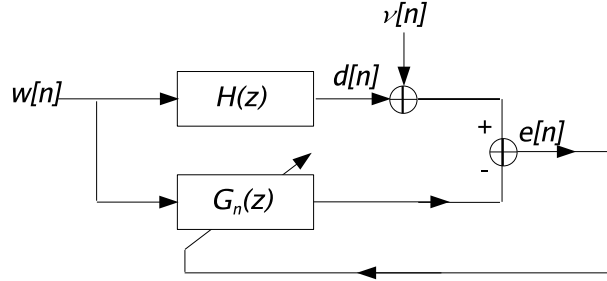


Figure 5.2: Identifying the impulse response of an unknown system using adaptive filtering

unknown system. The output from the unknown system is used as the desired signal $d[n]$:

$$d[n] = (h * w)[n]$$

$$e[n] = d[n] + \nu[n] - (g_n * w)[n]$$

$$(h * w)[n] \triangleq \sum_{k=-\infty}^{\infty} h[k]w[n - k]$$

where $\nu[n]$ is noise signal (usually white) that models the noise in the system, and $(*)$ denotes the discrete convolution operator. It is clear from the equations above that the variance of the error signal is minimized when $g \rightarrow h$ as $n \rightarrow \infty$ ($e[n]$ is reduced to white noise sequence).

The unknown system response does not necessarily have to be a time-domain filter. In fact, in the case of beamforming, the unknown system to be identified is a spatial filter. The algorithm proceeds in a similar manner. For example, a receiver Rx estimating the complex channel response at each antenna from a remote transmitter Tx using the system identification framework would follow these steps¹⁶:

- 0) Rx locally generates a random sequence $\{w[n]\}$ (making the samples of $w[n]$ *i.i.d* usually results in faster convergence)¹⁷.
- 1) Tx sends a known pilot $x[n]$. The choice of $x[n]$ does not matter much since the same signal is received at the different antennas at the receiver. Furthermore, we assume that if the transmitter has multiple antennas, the transmit beamforming weight vector remains fixed throughout this process.

¹⁶Without loss of generality, we assume that the channels are narrow band or single tap (flat fading) channels. A large delay spread (frequency selective) channel can be converted into a group of smaller narrow band subchannels (e.g. OFDM), where the complex channel response is flat fading.

¹⁷ $w[n]$ can be sampled from any distribution (e.g. Gaussian, Bernoulli, etc.). The choice will depend on several factors such the set of available beamforming weights, which determines both resolution and the dynamic range. For example, when the beamformers have low resolution, a Bernoulli distribution maybe preferable since it takes only two values (e.g. 1, -1). However, as we show in Section 5.4, using this adaptive filtering based framework in the presence of interference will require much higher resolution for the beamforming weights. We note, however, that the accuracy of channel estimation does not depend on the resolution of the beamforming weights, but instead depends on the resolution of the A/D component.

- 2) Rx sets the antenna weights (beamforming weight vector) to a block of length N (number of antennas) of the random sequence $\{w[n]\}$.
- 3) At each time step n , the block of $\{w[n]\}$ assigned to the antenna weights is shifted to the right by one sample. Thus the output of the array at the receiver $y[n]$ mimics the time domain convolution of the sequence $\{w[n]\}$ with an N tap filter (channel response \mathbf{h})¹⁸.
- 4) The received signal at the output of the array $y[n] = \mathbf{w}^H[n]\mathbf{h}[n]x[n] + \nu[n]$ is used by the adaptive filter as the desired signal ($y[n]$ is the output from the A/D after aggregation). The error signal $e[n]$ is used to update the N tap filter g_n .

These steps are summarized in Figure 5.3. The figure shows how the channel vector from a transmitter to a multi-antenna receiver can be estimated by posing the problem in the system identification framework. The complex vector \mathbf{h} is the unknown channel response from the transmitter to each antenna in the array¹⁹. The adaptive filter can be implemented by one of the standard algorithms (e.g. LMS or RLS). When the step size in the filter is chosen appropriately, the filter coefficients $\mathbf{g}[n]$ will approximate \mathbf{h} , the unknown channel²⁰. Observe that while the beamforming weights may be applied in the RF or analog domains, the computation of those weights occurs in the digital domain.

5.2.1 Performance evaluation

In this Section, we analyze the performance (both speed and accuracy) of the adaptive filtering based beamformer. We also consider the impact of noise and channel variation. For simplicity, we will focus on the Normalized LMS (NLMS) algorithm. However, the results extend to a generic adaptive filter.

Let us begin by deriving the expression for the running time under NLMS. Recall from Equation 5.6 in Section 5.1.1, the convergence time of the LMS algorithm τ :

$$\tau = \frac{1}{\mu} \frac{\lambda_{max}}{\lambda_{min}}$$

Since the elements of the sequence $\{w[n]\}$ are chosen *i.i.d.*:

$$\begin{aligned} \overline{\mathbf{R}}_{ww} &\sim \mathbf{R}_{ww} = \sigma_w^2 I_N \\ &\Rightarrow \frac{\lambda_{max}}{\lambda_{min}} \approx 1 \end{aligned}$$

¹⁸It is not strictly necessary for a spatial filter to simulate a time domain filter. For example, new sets of N weights can be generated independently at each time step n and the subsequent algorithm will still robustly estimate the channel. In fact, we show in Section 5.4 that in order to cancel a known interferer, emulating a time domain convolution may no longer be an option.

¹⁹In a rich-scattering multipath environment, the channel responses usually follow a Rayleigh distribution (i.e. the channel coefficients are *i.i.d.* complex Gaussian random variables). In a clear line of sight environment, the channel coefficients are functions of the array geometry and the direction of arrival.

²⁰because $y[n]$ is a noisy version of the true desired signal $d[n]$ (due to the presence of the white (thermal) noise signal $\nu[n]$), g_n will be a close, not an exact, estimate of h

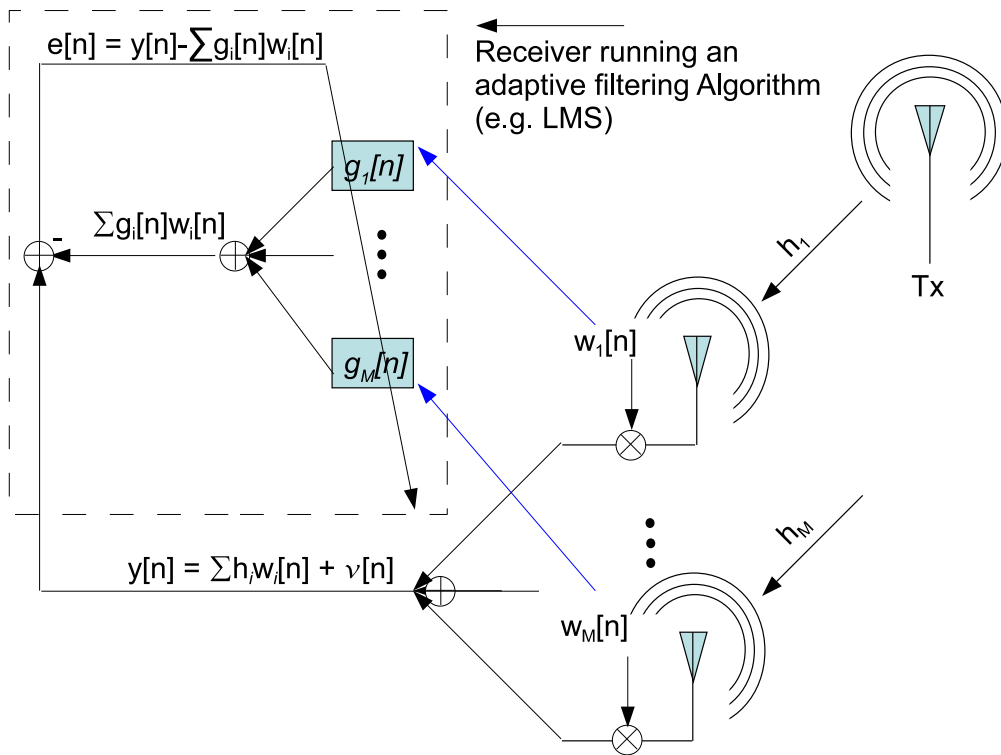


Figure 5.3: The transmitter sends a known signal, and the multi-antenna receiver employs an adaptive filter to estimate the optimal beamforming weights.

where $\sigma_w^2 = E[w[n]^*w[n]]$ (for simplicity we choose $\sigma_w^2 = 1$), and I_N is the identity matrix of size N . The step size μ under NLMS is chosen as:

$$\mu_{NLMS} = \frac{\mu'}{\epsilon + \|\mathbf{w}[n]\|^2}$$

where μ' is a constant (usually $\mu' < 1$) that is independent of the regression data $\{w[n]\}$, $\mathbf{w}[n]$ is a vector of size N of the regression data at time n , ϵ is a positive number to avoid division by zero or by a small number when $\|\mathbf{w}[n]\|^2$ is close to zero²¹. Since $E[\|\mathbf{w}[n]\|^2] = \sigma_w^2 N = N$:

$$\begin{aligned} \mu_{NLMS} &\sim \frac{1}{N} \\ \Rightarrow \tau_{NLMS} &\sim \frac{1}{\mu_{NLMS}} = O(N) \end{aligned}$$

Therefore, the number of steps required for convergence is proportional to the number of antennas N . This result is not surprising. Another more intuitive way to think about this is that since we are solving for the N coefficients of unknown spatial filter, we require at least N unique equations (samples) to solve for these unknown coefficients²². Therefore, the linear running time is fundamental to this architecture, and not specific to a particular algorithm, since we can only capture one sample at a time. However, different adaptive algorithms can differ by a constant factor as well as their performance in the presence of noise, channel variation, finite resolution, and other factors.

Figure 5.4a shows the fraction of the maximum beamforming power achieved by the LMS algorithm as a function of the number of iterations (normalized by the number of antennas N) for different values of N . The fraction of maximum power is a metric of the alignment between the channel estimate and the actual channel. The graph shows that LMS achieves at least 90% of the maximum power after approximately $3N$ iterations, independent of the value of N . Figure 5.4b shows the same results for RLS. While also linear in N , it achieves 90% of the maximum power much faster than LMS (after approximately N iterations) at the expense higher computational complexity.

The impact of noise

The signal to noise ratio (SNR) at each individual antenna is not boosted by the beamforming (signal combining) gain. Therefore, it is critical for beamforming systems that the adaptive algorithm performs well under low SNRs. In Figure 5.5a, we add noise to the system to analyze the performance degradation as a function of the SNR. The graph shows that even when the signal is below the noise level prior to beamforming, the LMS estimate still achieves a significant fraction of the maximum beamforming gain. The performance in the presence of noise can be further improved by decreasing the step size of the LMS algorithm, which leads to better noise averaging and more accurate convergence²³. Figure 5.5b, shows

²¹This why the algorithm is sometimes referred to as ϵ -NLMS.

²²More than N equations are usually required in practice since the samples are noisy.

²³ A small step size will also lead to slower response to channel variations and slower convergence. Thus, there is an inherent trade off in the choice of the step size.

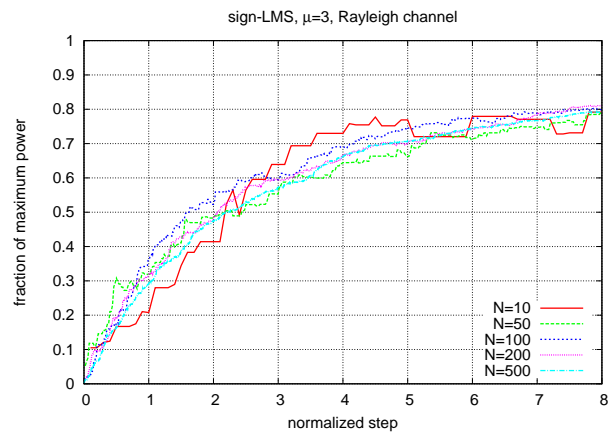
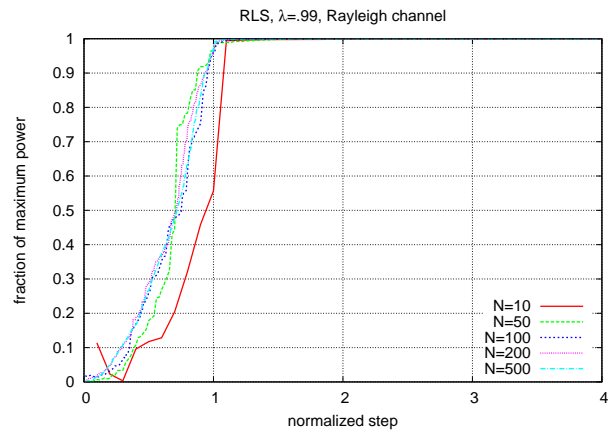
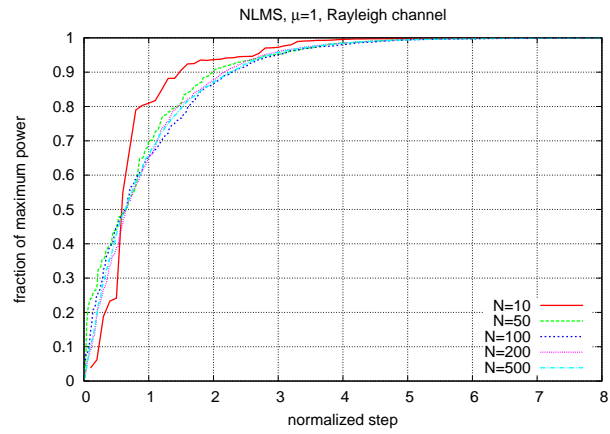


Figure 5.4: (a) NLMS convergence speed for different values of N . (b) RLS convergence speed for different values of N . (c) sing-LMS convergence speed for different values of N . The x-axis on the plots is the number of iterations normalized by the number of antennas N , and on the y-axis is fraction of the maximum beamforming gain.

that almost 70% of the beamforming can be achieved at -5dB SNR when the step size is decreased from $\mu = 1$ to $\mu = .1$. Compared to Figure 5.4, we see that the overall impact of noise is slower convergence²⁴.

The impact of channel variation

Time-varying channels is another factor that often degrades the performance of adaptive algorithms. Channel variation can occur due to the mobility of either or both nodes in a wireless link or due to the mobility of the reflectors and scatterers²⁵. In a rich multipath environment (a Rayleigh channel), the variation is independent across the different channel coefficients and is known as the Doppler rate or Doppler spread [59]. In a line of sight (LOS) environment, the channel variation results from a change in the angle of arrival (AOA). In this case, the channel coefficients are functions of the angle of arrival. Figure 5.7a shows the performance of LMS in a Rayleigh channel with nonzero Doppler spreads. The figure shows that the LMS still converges even when the channel variation rate (the Doppler spread) is as large as 1% of the symbol rate (filter update rate). However, adapting to fast channel variations requires a relatively large step size μ , which in the presence of low SNRs leads to poor performance. This problem can be addressed with the differential MMSE framework discussed in Section 5.5.

Quantization

While the performance of the LMS algorithm does not depend on the resolution of the beamforming weights, it does depend on the resolution of the aggregate signal (A/D converter). Low resolution A/D converters are popular due to their low complexity. However, they add more quantization noise that usually slows down the adaptive filter. There are variants of LMS for dealing with quantization. By far, the most popular is sign-LMS; sign-LMS is the simplest of all since it quantizes the output to a single bit (the sign bit). It's popular due to its relatively low computational complexity. Our simulation results show that sign-LMS exhibits similar performance to NLMS (linear in the number of antennas) with a slight performance hit due to the quantization noise. The results for sign-LMS are repeated in Figures 5.4c, 5.6ab, and 5.7b.

5.3 Null-steering

Conventional digital beamforming architectures with multiple inputs enables computing the optimum beamforming weights, which also implicitly place nulls towards interferers when they are present. This, however, is not possible with a single input architecture (e.g. RF based). Under these architectures, each channel response vector has to be estimated separately (this can only happen sequentially) before the optimum beamforming weights can be

²⁴In addition to reducing the step size, we can also reduce the LMS update rate, and thus each sample will involve more noise averaging.

²⁵Sometimes channels can vary due to temperature and other atmospheric changes that cause variations in the air density. However, this type of variation is usually very slow compared to ones caused by mobility.

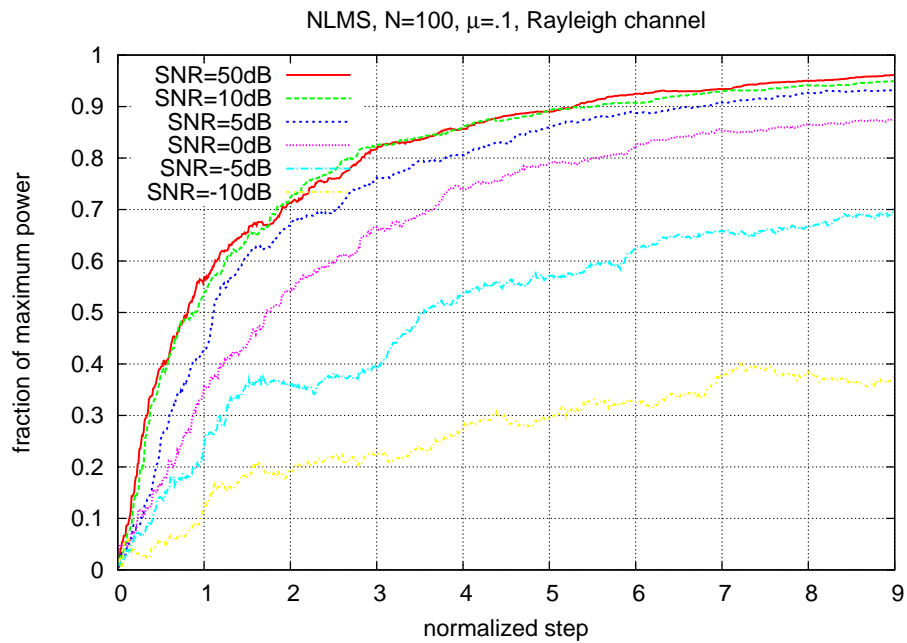
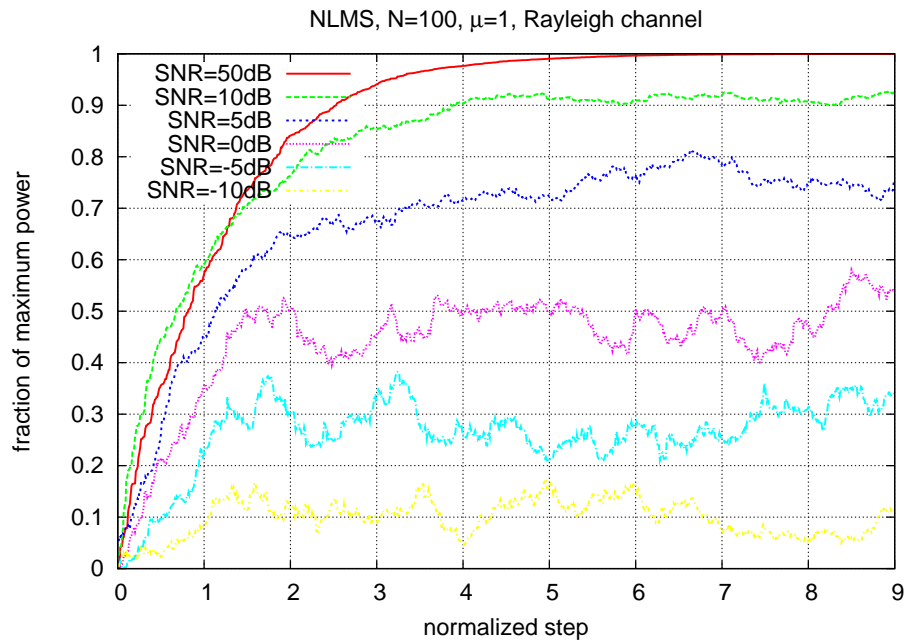


Figure 5.5: (a) NLMS convergence for different SNRs, with a step size $\mu = 1$. (b) NLMS convergence for different SNRs, with a step size $\mu = .1$.

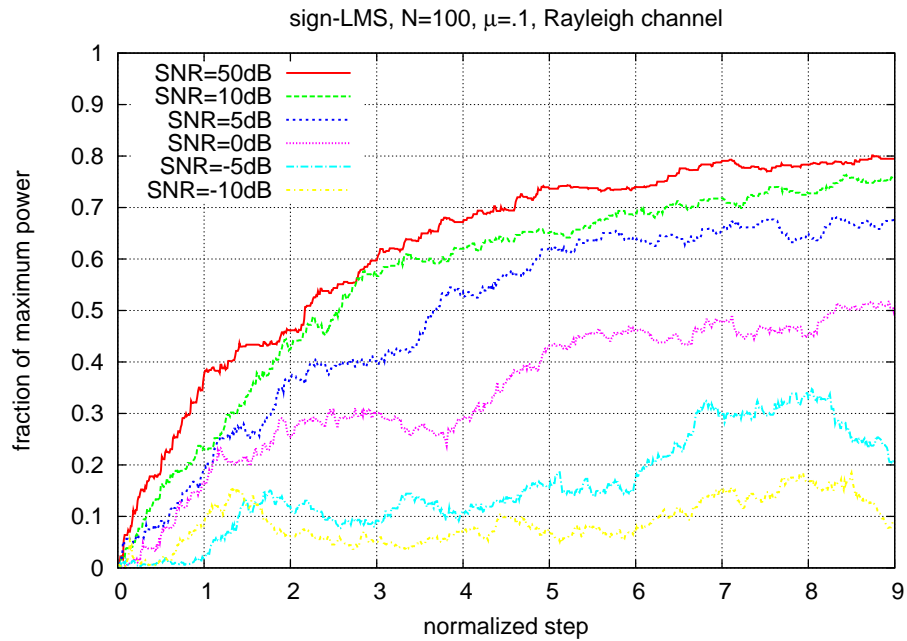
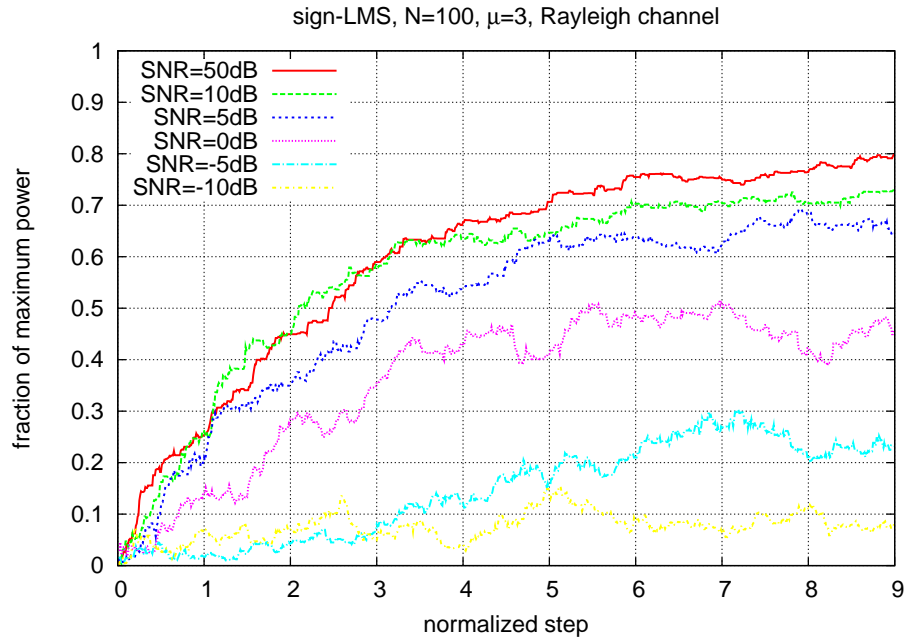


Figure 5.6: (a) sign-LMS convergence for different SNRs, with a step size $\mu = 3$. (b) sign-LMS convergence for different SNRs, with a step size $\mu = .1$.

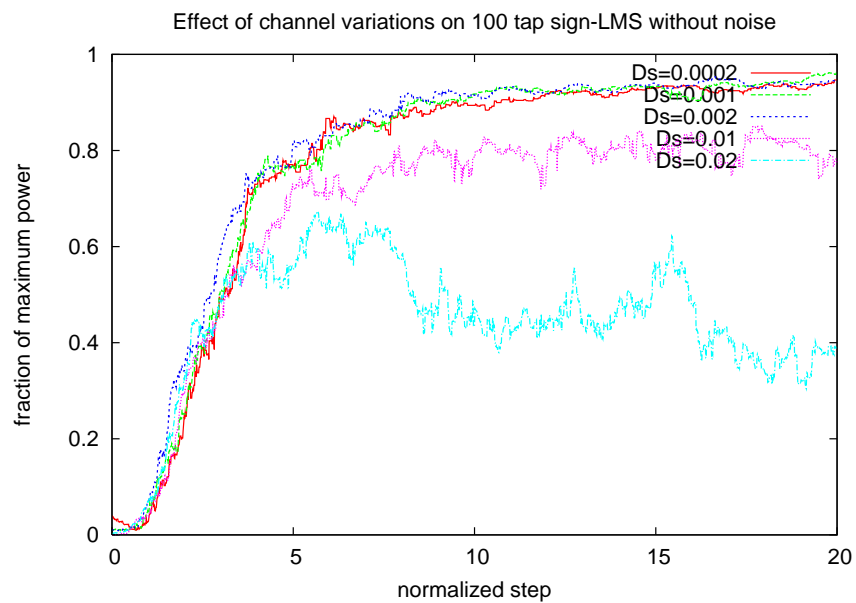
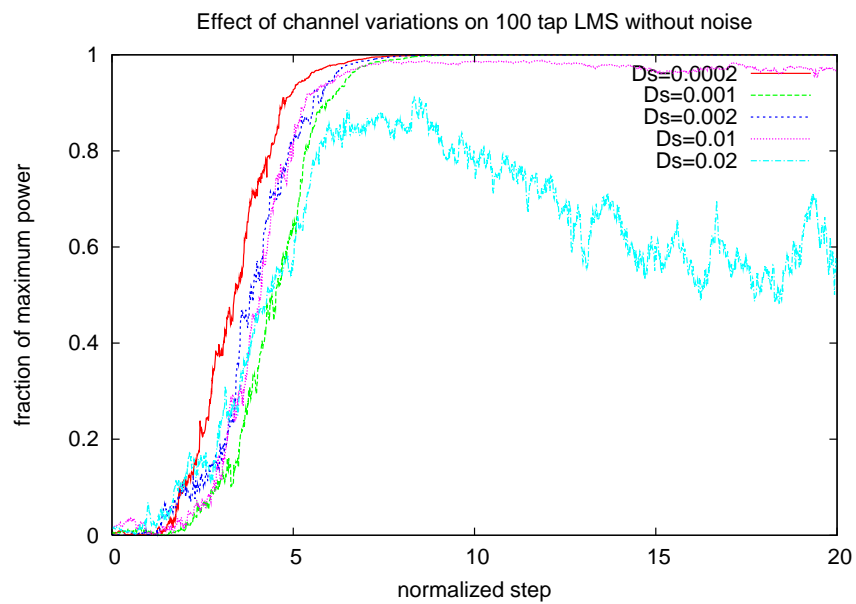


Figure 5.7: (a) NLMS convergence for different Doppler rates (as fractions of the symbol rate). (b) sign-LMS convergence for different Doppler rates (as fractions of the symbol rate).

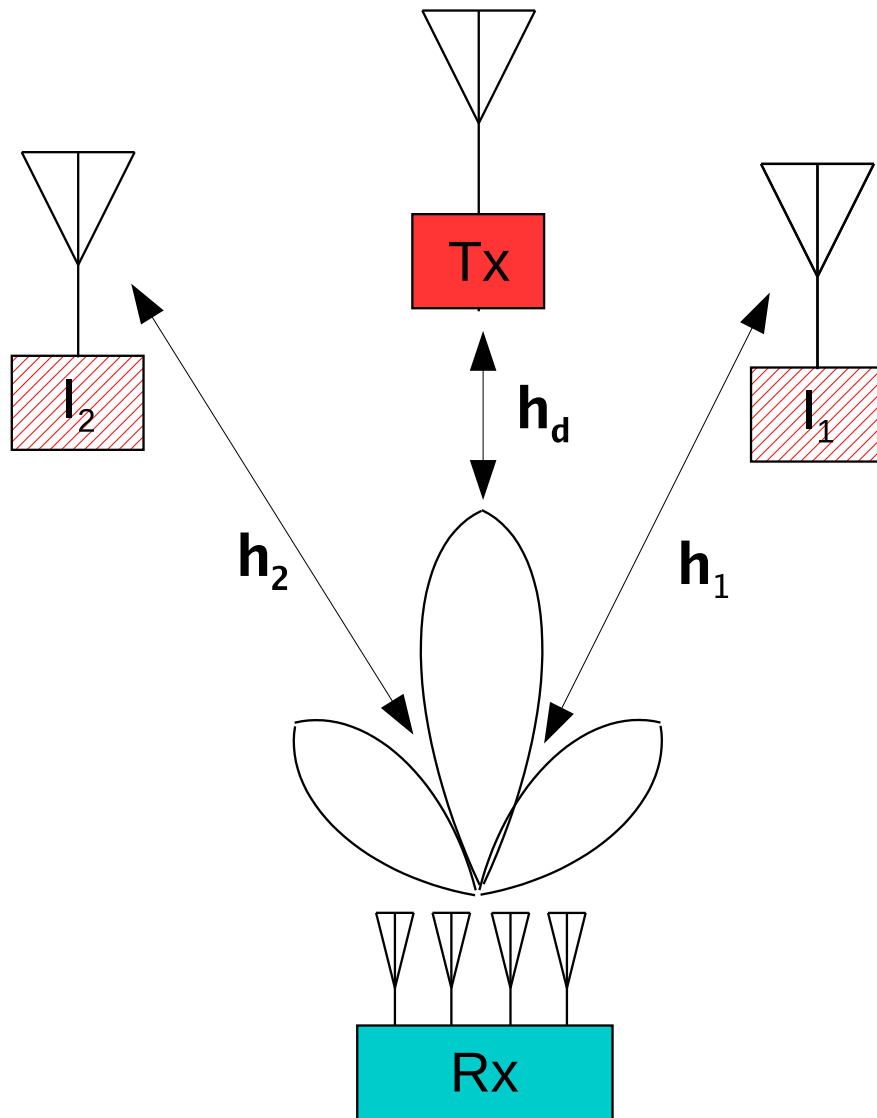


Figure 5.8: An adaptive array receiver must independently estimate the channels from all the transmitters before forming nulls in directions of interfering signals.

computed. When the estimates of all the channel response vectors are available, the beamforming weights can be optimized for different criteria depending on the application (e.g. communication, locationing). In this chapter we focus on maximizing the signal to noise and interference ratio (SINR). The Minimum Variance Distortionless Response (MVDR) beamformer is a popular technique for achieving the maximum SINR [33]. MVDR works as follows:

- 1) The received signal at the output of the adaptive array is a composed of the signal from the desired transmitter as well as the signal from the interferers (see Figure 5.8):

$$y[n] = \mathbf{c}^H (\mathbf{h}_d x_d[n] + \sum_{i=1}^K \mathbf{h}_i x_i[n] + \mathbf{v}[n])$$

where $y[n]$ is the received sequence, \mathbf{c} is the complex beamforming weight vector of size N ($\|\mathbf{c}\| = 1$), \mathbf{h}_d is the complex spatial channel response from the desired transmitter (a vector of size N), $x_d[n]$ is the signal from the desired transmitter, \mathbf{h}_i is the complex spatial channel response from the i^{th} interferer (a vector of size N), $x_i[n]$ is the signal from the i^{th} interferer, $K < N$ is the number of interferers, and $\mathbf{v}[n]$ is a vector of white signals (with variance σ_v) that models the noise at each antenna at the receiver.

- 2) The receiver selects the beamforming weight vector \mathbf{c} to maximize the SINR at the output of the array defined as follows:

$$\begin{aligned} \text{SINR}(\mathbf{c}) &= \frac{|\mathbf{c}^H \mathbf{h}_d|^2}{\sum_{i=1}^K |\mathbf{c}^H \mathbf{h}_i|^2 + E[|\mathbf{c}^H \mathbf{v}[n]|^2]} \\ \Rightarrow \mathbf{c}_{opt} &= \operatorname{argmax}_{\mathbf{c}} \text{SINR}(\mathbf{c}) \quad \text{s.t. } \|\mathbf{c}\| = 1 \end{aligned}$$

The solution to this optimization problem is the Minimum Variance Distortionless Response and takes the following form:

$$\mathbf{c}_{opt} = \frac{\mathbf{R}_{I+N}^{-1} \mathbf{h}_d}{\|\mathbf{R}_{I+N}^{-1} \mathbf{h}_d\|}, \quad \mathbf{R}_{I+N} \triangleq \sum_{i=1}^K \mathbf{h}_i \mathbf{h}_i^H + \sigma_v^2 \mathbf{I}_N \quad (5.7)$$

\mathbf{R}_{I+N} is the spatial interference plus noise correlation matrix (\mathbf{I}_N is an identity matrix of dimension N).

The MVDR beamformer maximizes the SINR. However, in practice, the beamformer will not have access to the exact values of \mathbf{h}_d and $\mathbf{h}_1, \dots, \mathbf{h}_K$. Instead, it has access only to noisy estimates of these quantities $\hat{\mathbf{h}}_d$ and $\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_K$ produced by the LMS algorithm. The resulting SINR will degrade because of the mismatch between the actual channel responses and the estimates²⁶. However, the simulation results in Section 5.2.1 show the channel estimates improving with SNR. That means that the nulls placed in the directions of large interferers will be deeper than those in the direction of small interferers, which is critical

²⁶In Chapter 6, we show that the resulting SINR is more sensitive to mismatches in the interfering channel responses $\mathbf{h}_1, \dots, \mathbf{h}_K$ than the desired channel response \mathbf{h}_d .

since larger interferers have more impact on the SINR. In fact, using this method, the amount of interference rejection is directly proportional to the input SNR (i.e. SNR prior to beamforming)²⁷. The simulation results in Figure 5.9 confirm this observation, which will be proven more rigorously in Chapter 6. In Figure 5.9a, we simulated the interference rejection of 4 interfering signals whose channels were estimated using the LMS algorithm. On the x-axis, we vary the input SNR of one of the interferers (the SNR at which the adaptive filter operates while estimating the channel), and on the y-axis, we plot the output SNR of that interferer after beam-nulling (The different curves represent different number of antennas). First, we see that the interference rejection does not depend on the number of antennas as long as $N > K$. Second, the curves are relatively flat versus the input SNR. This means that the interference rejection is directly proportional to the input SNR. In Figure 5.9b, we repeated the same experiment with quantization noise added to the beamforming weights. The Figure shows that quantization noise becomes the limiting factor at high input SNRs²⁸.

The interference rejection or null depth can be improved by reducing the step size μ in the LMS algorithm (i.e. averaging the results for longer periods). We note, however, that since the channels are estimated one at a time, the process is relatively slow, which means it can tolerate very little channel variations. We showed in Figure 5.7 that a large fraction of the signal power (beamforming gain) can be captured by the LMS algorithm in the presence of fast fading. However, beam-nulling and interference rejection are more sensitive to errors than beamforming as demonstrated in Chapter 6. Figure 5.10 shows the interference rejection for different channel variation (Doppler) rates. The x-axis denotes “time” or the progression of the LMS algorithm. For clarity, we only plot the data after the algorithm reaches a stable state. The results show that even moderate Doppler rates can have a severe impact on interference rejection. The requirements are even more strict for higher data-rates that require higher average SINRs. As result, this architecture is more suitable for fixed wireless systems than for mobile systems. In Section 5.6, we discuss some techniques that utilize hybrid and multibeam architectures to enhance the performance.

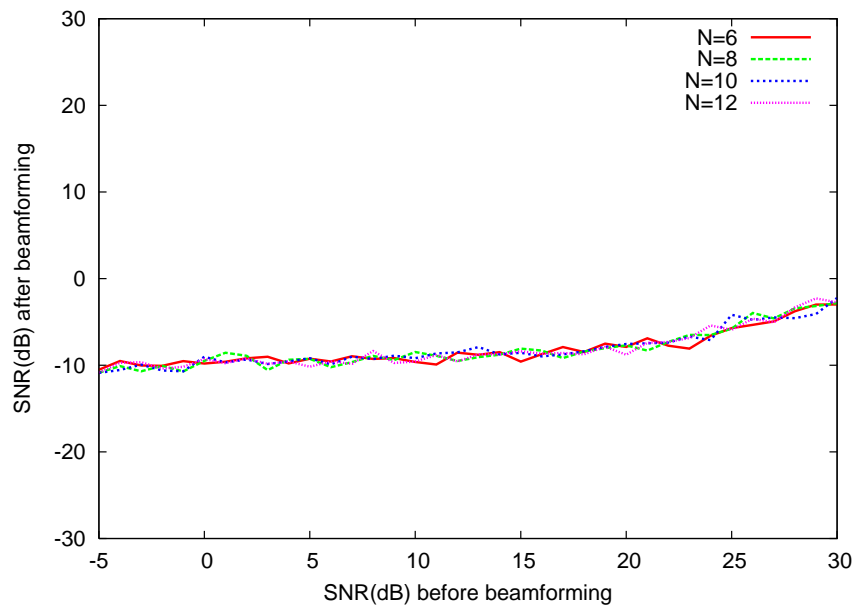
5.4 Adaptive filtering with interference cancellation

In most practical circumstances, there will be other radios that are transmitting at the same time and in the same frequency band. It is usually hard to coordinate the transmission of the different radios especially if the bands are unlicensed. These other radios thus generate interference at the adaptive array, which impedes its ability to estimate the desired channel response(s). One way to deal with these interfering signals is to treat them like noise, and average over longer periods. This, however, can introduce unacceptable delays. Furthermore, unlike noise, the interfering signals at each antenna are spatially correlated. We can take advantage of this to minimize the effects of interference without introducing unnecessary delays (assuming we can estimate the interfering channels prior to the desired channel) by modifying the original algorithm as follows:

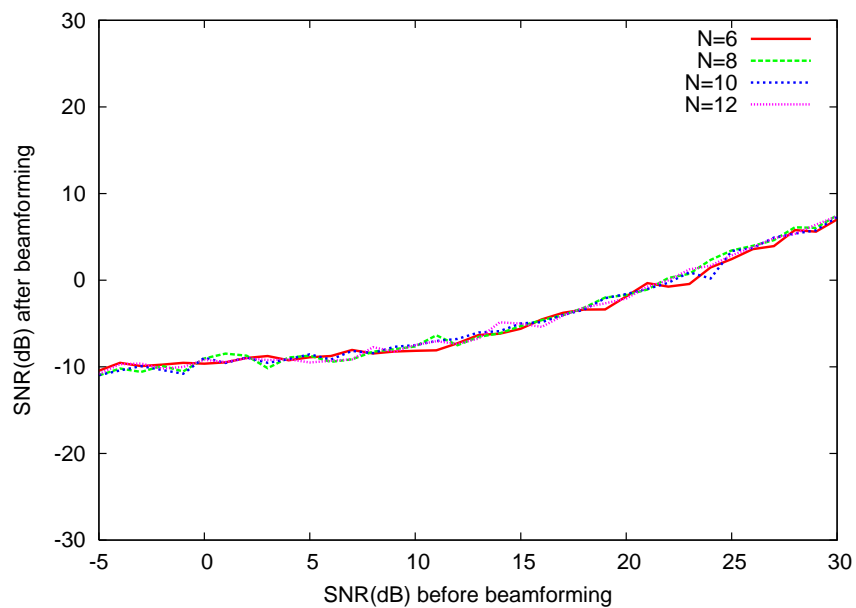
- 0) Let \mathbf{h}_d be the channel response vector to be estimated, and let $\mathbf{h}_1, \dots, \mathbf{h}_K$ for $K < N$

²⁷The SNR referred to here is the interfering signal to noise ratio.

²⁸In Chapter 7, we introduce algorithms for mitigating quantization noise.



(a)



(b)

Figure 5.9: (a) The SNR of the interferer at the output of array (post-beamforming) as a function of the SNR of the interferer at the input of the array (pre-beamforming), for a time-invariant channel and a total of $K = 4$ interferers. The curves represent different numbers of antennas at the receiver. (b) An identical experiment with quantization noise added to the beamforming weights.

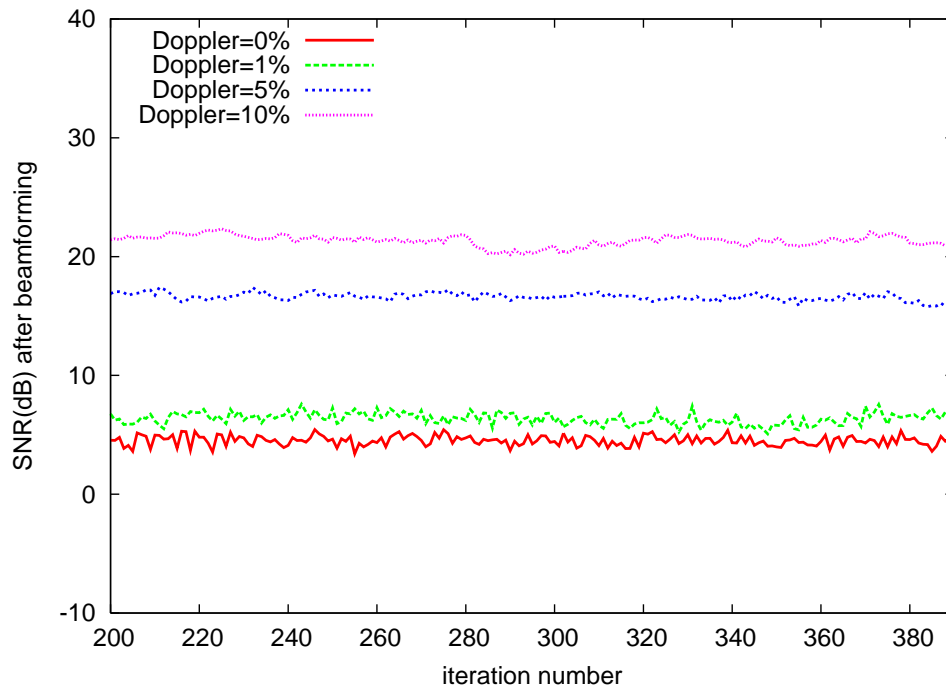


Figure 5.10: The interferer SNR after beamforming as a function of time, for a time-varying channel. The SNR before beamforming is 30dB. The curves represent different Doppler rates as a percentage of the LMS update rate. Quantization noise is also added to the beamforming weights.

be the channel response vectors from K different interfering radios that the adaptive array receiver Rx has already estimated.

- 1) At each step n , Rx generates a random vector $\mathbf{w}[n]$ of length N as described in Section 5.2.
- 2) Instead of applying $\mathbf{w}[n]$ directly to the antenna weights, $\mathbf{w}[n]$ is decomposed into two orthogonal components²⁹, one that is parallel and one orthogonal to the subspace spanned by $\mathbf{h}_1, \dots, \mathbf{h}_K$:

$$\begin{aligned}\mathbf{w}[n] &= \mathbf{w}^\perp[n] + \mathbf{w}^\parallel[n] \\ \mathbf{w}^\perp[n]^H \mathbf{w}^\parallel[n] &= \mathbf{w}^\perp[n]^H \mathbf{h}_1 = \dots = \mathbf{w}^\perp[n]^H \mathbf{h}_K = 0\end{aligned}$$

- 3) The adaptive filter framework in Section 5.2 is used with the new input sequence $\{\mathbf{w}^\perp[n]\}$.

Since the training weight vectors are chosen orthogonal to the subspace spanned by the interfering vectors $\mathbf{h}_1, \dots, \mathbf{h}_K$, the interference disappears at the output of the array and its impact on the channel estimation process becomes negligible. However, even though the modified version of the adaptive filtering algorithm (with interference cancellation) has a very similar performance to the conventional algorithm described Section 5.2, the algorithm does not converge to the desired channel response \mathbf{h}_d , but instead converges to \mathbf{h}_d^\perp (the component of \mathbf{h}_d that is orthogonal to the subspace spanned by $\mathbf{h}_1, \dots, \mathbf{h}_K$). This is not surprising since \mathbf{h}_d^\parallel (the component of \mathbf{h}_d that is parallel to the subspace spanned by $\mathbf{h}_1, \dots, \mathbf{h}_K$) also gets canceled. Simulation results plotted in Figure 5.11 also confirm this observation³⁰.

If this scheme is used to estimate the interference channels $\mathbf{h}_1, \dots, \mathbf{h}_K$ (for $K < N$) one at a time as described in Section 5.3. At each time, we estimate the orthogonal component of each channel vector. The resulting estimates $\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_K$ will approximately form an orthogonal basis of the set of channel vectors $\mathbf{h}_1, \dots, \mathbf{h}_K$. The process is very similar to computing the Gram-Schmidt basis. Since the estimated vectors span the same subspace as the actual interfering vectors, they contain the necessary and sufficient information to form nulls and suppress the interference. In Figure 5.12, we repeated the same simulation in Figure 5.9 using channel estimation with interference cancellation. The interference suppression levels achieved are very similar in both cases.

²⁹These components can be computed via projection.

³⁰The interference cancellation feature imposes additional restrictions on the choice of the training weights. In the absence of interference, the training weights can be as simple as a sequence of independent Bernoulli random variables (i.e. $\{-1, 1\}^N$ or $\{0, 1\}^N$). This property was important because it meant that the channel estimation algorithm is not impacted by the resolution of the beamforming weights. However, when the training sequence is projected onto the subspace orthogonal to the interference subspace, the set of new weights take on a range of values that is continuous. In this case, the resolution of the beamforming weights begin to matter; finite resolution will result in interference leaking into the system, which raises the noise floor and degrades the performance. A full analysis as well as techniques for improving the performance will be presented in Chapters 6 and 7.

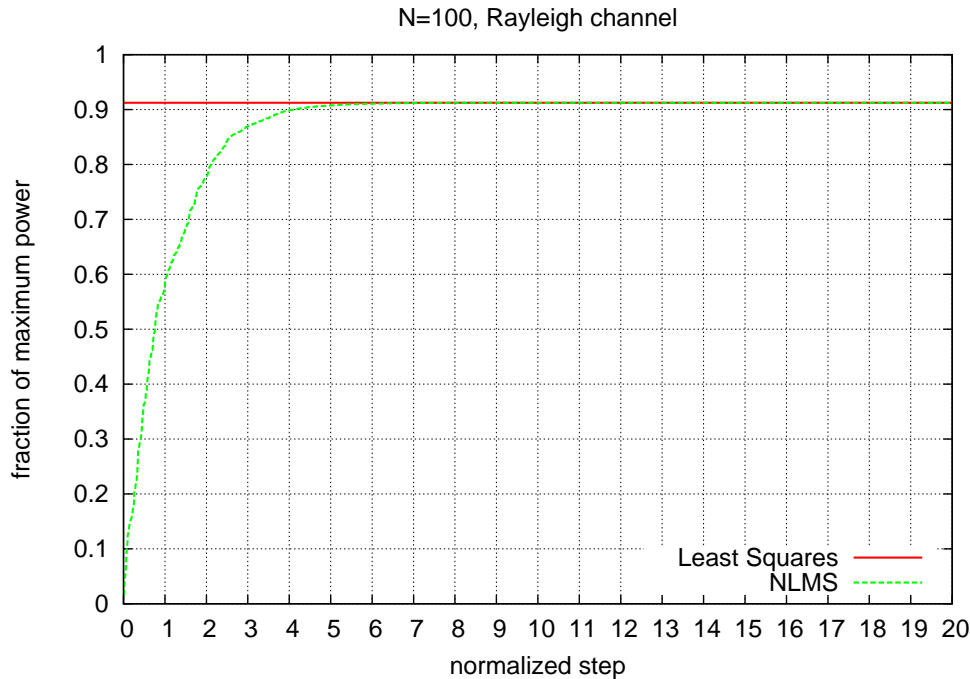


Figure 5.11: The LMS algorithm converges to the component of the channel response that is orthogonal to the interference.

5.5 Frequency synchronization

We showed in Section 5.2 that our adaptive filtering framework is robust to noise and slow fading. However, most standard adaptive filters (e.g. LMS, RLS) require that both ends of the link be locked in frequency to guarantee convergence. Even small synchronization errors or frequency offsets can cause the filter to oscillate. In practice, this requirement would be a significant limitation, since it is difficult to frequency lock prior to beamforming, when the SNR is much lower than after beamforming. To address this problem, we use the differential minimum mean square error (DMMSE) adaptive filtering framework [31]. The DMMSE framework tracks the ratio between two consecutive symbols instead of their absolute values. The intuition behind this is that even in the presence of fast fading and frequency offsets, the channel is static over two consecutive symbols, and thus their ratio remains constant. Therefore, the algorithm need not explicitly track the channel variations.

One approach for dealing with time-variation and frequency offsets is to change the optimization criterion for the adaptive filter. The new criterion, known as the differential MMSE or DMMSE criterion, is a variation of the MMSE and was first introduced in [31] and is defined as:

$$\mathbf{c}_{DMMSE} \triangleq \arg \min_{\mathbf{c}} |(\mathbf{c}^H \mathbf{x}[n-1])d[n] - (\mathbf{c}^H \mathbf{x}[n])d[n-1]|^2, \quad \text{s.t. } E[|\mathbf{c}^H \mathbf{x}[n]|^2] = 1 \quad (5.8)$$

The intuition behind this criterion is that even the most rapidly varying channels can be assumed to be fairly constant over two consecutive symbols. Therefore, by minimizing the

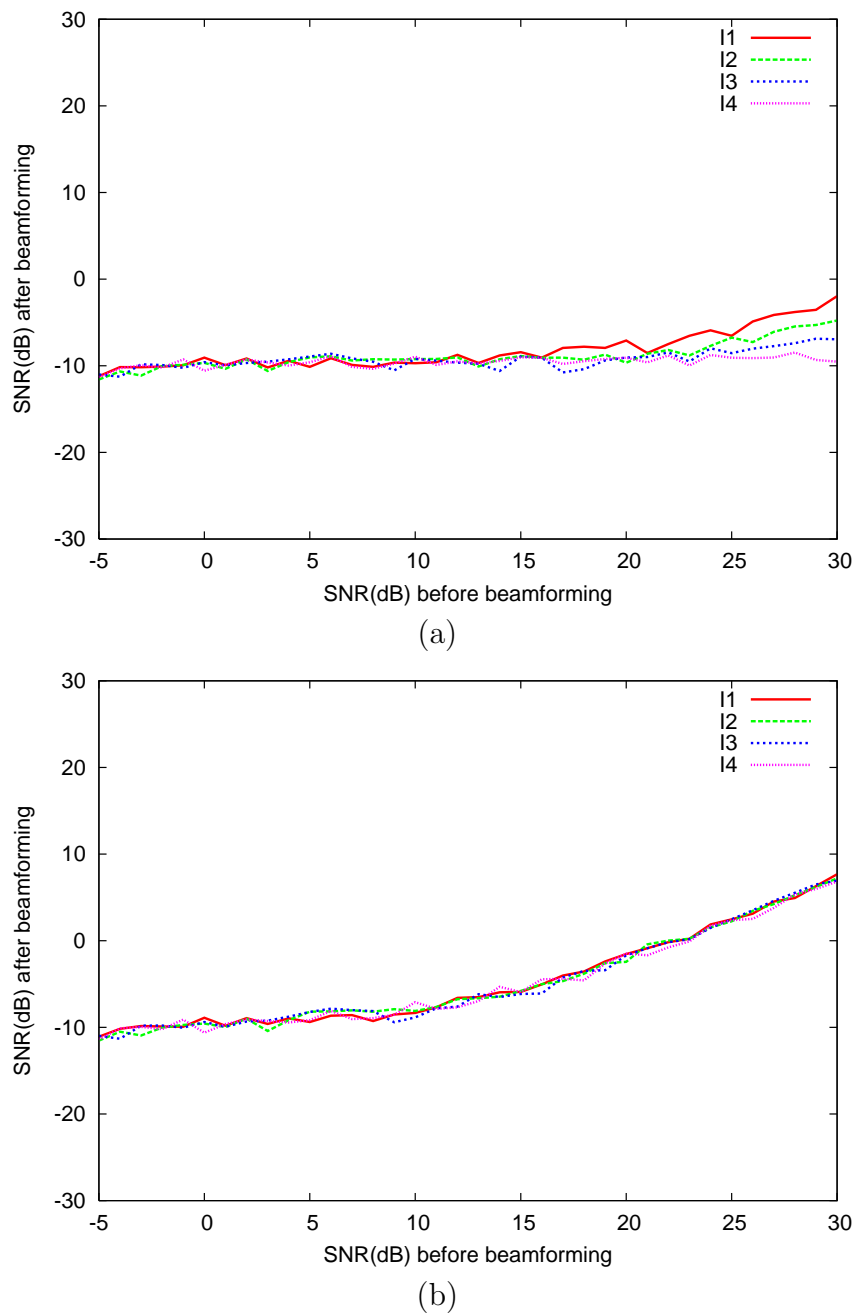


Figure 5.12: (a) The SNR of 4 different interferers, when the Gram-Schmidt method is used to iteratively estimate the channels, as a function of the input SNR. (b) An identical experiment with quantization noise added to the beamforming weights.

error in tracking the *ratio* of two consecutive symbols, the algorithm is spared the task of explicitly tracking the channel variations. Using this definition, the derivation of Equation 5.8 is straight forward. We start by rewriting the error signal $e_{DMMSE}[n]$ as the difference between ratios:

$$e_{DMMSE}[n] = \frac{d[n]}{d[n-1]} - \frac{\mathbf{c}^H \mathbf{x}[n]}{\mathbf{c}^H \mathbf{x}[n-1]} \quad (5.9)$$

Since we want to avoid the trivial solution $\mathbf{c} = 0$, we restrict the amplitude of \mathbf{c} such that $|\mathbf{c}^H \mathbf{x}[n-1]|^2 = 1$. Also, in order to avoid dividing by 0 or a small number, we can rewrite 5.9 as

$$\begin{aligned} \Rightarrow e_{DMMSE}[n] &= \frac{1}{\mathbf{c}^H \mathbf{x}[n-1]} \left(\frac{d[n](\mathbf{c}^H \mathbf{x}[n-1])}{d[n-1]} - \mathbf{c}^H \mathbf{x}[n] \right) \\ &= \frac{1}{d[n-1]\mathbf{c}^H \mathbf{x}[n-1]} (d[n](\mathbf{c}^H \mathbf{x}[n-1]) - d[n-1](\mathbf{c}^H \mathbf{x}[n])) \\ \Rightarrow e_{DMMSE}[n] &= d[n](\mathbf{c}^H \mathbf{x}[n-1]) - d[n-1](\mathbf{c}^H \mathbf{x}[n]) \end{aligned}$$

Since both $d[n]$ and $\mathbf{c}^H \mathbf{x}[n-1]$ are fixed, they will not affect optimization, and thus can be ignored. Therefore, the differential mean square error criterion becomes:

$$\xi_{DMMSE}[n] = |e_{DMMSE}[n]|^2 \quad \text{s.t.} \quad |\mathbf{c}^H \mathbf{x}[n-1]|^2 = 1$$

which is similar to Equation 5.8. The standard adaptive filters (e.g. LMS and RLS) can be easily modified for this criterion. The update equations can be derived similar to Section 5.1.1 by taking the gradient of $\xi_{DMMSE}[n]$. DMMSE based adaptive filters are shown to converge to the scaled version of the optimum MMSE solution [31].

We can implement both LMS and RLS differentially, in order to improve their performance in the presence of frequency offsets and fast channel variations. Figure 5.13 compares the performance of DMMSE-based algorithms and conventional MMSE-based algorithms for different frequency offsets. The graph shows that the DMMSE can achieve more than 60% of the maximum beamforming gain with offsets as large as 15% of the symbol rate. On the other hand, LMS starts oscillating when the offset is more than 1% of the symbol rate.

5.6 Multi-beam (multi-receiver) arrays

Receivers that are equipped with multiple digital baseband signal paths, as described in Section 3.1.1, offer a variety of means to enhance performance. First, the basebands can be used to increase the number of training samples for the adaptive algorithm by processing multiple random beamforming vectors simultaneously. Second, a fraction of the baseband processors can be dedicated to channel estimation, while others can be used to receive data. This scheme allows the array to continue refining the channel estimates after the pilot phase (by using the decoded data as a reference signal). Finally, having multiple basebands enables the beamformer to take advantage of conventional digital beamforming algorithms, which are usually more robust and accurate, to provide better isolation between different beams in a multibeam system and more accurate null-steering. The reason this can be accomplished is that the beamforming weights can be applied with a much higher resolution in the digital

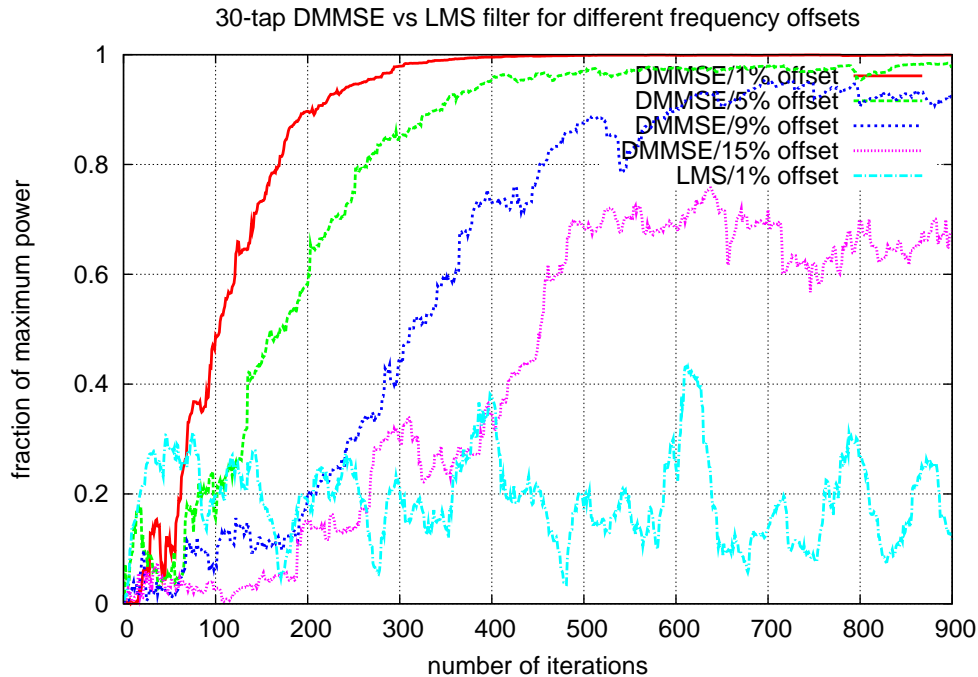


Figure 5.13: DMMSE vs LMS in the presence of frequency offsets (as fractions of the symbol rate).

domain than the analog domain. The use of digital beamforming techniques takes the following general form:

- 1) Let M be the number of digital inputs (ADCs) at the adaptive array receiver ($M < N$), and $\mathbf{y}[n]$ be a complex vector of size M denoting the received (digitized) signal at each of these inputs:

$$\mathbf{y}[n] = \mathbf{f}_d d[n] + \sum_{i=1}^K \mathbf{f}_i d_i[n] + \mathbf{v}[n]$$

where $\mathbf{v}[n]$ is additive white noise, $\mathbf{f}_d, \mathbf{f}_1, \dots, \mathbf{f}_K$ represent the combined channel responses of the desired transmitter (Tx) and interferers at each digital input at the receiver respectively (K is the number of interferers)³¹, $d[n]$ is the received signal from the desired transmitter³² and $d_1[n], \dots, d_K[n]$ are the interfering signals.

- 2) The digital beamforming weight vector \mathbf{u} is chosen to maximize the output SINR given by (assuming $\|\mathbf{u}\| = 1$):

$$\text{SINR}(\mathbf{u}) = \frac{|\mathbf{u}^H \mathbf{f}_d|^2}{\sum_{i=1}^K |\mathbf{u}^H \mathbf{f}_i|^2 + \sigma_v^2}$$

³¹ \mathbf{f}_i represents the combined gain of the channel \mathbf{h}_i and the analog beamforming weight vectors $\mathbf{w}_1, \dots, \mathbf{w}_M$ (a set of analog beamforming weights per digital input). We assume that $\mathbf{w}_1, \dots, \mathbf{w}_M$ are estimated as described in Section 5.2, and remain fixed when the digital beamforming algorithm is running.

³² $d[n]$ can be a known training signal or obtained in a decision directed manner

- 3) The adaptive filter implicitly computes SINR maximizing weights using only knowledge of the desired signal $d[n]$.

The optimum beamforming vector can be computed using the MMSE criterion or DMMSE if channels vary quickly and/or frequency offsets are large [8]. We simulated the performance of both the LMS and DMMSE digital beamforming algorithms. In both cases, we simulated an adaptive array receiver with $M = 8$ digital inputs and $K = 5$ interfering signals. The desired pilot signal power is 20dB below each interferer (thus 27dB below the total interference power, given that there are 5 interferers) and 10dB above the noise floor. Therefore, the pre-beamforming SINR is -27dB, and the maximum post-beamforming SINR, achieved when all interferers are perfectly canceled, is about 19dB. Figure 5.14a shows the post beamforming SINR under the DMMSE algorithm gets very close to the maximum achievable SINR when the channel is time-invariant. Figure 5.14b shows that the interference suppression under DMMSE is a lot more robust than NLMS in the presence of time-varying channels and frequency offsets. In the figure, the SINR under DMMSE is higher than NLMS even though DMMSE was simulated under channels with more than 10 times the Doppler spread and a frequency offset that is 20% of the symbol rate³³.

5.7 Transmit beamforming

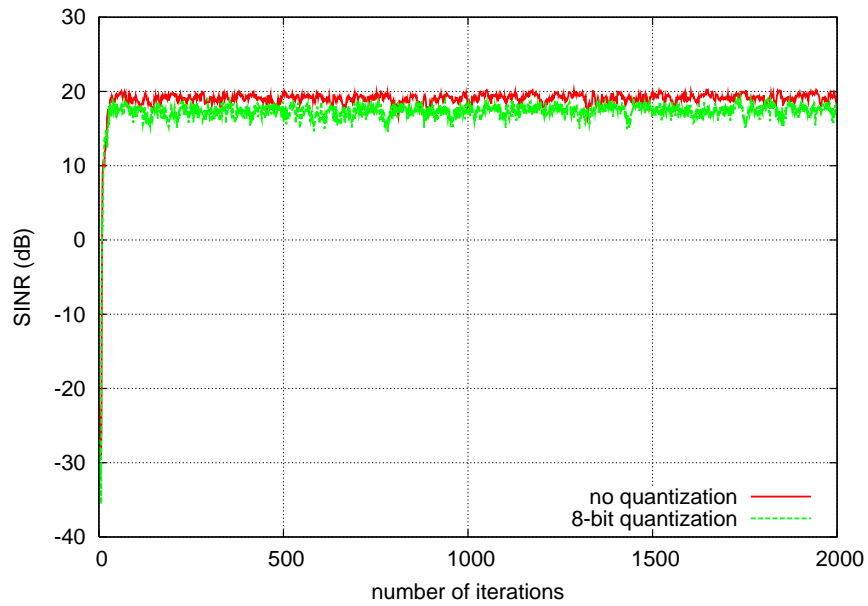
In practice, the optimum array weights for transmit beamforming can be directly computed from the optimum receive weights if there is channel reciprocity. Channel reciprocity holds when the uplink (transmit) and the downlink (receive) are on the same frequency band, or if the environment is line of sight. In other cases, the two channel responses must be estimated independently. In order to estimate the optimum transmit beamforming weights, the same adaptive framework can be used. However, in this case, the array transmits the pilot weights on the “uplink” and receives the feedback from the remote receiver on the “downlink” [8]. This scenario is shown in Figure 5.15. Note that the remote receiver does not need to know the random sequence used by the transmitter array. It simply quantizes the received signal and sends it back on the downlink.

In general, receive beamforming is usually more desirable because it does not require explicit feedback from the target node, which incurs additional latency and overhead³⁴. On the other hand, taking advantage of channel reciprocity requires very careful calibration between the transmitters and receivers on the array. Any mismatch between the transmit and receive circuits can significantly degrade the transmit beamforming performance, especially null-steering³⁵. In practice, this is implemented with an external circuit that injects (and extracts) the same signal into each transmit and receive pair in order to synchronize their phase and amplitude responses. This process is repeated periodically in order to track

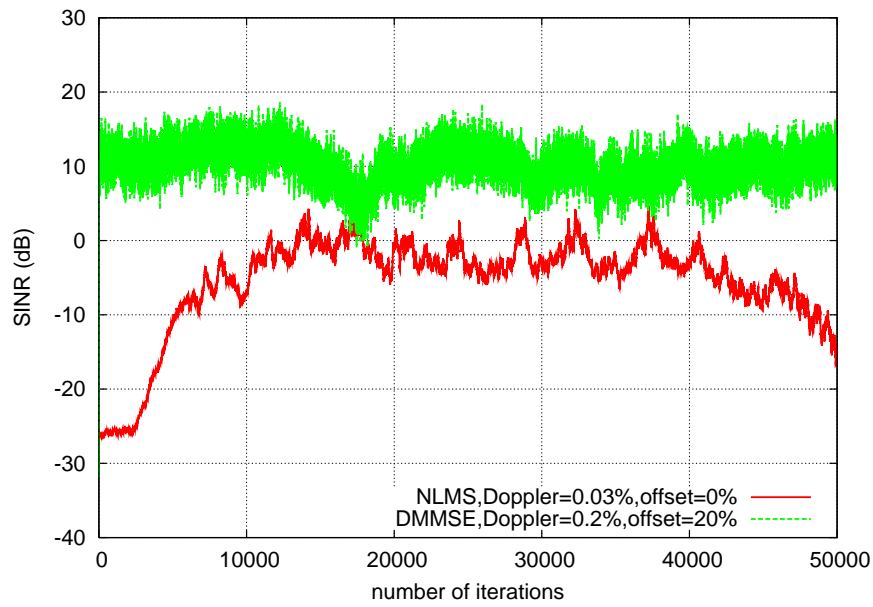
³³The frequency accuracy of most crystal oscillators found in commodity radios is of the order of tens of parts per million (ppm). This results in an average frequency offset of the order of tens of KHz when the carrier frequency is of the order of 1GHz. If the symbol rate is 1MHz, then the frequency offset will be several percent as a fraction of the symbol rate.

³⁴The extra latency can be very costly when the channels are time varying.

³⁵Chapter 6 provides a more rigorous analysis on the impact of phase and amplitude mismatch.



(a)



(b)

Figure 5.14: (a) SINR as a function of time for a time-invariant channel, with the DMMSE algorithm used at the receiver. (b) The SINR as a function of time for time varying channels. For the DMMSE algorithm, the Doppler spread is 0.2% of the symbol rate and the frequency offset is 20%. For the NLMS algorithm the Doppler spread is 0.03% of the symbol rate with no frequency offset. In both Figures, the number of digital inputs at the receiver is $M = 8$, the number of interferers $K = 5$, and the average desired pilot signal power at the input is 10dB above the noise level while each interfering signal is 30dB above the noise level.

thermal variations³⁶.

5.8 Conclusion

In this chapter, we considered the problem of beamforming for RF and hybrid multichannel radio architectures. We proposed a framework based on both conventional and advanced adaptive filtering techniques for robust adaptation in adverse conditions. We showed that using a system identification formulation of adaptive filtering, we can efficiently estimate the beamforming weights even in noisy environments. The estimation of beamforming weights can also occur in the presence of known interferers by appropriately choosing the training weights. We also showed that by adopting the differential MMSE criterion, the beamforming framework can significantly increase its resiliency towards synchronization errors and fast fading channels. The RF beamforming framework can be further augmented with digital beamforming (for hybrid architectures) to further enhance the performance (especially for interference cancellation).

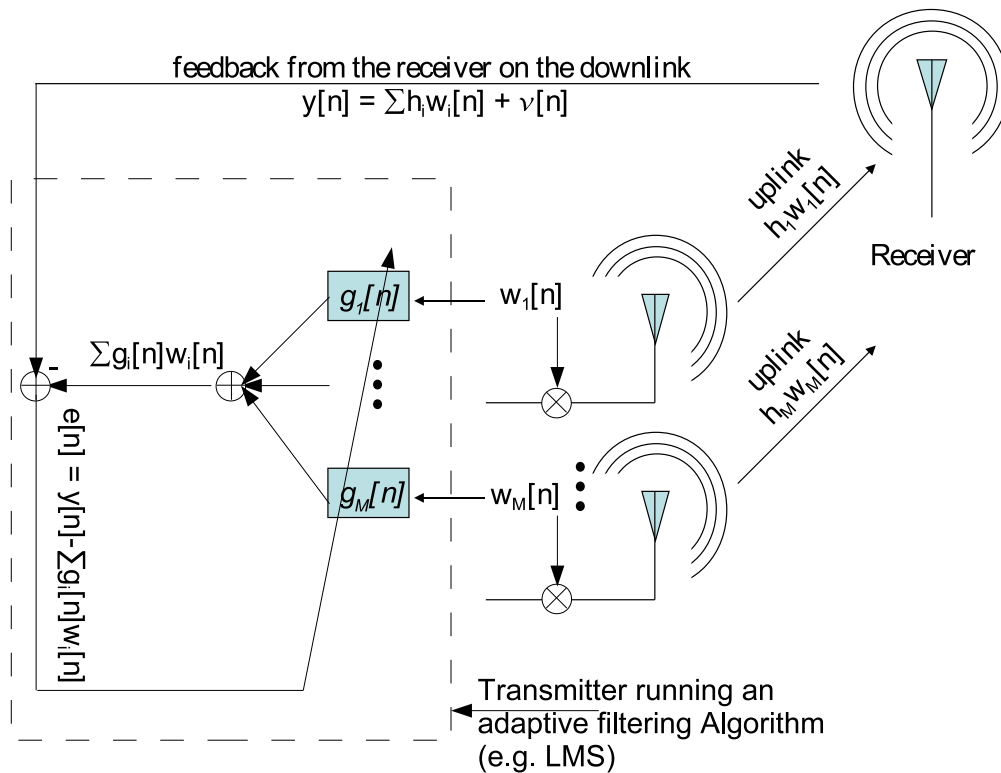


Figure 5.15: Transmit beamforming scenario: 1) A multi-antenna transmitter sends a known signal on the uplink. 2) The receiver sends back a channel estimate on the downlink. 3) The transmitter uses the estimate as input to an adaptive filter.

³⁶Thermal variations are usually very slow, and thus, calibration does not need to be performed very often.

Chapter 6

Beamforming performance with weight errors

Adaptive array systems improve the reliability of a wireless communication link by steering the beam in the direction of the desired signal (beamforming) and away from interfering signals by placing nulls in their directions (beam-nulling or null-steering), and thus maximizing the signal to noise and interference ratio (SINR). In order to achieve the maximum SINR, however, very precise control over the complex weights (both phases and amplitudes) applied to the signals at each element in the array is required. However, various factors such as finite resolution, noise, mismatch in circuit elements, and channel uncertainty limit the precision that can be achieved in practice. Many of these error sources are random, and cannot be compensated for using pre-calibration or adaptive signal processing techniques. These errors will degrade the maximum achievable SINR by degrading the strength of the desired signal and the interference rejection. In this chapter, we examine the impact of phase and amplitude errors on the array performance. Section 6.1 analyzes the beamforming gain (desired signal power) as a function of phase errors. In this section, we derive expressions for the expected (average) degradation in the array gain and establish lower bounds for the worst case scenario. We show that the beamforming gain is robust to errors in the beamforming weight vector, and the average and worst case loss are both independent of the number of antennas. Section 6.2 introduces amplitude errors and focuses on interference rejection. Like the degradation in the array gain, the degradation in interference rejection depends only on the magnitude of errors in the beamforming weights and is independent of the number of antennas. However, unlike the array gain, the interference rejection is a lot more sensitive to errors in the beamforming weight vector. In both sections, we provide both mathematical proofs and simulation results that characterize the array performance as a function of errors in the beamforming weight vector.

The work presented in this chapter was done in collaboration with Mark Johnson, Raghuraman Mudumbai, and Upamanyu Madhow [6, 7].

6.1 Beamforming

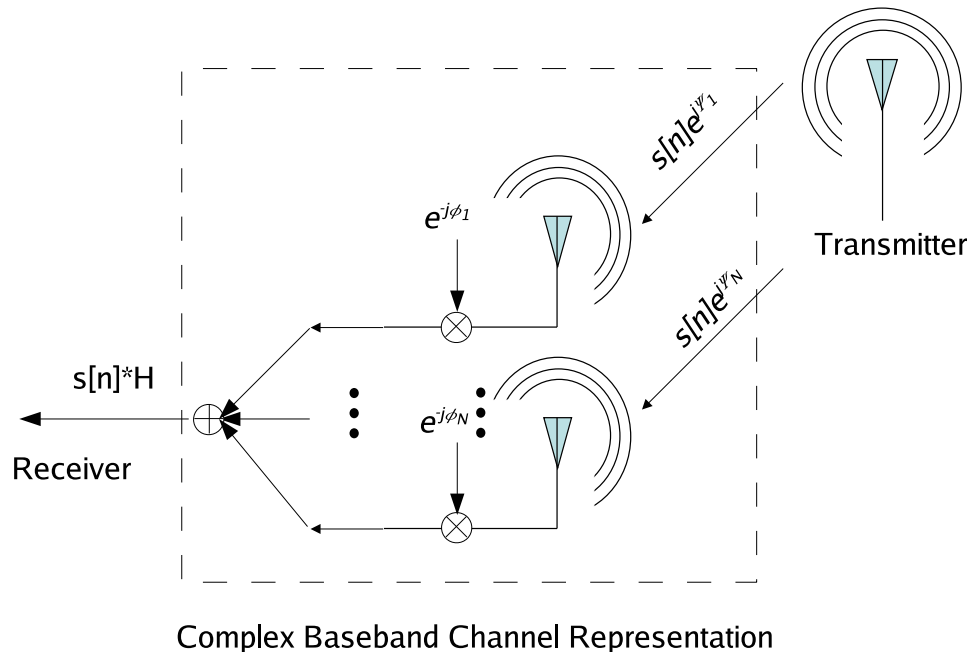


Figure 6.1: A communication system with an N -antenna phased array receiver and a single antenna transmitter, which transmits a narrowband signal $s[n]$. $s[n]$ arrives at antenna i at the receiver with a phase-shift ψ_i . The receiver applies a phase shift of ϕ_i at each antenna and sums the signals.

Consider the N -antenna beamforming receiver shown in Figure 6.1. A single antenna transmitter transmits a narrowband signal $s[n]$. The transmitted signal arrives at different antenna elements with different delays, which for narrowband signals can be approximated with phase shifts [58]. Let $e^{j\psi_i}$ be the normalized channel response at antenna i , which represents the phase delay¹. Antenna i will then apply a phase-shift ϕ_i to the incoming signal. Therefore, the overall complex (baseband) channel response H at the output of the array is given by:

$$H = \sum_{i=1}^N e^{j(\psi_i - \phi_i)}$$

The maximum gain $|H_{max}|^2 = N^2$ is realized by choosing $\phi_i^{opt} = \psi_i$.

¹for simplicity, we assume that signal arrives at the different receivers in the array with the same amplitude. This assumption is true for an array of antennas with uniform beam patterns in a line of sight environment. In practice, however, the signal will arrive at different antennas with different amplitudes due to channel fading and non-uniform antenna patterns. In Section 6.2, we extend the analysis to include amplitude variation as well.

6.1.1 Analysis of Imperfect Phase Shifts

Phase errors are usually difficult to predict and can be time varying and are best modeled as random variables:

$$\hat{\phi}_i = \phi_i^{opt} + \delta_i$$

We will assume that $\delta_i \sim U[-\delta_{max}, \delta_{max}]$, where $0 \leq \delta_{max} \leq \pi$ is an upper bound on the amplitude of phase deviation². Furthermore, we assume that the errors are *i.i.d* across different antennas. In this case the channel response at the output of the array becomes:

$$\hat{H}_{max} = \sum_{i=1}^N e^{j\delta_i} = \sum_{i=1}^N \cos(\delta_i) + j \sum_{i=1}^N \sin(\delta_i)$$

Our goal is to characterize the effect of the phase errors on the square magnitude of the channel response $E[|\hat{H}_{max}|^2]$. To simplify the analysis, let $X_i = \cos(\delta_i)$ and $Y_i = \sin(\delta_i)$:

$$\begin{aligned} \mu_X = E[X_i] &= E[\cos(\delta_i)] = \frac{1}{2\delta_{max}} \int_{-\delta_{max}}^{\delta_{max}} \cos(x) dx = \frac{1}{\delta_{max}} \int_0^{\delta_{max}} \cos(x) dx = \frac{\sin(\delta_{max})}{\delta_{max}} \\ \mu_{X^2} = E[X_i^2] &= \frac{1}{2\delta_{max}} \int_{-\delta_{max}}^{\delta_{max}} \cos^2(x) dx = \frac{1}{\delta_{max}} \int_0^{\delta_{max}} \cos^2(x) dx \\ &= \frac{1}{2\delta_{max}} \int_0^{\delta_{max}} (1 + \cos(2x)) dx = \frac{1}{2} + \frac{\sin(2\delta_{max})}{4\delta_{max}} \\ \mu_Y = E[Y_i] &= E[\sin(\delta_i)] = 0 \quad (\text{by symmetry}) \\ \mu_{Y^2} = E[Y_i^2] &= \frac{1}{2\delta_{max}} \int_{-\delta_{max}}^{\delta_{max}} \sin^2(x) dx = \frac{1}{\delta_{max}} \int_0^{\delta_{max}} \sin^2(x) dx \\ &= \frac{1}{2\delta_{max}} \int_0^{\delta_{max}} (1 - \cos(2x)) dx = \frac{1}{2} - \frac{\sin(2\delta_{max})}{4\delta_{max}} \end{aligned}$$

Now, we can rewrite the expression for the channel response as:

$$\begin{aligned} \hat{H}_{max} &= \sum_{i=1}^N X_i + j \sum_{i=1}^N Y_i \\ \Rightarrow |\hat{H}_{max}|^2 &= \left(\sum_{i=1}^N X_i \right)^2 + \left(\sum_{i=1}^N Y_i \right)^2 = \sum_{k=1}^N \sum_{l=1}^N (X_k X_l + Y_k Y_l) \\ \Rightarrow E[|\hat{H}_{max}|^2] &= \sum_{k=1}^N \sum_{l=1}^N (E[X_k X_l] + E[Y_k Y_l]) \\ E[X_k X_l] &= \begin{cases} E[X_k]E[X_l] = \mu_X^2 & \text{when } k \neq l \text{ (using independence)} \\ E[X_k^2] = \mu_{X^2} & \text{when } k = l \end{cases} \end{aligned}$$

²We assume a uniform distribution to simplify the calculations. Note that no assumptions were made regarding the geometry of the array or the direction of arrival, so the result holds for an arbitrary array.

$$\begin{aligned}
E[Y_k Y_l] &= \begin{cases} E[Y_k]E[Y_l] = \mu_Y^2 & \text{when } k \neq l \text{ (using independence)} \\ E[Y_k^2] = \mu_{Y^2} & \text{when } k = l \end{cases} \\
&\Rightarrow E[|\hat{H}_{max}|^2] = (N^2 - N)(\mu_X^2 + \mu_Y^2) + N(\mu_{X^2} + \mu_{Y^2}) \\
&= (N^2 - N) \left(\frac{\sin^2(\delta_{max})}{\delta_{max}^2} \right) + N = (N^2) \left(\frac{\sin^2(\delta_{max})}{\delta_{max}^2} \right) + N \left(1 - \frac{\sin^2(\delta_{max})}{\delta_{max}^2} \right)
\end{aligned}$$

If we normalize $E[|\hat{H}_{max}|^2]$ by dividing by the maximum value $|H_{max}|^2 = N^2$, we obtain:

$$\begin{aligned}
\Phi_N(\delta_{max}) &= \frac{E[|\hat{H}_{max}|^2]}{N^2} = \frac{\sin^2(\delta_{max})}{\delta_{max}^2} + \frac{1}{N} \left(1 - \frac{\sin^2(\delta_{max})}{\delta_{max}^2} \right) \\
&\Rightarrow \Phi(\delta_{max}) = \lim_{N \rightarrow \infty} \Phi_N(\delta_{max}) = \frac{\sin^2(\delta_{max})}{\delta_{max}^2}
\end{aligned}$$

Thus, the loss in gain due to errors in the phase shifts is bounded by a quantity that depends only on the distribution of the errors, and not on the number of antennas.

A second method of proving a lower bound on the array gain is by using the mean of the random variable, which is often easier to compute, instead of the mean of the square of the random variable. By Jensen's inequality, the square of the mean of a random variable is less than or equal to the mean of its square:

$$E[X]^2 \leq E[X^2]$$

for any random variable X . More generally:

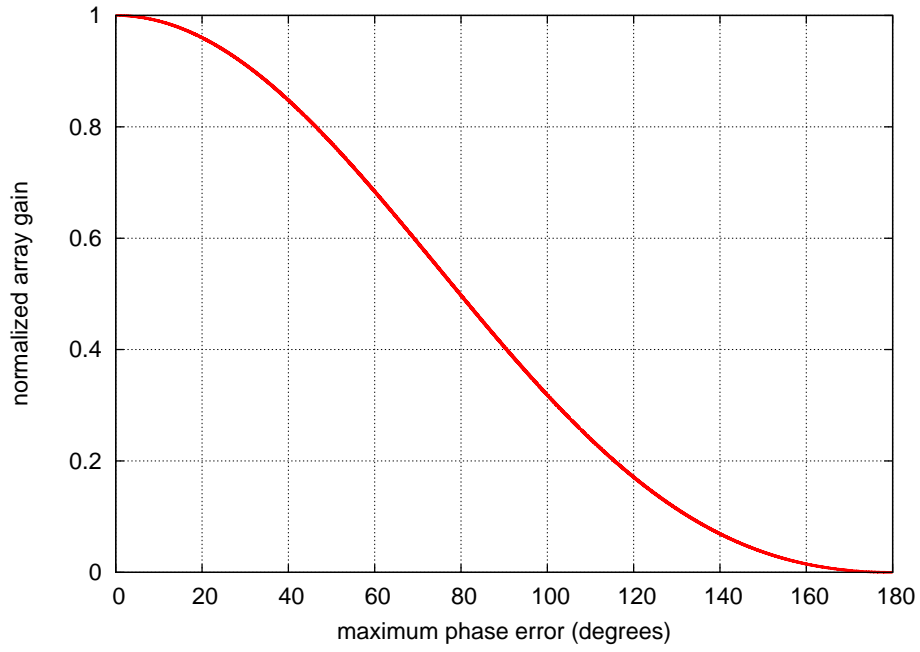
$$f(E[X]) \leq E[f(X)] \quad \text{when } f(\cdot) \text{ is a convex function.}$$

Using this fact, and the expected value of the channel response, we see that:

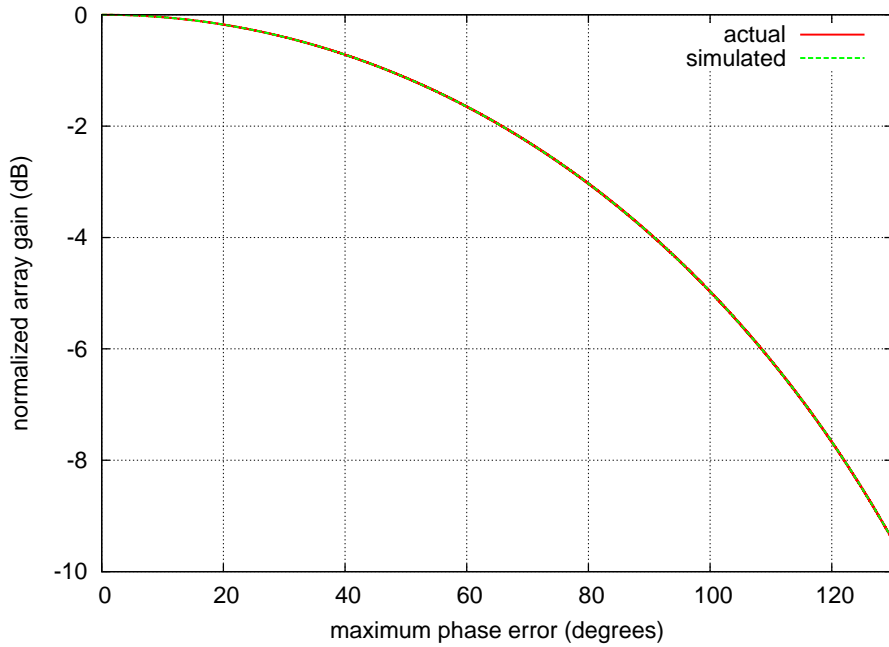
$$\begin{aligned}
E[\hat{H}_{max}] &= E \left[\sum_{i=1}^N X_i + j \sum_{i=1}^N Y_i \right] = N\mu_X = N \frac{\sin(\delta_{max})}{\delta_{max}} \\
&\Rightarrow E[|\hat{H}_{max}|^2] \geq E[\hat{H}_{max}]^2 = N^2 \left(\frac{\sin(\delta_{max})}{\delta_{max}} \right)^2 \\
\Phi_N(\delta_{max}) &= \frac{E[|\hat{H}_{max}|^2]}{N^2} \geq \left(\frac{\sin(\delta_{max})}{\delta_{max}} \right)^2 = \Phi(\delta_{max})
\end{aligned}$$

6.1.2 Simulation Results

Figure 6.2(a) shows a plot of $\Phi(\delta_{max})$ for $0 \leq \delta_{max} \leq \pi$. Figure 6.2(b) shows the same function in dB scale. Figure 6.2(b) also shows that the calculated array gain closely matches simulation results. Figure 6.3 shows that the actual distribution of the phase errors has little impact on the loss in array gain. Notice that using a single bit of phase resolution corresponds to $\delta_{max} = 90^\circ = \frac{\pi}{2}$, and $\Phi(\frac{\pi}{2}) = \left(\frac{2}{\pi}\right)^2 \approx .4 \approx -3.9dB$. We expect the bound to become tighter as N increases, due to the law of large numbers. The graphs in Figures 6.4/6.5 show the loss in array gain as a result of quantizing the phase to one and two bits



(a)



(b)

Figure 6.2: (a) The normalized array gain $\Phi(\delta_{max})$ as a function of the maximum phase error δ_{max} . (b) The normalized array gain in dB scale, $10 \log \Phi(\delta_{max})$. The plot shows both the calculated gain and the simulated gain for a 10000 element array.

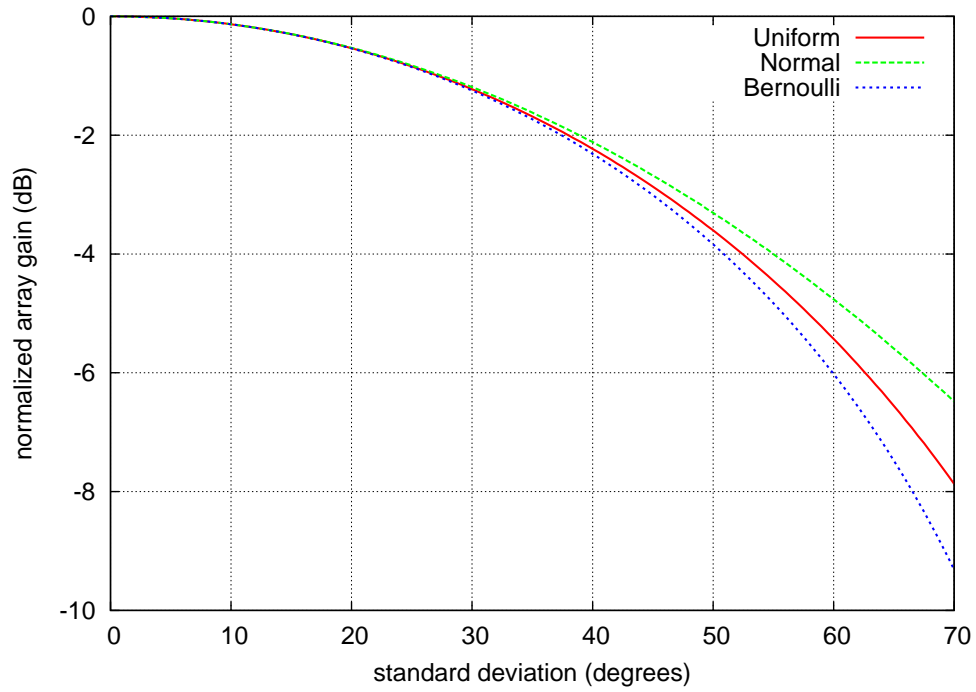


Figure 6.3: Simulated normalized array gain (dB) for a 10000 element array for different phase error distributions (Uniform, Gaussian, and Bernoulli). For a uniform distribution, the standard deviation is $\sigma_\delta = \delta_{max}/\sqrt{3}$.

for different array sizes. Figures 6.6ab show that quantization does not increase the width of the main lobe. Also, notice that when $\delta_{max} = 180^\circ = \pi$, which corresponds to completely randomizing the phase of each antenna, the normalized array response $\Phi_N(\pi) = \frac{1}{N}$, which reduces the array gain to that of an omni-directional antenna. So a simple way of creating an omni-like beam pattern without reducing the radiated power is to choose the phases randomly.

6.1.3 Worst case performance analysis

In Sections 6.1.1 and 6.1.2 we analyzed the average (expected) array gain in the presence of phase errors. In this section, we establish a lower bound on the absolute worst case scenario. Unlike the average case scenario, the worst scenario is a lot more difficult to analyze especially if the errors are random and cannot be predicted or controlled by the system. In this case, the gain can take on any value regardless of how unlikely the event might be. For example, if the phase error, or more generally, the weight error can take on any arbitrary value (e.g. a Gaussian distribution), then the gain be arbitrarily small with a nonzero probability. Therefore, to simplify the analysis, we will assume bounded and controlled phase errors, which is the case for quantization errors. In the extreme case, where the beamforming weights can take only a single value (e.g. 1), the worst case gain can be as low as zero (occurs when the channel response vector $[e^{j\psi_1}, \dots, e^{j\psi_N}]^\top$ is orthogonal to the $[1, 1, \dots, 1]^\top$). We will show that as long as the beamforming weights can take at least two distinct values, then the array gain in the worst case will still be a constant fraction of the maximum gain independent of the number of antennas.

Let a_0 and $a_1 \neq a_0$ be the two possible complex values for the beamforming weights. We want to show that the ratio of the lower bound minimum gain to the maximum gain is proportional to $|a_1 - a_0|^2$. In order to do that, let us first consider two special cases. In Case 1, we can only turn antennas on or off (i.e. the possible beamforming weights are 1 or 0). In Case 2, we assume that the antennas are always on and the possible beamforming weights are $a_0 = 1$ and $a_1 = -1$.

Let $\Psi = \{\psi_1, \dots, \psi_N\}$ be the set of phase responses at the N antennas such that $0 \leq \phi_i < 2\pi \quad \forall_{1 \leq i \leq N}$. If we divide the interval $[0, 2\pi]$ into 8 equal and contiguous sub-intervals of length $\pi/4$, then we know that one of these intervals will contain at least $N/8$ of the phase responses³. Let $\Psi_{max=\pi/4} \subset \Psi$ be the subset of phase responses (or antennas) that fall in that interval. If we choose a beamforming vector \mathbf{w} such that $\mathbf{w}_i = 1$ if $\psi_i \in \Psi_{max=\pi/4}$ and $\mathbf{w}_i = 0$ otherwise (Case 1), then the power at the output of the array:

$$\begin{aligned} E[|H|^2] &= \left| \sum_{\psi \in \Psi_{max=\pi/4}} e^{j\psi} \right|^2 = \left| \sum_{\psi \in \Psi_{max=\pi/4}} \cos(\psi) \right|^2 + \left| \sum_{\psi \in \Psi_{max=\pi/4}} \sin(\psi) \right|^2 \\ &\geq \left| \sum_{\psi \in \Psi_{max=\pi/4}} \cos(\psi) \right|^2 \geq (N/8)^2 \cos(\pi/4)^2 = N^2/128 \end{aligned}$$

³The worst case scenario occurs when the phases are uniformly distributed, in which case each interval will contain either $\lfloor N/8 \rfloor$ or $\lceil N/8 \rceil$. This is sometimes referred to as the Pigeonhole Principle [49].

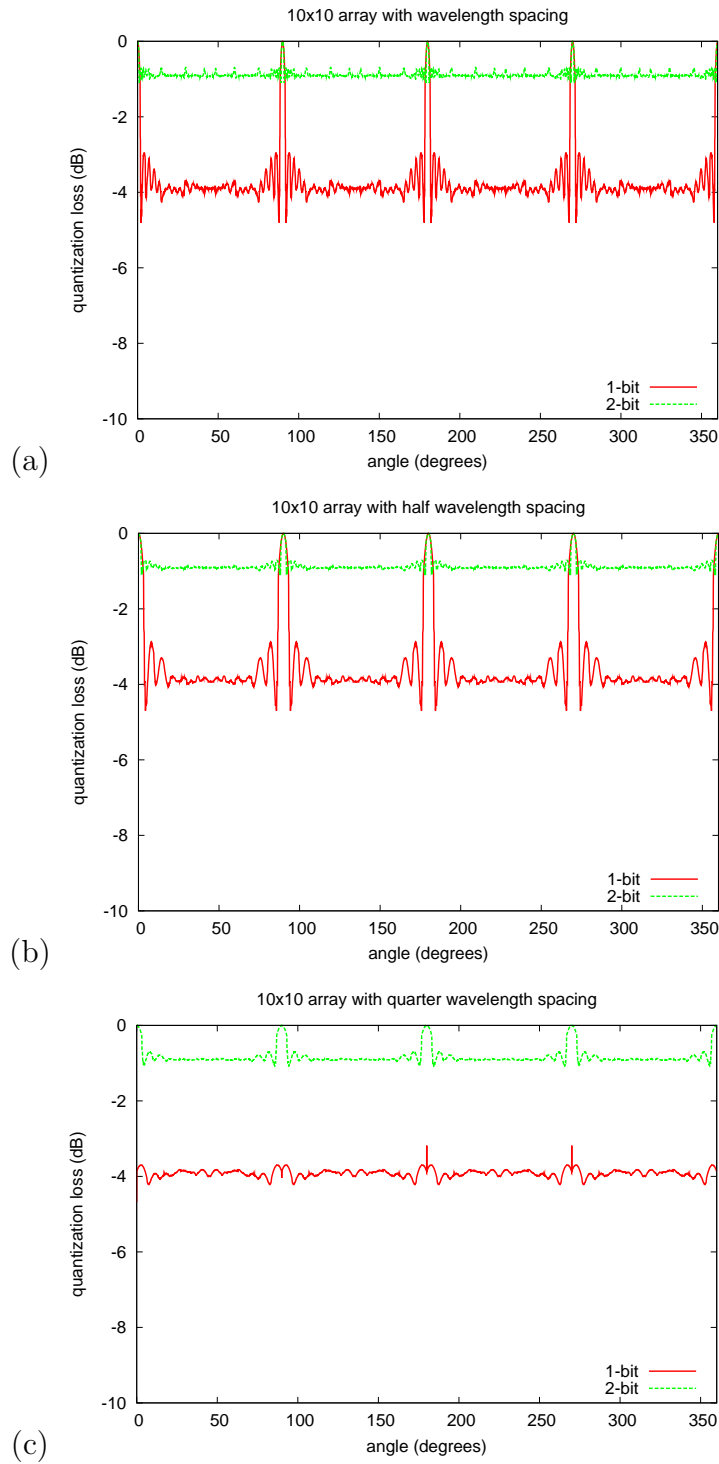


Figure 6.4: (a)-(f) Array loss as a result of phase error (quantization) for uniform 10×10 square array with different inter-element spacings.

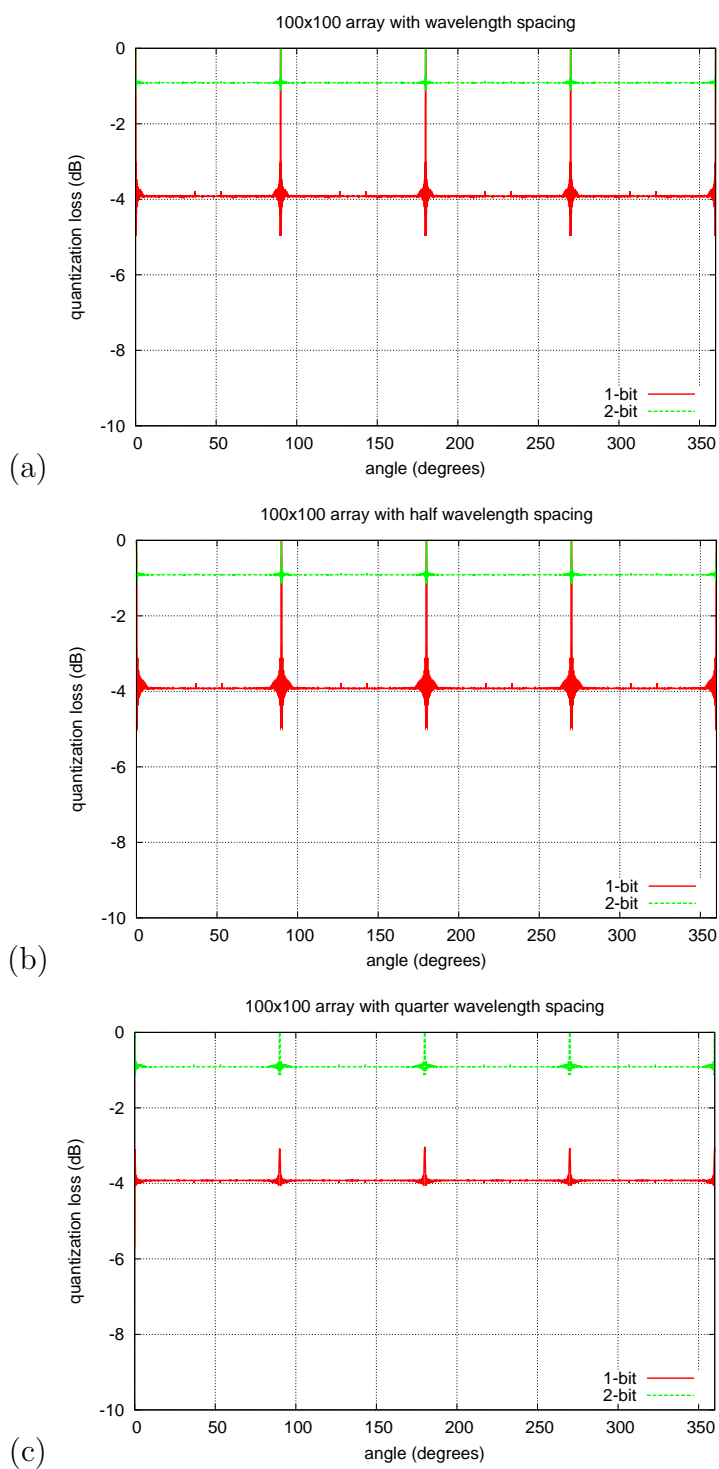
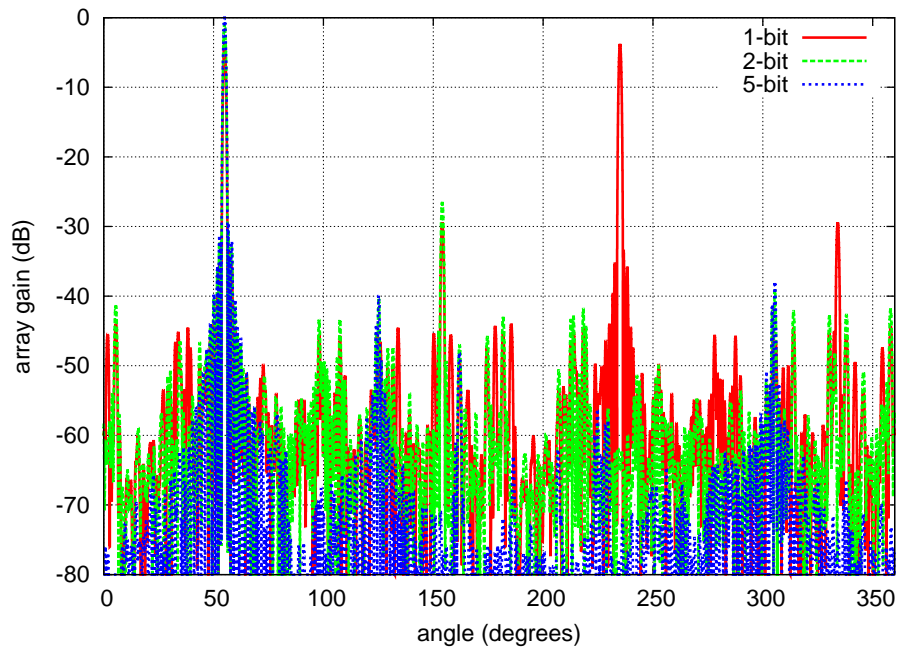
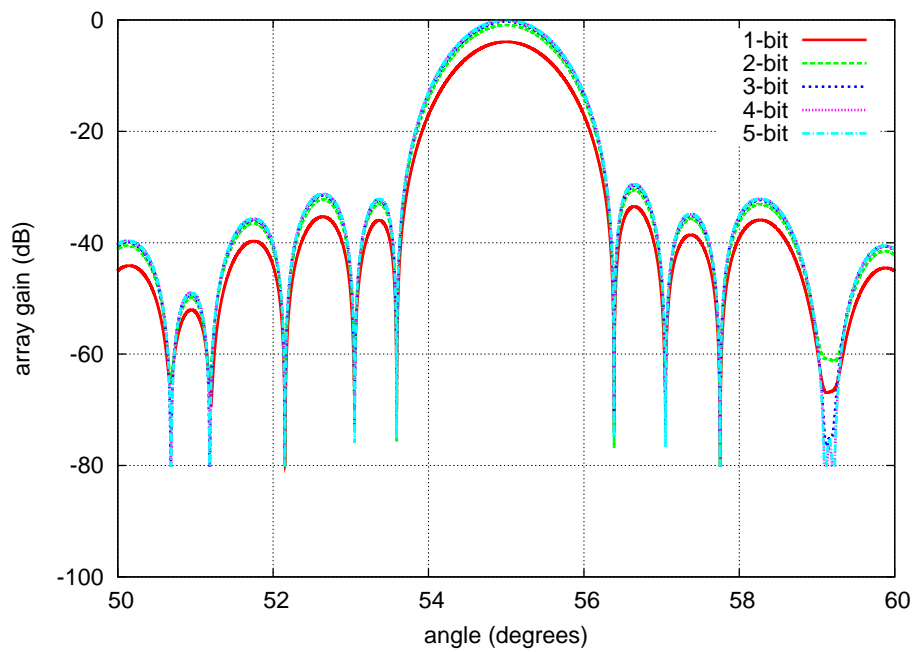


Figure 6.5: (a)-(f) Array loss as a result of phase error (quantization) for uniform 100×100 square array with different inter-element spacings.



(a)



(b)

Figure 6.6: (a)-(b) 2-dimensional horizontal beam pattern of a 100x100 array with $\lambda/2$ spacing (steered towards 55 degrees) for different phase resolutions.

$$\Rightarrow E[|H|^2]/G_{max} \geq 1/128$$

Therefore, the ratio between the minimum and the maximum gain can be bounded from below by a constant that is independent of the number of antennas N . Therefore, the array gain will continue to grow linearly with N independent of the phase resolution. The argument can be extended to the second case (1 and -1 complex weights) by making the following observation. When we have two complex signals with phase difference (in absolute value) that is $\pi/2$ or less, then the magnitude of the sum will be at least as large as the magnitude of the largest of the two. Let $c_{max=\pi/4}$ be the sum of signals in the set $\Psi_{max=\pi/4}$ (i.e. $\sum_{\psi \in \Psi_{max=\pi/4}} e^{j\psi}$).

Since we can shift the phase of each signal by π (using $a_1 = -1$), we can guarantee that every complex signal, before being added, will have a phase difference no larger than $\pi/2$ with $c_{max=\pi/4}$. Therefore, the minimum gain will grow linearly with N .

Now we can consider the general case where a_0 and a_1 are arbitrary numbers. Let us define the following variables:

$$\Delta = |a_0 - a_1|, \quad X = \sum_{\psi \in \Psi_{max=\pi/4}} e^{j\psi} = \alpha N, \quad Y = \sum_{\psi \notin \Psi_{max=\pi/4}} e^{j\psi} = -\beta N$$

Where α and β are complex constants. We showed that $|\alpha|^2$ can be lower bounded by a positive constant that is independent of N . We will also assume without loss of generality that $a_0 = 1$, $|a_1| \leq 1$, and $|\alpha| \geq |\beta|^4$. Therefore, the total gain can be expressed in terms of these variables⁵:

$$G = E[|H|^2] = |a_i X + a_j Y|^2 = |a_i \alpha - a_j \beta|^2 N^2 = |\alpha|^2 |a_i - a_j \frac{\beta}{\alpha}|^2 N^2 = |\alpha|^2 |a_i - a_j \gamma|^2 N^2, \quad \text{where } \gamma = \frac{\beta}{\alpha}, \text{ and } 0 \leq i, j \leq 1$$

We want to show that we can choose a_i and a_j such that we can always guarantee that the term $|\alpha|^2 |a_i - a_j \gamma|^2$ cannot fall below a constant number $g_{min} = \alpha^2 \Delta^2 / 4$. This will prove that the gain, in the worst case, can be bounded from below by a constant that is independent of the number of antennas N . We have two cases. In the first case, $|1 - \gamma| \geq \Delta/2$. In this case, choosing $a_i = a_j = a_0 = 1$ is sufficient to guarantee that the gain does not fall below the desired lower bound g_{min} . In the second case, $|1 - \gamma| < \Delta/2$. If we choose $a_i = a_0 = 1$ and $a_j = a_1$, then:

$$\begin{aligned} \alpha(1 - a_1 \gamma) &= \alpha(1 - a_1 + a_1 - a_1 \gamma) = \alpha(1 - a_1 + a_1(1 - \gamma)) \\ \Rightarrow |\alpha(1 - a_1 \gamma)| &\geq |\alpha| (|1 - a_1| - |a_1| |1 - \gamma|) \quad \text{using the Triangle Inequality} \\ &\Rightarrow |\alpha(1 - a_1 \gamma)| \geq |\alpha| (\Delta - \frac{\Delta}{2}) = |\alpha| \frac{\Delta}{2} \\ \Rightarrow |\alpha(1 - a_1 \gamma)|^2 &= |\alpha|^2 |1 - a_1 \gamma|^2 \geq \alpha^2 \frac{\Delta^2}{4} = g_{min} \end{aligned}$$

This completes the proof. Notice that $g_{min} \rightarrow 0$ as $\Delta \rightarrow 0$ (or $a_1 \rightarrow a_0$), which is what we would expect.

⁴We can always reassign the variables to ensure that these conditions hold.

⁵To simplify the proof, we will assume that all complex signals that make up the variable X will be scaled by the same complex beamforming weight a_i , and all complex signals that make up the variable Y will be scaled by the same complex beamforming weight a_j . a_i and a_j need not be equal.

6.2 Beam-nulling

In Section 6.1, we analyzed the impact of phase errors on the beamforming gain, and showed that the gain is very robust in the presence of errors. In this section, we analyze how the errors impact the interference power, the other component in the SINR⁶.

In general, adaptive array systems minimize interference by placing beam nulls in the directions of interfering signals. Controlling the locations of the nulls in the beampattern will usually require modifying both the phases and amplitudes of the complex beamforming weights, whether the wireless channel is Rayleigh or LOS. In this case, we need to account for both phase and amplitude errors. Analyzing the combined effect of phase and amplitude errors is easier when we consider the problem in the spatial domain where the optimal complex beamforming weights and channel responses can be represented as complex vectors in the N -dimensional Euclidean space, where N is the number of antennas in the array. Let us assume that we have $K + 1$ complex spatial channel response vectors: a desired vector \mathbf{h}_d corresponding to the direction of the desired signal⁷, and K interfering vectors $\mathbf{h}_i \forall 1 \leq i \leq K$ corresponding to the directions of K interfering signals.

$$\mathbf{h}_d = [\alpha_{1d}e^{j\beta_{1d}}, \dots, \alpha_{Nd}e^{j\beta_{Nd}}]^\top$$

$$\mathbf{h}_i = [\alpha_{1i}e^{j\beta_{1i}}, \dots, \alpha_{Ni}e^{j\beta_{Ni}}]^\top \quad \forall 1 \leq i \leq K$$

The incoming signal at the input of the array $\mathbf{y}[n]$ is the sum of the desired signal and interference and noise:

$$\mathbf{y}[n] = \mathbf{h}_d d[n] + \sum_{i=1}^K \mathbf{h}_i d_i[n] + \mathbf{v}[n]$$

where $d[n]$ is the desired signal, $d_i[n]$ is interfering signal i , and $\mathbf{v}[n]$ is the white noise vector at the receiver (the variance of each component of $\mathbf{v}[n]$ is σ_v^2). For simplicity, we shall assume that the desired and interfering signals have the same power. Using beamforming weights \mathbf{w} (without loss of generality, we can restrict $|\mathbf{w}| = 1$), the signal at the output of the array will be $\mathbf{w}^H \mathbf{y}[n]$. The output SINR is given by:

$$\text{SINR}_{out} = \frac{|\mathbf{w}^H \mathbf{h}_d|^2}{|\sum_{i=1}^K \mathbf{w}^H \mathbf{h}_i|^2 + \sigma_v^2}$$

where $(\cdot)^H$ denotes the complex conjugate transpose. Let $\mathbf{H}_I = [\mathbf{h}_1, \dots, \mathbf{h}_K]$ be the matrix whose columns are the interference vectors. Our goal is to achieve maximum interference rejection while retaining most of the desired signal. This can be achieved by choosing a beamforming weight vector \mathbf{w} that is the projection of the desired vector \mathbf{h}_d onto the subspace orthogonal to the column space of \mathbf{H}_I (or the null-space of \mathbf{H}_I^\top , which is also known as the left nullspace of \mathbf{H}_I), as described in [53]:

$$\mathbf{w}_{opt} = \mathbf{w}_{projection} = \mathbf{h}_d - \mathbf{H}_I(\mathbf{H}_I^H \mathbf{H}_I)^{-1} \mathbf{H}_I^H \mathbf{h}_d$$

⁶We will assume that the system is interference limited (i.e. the interference power is much larger than the noise power).

⁷In this chapter, we denote scalars in lower case, vectors in bold lower case, and matrices in bold upper case.

The projection-based beamformer is sometimes referred to as the zero-forcing (ZF) beamformer since it completely rejects all the interference (along with any component of the desired signal that lies in the interference subspace). We can see that rejecting all the interfering signals is only possible when the left nullspace is non-empty, which is guaranteed when $K < N$. The projection-based beamformer maximizes the signal to interference ratio (SIR), but not necessarily the SINR since it does not account for noise⁸. Optimizing the output SINR leads to the Minimum Variance Distortionless Response (MVDR) beamformer [33]. If we define the noise+interference correlation matrix \mathbf{R}_{N+I} as:

$$\mathbf{R}_{N+I} = \sum_{i=1}^K \mathbf{h}_i \mathbf{h}_i^H + \sigma_\nu^2 I_N$$

where I_N is the $N \times N$ identity matrix, then the output SINR can be maximized by choosing \mathbf{w}_{opt} :

$$\mathbf{w}_{opt} = \mathbf{w}_{MVDR} = \frac{\mathbf{R}_{N+I}^{-1} \mathbf{h}_d}{\mathbf{h}_d^H \mathbf{R}_{N+I}^{-1} \mathbf{h}_d}$$

The denominator is a normalizing factor. When the interference power is much larger than the noise power, both projection and MVDR yield virtually identical results.

6.2.1 Analysis of Errors in Weights

In this chapter, we focus on interference limited systems (i.e. when interference power is much larger than the noise floor), in which case both projection and MVDR yield similar results. Therefore, without loss of generality, we will assume that an optimum beamformer \mathbf{w}_{opt} is computed using projection, and $\hat{\mathbf{w}}_{opt}$ takes into account both phase and amplitude errors:

$$\begin{aligned} \mathbf{w}_{opt} &= [\alpha_{1w} e^{j\beta_{1w}}, \dots, \alpha_{Nw} e^{j\beta_{Nw}}]^\top \\ \hat{\mathbf{w}}_{opt} &= [\alpha_{1w} (1 + \epsilon_1) e^{j(\beta_{1w} + \delta_1)}, \dots, \alpha_{Nw} (1 + \epsilon_N) e^{j(\beta_{Nw} + \delta_N)}]^\top \end{aligned}$$

where $\epsilon_i \forall_{1 \leq i \leq N}$ are *i.i.d* zero mean real random variables with variance $E[\epsilon_i^2] = \sigma_\epsilon^2$, and $\delta_i \forall_{1 \leq i \leq N}$ are *i.i.d* zero mean real random variables with variance $E[\delta_i^2] = \sigma_\delta^2$. We also assume that the phase and amplitude errors are independent of each other. Furthermore, we scale the weights so that \mathbf{w}_{opt} has unit norm (i.e. $\sum_{i=1}^N \alpha_{iw}^2 = 1$).

The phase and amplitude errors result in $\hat{\mathbf{w}}_{opt}$ deviating from \mathbf{w}_{opt} by an angle θ . Note that θ is an angle in N -dimensional space, and does not correspond to a physical direction. This deviation will result in a reduction in the signal strength in the desired direction as well as an increase in the interference power, since $\hat{\mathbf{w}}_{opt}$ will no longer be orthogonal the interference subspace. The desired power is proportional to $\cos(\theta)$, and the increase in interference (leakage) is proportional to $\sin(\theta)$ (see Figure 6.7). For small angles θ , we can use the standard approximations⁹ $\sin(\theta) \approx \theta$ and $\cos(\theta) \approx 1$. Thus, we can characterize

⁸When interference power is near or below the noise floor, complete interference rejection may not result in the maximum SINR since it might also unnecessarily reduce the power of the desired signal (The component that lies in the interference subspace).

⁹This explains why nulls are more sensitive than peaks to phase and amplitude errors, since $\sin(\theta)$ changes more rapidly than $\cos(\theta)$ when θ is small.

the effect of phase and amplitude errors on beam nulls by considering how the mean square angle $\sigma_{\theta}^2(\sigma_{\delta}, \sigma_{\epsilon}, N) = E[\theta^2]$ behaves as a function of σ_{δ} , σ_{ϵ} , and N .

- Interference/interference subspace
- Desired signal
- Optimum beamforming vector
- - Distorted beamforming vector

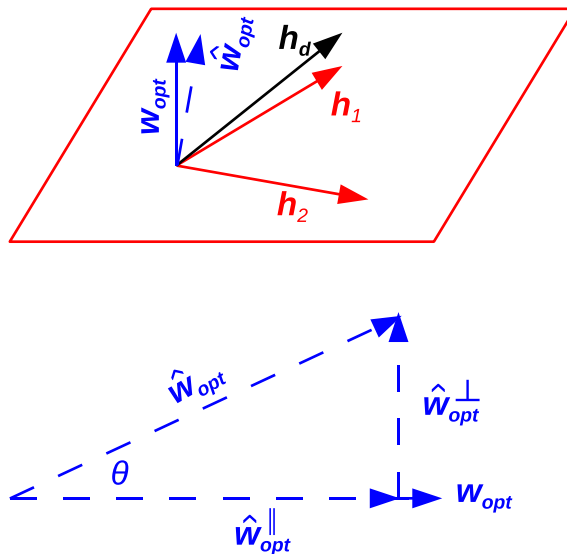


Figure 6.7: The optimum beamforming vector \mathbf{w}_{opt} can be viewed as a projection of the desired signal onto the subspace orthogonal to the interference subspace. The distorted beamforming vector $\hat{\mathbf{w}}_{opt}$ can be decomposed into two orthogonal components: $\hat{\mathbf{w}}_{opt} = \hat{\mathbf{w}}_{opt}^{\perp} + \hat{\mathbf{w}}_{opt}^{\parallel}$. $\hat{\mathbf{w}}_{opt}^{\parallel}$, which is parallel to \mathbf{w}_{opt} , represents the potential loss in beamforming gain, and is proportional to $\cos(\theta)$. $\hat{\mathbf{w}}_{opt}^{\perp}$, which is orthogonal to \mathbf{w}_{opt} , represents the potential leakage into the interference subspace, and is proportional to $\sin(\theta)$.

If we assume that the phase and amplitude variations are small, and given that \mathbf{w}_{opt} is unit norm, then we can approximate the error angle θ with the length of the error vector:

$$\Delta \mathbf{w} = \mathbf{w}_{opt} - \hat{\mathbf{w}}_{opt} = [\alpha_{1w} e^{j\beta_{1w}} (1 - (1 + \epsilon_1) e^{j\delta_1}), \dots, \alpha_{Nw} e^{j\beta_{Nw}} (1 - (1 + \epsilon_N) e^{j\delta_N})]^T$$

We can further simplify the above expression using the approximations $\cos(\delta_i) \approx 1$, $\sin(\delta_i) \approx \delta_i$, and $\delta_i \epsilon_i \approx 0$.

$$\begin{aligned} \Delta \mathbf{w} &= [\alpha_{1w} e^{j\beta_{1w}} (1 - 1 - \epsilon_1 - j\delta_1), \dots, \alpha_{Nw} e^{j\beta_{Nw}} (1 - 1 - \epsilon_N - j\delta_N)]^T \\ &= [-\alpha_{1w} e^{j\beta_{1w}} (\epsilon_1 + j\delta_1), \dots, -\alpha_{Nw} e^{j\beta_{Nw}} (\epsilon_N + j\delta_N)]^T \\ \Rightarrow \|\Delta \mathbf{w}\|^2 &= (\Delta \mathbf{w})^H (\Delta \mathbf{w}) = \sum_{i=1}^N \alpha_{iw}^2 (\epsilon_i^2 + \delta_i^2) \end{aligned}$$

By taking the expectation of this expression:

$$\begin{aligned}\sigma_\theta^2 &\approx E[\|\Delta \mathbf{w}\|^2] = E\left[\sum_{i=1}^N \alpha_{iw}^2 (\epsilon_i^2 + \delta_i^2)\right] = \sum_{i=1}^N \alpha_{iw}^2 (E[\epsilon_i^2] + E[\delta_i^2]) \\ \sigma_\theta^2 &\approx \sum_{i=1}^N \alpha_{iw}^2 (\sigma_\epsilon^2 + \sigma_\delta^2) = (\sigma_\epsilon^2 + \sigma_\delta^2) \sum_{i=1}^N \alpha_{iw}^2 = \sigma_\epsilon^2 + \sigma_\delta^2\end{aligned}\quad (6.1)$$

As we can see, the mean square error angle σ_θ^2 is equal to the sum of the mean square phase error σ_δ^2 and the mean square amplitude error σ_ϵ^2 . The key conclusion that we draw from this result is that the angle error is independent of N , the number of antennas¹⁰.

In many phased and adaptive array systems, phase and amplitude control is achieved by independently modifying the real (in-phase) and imaginary (quadrature) components of the signal. The results are similar if we consider the error in the real and imaginary parts of the components of the beamforming vector. The proof would proceed in a similar fashion:

$$\mathbf{w}_{opt} = [I_{1w} + jQ_{1w}, \dots, I_{Nw} + jQ_{Nw}]^\top, \quad \|\mathbf{w}_{opt}\|^2 = \sum_{i=1}^N (I_{iw}^2 + Q_{iw}^2) = 1$$

$$\hat{\mathbf{w}}_{opt} = [I_{1w}(1 + \epsilon_{I_1}) + jQ_{1w}(1 + \epsilon_{Q_1}), \dots, I_{1w}(1 + \epsilon_{I_N}) + jQ_{1w}(1 + \epsilon_{Q_N})]^\top$$

where $\epsilon_{I_i} \forall 1 \leq i \leq N$ are *i.i.d* zero mean real random variables with variance $E[\epsilon_{I_i}^2] = \sigma_{\epsilon_I}^2$, and $\epsilon_{Q_i} \forall 1 \leq i \leq N$ are *i.i.d* zero mean real random variables with variance $E[\epsilon_{Q_i}^2] = \sigma_{\epsilon_Q}^2$. We assume that the errors in the real parts and the imaginary parts are independent of each other. It is also reasonable to assume that $\sigma_{\epsilon_I}^2 = \sigma_{\epsilon_Q}^2$, which is very common in practice.

$$\begin{aligned}\Delta \mathbf{w} &= \mathbf{w}_{opt} - \hat{\mathbf{w}}_{opt} = \\ &[I_{1w}(1 - (1 + \epsilon_{I_1})) + jQ_{1w}(1 - (1 + \epsilon_{Q_1})), \dots, I_{Nw}(1 - (1 + \epsilon_{I_N})) + jQ_{Nw}(1 - (1 + \epsilon_{Q_N}))]^\top \\ &= [I_{1w}\epsilon_{I_1} + jQ_{1w}\epsilon_{Q_1}, \dots, I_{Nw}\epsilon_{I_N} + jQ_{Nw}\epsilon_{Q_N}]^\top \\ &\Rightarrow \|\Delta \mathbf{w}\|^2 = \Delta \mathbf{w}^H \Delta \mathbf{w} = \sum_{i=1}^N (I_{iw}^2 \epsilon_{I_i}^2 + Q_{iw}^2 \epsilon_{Q_i}^2) \\ &\Rightarrow \sigma_\theta^2 \approx E[\|\Delta \mathbf{w}\|^2] = E\left[\sum_{i=1}^N (I_{iw}^2 \epsilon_{I_i}^2 + Q_{iw}^2 \epsilon_{Q_i}^2)\right] = \sum_{i=1}^N (E[I_{iw}^2]E[\epsilon_{I_i}^2] + E[Q_{iw}^2]E[\epsilon_{Q_i}^2]) = \\ &\quad \sum_{i=1}^N (E[I_{iw}]^2 \sigma_{\epsilon_I}^2 + E[Q_{iw}^2] \sigma_{\epsilon_Q}^2)\end{aligned}$$

We assume without loss of generality that $\sum_{i=1}^N E[I_{iw}^2] = \sum_{i=1}^N E[Q_{iw}^2] = \frac{1}{2}$

$$\Rightarrow \sigma_\theta^2 \approx \sum_{i=1}^N (I_{iw}^2 \sigma_{\epsilon_I}^2 + Q_{iw}^2 \sigma_{\epsilon_Q}^2) = \sigma_{\epsilon_I}^2 \sum_{i=1}^N E[I_{iw}^2] + \sigma_{\epsilon_Q}^2 \sum_{i=1}^N E[Q_{iw}^2] = \frac{\sigma_{\epsilon_I}^2 + \sigma_{\epsilon_Q}^2}{2}\quad (6.2)$$

The result is very similar to the expression in Equation 6.1.

¹⁰The power leakage into the interference subspace is independent of the number of antennas. However, increasing the number of antennas can still increase the output SINR (peak to null ratio) by increasing the power gain of the desired signal. Increasing the number of antennas also increases the degrees freedom necessary to null more interferers.

6.2.2 Simulation Results

The simulation results shown in Figure 6.8 verify the results in Equations 6.1 and 6.2. The variance of the weights errors $E[\|\Delta\mathbf{w}\|^2]$ is shown on the x-axis in Figures 6.8(a)-(d). The variance of the error angle $\sigma_\theta^2 = E[\theta^2]$ is shown on the y-axis. We have injected random noise into different components of the complex weight (real, imaginary, phase, and amplitude) separately, represented by the different curves. The different curves show almost exactly the same linear relationship between σ_θ^2 and the real (in-phase) error (σ_I^2), imaginary (quadrature) error (σ_Q^2), phase error (σ_δ^2), and amplitude error (σ_c^2) as predicted in Equations 6.1 and 6.2. Furthermore, the constant of proportionality is approximately 1 (consistent with the equations). We repeated the simulations for different values of N (Figures 6.8(b)-(d)), and the results are the same. This shows that the mean square error angle only depends on the magnitude of the errors in the beamforming weight vector and not the number of antennas, and that the different quantities σ_δ^2 , σ_c^2 , $\sigma_I^2/2$ and $\sigma_Q^2/2$ contribute equally to the overall error angle¹¹.

We repeated the same simulations in Figure 6.9 with the interference rejection or leakage (average interferer power after applying the beamforming weights in the presence of errors) on the y-axis. The results are virtually identical to Figure 6.8. This shows that the average interference rejection can be accurately represented by the error angle θ for small values of θ . This means that the depth of beam nulls is limited by the accuracy of the beamforming weights, and is independent of the size of the array N and the number of interferers K , as long as $N > K$ ¹².

For completeness, we also simulated the impact of general errors on the desired signal (beamforming) gain (Figures 6.10, 6.11, and 6.12). These figures show that similar to null-steering, the loss in beamforming depends on the errors in the weight vector, regardless of whether the errors are injected directly into the real and imaginary components (Cartesian representation) or phases and amplitudes (polar representation), and is independent of the number of antennas N . The figures also show that for a small error angle θ , the normalized beamforming gain can be approximated by $\cos(\theta)^2$.

6.3 Conclusion

In this chapter, we analyzed the array performance loss due to errors in the beamforming weights. We began by considering a beamforming system, which maximizes the gain in a desired direction. We derived an expression for the loss in gain due to uniform phase errors, and provided simulations that validate this result. We showed that beamforming is robust to phase errors. We also derived a lower bound on worst case scenario, and showed that it is independent of the number of antennas. Then, we considered a beam-nulling system, which rejects interfering signals. We analyzed the effect of uniform amplitude and phase errors, and again provided numerical simulations. We showed that the interference rejection is directly proportional to the magnitude of the errors in the weights, and is independent

¹¹ We have assumed that the different errors (real, imaginary) and (phase, amplitude) and independent of each other and independent across antennas.

¹²Even though the depth of the null is independent of N , the desired signal power (beamforming gain) grows linearly with N , and thus, the peak to null (signal to interference) ratio grows linearly with N .

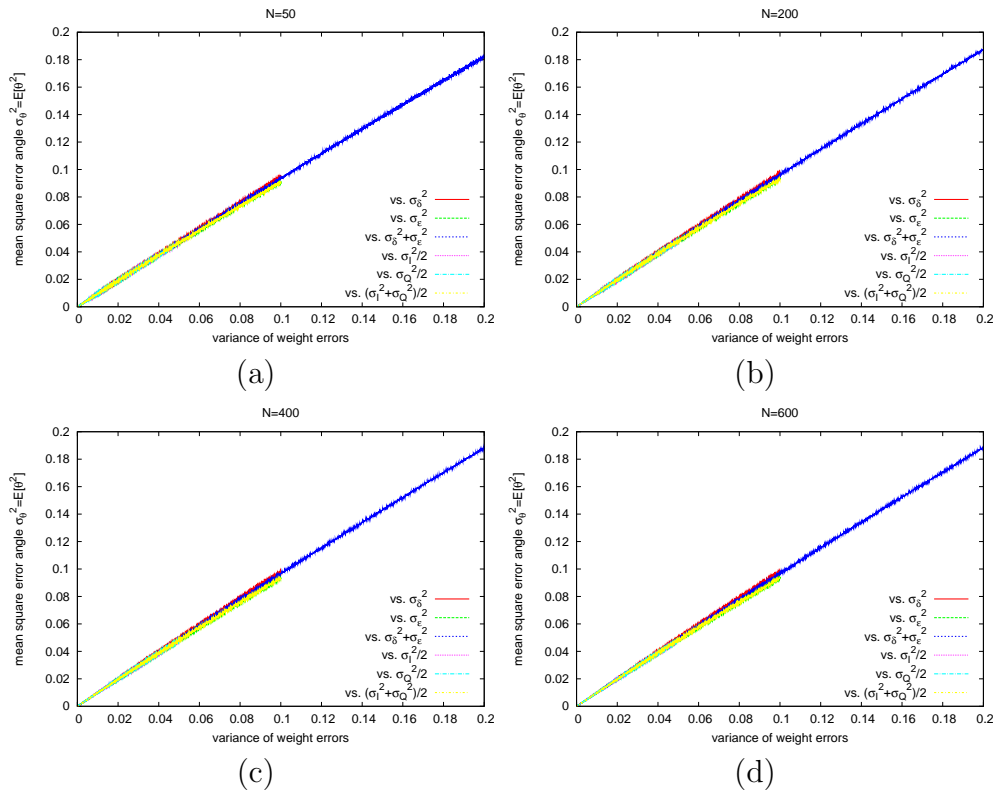


Figure 6.8: Simulated relationship between weight errors (both phase (σ_δ^2), amplitude (σ_ϵ^2), real (σ_I^2), and imaginary (σ_Q^2)) on the x-axis and the mean square error angle $\sigma_\theta^2 = E[\theta^2]$ on the y-axis: (a) $N=50$. (b) $N=200$. (c) $N=400$. (d) $N=600$.

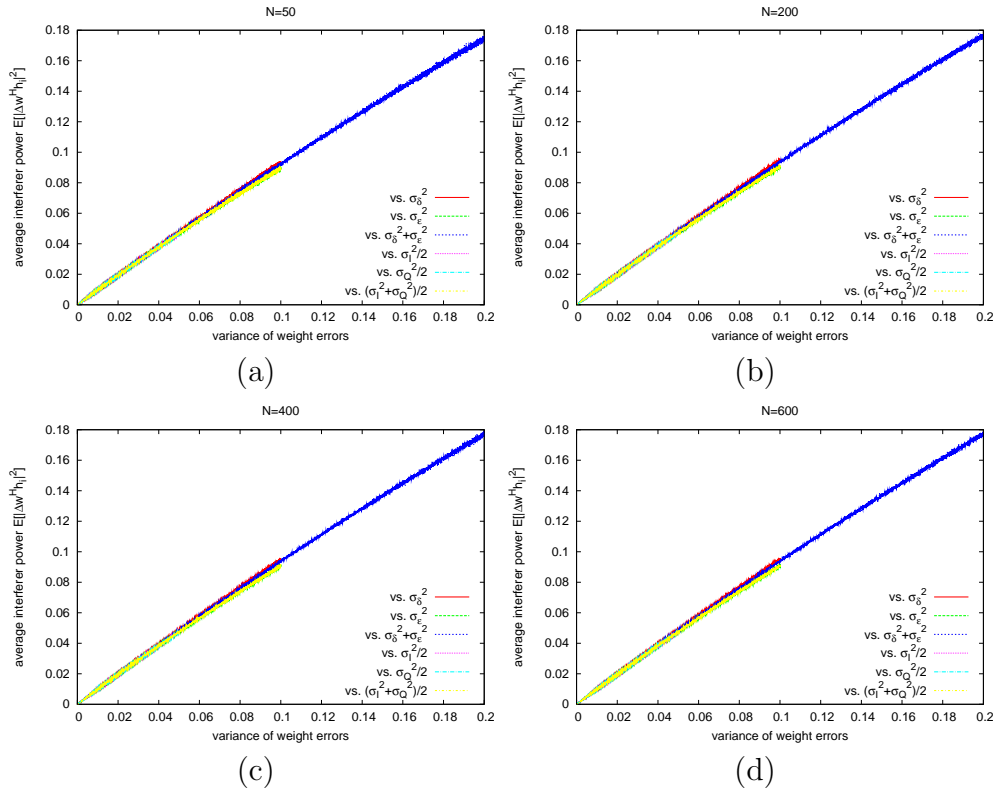


Figure 6.9: Simulated relationship between weight errors (both phase (σ_δ^2), amplitude (σ_ϵ^2), real (σ_I^2), and imaginary (σ_Q^2)) on the x-axis and the average interference rejection or leakage $E[|\Delta \mathbf{w}^H \mathbf{h}_i|^2]$ on the y-axis: (a) $N=50$. (b) 200. (c) $N=400$. (d) $N=600$.

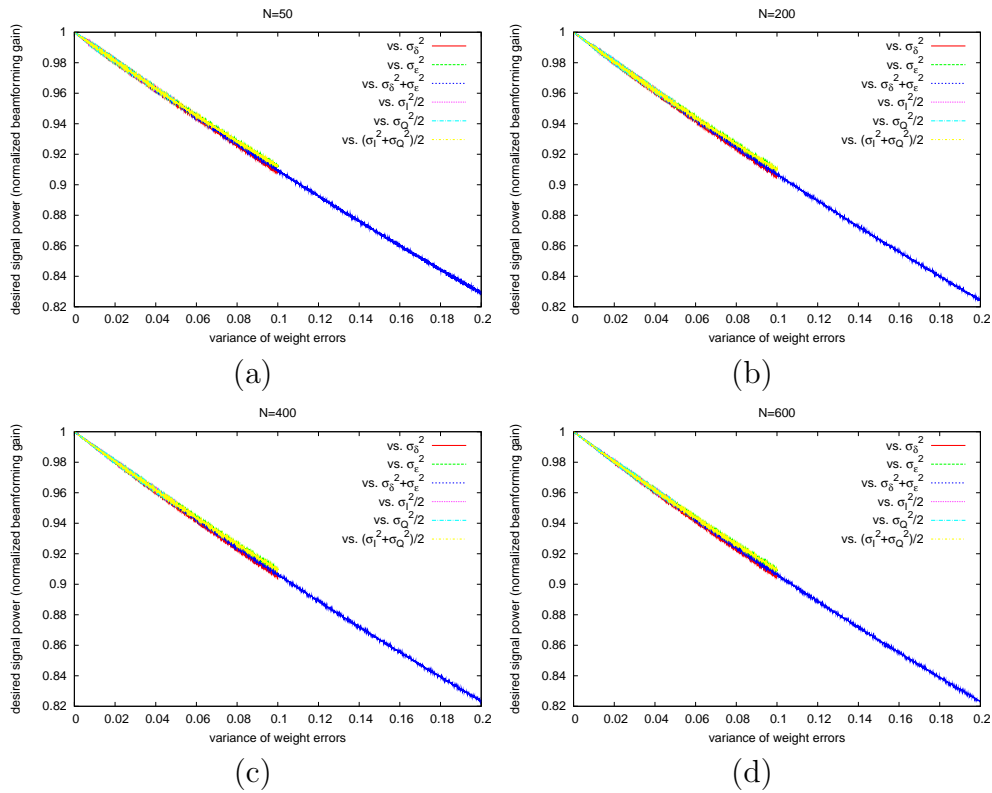


Figure 6.10: Simulated relationship between weight errors (both phase (σ_δ^2), amplitude (σ_ϵ^2), real (σ_I^2), and imaginary (σ_Q^2)) on the x-axis and the average desired signal power (beamforming gain) on the y-axis: (a) $N=50$. (b) 200. (c) $N=400$. (d) $N=600$.

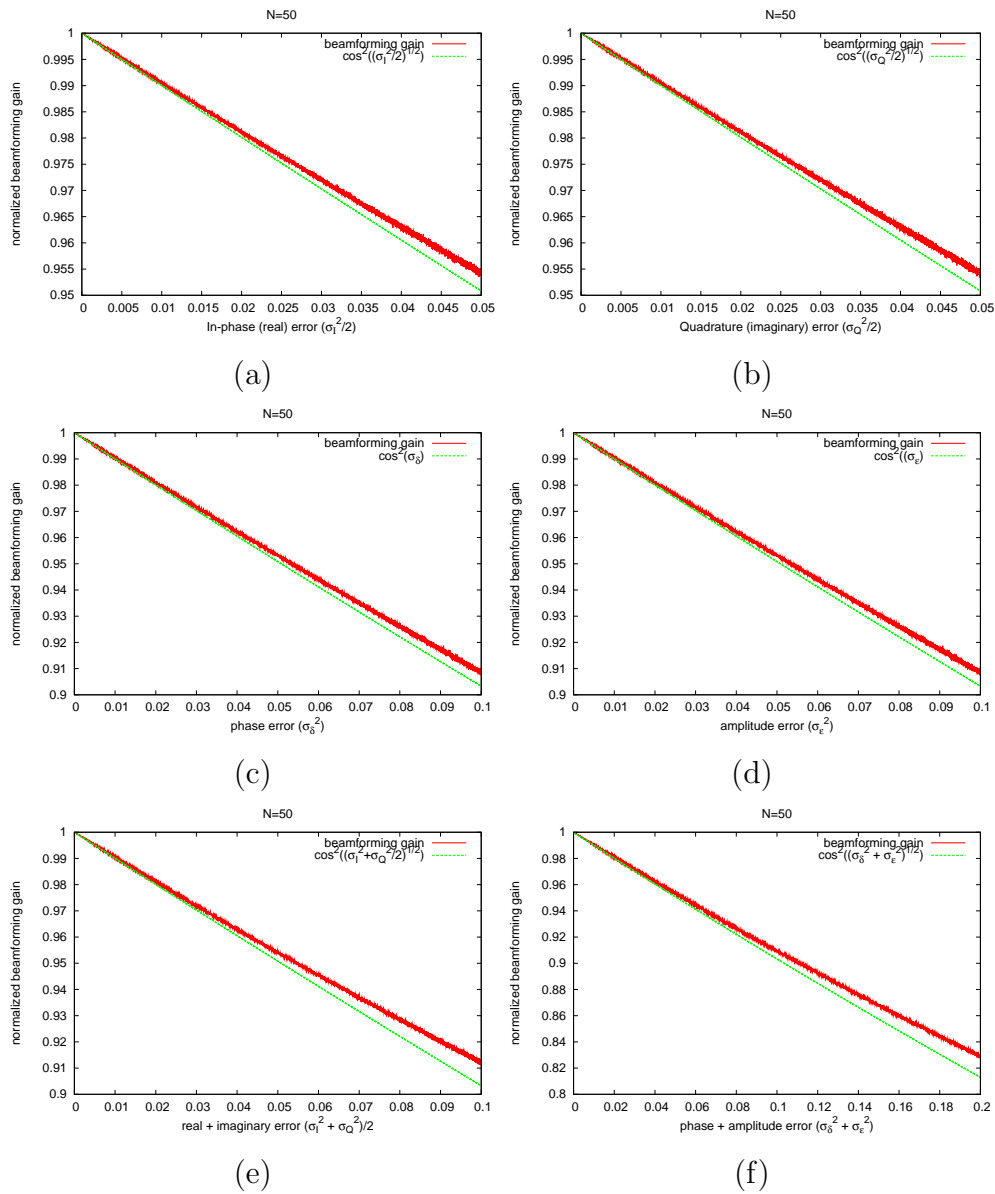


Figure 6.11: Simulated beamforming gain and cosine of the error angle $E[\cos(\theta)^2]$ versus weight errors ($N=50$): (a) Errors in the real part σ_I^2 . (b) Errors in the imaginary part σ_Q^2 . (c) Phase errors σ_δ^2 . (d) Amplitude errors σ_ϵ^2 . (e) Real and imaginary errors $(\sigma_I^2 + \sigma_Q^2)/2$. (f) Phase and amplitude errors $\sigma_\delta^2 + \sigma_\epsilon^2$.

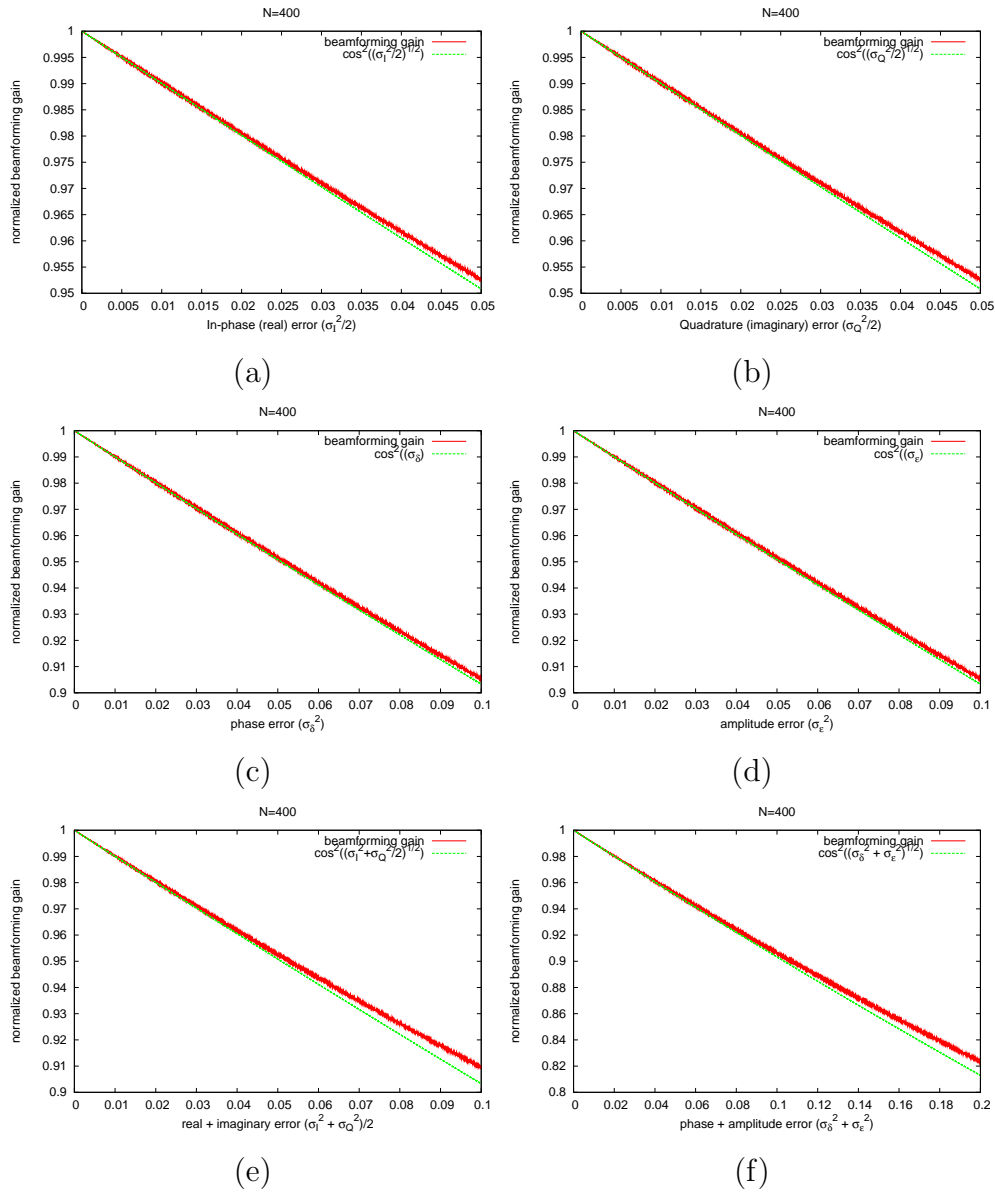


Figure 6.12: Simulated beamforming gain and cosine of the error angle $E[\cos(\theta)^2]$ versus weight errors ($N=400$): (a) Errors in the real part σ_I^2 . (b) Errors in the imaginary part σ_Q^2 . (c) Phase errors σ_δ^2 . (d) Amplitude errors σ_ϵ^2 . (e) Real and imaginary errors $(\sigma_I^2 + \sigma_Q^2)/2$. (f) Phase and amplitude errors $\sigma_\delta^2 + \sigma_\epsilon^2$.

of the number of antennas, assuming that there are more antennas than interferers. A key conclusion of this chapter is that interference rejection (null-steering) is a lot more sensitive to errors than beamforming (peak-steering), and in both cases, the performance degradation is independent of the number of antennas.

Chapter 7

Error mitigation with vector quantization

In Chapter 6, we examined how the errors in the beamforming weights affect the signal to noise and interference ratio (SINR) at the output of a beamformer. We showed that the desired signal component of the SINR is largely insensitive to most phase and amplitude errors that occur in practice. However, the interference component is directly proportional to the magnitude of phase and amplitude errors, and is independent of the number of antennas when those errors are uncorrelated. This high sensitivity of beam-nulling (interference rejection) to errors in the beamforming system can be detrimental for wireless systems that rely on spatial reuse to increase their capacity. In this chapter, we provide a more in-depth analysis for one of these sources of error: those that arise from quantizing the beamforming weight vectors. We focus on quantization errors for two reasons. First, whereas most other sources of error (e.g. noise, jitter) are unpredictable, difficult to control and measure, time-varying and independent across antennas, quantization errors can be predicted and shaped by the beamformer. Second, quantization is usually the dominant source of errors in analog beamforming architectures discussed in Chapter 3 where it is difficult to build high resolution accurate beamformers especially at high frequencies¹.

The chapter is organized as follows. In Section 7.1, we describe the basic quantization problem. In Section 7.2, we present two popular beamforming schemes and analyze their performance under the naive scalar quantization approach. We show that in both cases, the signal to interference ratio (SIR) is proportional to the ratio of the antennas to the interferers (N/K), while the average interferer power is independent of both N and K (consistent with the results derived in Chapter 6). In Section 7.3, we shift the focus to vector quantization techniques. We first derive a lower bound on the achievable SIR under vector quantization, and show that it is quadratic in the number of antennas N (as opposed to linear in the case of scalar quantization). Next, we present several computationally efficient

The work presented in this chapter was done in collaboration with Mark Johnson, Raghuraman Mudumbai, and Upamanyu Madhow [7, 9].

¹The LO beamforming architecture presents a bigger challenge because it lacks control over the amplitude of the signal.

vector quantization algorithms, based on greedy coordinate descent (GCD), that achieve this lower bound. We analyze their performance in terms of both achievable SIRs and computational complexity. We show that the achievable SIRs are proportional to $(N/K)^2$ (compared to N/K for scalar quantization). These algorithms better exploit the available degrees of freedom to cancel interference without much sacrifice in computational efficiency as we show. Finally, we examine the optimality of GCD based algorithms by comparing them with the optimum method based on exhaustive search. Our simulation results show that the SIR under optimum quantization can grow exponentially in the number of antennas N . This indicates that although the vector quantization techniques proposed in this chapter result in significant performance gains, they are still far from optimal, and that a computationally tractable algorithm that achieves the upper bound remains an open problem. The results in this chapter are based on a combination of mathematical proofs and simulation results. Even though we focus on the Rayleigh channel model because it is easier to analyze, we also present simulation results based on the line of sight (LOS) channel model (the other extreme), and show that the results are largely consistent.

7.1 Quantization problem statement

We begin by considering an N -antenna receive beamforming system with the goal of maximizing the array gain in the direction of a desired transmitter with respect to the interfering signals². For simplicity and without loss of generality, we assume that the desired transmitter and the interferers have single antennas. Let $\mathbf{h}_d \triangleq [h_d[1], h_d[2], \dots, h_d[N]]^\top$ be the complex baseband channel response vector from the desired transmitter to the N -antenna receive array³. We similarly denote the channel gain from the l^{th} interferer to the N -antenna receive array as $\mathbf{h}_l \triangleq [h_l[1], h_l[2], \dots, h_l[N]]^\top$ for $1 \leq l \leq K$, where K is the number of interferers ($K < N$). We assume that the components of $\mathbf{h}_d, \mathbf{h}_1, \dots, \mathbf{h}_K$ are drawn from zero-mean *i.i.d.* complex Gaussian distributions, i.e. Rayleigh fading [59]. We assume that the variance of the channel gains is unity (i.e. $E[|h_d[i]|^2] = E[|h_1[i]|^2] = \dots = E[|h_K[i]|^2] = 1 \quad \forall_{1 \leq i \leq N}$). This setup can be readily generalized to interferers with different power levels and to Rician or Line-of-Sight (LOS) channels, and we will present simulation results to show that the same ideas extend to more general setups⁴.

Let $\mathbf{w} = [w[1], \dots, w[N]]^\top$ be the N -element spatial filter (beamforming vector) used by the receiver to combine the signals from all N antennas. In this case, the SINR at the output of the receiver (beamformer):

$$SINR(\mathbf{w}) = \frac{|\mathbf{w}^H \mathbf{h}_d|^2}{\sum_{i=1}^K |\mathbf{w}^H \mathbf{h}_i|^2 + |\mathbf{w}^H \mathbf{v}|^2}$$

²A similar analysis also applies to a transmit array that seeks to minimize its interference at given locations. We focus on the receive array for clarity.

³In this chapter, we denote scalars in lower case, vectors in bold lower case, and matrices in bold upper case.

⁴Without loss of generality, we assume a narrow band system or a flat fading channel (i.e. a channel response that can be represented by a single tap). A frequency-selective (multi-tap) channel can always be transformed into a group of parallel narrow band subchannels, where each subchannel can be processed independently (e.g. OFDM).

$(\cdot)^H$ denotes the conjugate transpose operation, and \mathbf{v} is an N -element vector denoting the complex white noise signal generated at each antenna element at the receiver (e.g. thermal noise). We will assume that the system is interference limited (i.e. the interference component is much larger than the noise). In this case, we can ignore the noise and focus on the signal to interference ratio (SIR):

$$SIR(\mathbf{w}) = \frac{|\mathbf{w}^H \mathbf{h}_d|^2}{\sum_{i=1}^K |\mathbf{w}^H \mathbf{h}_i|^2}$$

Since the beamforming weight vector \mathbf{w} appears in both the numerator and the denominator, the SIR does not depend on $\|\mathbf{w}\|$. Also, let B be the number of available quantization bits (i.e. the I (in-phase or real) and Q (quadrature or imaginary) components of each element $w[i]$, $\forall_{1 \leq i \leq N}$ of the weight vector \mathbf{w} are each quantized to 2^B discrete levels). Therefore, there are $N_w = 2^{2BN}$ total beamforming weight vectors available from which to choose⁵. Let \mathcal{W} be the set of these discrete weight vectors. We can then formulate the problem of computing the optimal weight as an SIR maximization problem:

$$\mathbf{w}_{opt} \triangleq \arg \max_{\hat{\mathbf{w}} \in \mathcal{W}} \frac{|\hat{\mathbf{w}}^H \mathbf{h}_d|^2}{\sum_{i=1}^K |\hat{\mathbf{w}}^H \mathbf{h}_i|^2}$$

Our goal in this chapter, is to compare different strategies for choosing $\hat{\mathbf{w}}$ in terms of both the SIR performance and computational complexity. In order to do so, we need a good metric for evaluating SIR performance. If one algorithm outperforms another for every possible realization of the desired and interference channels, then it is clearly better. However, such a definition is too stringent to be useful, and in practice, we may have to settle for comparing some averages. At first glance, it seems as if the natural choice would be to compare the average SIR, i.e. $E[SIR]$; however this measure suffers from some fundamental disadvantages. First of all, it is rather unwieldy analytically. Secondly, the average can be dominated by a small number of channel realizations where the SIR becomes very large. We therefore choose a different measure: the ratio of average signal power to average interference power:

$$\overline{SIR}(\mathbf{w}) = \frac{E[|\hat{\mathbf{w}}^H \mathbf{h}_d|^2]}{\sum_{i=1}^K E[|\hat{\mathbf{w}}^H \mathbf{h}_i|^2]}$$

7.2 Scalar quantization of beamforming weights

In this section, we analyze techniques based on scalar quantization. In scalar quantization, each component of a beamforming vector is quantized separately to the nearest discrete level [47]. For simplicity, we assume that the quantization is uniform, and the quantization

⁵ Note that some of these weight vectors may be scaled versions of others. Therefore, N_w is actually an upper-bound on the number of distinct weight vectors available to choose from. Also, since the SIR is independent of the magnitude of the beamforming weight vector, we can assume without loss of generality that the 2^B discrete levels are the 2^B equal and non-overlapping subintervals between $[-1, 1]$ (i.e. uniform quantization).

intervals are obtained by dividing the interval $[-1, 1]$ into 2^B equal and contiguous subintervals⁶. The size of the intervals is not relevant to this problem since we can always scale the beamforming vector without changing the SIR. As a result, scalar quantization can be implemented by first dividing the vector by its real or imaginary components with the largest absolute value (to ensure that all components lie in the $[-1, 1]$ interval), and then quantizing each component (real and imaginary) separately. We shall denote this operation by the operator $Q_B(\cdot)$, and thus $\mathbf{w}_q = Q_B(\mathbf{w})$ represents the B-bit quantized (and possibly normalized) version of the vector \mathbf{w} . We will also denote the quantization error vector by $\Delta\mathbf{w} = \mathbf{w} - Q_B(\mathbf{w})$. In Chapter 5, we showed that the ability of the beamformer to estimate the channel responses does not depend on the accuracy of the beamforming weights. Therefore, we will assume that the beamformer has full knowledge of the channel responses $\mathbf{h}_d, \mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K$. We will now consider two strategies for choosing \mathbf{w} to improve the SIR and analyze how they are impacted by scalar quantization.

7.2.1 Spatial matched filter

A simple and straight forward choice for the beamforming weight vector is $\mathbf{w}_{\text{mf}} = \frac{\mathbf{h}_d}{\|\mathbf{h}_d\|}$, also known as the spatial matched filter or maximum ratio combiner (MRC). The spatial matched filter ensures the maximum attainable signal to noise ratio (SNR). The SIR under the matched filter is

$$\begin{aligned} \text{SIR}(\mathbf{w}_{\text{mf}}) &= \frac{|\mathbf{w}_{\text{mf}}^H \mathbf{h}_d|^2}{\sum_{i=1}^K |\mathbf{w}_{\text{mf}}^H \mathbf{h}_i|^2} = \frac{|\mathbf{h}_d^H \mathbf{h}_d|^2}{\sum_{i=1}^K |\mathbf{h}_d^H \mathbf{h}_i|^2} \\ \overline{\text{SIR}}(\mathbf{w}_{\text{mf}}) &= \frac{E[|\mathbf{h}_d^H \mathbf{h}_d|^2]}{\sum_{i=1}^K E[|\mathbf{h}_d^H \mathbf{h}_i|^2]} \\ E[|\mathbf{h}_d^H \mathbf{h}_d|^2] &= E\left[\left(\sum_{i=1}^N |h_d[i]|^2\right)^2\right] = E\left[\sum_{i=1}^N \sum_{j=1}^N |h_d[i]|^2 |h_d[j]|^2\right] \\ &= \sum_{i=1}^N \sum_{j=1}^N E[|h_d[i]|^2 |h_d[j]|^2] = \sum_{i=1}^N \sum_{j=1, j \neq i}^N E[|h_d[i]|^2] E[|h_d[j]|^2] + \sum_{i=1}^N E[|h_d[i]|^4] \\ &= N^2 - N + 3N = N^2 + 2N \sim N^2 \\ E[|\mathbf{h}_d^H \mathbf{h}_i|^2] &= E\left[\left|\sum_{j=1}^N h_d^*[j] h_i[j]\right|^2\right] \\ &= E\left[\sum_{j=1}^N \sum_{k=1}^N h_d^*[j] h_i[j] h_d[k] h_i^*[k]\right] \\ &= \sum_{j=1}^N E[|h_d[j]|^2] E[|h_i[j]|^2] = N \end{aligned}$$

⁶In general, the number and size of each interval will depend on the analog and digital beamforming circuitry.

$$\Rightarrow \overline{\text{SIR}}(\mathbf{w}_{\text{mf}}) \sim \frac{N}{K}$$

Therefore, the average SIR is proportional to $\frac{N}{K}$. The linear growth with the number of antennas is due to the beamforming gain which boosts the desired signal power by a factor N , but leaves the interference power almost unchanged because the beamforming weight vector is chosen independent of the channel responses of the interfering signals. Now let's look at the impact of quantization on the matched filter beamformer. Let $\mathbf{w}_{\text{qmf}} = Q_B(\mathbf{w}_{\text{mf}})$ be the matched filter quantized to B bits. The interference should not be affected by the quantization since the beamforming vector is still independent of the channel responses of the interfering vectors (i.e. $E[|\mathbf{w}_{\text{qmf}}^H \mathbf{h}_i|^2] = E[|\mathbf{w}_{\text{mf}}^H \mathbf{h}_i|^2]$). In Chapter 6, we showed that the desired signal (beamforming) gain is robust to errors in the beamforming weights (the gain degradation is proportional to the cosine of the error angle). Therefore, even for $B = 1$, the loss in beamforming is small. Therefore, the quantized matched filter yields similar gain in the SIR as the matched filter:

$$\overline{\text{SIR}}(\mathbf{w}_{\text{qmf}}) \simeq \overline{\text{SIR}}(\mathbf{w}_{\text{mf}}) \sim \frac{N}{K} \quad (7.1)$$

In order to support the results derived in this section, we simulated the relationship between the average SIR, as well as its subcomponents (desired signal and interference power), and the number of antennas N , the number of interferers K , and the number of quantization bits B under the spatial matched filter beamformer. Figure 7.1a shows that the average SIR, when normalized by the number of antennas N , remains constant as a function of N for different quantization levels ($B = 1 - 6$). Figure 7.1b shows that the average SIR, when multiplied by the number of interferers K , remains constant as a function of K . Both these results are consistent with Equation 7.1. In Figure 7.2, we simulated the SIR while keeping the ratio N/K constant. In all the Figures 7.1/7.2, however, we see that using more bits of resolution slightly improves the SIR by a constant factor independent of both N and K . Figures 7.3/7.4 show that this gain in the SIR is a result of improved beamforming gain and not interference cancellation (Figures 7.5/7.6). In Figures 7.3/7.4, the SIR is replaced by the average desired signal power on y-axis, and the results show a linear growth with N , a slight improvement with B , and an independence of K . Figures 7.5/7.6 show a constant average interferer power independent of N, K , and B . Therefore, the improvement in average SIR obtained from increasing N and B is due to the desired signal, while the degradation from increase in K is due to the interference.

The results in Figures 7.1/7.2, 7.3/7.4, and 7.5/7.6 were obtained using the Rayleigh channel model. We repeated the same simulations under the LOS channel. In Figures 7.7/7.8, 7.9/7.10, and 7.11/7.12, we simulated a uniform linear array with $\lambda/2$ spacing, and randomized the azimuth angle of arrival of each transmitter. In Figures 7.13/7.14, 7.15/7.16, and 7.17/7.18, we simulated a uniform square array with $\lambda/2$ spacing, and randomized the direction of arrival (in 3 dimensions) of each transmitter. These results show that the performance of the quantized spatial matched filter is consistent with with Equation 7.1, and is not sensitive to the channel or the quantization resolution.

In addition to being robust to quantization, notice that the matched filter beamformer requires only computing and quantizing the complex conjugate of the desired channel vector

\mathbf{h}_d and is independent of the interfering channel vectors, and hence it has a low computational complexity $O(N)$.

7.2.2 Zero-Forcing beamformer

In Chapter 6, we showed that complete interference cancellation (nulling) can be achieved by using a beamforming weight vector \mathbf{w}_{zf} that is the projection of \mathbf{h}_d onto the vector subspace orthogonal to the subspace spanned by the interfering vectors $\mathbf{h}_1, \dots, \mathbf{h}_K$, a technique sometimes referred to as zero forcing (ZF) in the signal processing literature. We also showed that in the presence of errors in the ZF beamformer will result in interference "leakage" that is directly proportional to the variance of the errors independent of the number of antennas N . Since quantization is only a special type of errors, the same derivation applies. Since the errors will be roughly uniform in the interval $[-2^{-B}, 2^{-B}]$ (within a constant factor), the variance of the quantization noise σ_w will be proportional to 2^{-2B} , and thus the power of a single interferer will also be proportional to 2^{2B} . Therefore, the total interference power will be proportional to $2^{-2B}K$. Since the desired signal grows linearly with N (beamforming gain), the average SIR at the output of the quantized ZF beamformer $\mathbf{w}_{\text{qzf}} = Q_B(\mathbf{w}_{\text{zf}})$

$$\overline{\text{SIR}}(\mathbf{w}_{\text{qzf}}) = \frac{E[|\mathbf{w}_{\text{qzf}}^H \mathbf{h}_d|^2]}{\sum_{i=1}^K E[|\mathbf{w}_{\text{qzf}}^H \mathbf{h}_i|^2]} \sim \frac{N}{K} 2^{2B} \quad (7.2)$$

This result is also supported by the simulation results we present in this section. Unlike the spatial matched filter, the zero forcing beamformer is a function of both the desired signal and the interfering channel responses. While the SIR at the output of the beamformer grows linearly with the ratio N/K (Figure 7.20(a) shows a constant SIR when N/K is constant), the relationship with both N and K individually is a little more complicated. The reason is that when \mathbf{h}_d is projected onto the subspace orthogonal to the interference subspace, we lose the component of \mathbf{h}_d that lies in the interference subspace (\mathbf{h}_d^{\parallel}). When $N \gg K$, $\|\mathbf{h}_d^{\parallel}\|$ will be negligible, and the SIR will be linear in N and $1/K$. However, when the interference subspace is larger than the orthogonal subspace ($K > N/2$), there will be significant gains in increasing N and reducing K . This is shown in Figures 7.19ab, and this gain in SIR is a result of increasing the desired signal power (Figures 7.21ab). Also, unlike the spatial matched filter, the zero forcing beamformer gets approximately 6dB boost in SIR for every additional quantization bit (Figures 7.19ab and 7.20ab). This gain in SIR comes mostly from interference rejection component of the SIR, and slightly from the desired signal (Figures 7.23ab, 7.24ab). While the average interferer power at the output of the beamformer decreases exponentially with the number of quantization bits B , like the spatial matched filter, it is largely independent of both N and K . The zero forcing beamformer also differs from the spatial matched filter in the dependence of the desired signal on the number of interferers. However, the loss in the desired signal power is insignificant compared to the large gain obtained from interference nulling.

The results in Figures 7.19/7.20, 7.21/7.22, and 7.23/7.24 were obtained using the Rayleigh channel model. We repeated the same simulations under the LOS channel. In Figures 7.25/7.26, 7.9/7.10, and 7.29/7.30, we simulated a uniform linear array with $\lambda/2$ spacing,

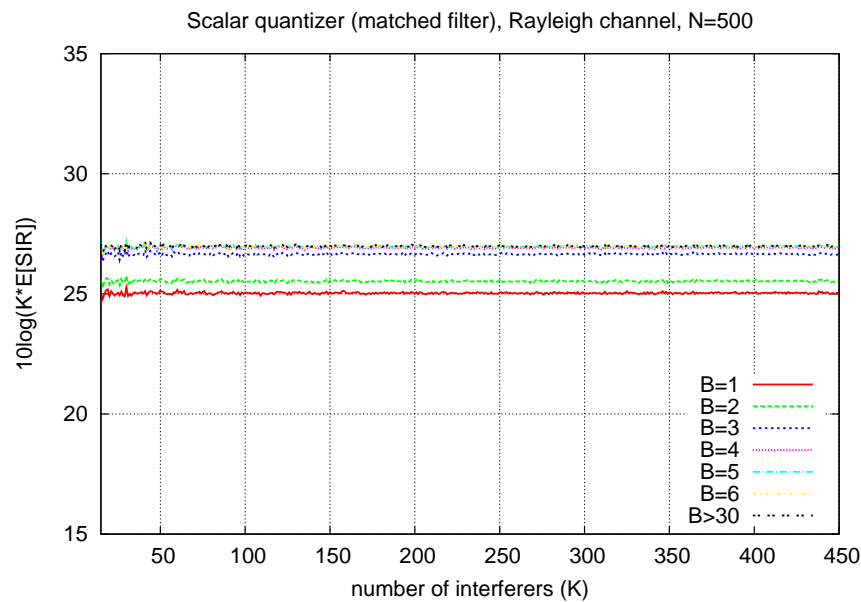
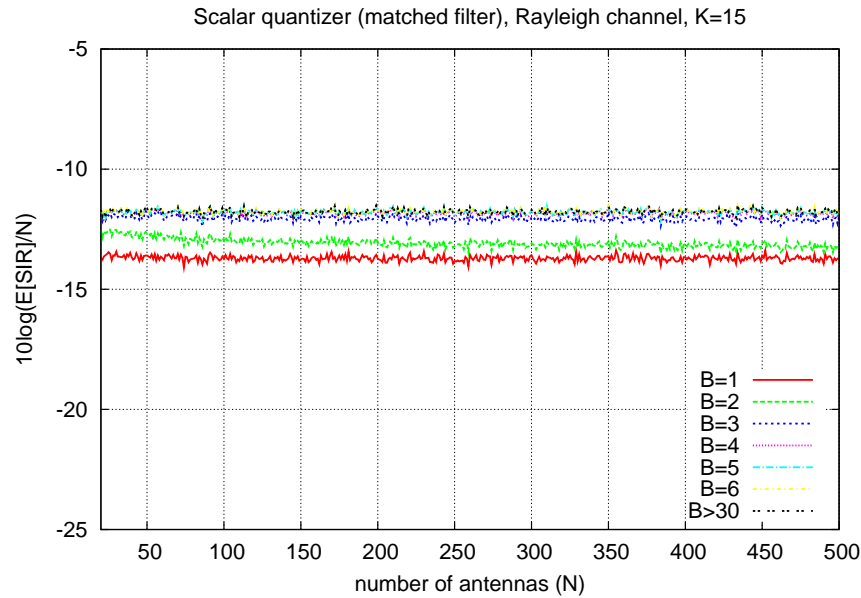


Figure 7.1: Simulated (averaged) SIR under the quantized spatial matched filter beamformer (QMF) in a Rayleigh channel. (a) SIR (normalized by N in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) SIR (multiplied by K in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

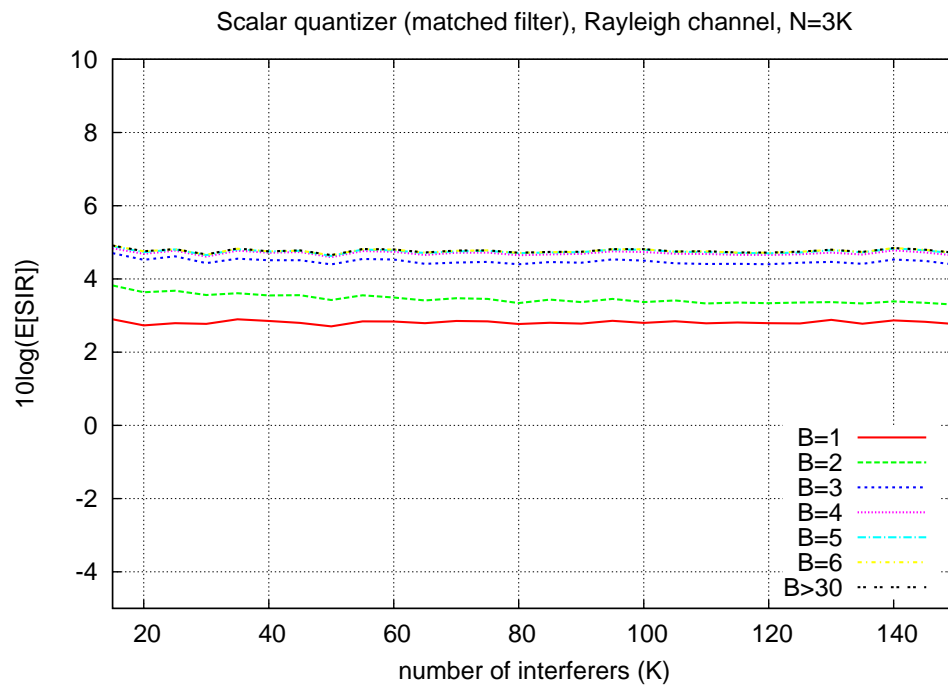


Figure 7.2: Simulated (averaged) SIR under the quantized spatial matched filter beamformer (QMF) in a Rayleigh channel. The SIR (in dB scale) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

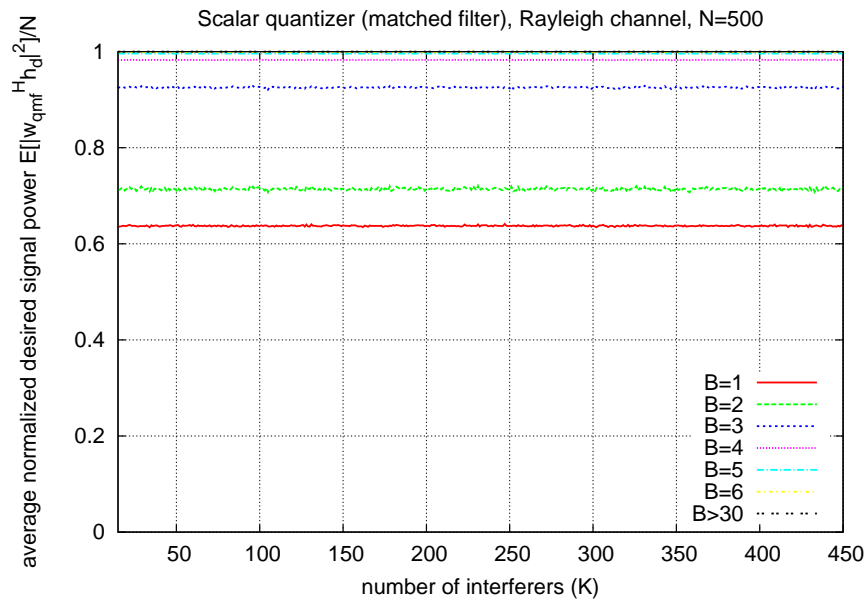
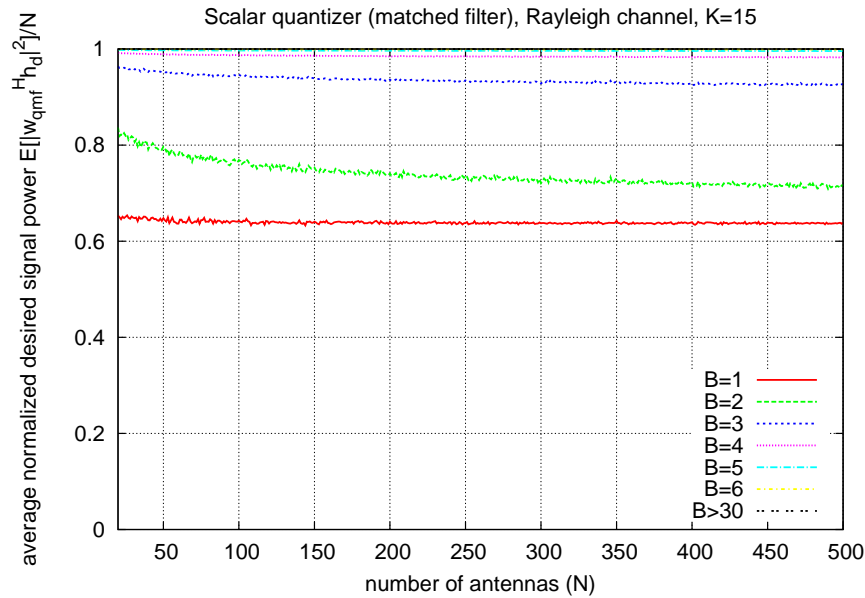


Figure 7.3: Simulated (averaged) desired signal under the quantized spatial matched filter beamformer (QMF) in a Rayleigh channel. (a) Desired signal (normalized by N) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Desired signal (normalized by N) as a function of the number of interferers K for different quantization levels B ($N = 150$).

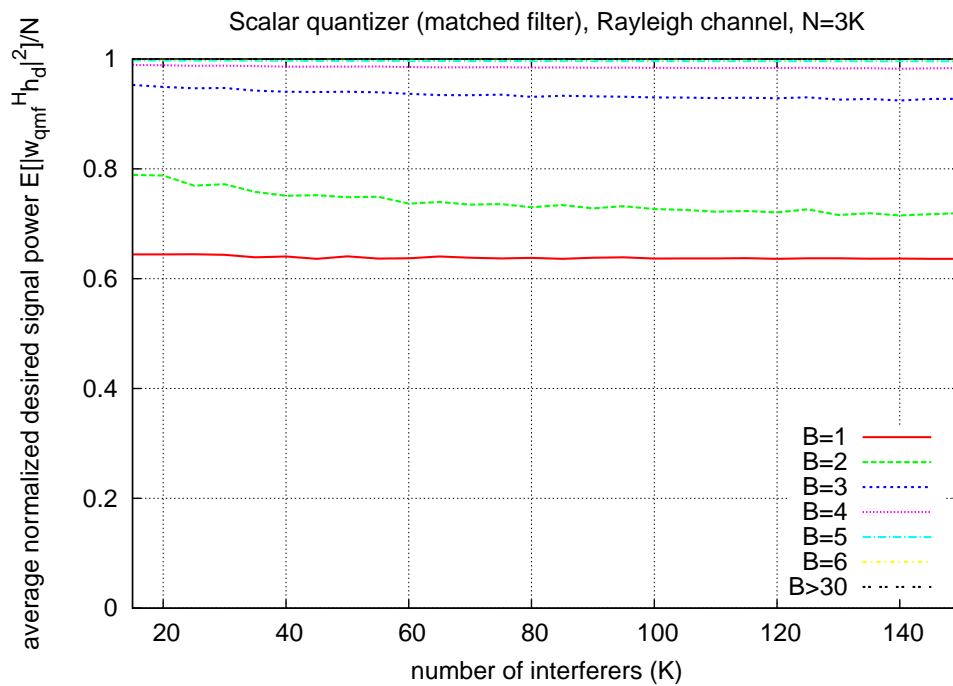


Figure 7.4: Simulated (averaged) desired signal under the quantized spatial matched filter beamformer (QMF) in a Rayleigh channel. The desired signal (normalized by N) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

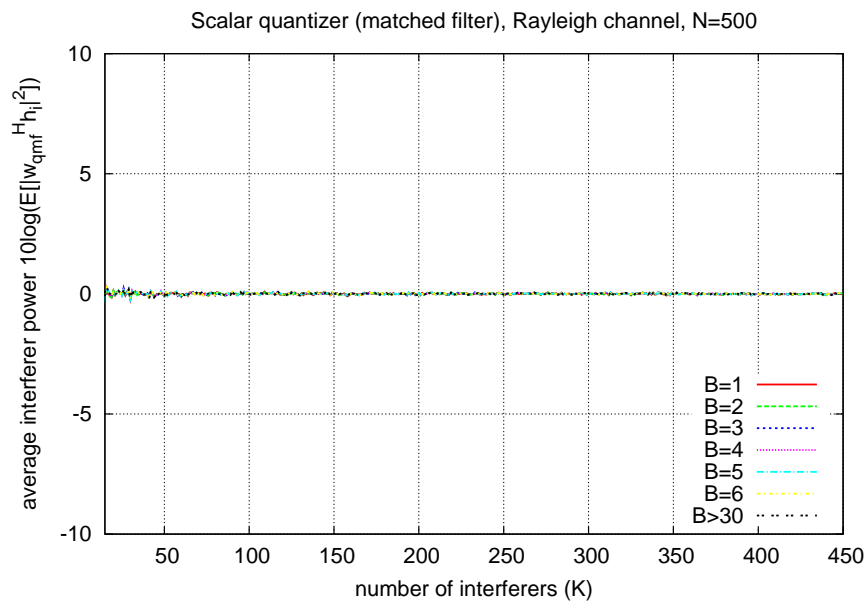
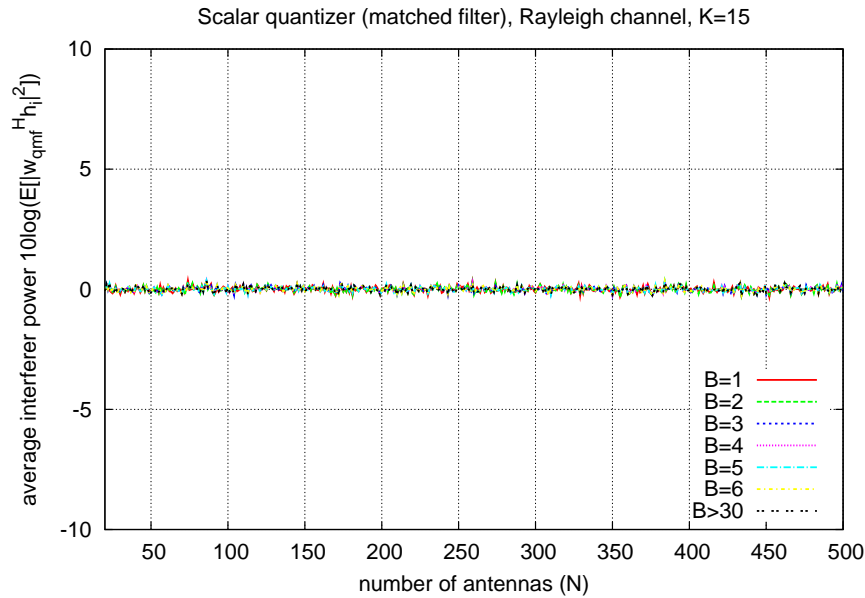


Figure 7.5: Simulated (averaged) power of a single interferer under the quantized spatial matched filter beamformer (QMF) in a Rayleigh channel. (a) Interferer power (in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

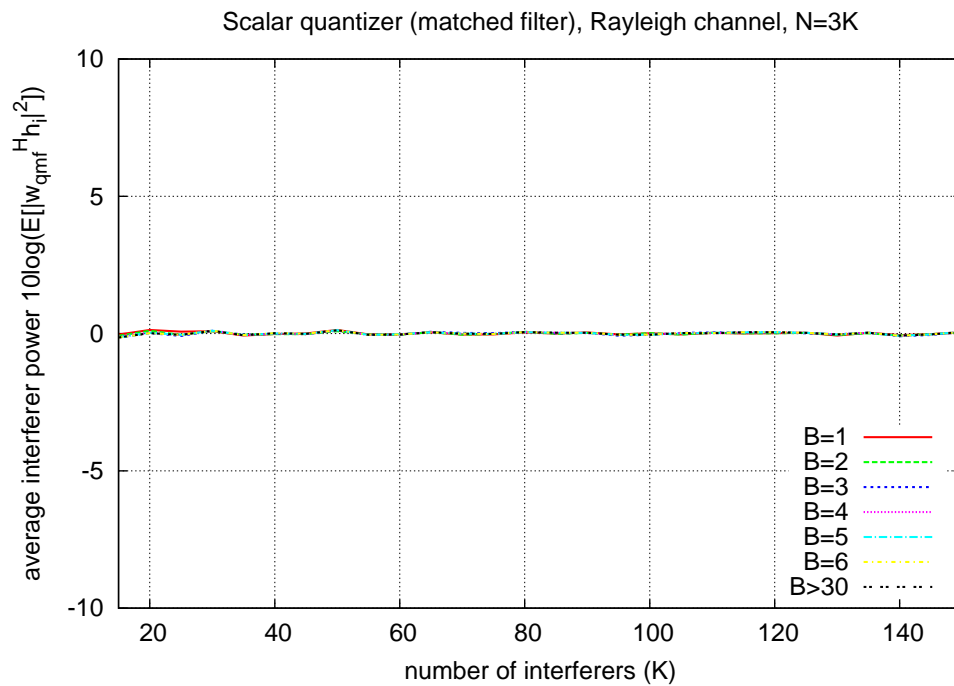


Figure 7.6: Simulated (averaged) power of a single interferer under the quantized spatial matched filter beamformer (QMF) in a Rayleigh channel. The interferer power (in dB scale) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

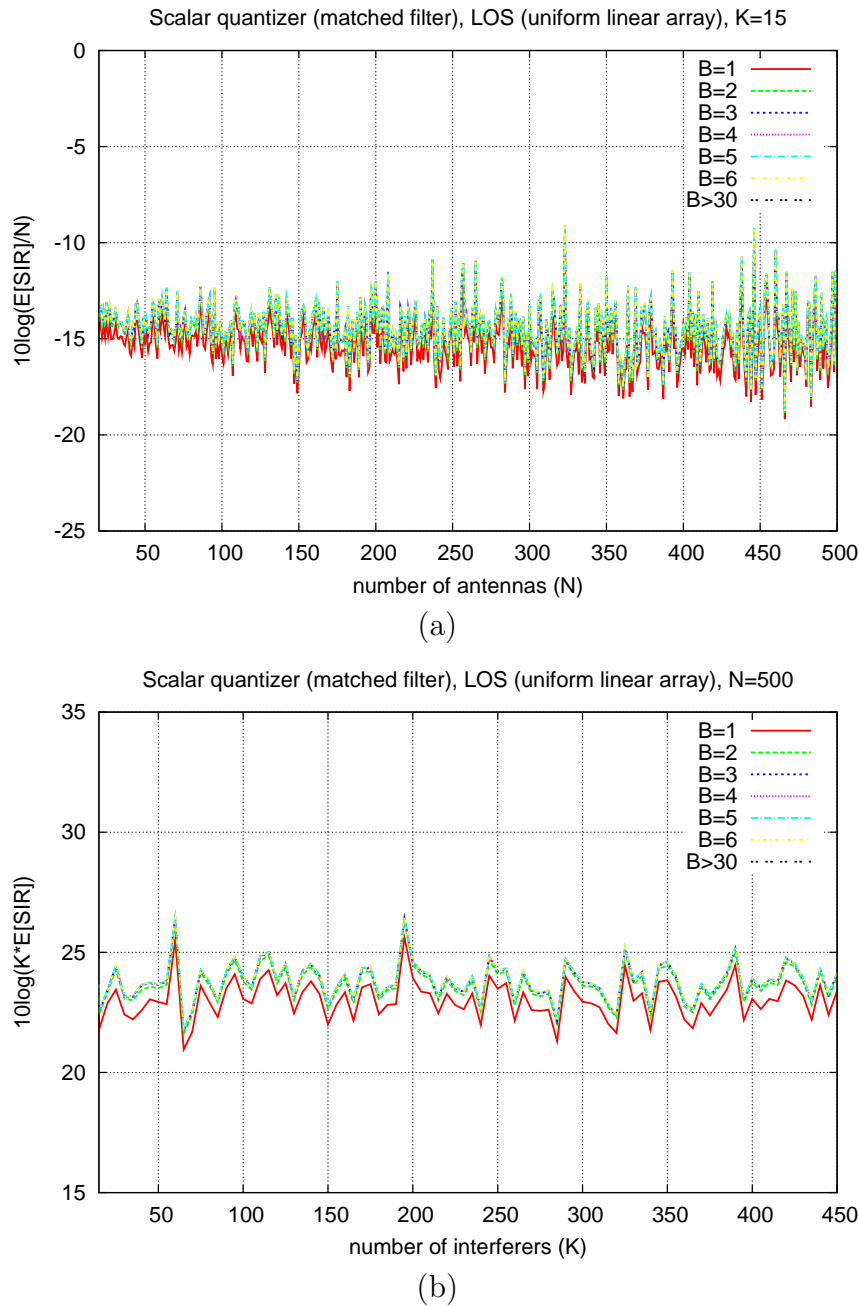


Figure 7.7: Simulated (averaged) SIR under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. (a) SIR (normalized by N in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) SIR (multiplied by K in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

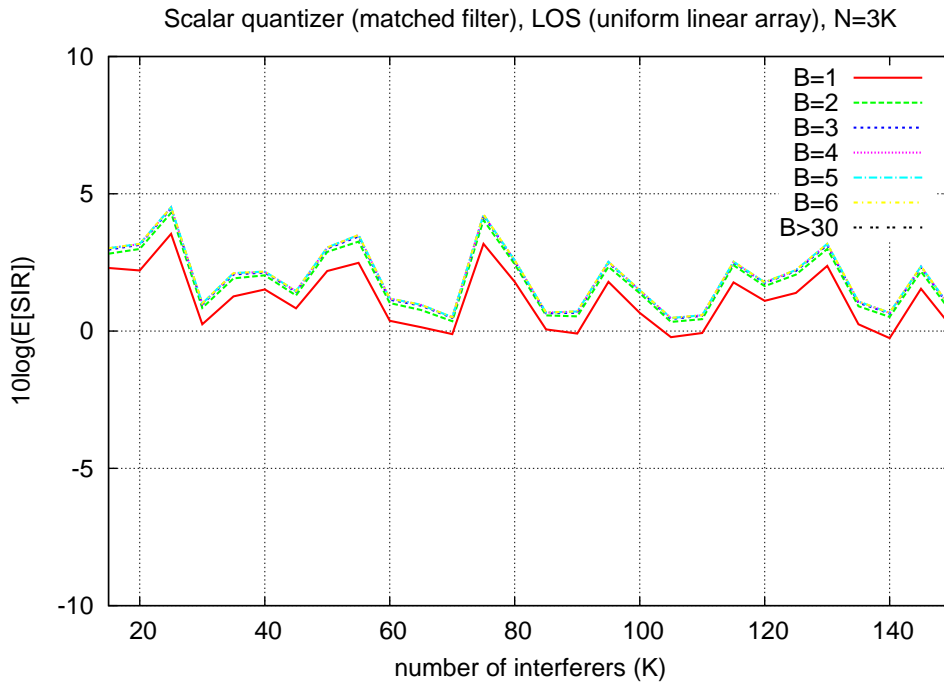


Figure 7.8: Simulated (averaged) SIR under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. The SIR (in dB scale) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

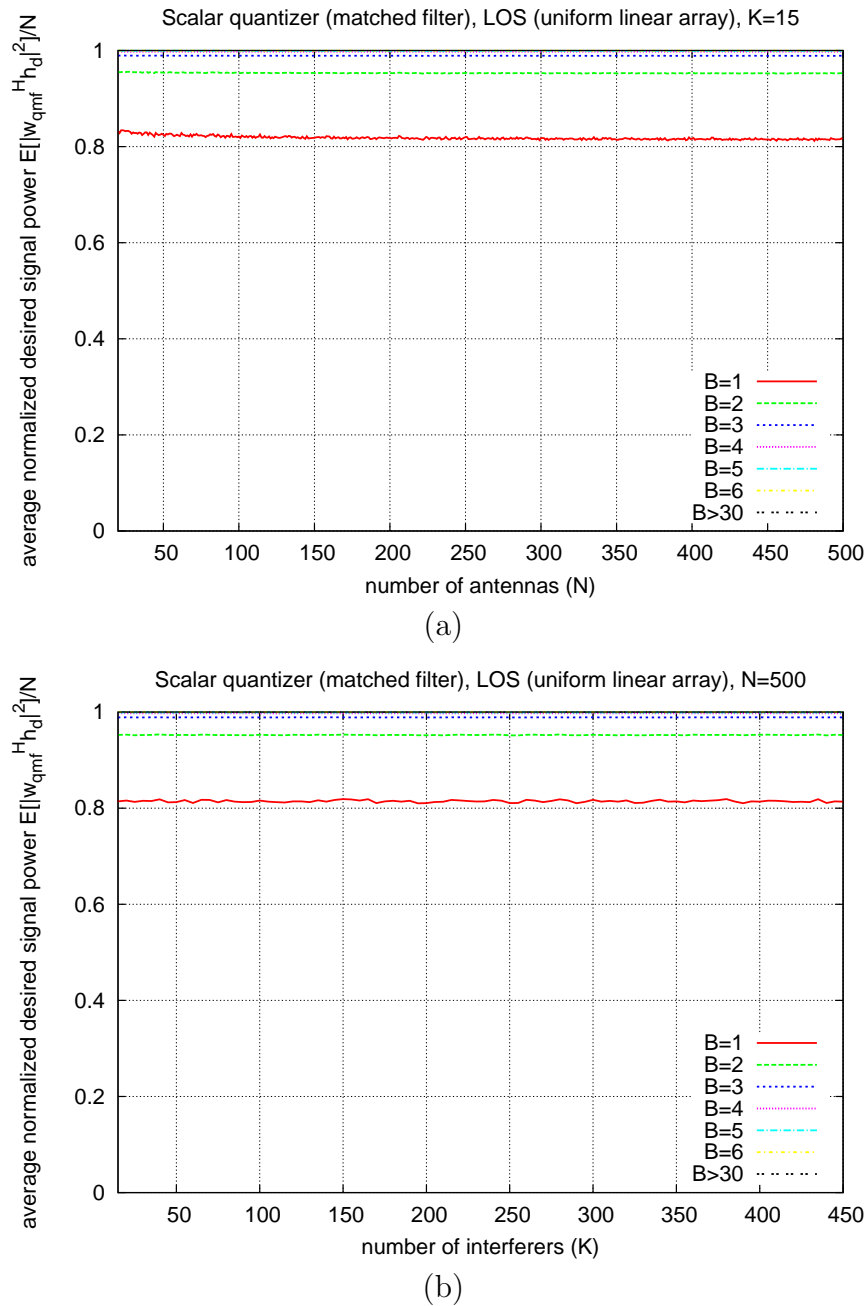


Figure 7.9: Simulated (averaged) desired signal under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. (a) Desired signal (normalized by N) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Desired signal (normalized by N) as a function of the number of interferers K for different quantization levels B ($N = 150$).



Figure 7.10: Simulated (averaged) desired signal under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. The desired signal (normalized by N) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

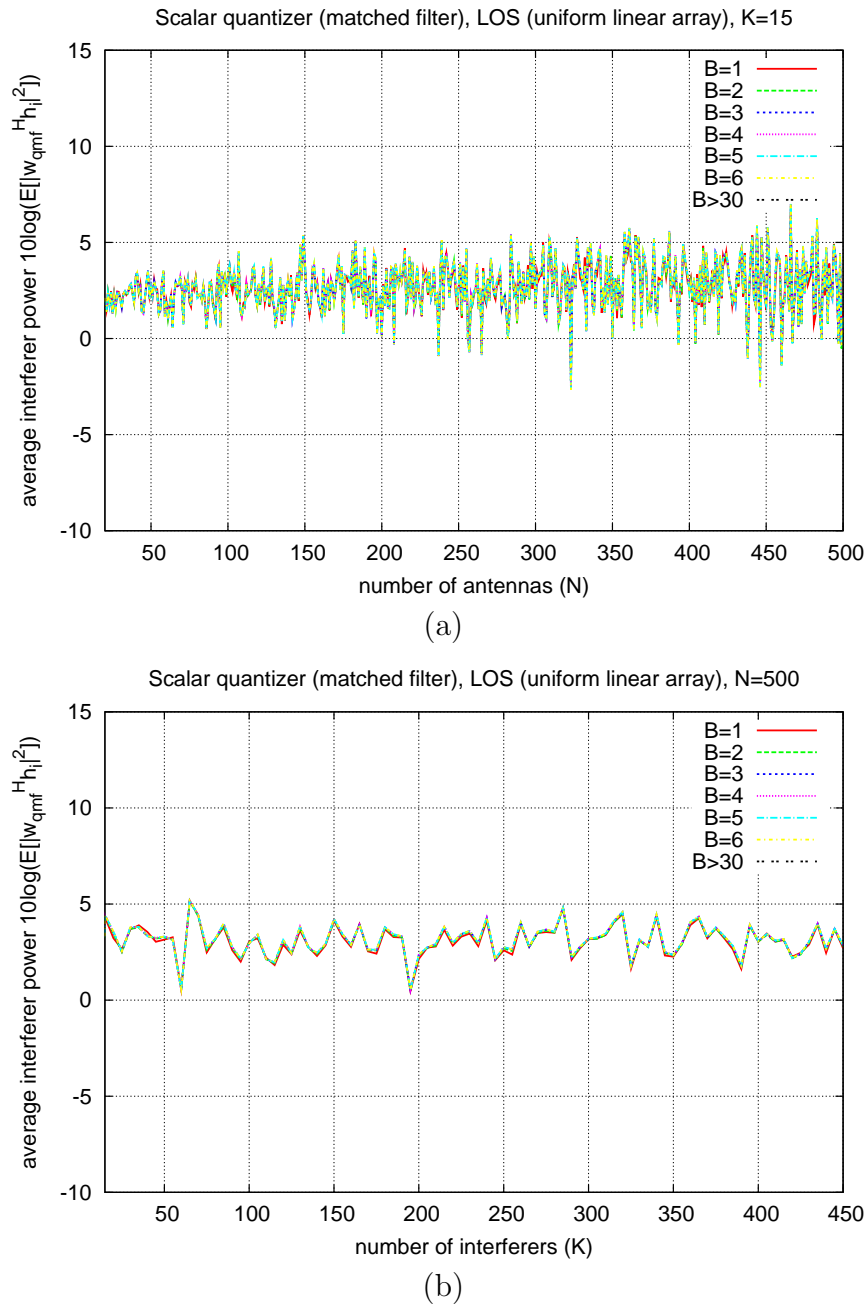


Figure 7.11: Simulated (averaged) power of a single interferer under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. (a) Interferer power (in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

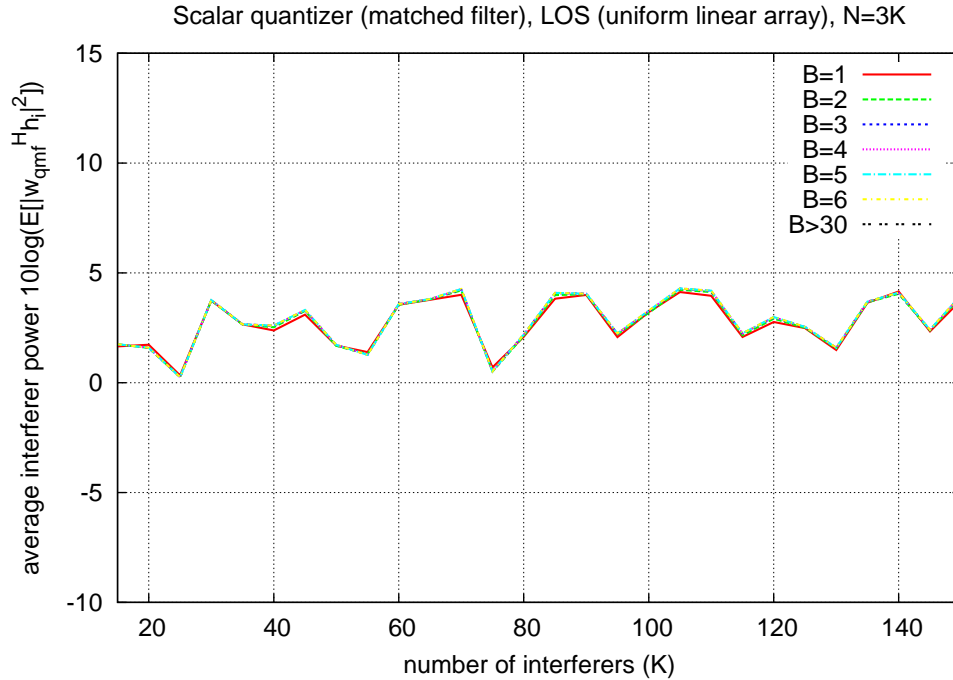


Figure 7.12: Simulated (averaged) power of a single interferer under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. The interferer power (in dB scale) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

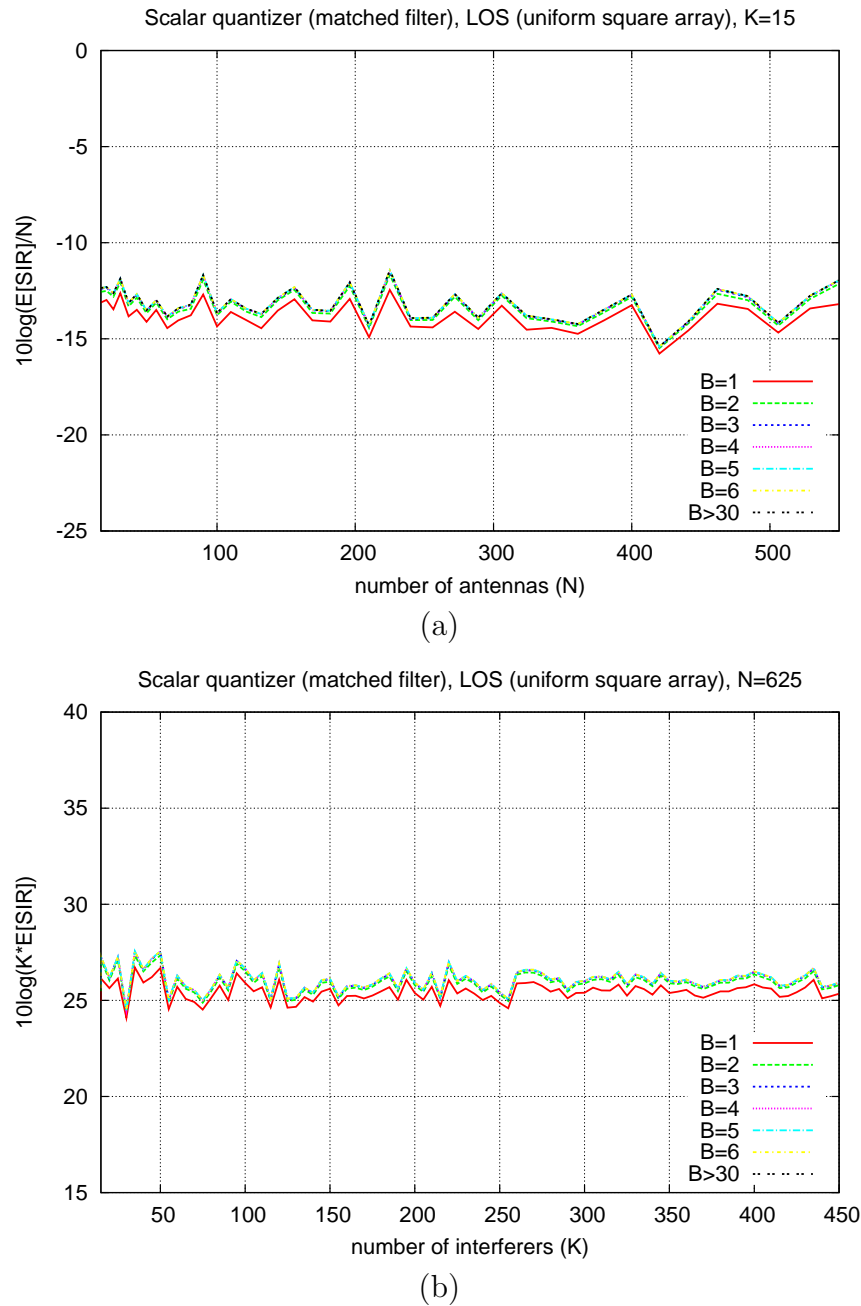


Figure 7.13: Simulated (averaged) SIR under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. (a) SIR (normalized by N in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) SIR (multiplied by K in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

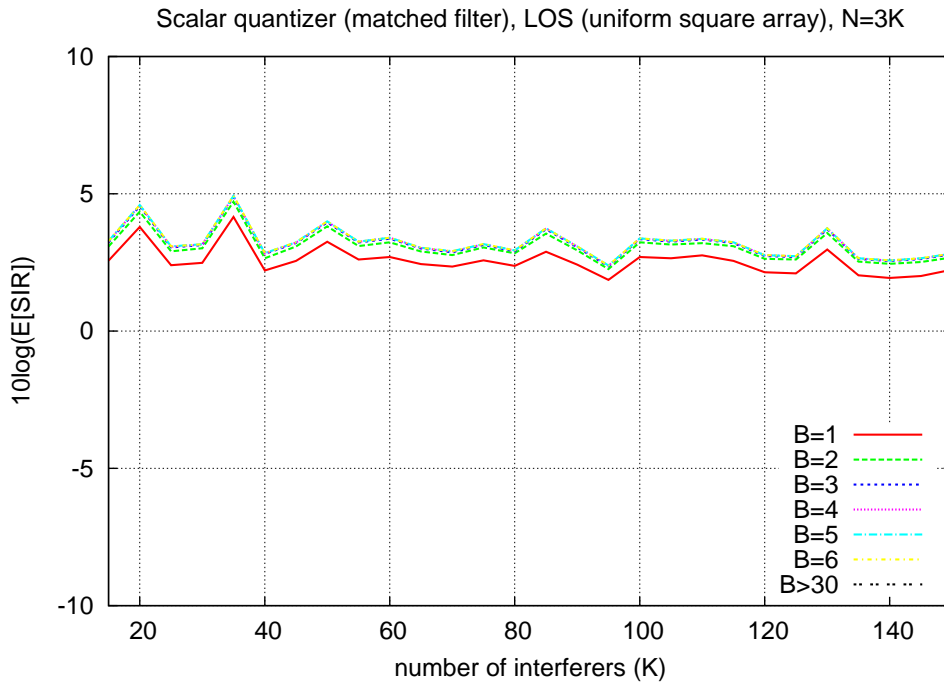


Figure 7.14: Simulated (averaged) SIR under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. The SIR (in dB scale) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

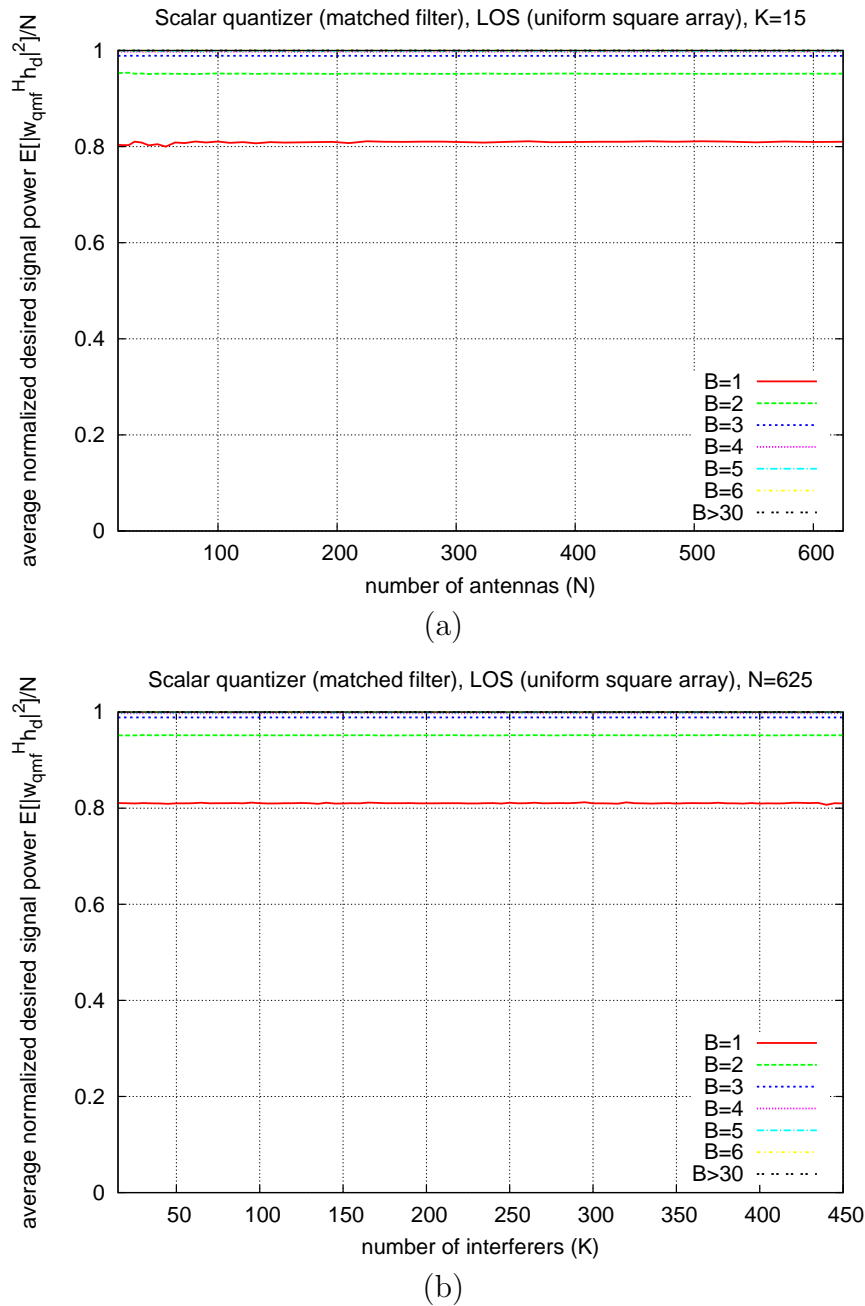


Figure 7.15: Simulated (averaged) desired signal under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. (a) Desired signal (normalized by N) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Desired signal (normalized by N) as a function of the number of interferers K for different quantization levels B ($N = 150$).

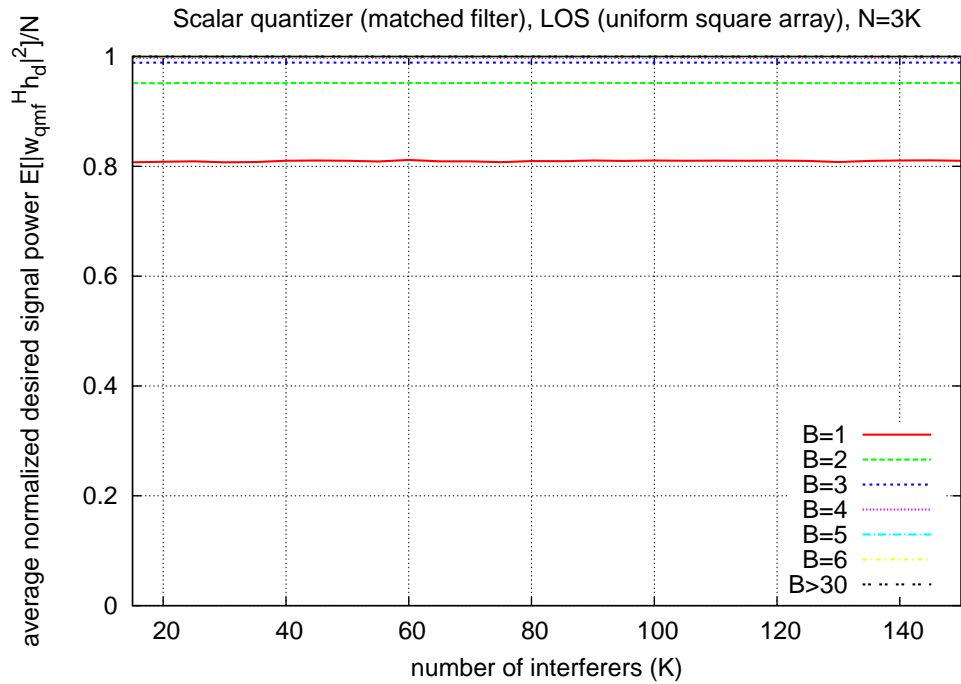


Figure 7.16: Simulated (averaged) desired signal under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. The desired signal (normalized by N) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

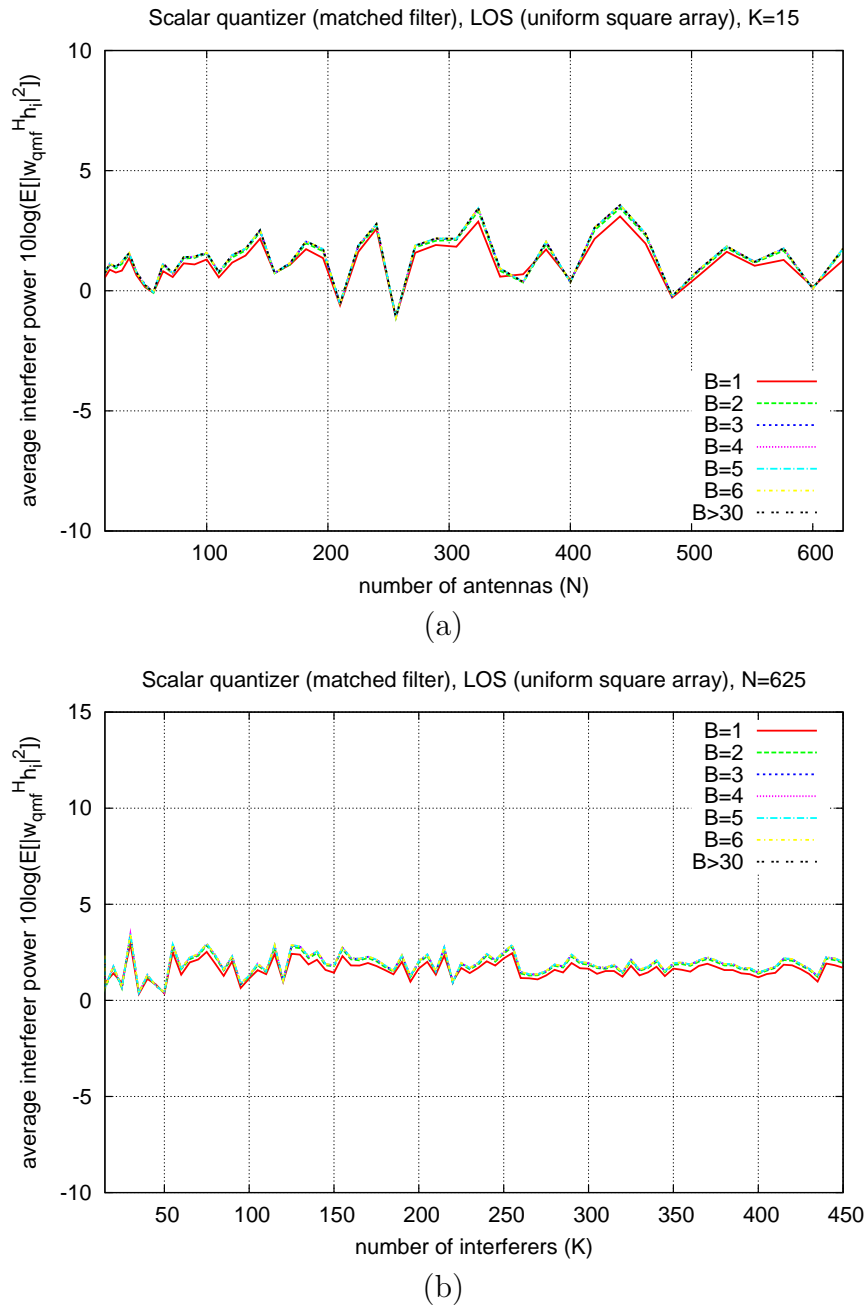


Figure 7.17: Simulated (averaged) power of a single interferer under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. (a) Interferer power (in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

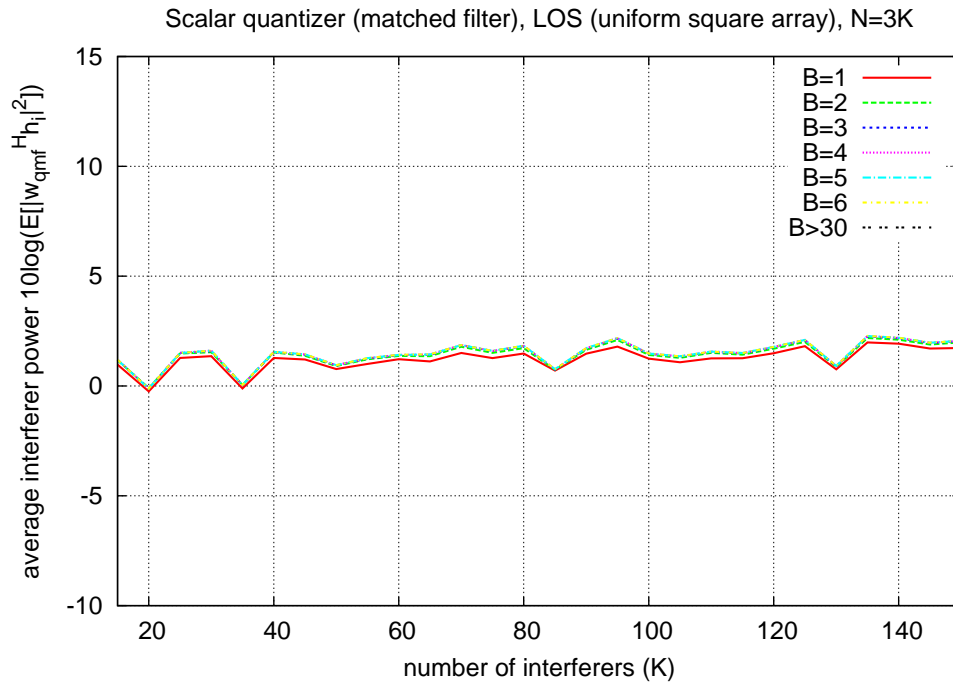


Figure 7.18: Simulated (averaged) power of a single interferer under the quantized spatial matched filter beamformer (QMF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. The interferer power (in dB scale) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

and randomized the azimuth angle of arrival of each transmitter. In Figures 7.31/7.32, 7.33/7.34, and 7.35/7.36, we simulated a uniform square array with $\lambda/2$ spacing, and randomized the direction of arrival (in 3 dimensions) of each transmitter. These results show that the performance of the quantized spatial matched filter is consistent with with Equation 7.2 independent of the channel model.

The quantized ZF beamformer improves the average SIR/interference rejection by a factor of 2^{2B} relative to the spatial matched filter. However, this improved SIR comes at the expense of increased computational complexity. Since the quantized ZF beamformer requires solving an $N \times K$ complex least squares problem, the overall computational complexity is $O(NK^2)$. In both cases, however, the interference leakage remains independent of the number of antennas N .

7.3 Vector quantization of antenna weights

In Section 7.2, we showed that a naive strategy of coefficient-by-coefficient quantization of the beamforming weights is not very efficient: the average SIR scales as N/K . That is, we obtain the combining gains for the desired signal, but increasing the number of degrees of freedom N does not lead to any improved scaling in the residual interference power at the output of the beamformer. In this section, we explore more sophisticated vector quantization strategies. In Section 7.3.1, we establish an analytical lower bound on the expected SIR improvement using vector quantization techniques. In Section 7.3.2, we propose three computationally efficient vector quantization schemes and analyze their performance, optimality, and complexity as a function of N , K , and B . Finally, in Section 7.3.3, we analyze the beamforming gain using vector quantization techniques.

7.3.1 An approximate lower bound on achievable SIR

In [7], we showed that the SIR that can be achieved by intelligently choosing the quantized weights has a lower bound that grows quadratically with the number of antennas N . The proof idea is very simple and intuitive. We can think of the problem of choosing the optimal weight vector \mathbf{w} as a problem of choosing the optimal error vector $\Delta\mathbf{w}$. Our goal is to show that when the components of $\Delta\mathbf{w}$ are chosen *dependently* of each other, the interference power can be made to decrease significantly with N (as opposed being independent of N when the components are chosen independently).

For simplicity, we will assume a single interferer or $K = 1$ (The extension to the more general case $K > 1$ is straight forward). Our goal is to minimize the term $|\Delta\mathbf{w}^H \mathbf{h}_1|^2$, when the coefficients of \mathbf{h}_1 are independent zero-mean complex Gaussian random variables with unit variance (i.e. Rayleigh channel coefficients). Let's start by considering a simpler problem. Suppose we simply wish to minimize $|\mathbf{u}^H \mathbf{h}_1|^2$, where \mathbf{u} takes all possible values in $\mathcal{U} = \frac{1}{\sqrt{N}}\{-1, +1\}^N$ (i.e. the elements of \mathcal{U} are binary vectors normalized to unit norm). Let $T(\mathbf{u}) \triangleq \mathbf{u}^H \mathbf{h}_1$. Note that $T(\mathbf{u}) \sim CN(0, 1)$ for any $\mathbf{u} \in \mathcal{U}$ since the coefficients of \mathbf{h}_1 are also assumed to be $CN(0, 1)$. Therefore, $T(\mathbf{u}_1), T(\mathbf{u}_2)$ are jointly complex Gaussian with covariance equal to $\mathbf{u}_1^H \mathbf{u}_2$. We now take a subset $\mathcal{U}_o \subset \mathcal{U}$ which forms an orthonormal basis in N dimensions: the Walsh-Hadamard codes [29]. Then $\{T(\mathbf{u}) : \mathbf{u} \in \mathcal{U}_o\}$ are N independent

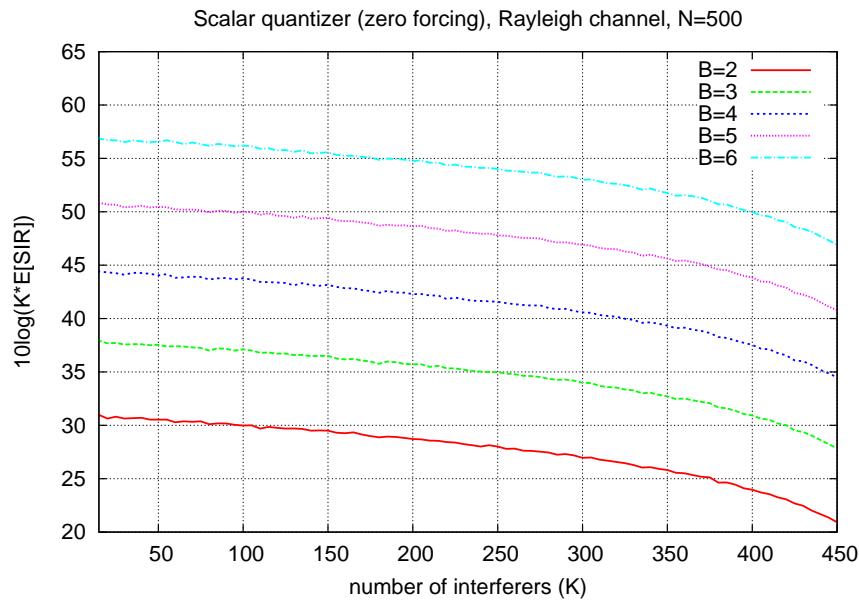
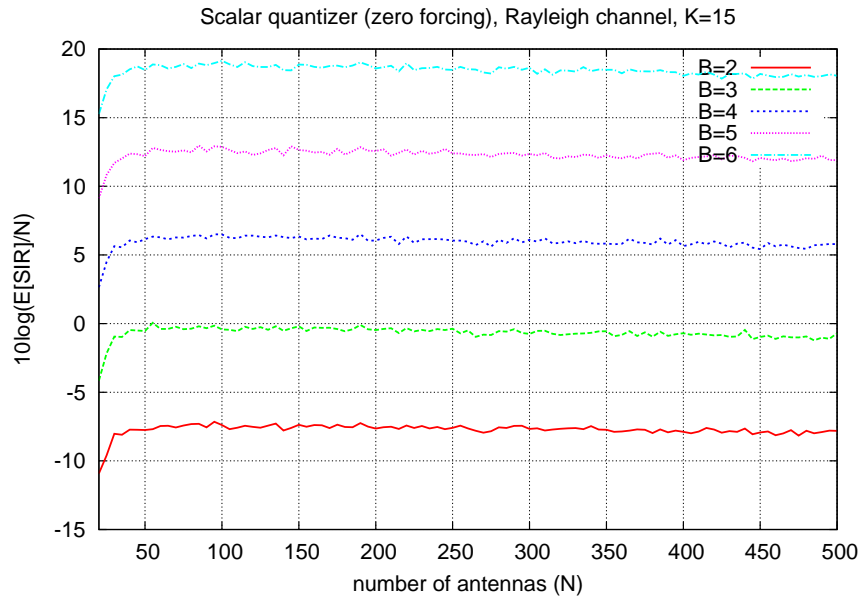


Figure 7.19: Simulated (averaged) SIR under the quantized zero-forcing beamformer (QZF) in a Rayleigh channel. (a) SIR (normalized by N in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) SIR (multiplied by K in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

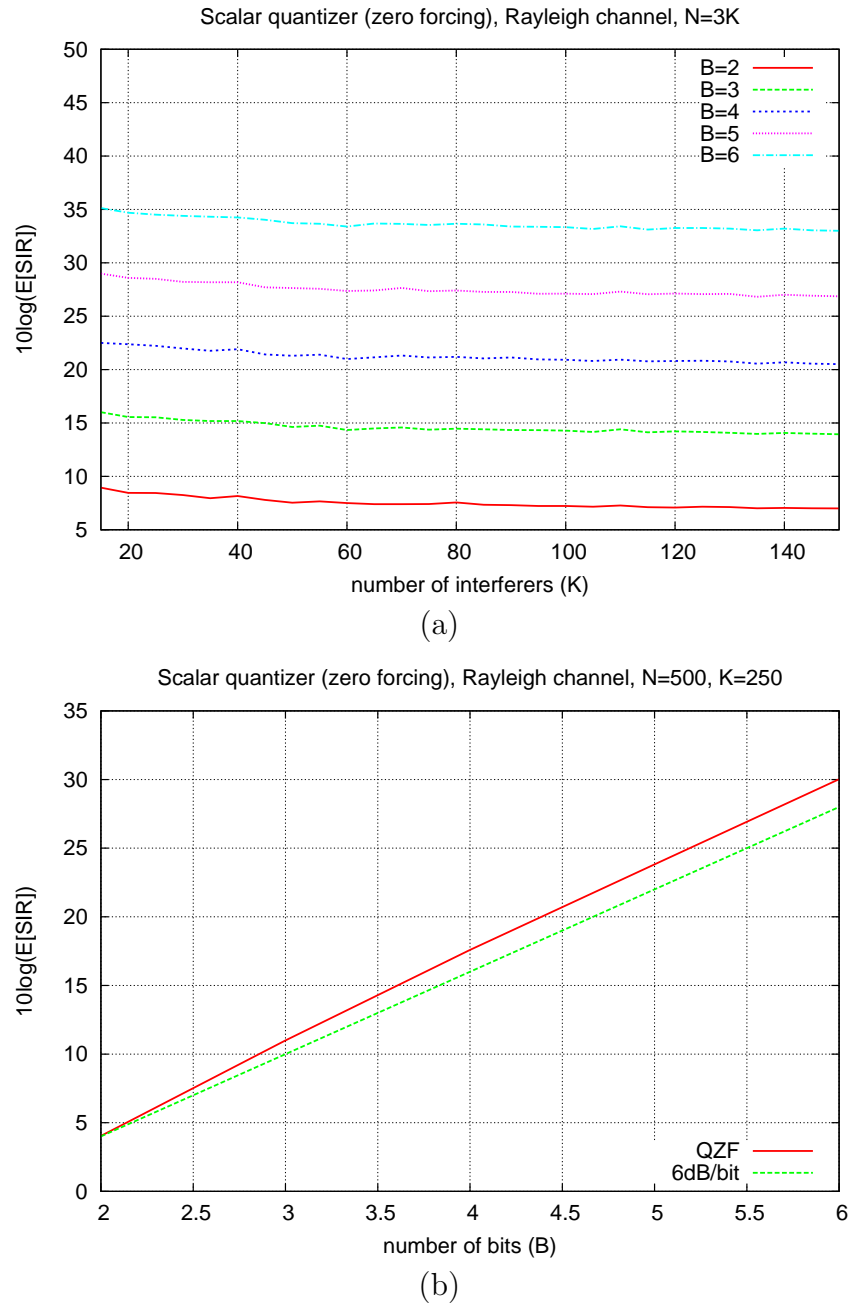


Figure 7.20: Simulated (averaged) SIR under the quantized zero-forcing beamformer (QZF) in a Rayleigh channel. (a) SIR (in dB scale) as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) SIR (in dB scale) as a function of the quantization level B for $N=500$ and $K=250$.

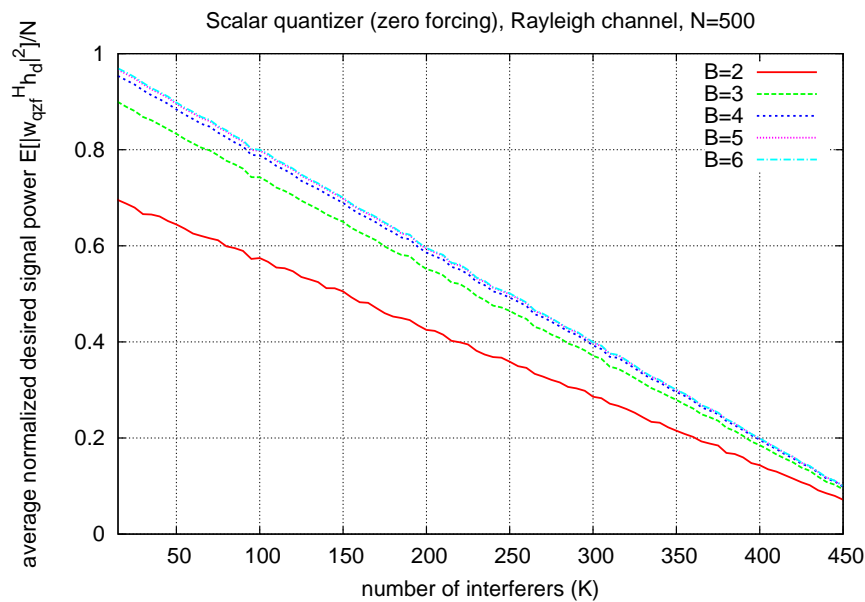
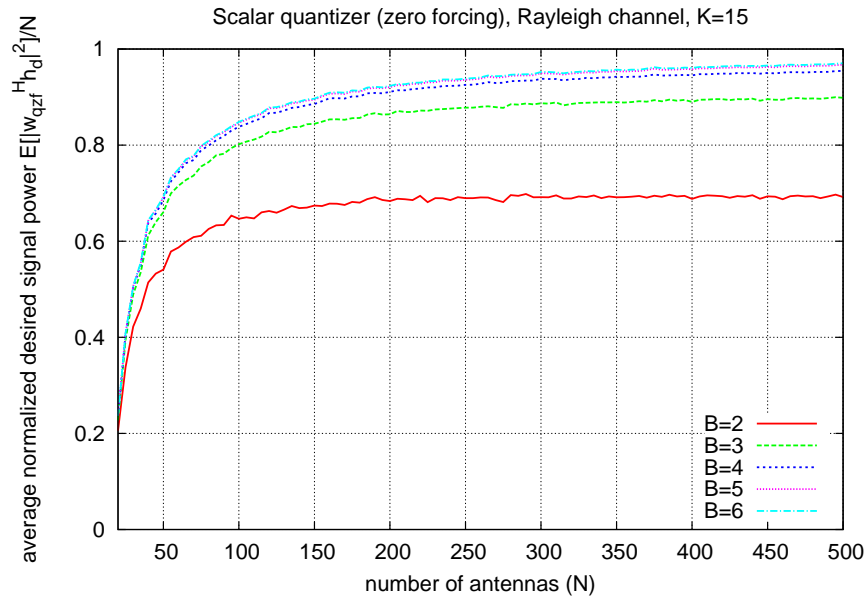


Figure 7.21: Simulated (averaged) desired signal under the quantized zero-forcing beam-former (QZF) in a Rayleigh channel. (a) Desired signal (normalized by N) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Desired signal (normalized by N) as a function of the number of interferers K for different quantization levels B ($N = 150$).

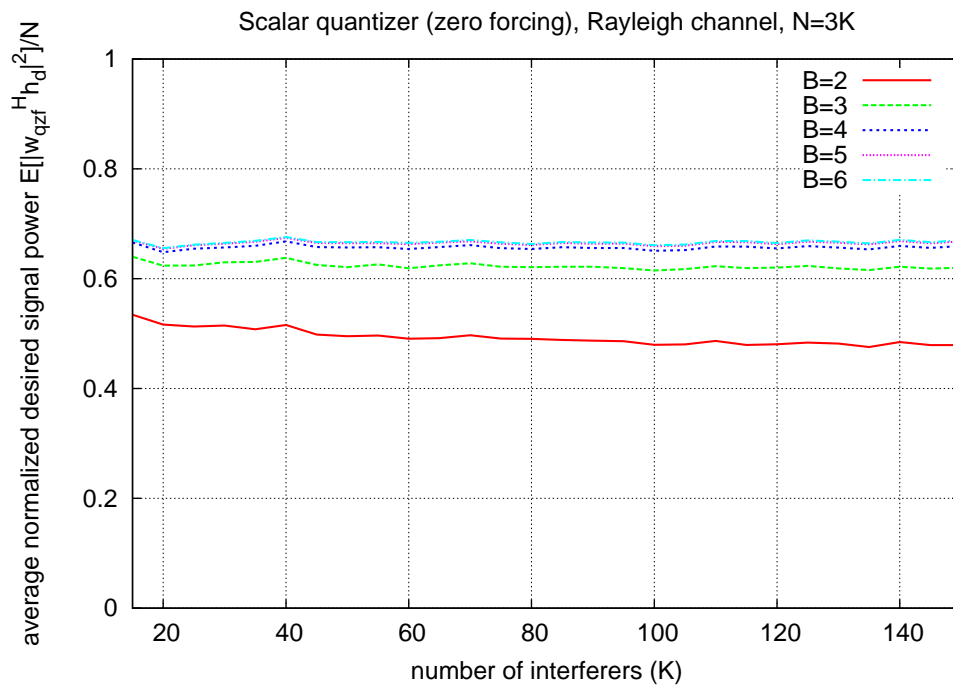


Figure 7.22: Simulated (averaged) desired signal under the quantized zero-forcing beamformer (QZF) in a Rayleigh channel. The desired signal (normalized by N) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

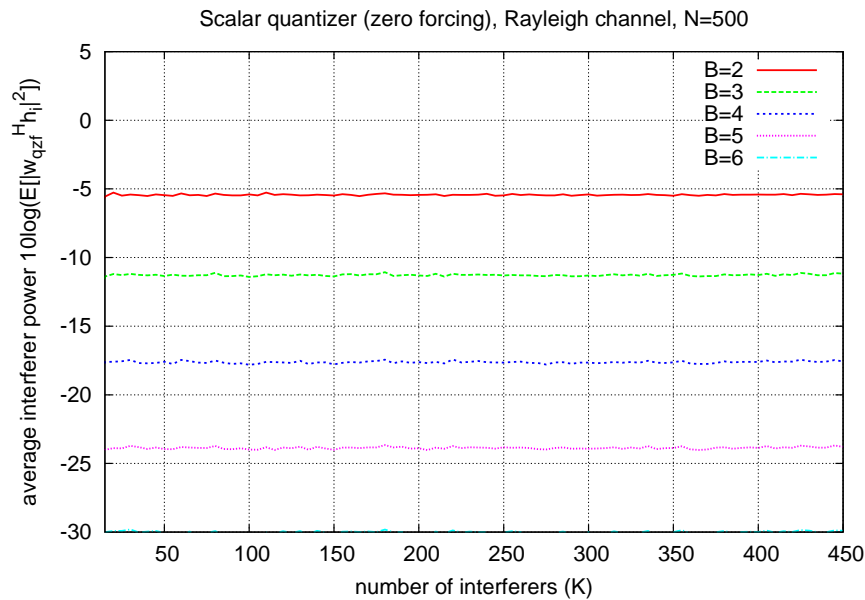
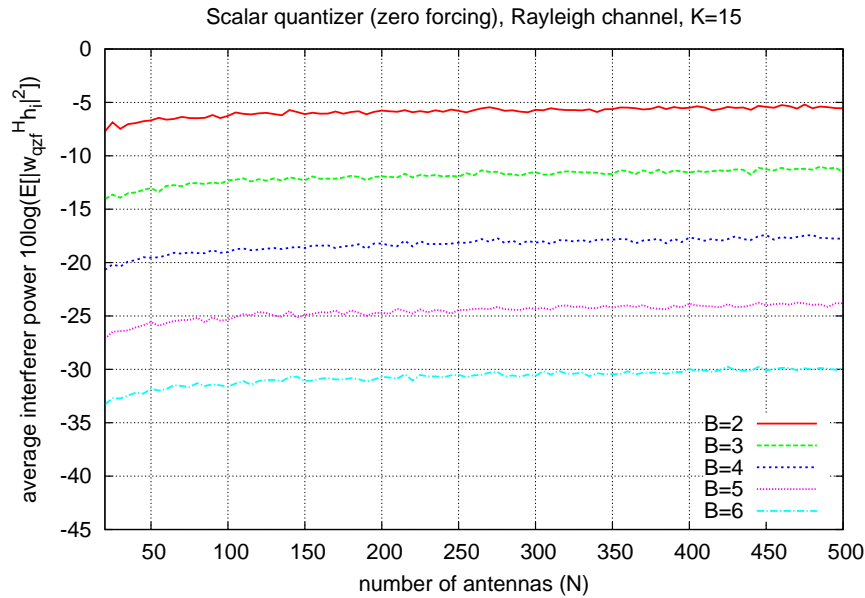


Figure 7.23: Simulated (averaged) power of a single interferer under the quantized zero-forcing beamformer (QZF) in a Rayleigh channel. (a) Interferer power (in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

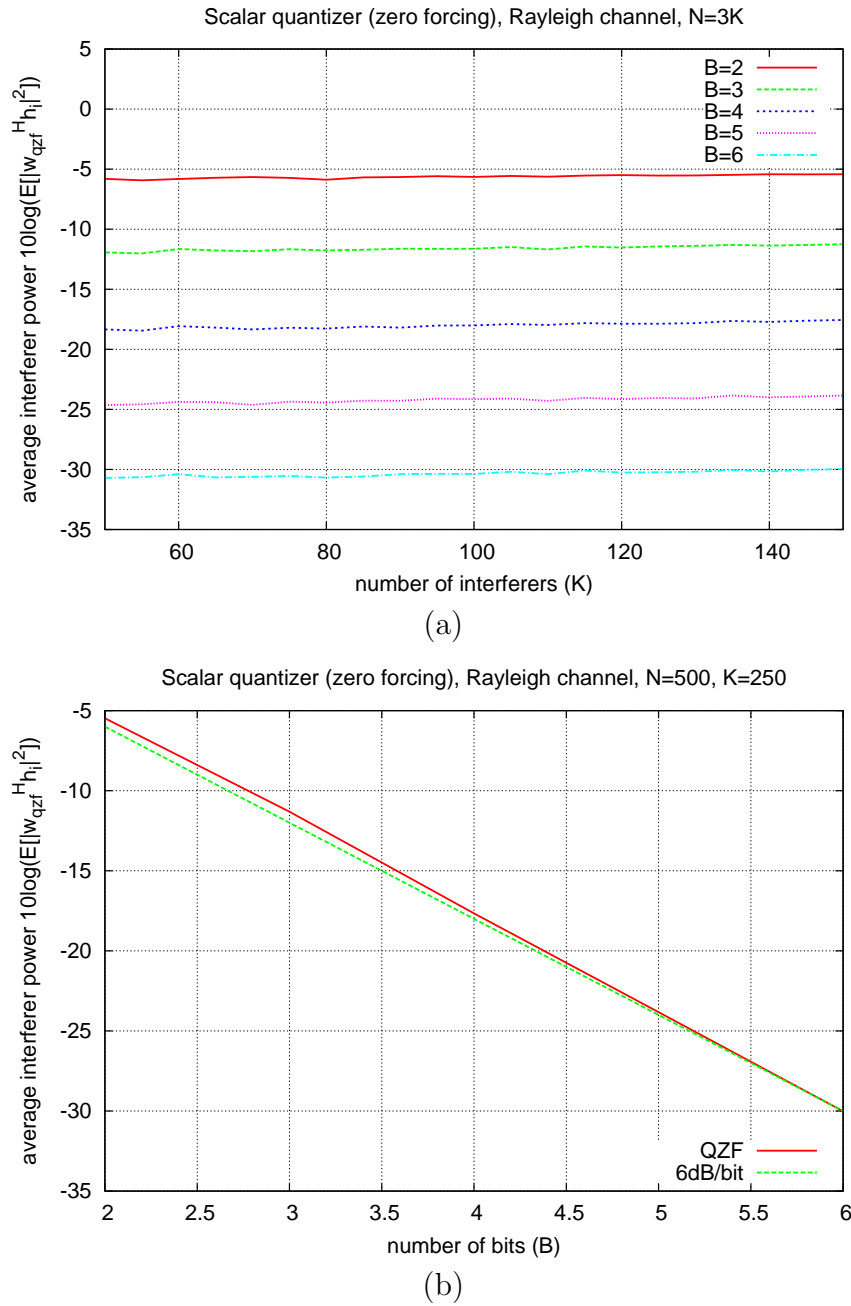


Figure 7.24: Simulated (averaged) power of a single interferer under the quantized zero-forcing beamformer (QZF) in a Rayleigh channel. (a) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Interferer power (in dB scale) as a function of the quantization level B for $N=500$ and $K=250$.

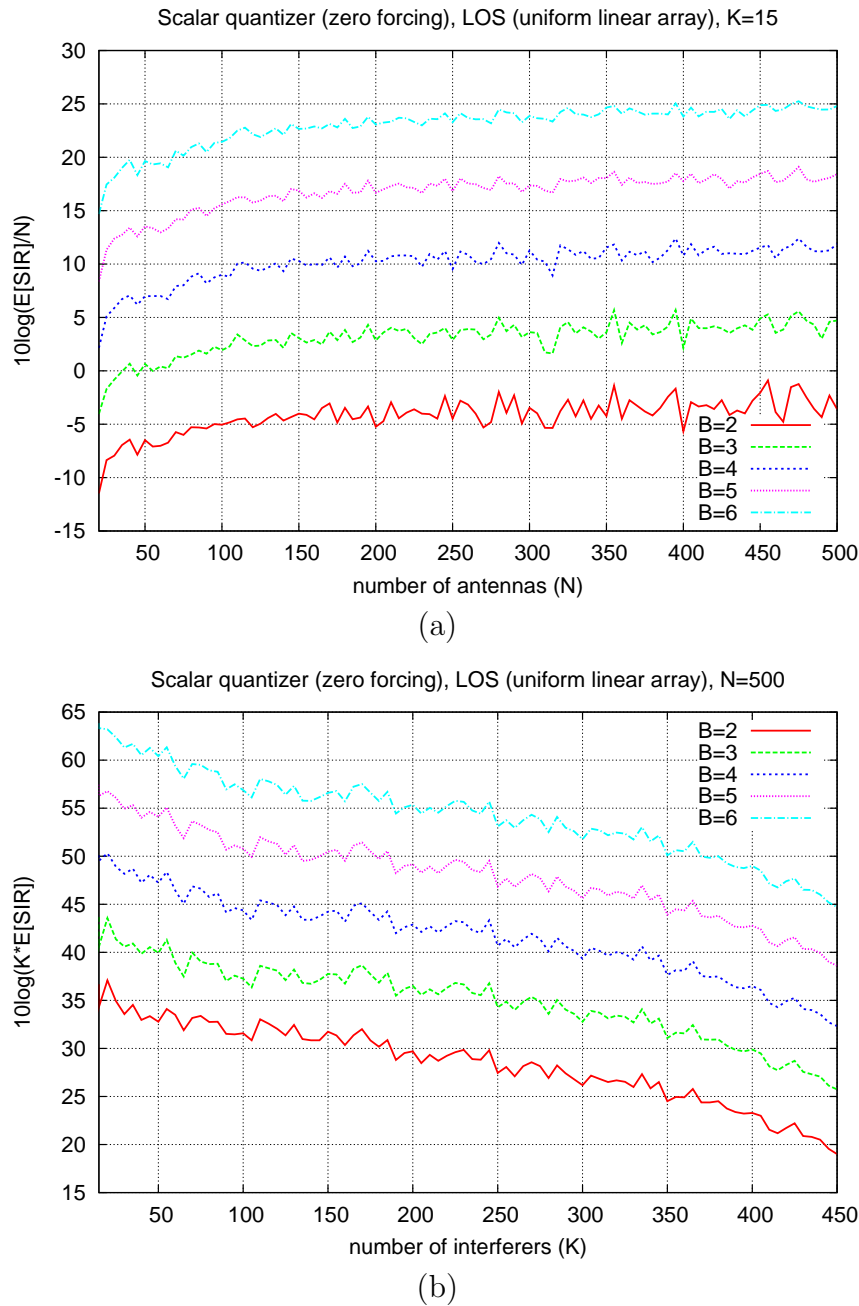


Figure 7.25: Simulated (averaged) SIR under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. (a) SIR (normalized by N in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) SIR (multiplied by K in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

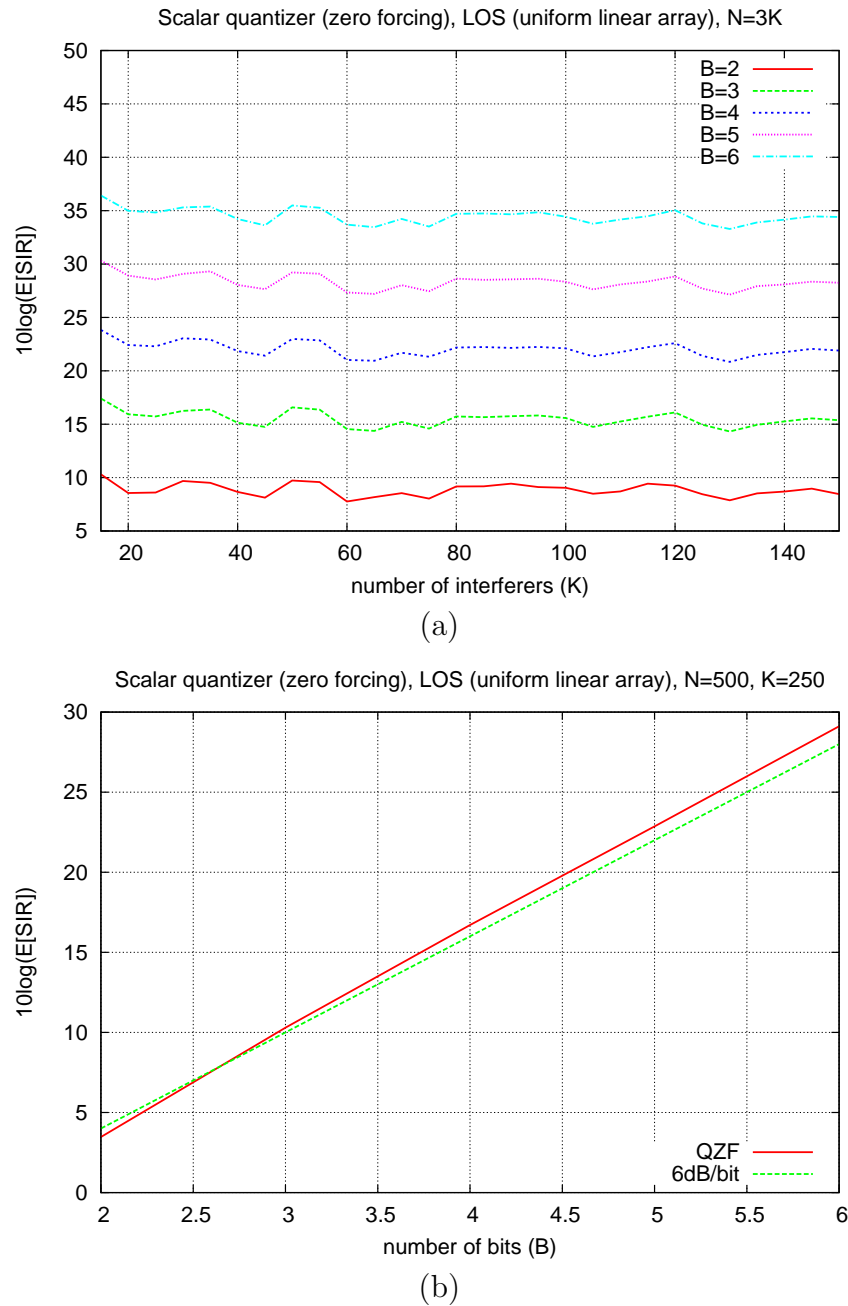


Figure 7.26: Simulated (averaged) SIR under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. (a) SIR (in dB scale) as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) SIR (in dB scale) as a function of the quantization level B for $N=500$ and $K=250$.

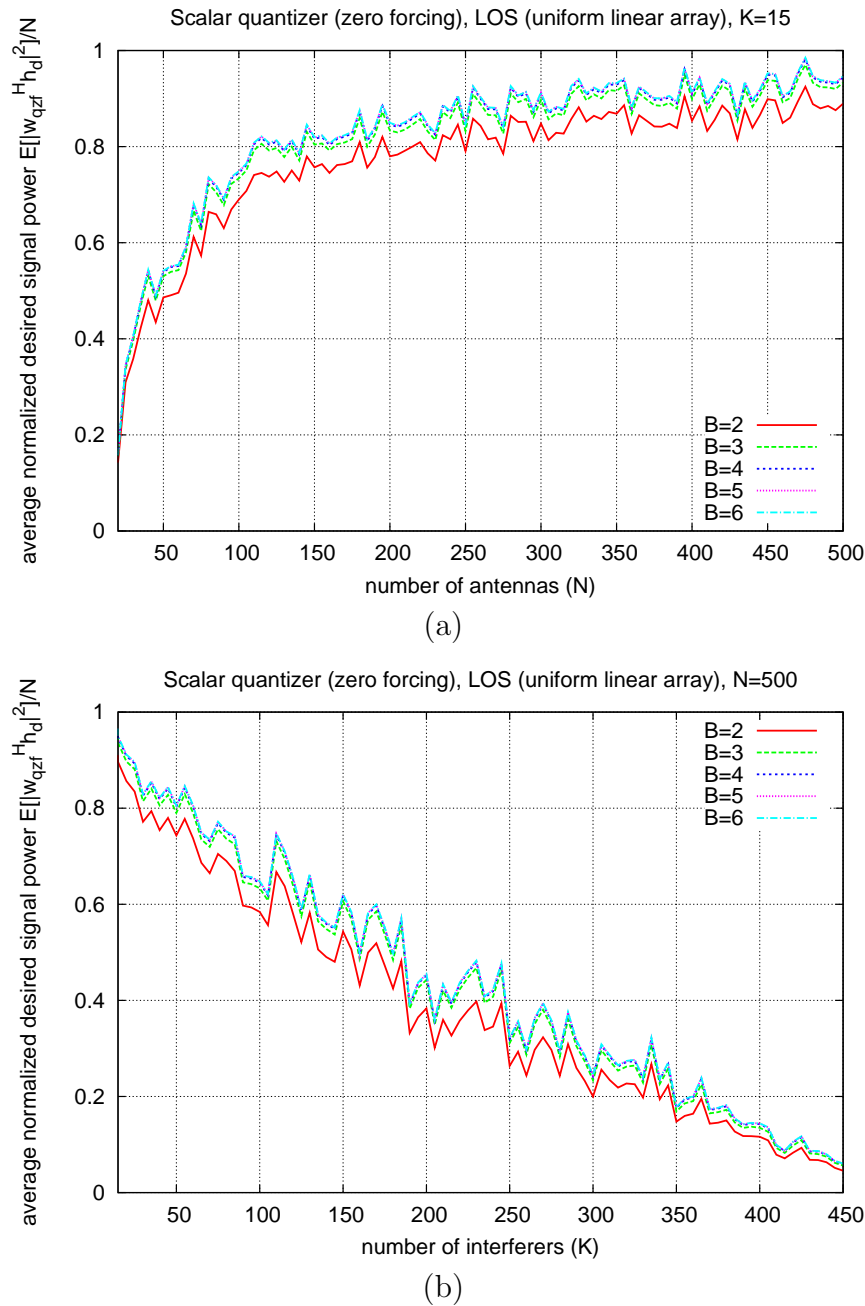


Figure 7.27: Simulated (averaged) desired signal under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. (a) Desired signal (normalized by N) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Desired signal (normalized by N) as a function of the number of interferers K for different quantization levels B ($N = 150$).

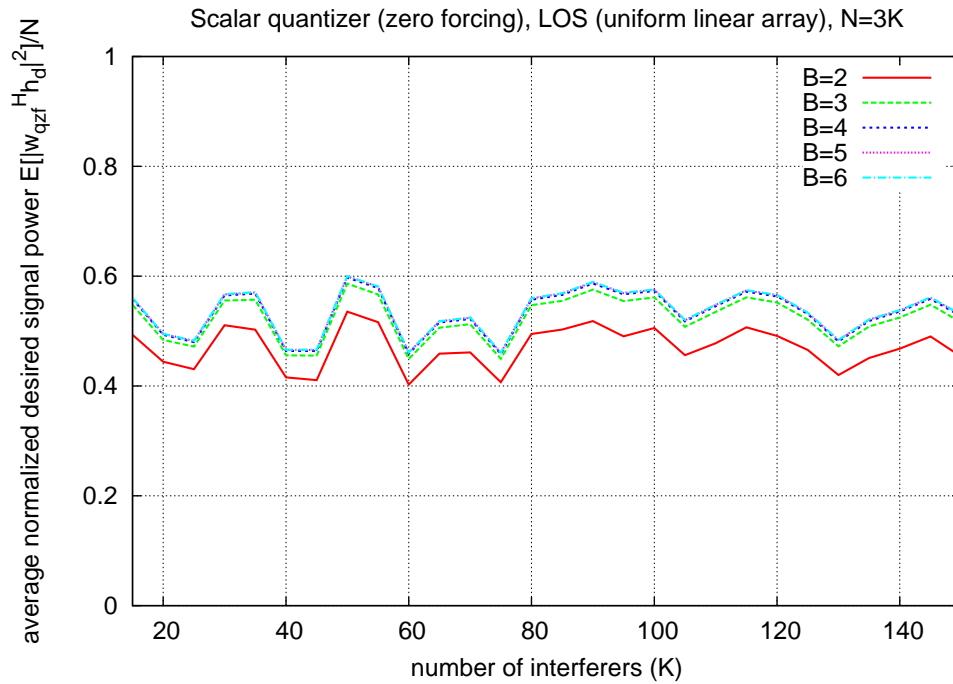


Figure 7.28: Simulated (averaged) desired signal under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. The desired signal (normalized by N) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

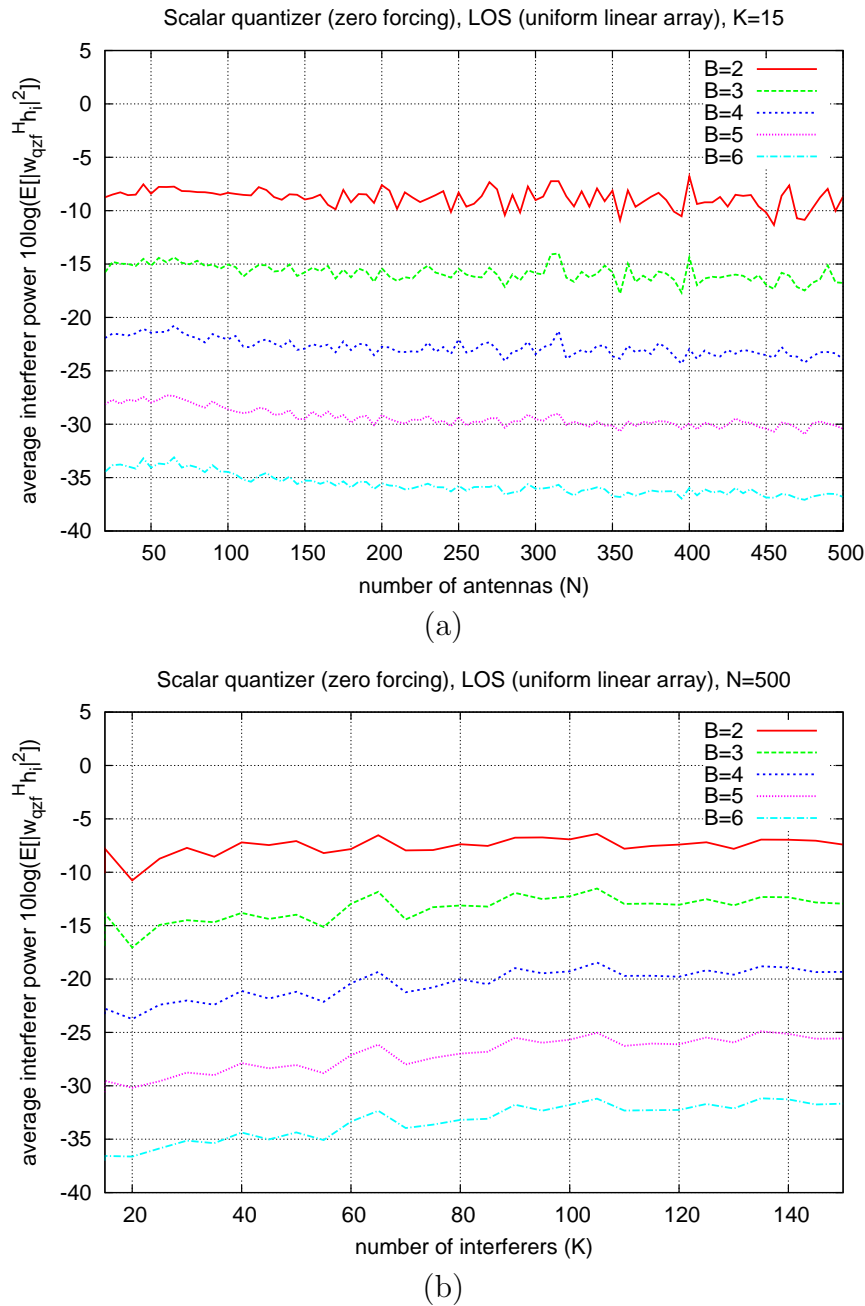


Figure 7.29: Simulated (averaged) power of a single interferer under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. (a) Interferer power (in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

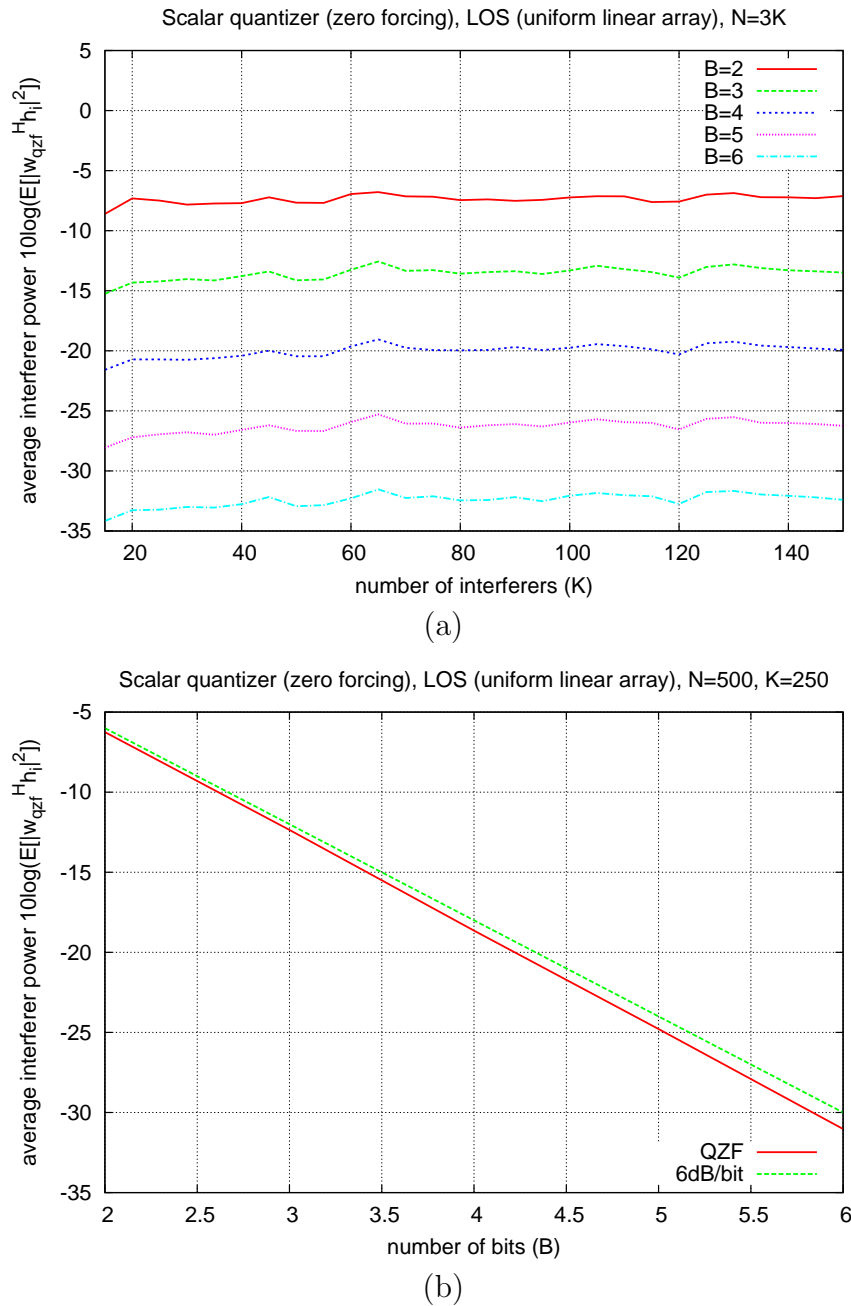


Figure 7.30: Simulated (averaged) power of a single interferer under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform linear array with $\lambda/2$ spacing, with all the antennas located on the xy -plane. Both the desired signal and the interferers are located on the xy -plane and their angle of arrival (azimuth) is chosen randomly at each iteration. (a) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Interferer power (in dB scale) as a function of the quantization level B for $N=500$ and $K=250$.

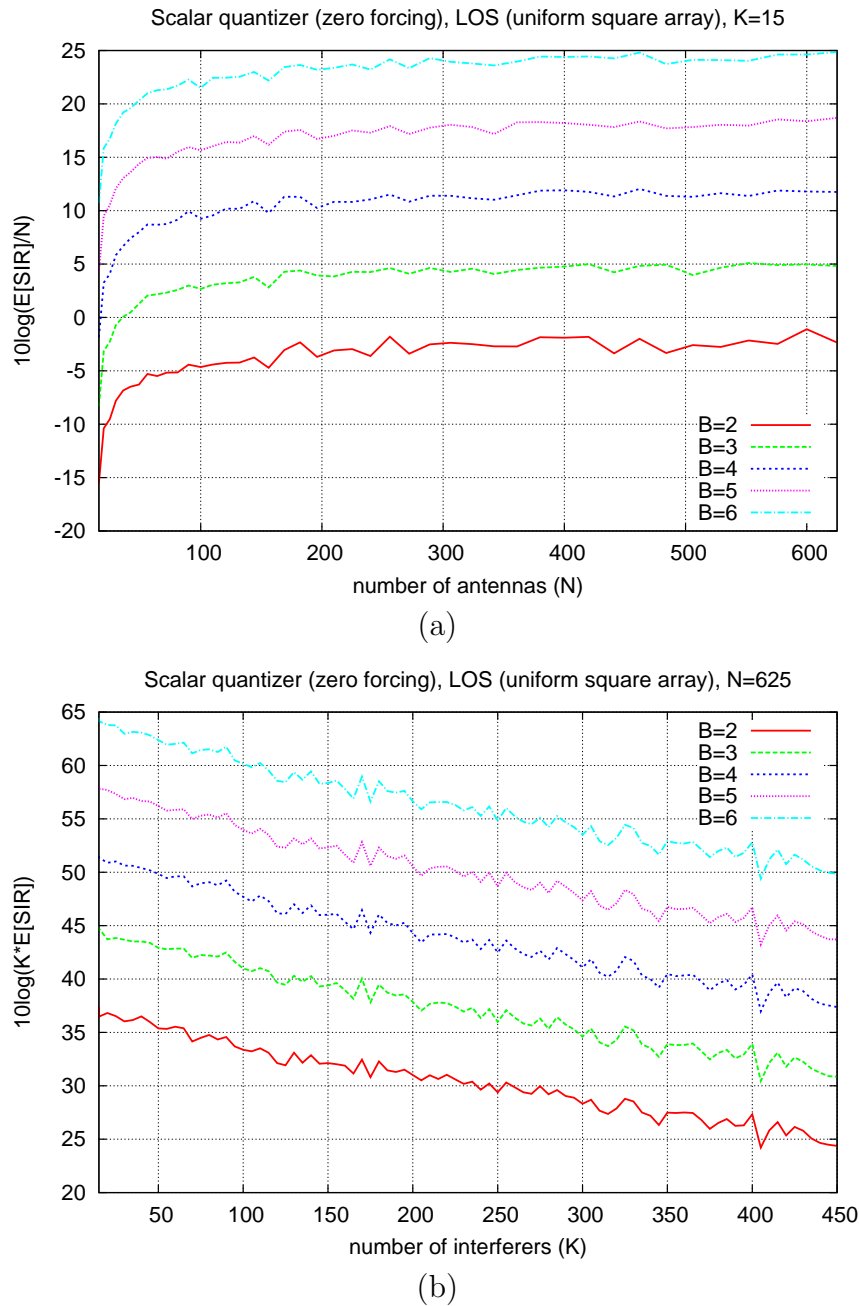


Figure 7.31: Simulated (averaged) SIR under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. (a) SIR (normalized by N in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) SIR (multiplied by K in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

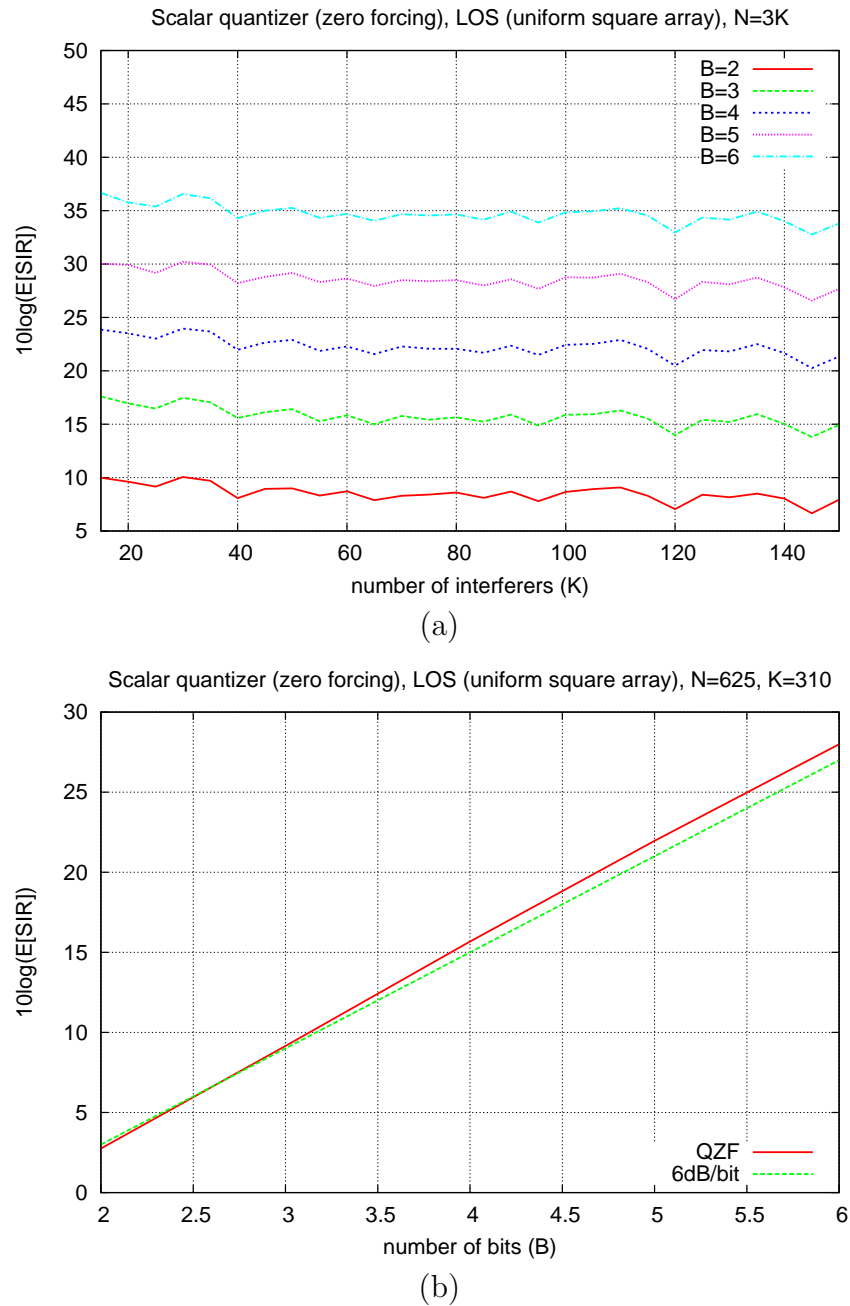


Figure 7.32: Simulated (averaged) SIR under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. (a) SIR (in dB scale) as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) SIR (in dB scale) as a function of the quantization level B for $N=500$ and $K=250$.

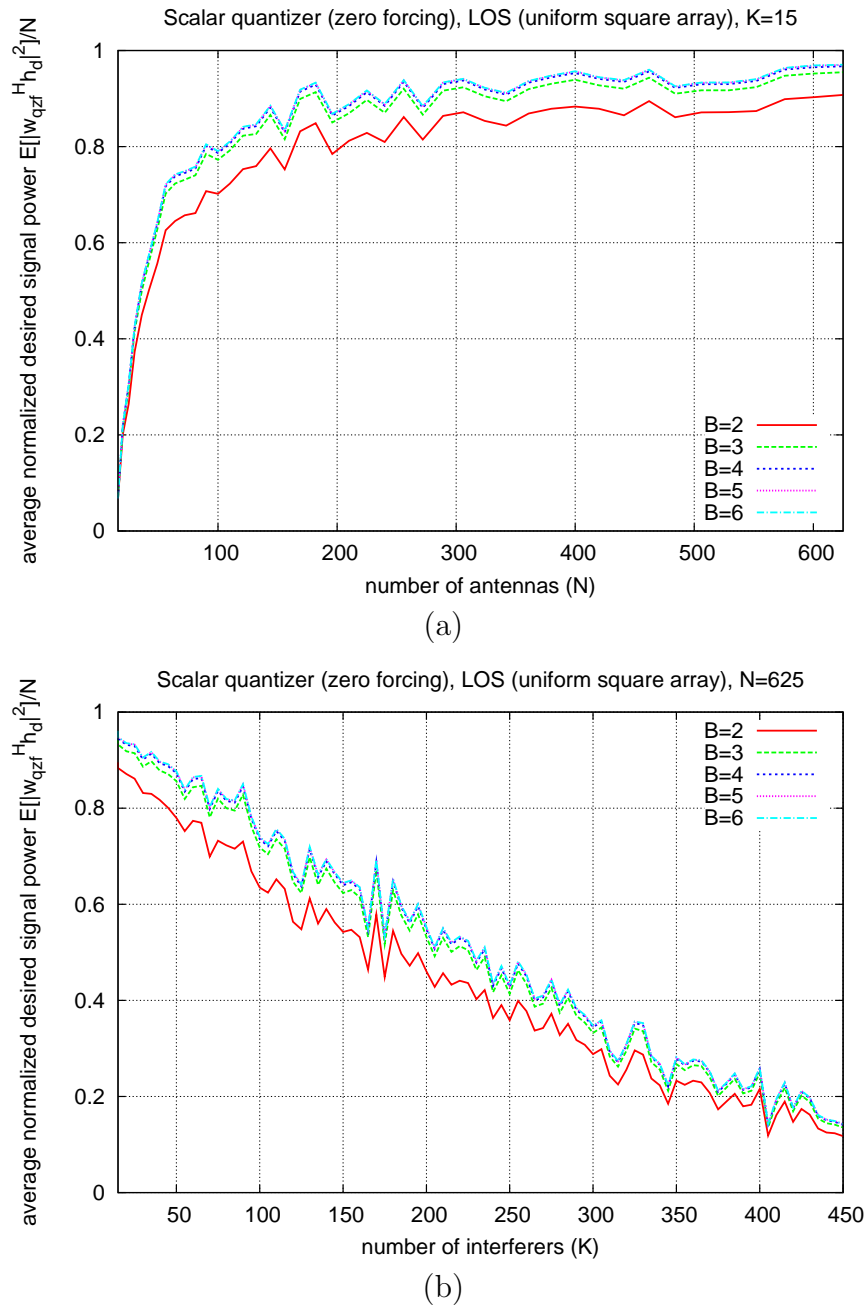


Figure 7.33: Simulated (averaged) desired signal under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. (a) Desired signal (normalized by N) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Desired signal (normalized by N) as a function of the number of interferers K for different quantization levels B ($N = 150$).

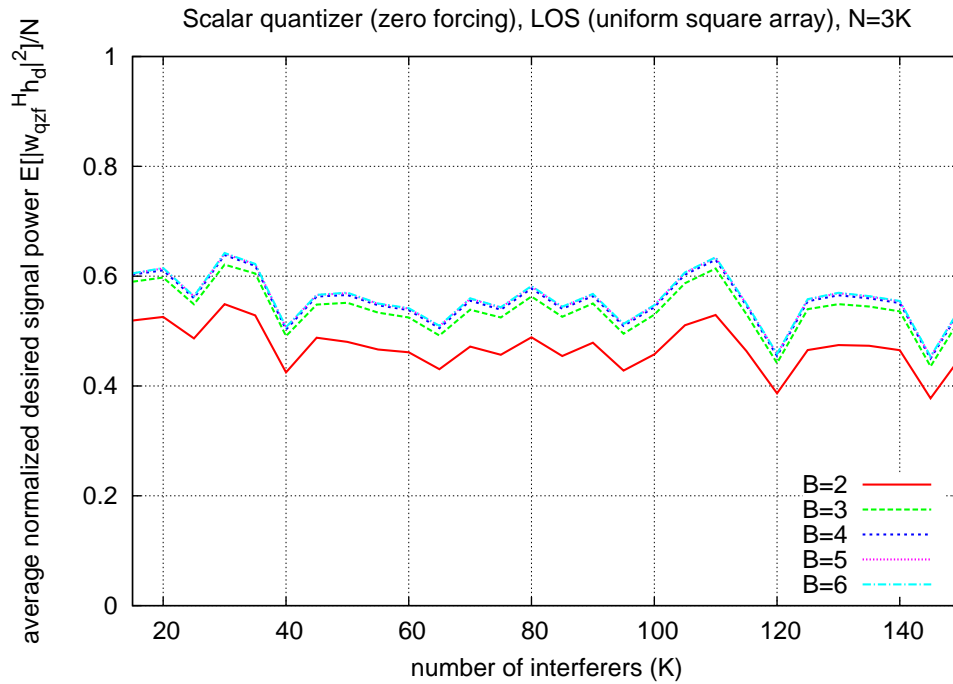


Figure 7.34: Simulated (averaged) desired signal under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. The desired signal (normalized by N) is simulated as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$.

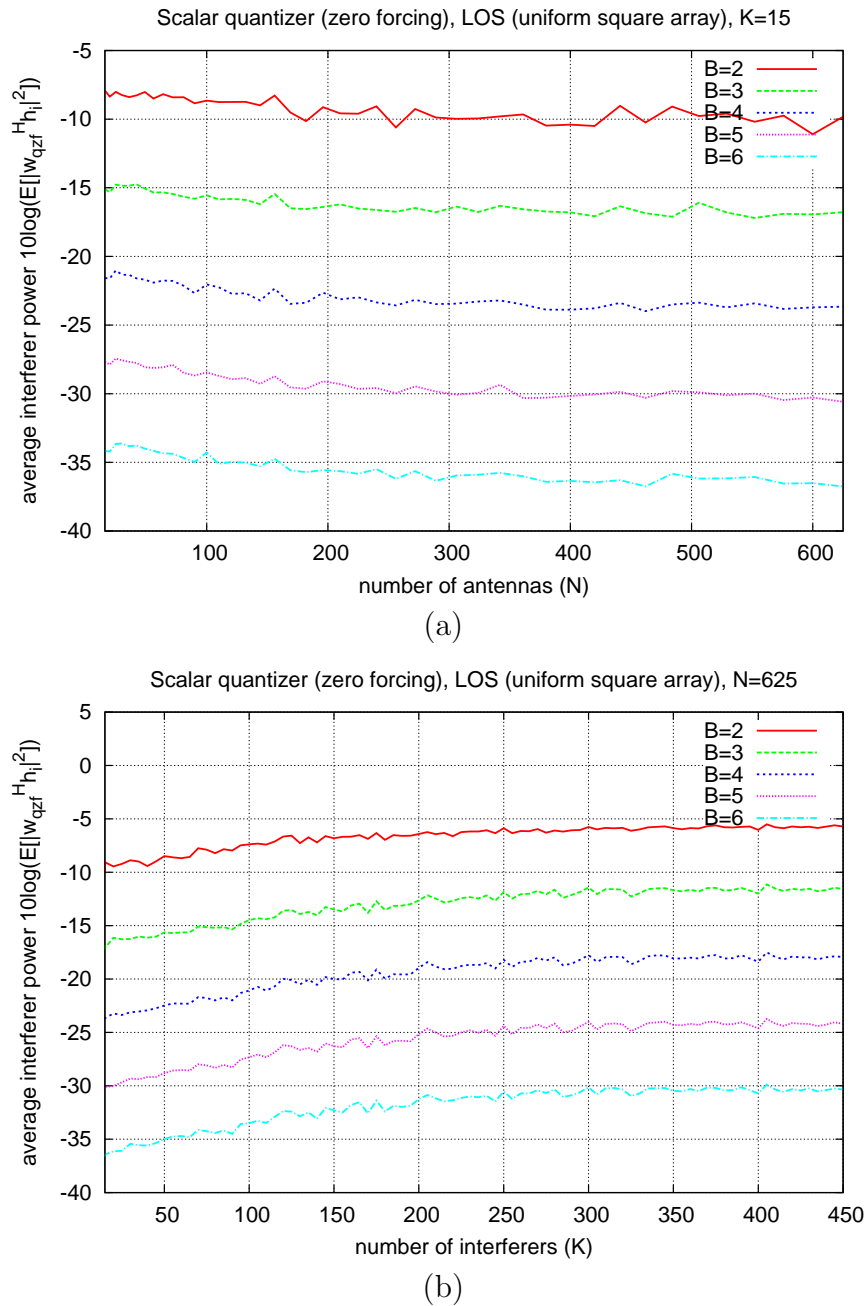


Figure 7.35: Simulated (averaged) power of a single interferer under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. (a) Interferer power (in dB scale) as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B ($N = 150$).

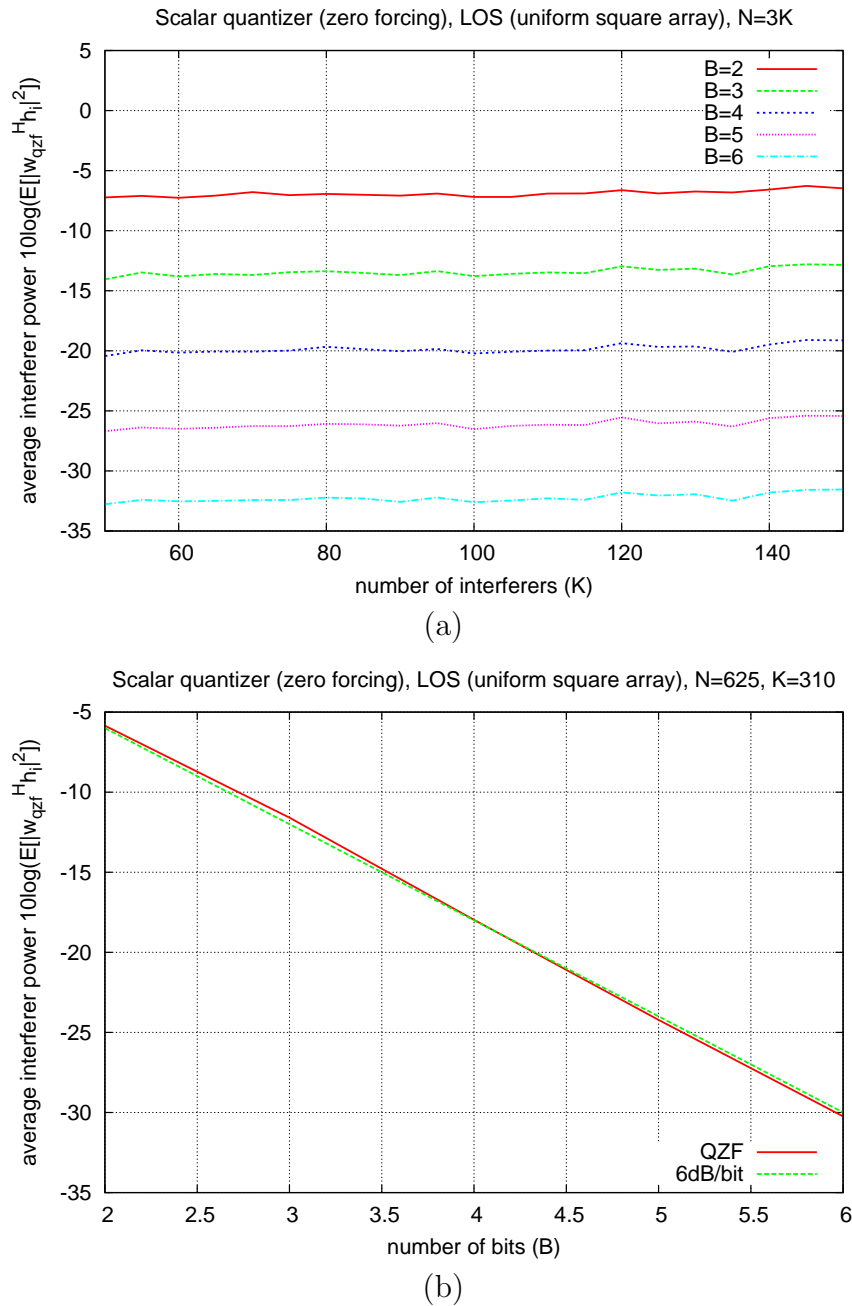


Figure 7.36: Simulated (averaged) power of a single interferer under the quantized zero-forcing beamformer (QZF) in an LOS channel. In this model, we used a uniform square array with $\lambda/2$ spacing. Both the desired signal and the interferers are placed randomly on the surface of a large sphere (in the far field of the array) centered around the origin at each iteration. (a) Interferer power (in dB scale) as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Interferer power (in dB scale) as a function of the quantization level B for $N=500$ and $K=250$.

Gaussian random variables. Since the square magnitude of a Gaussian random variable is exponentially distributed [59]⁷, the minimum interference power can be upper bounded as

$$P_{min} = \min_{\mathbf{u} \in \mathcal{U}} |T(\mathbf{u})|^2 \leq \min_{\mathbf{u} \in \mathcal{U}_o} |T(\mathbf{u})|^2 \quad (7.3)$$

It is easy to show that taking the minimum of N *i.i.d.* exponential random variables reduces the mean by a factor of N . Thus, for our simplified model, the interference power scales down at least as fast as $1/N$.

The original problem $T(\Delta \mathbf{w}) \equiv \Delta \mathbf{w}^H \mathbf{h}_1$ is very similar. For each coefficient of $\Delta \mathbf{w}$, if we restrict ourselves to two options: either round both I and Q coefficients up or round them both down, then there will be a total of 2^N choices for $\Delta \mathbf{w}$. There is a one-to-one correspondence between these choices and the binary vectors $\mathcal{U} = \frac{1}{\sqrt{N}}\{-1, +1\}^N$. The main difference is that the Walsh-Hadamard vectors no longer yield independent Gaussian random variables when we account for the distribution of the coefficients of $\Delta \mathbf{w}$. However, since \mathcal{U} is an N -dimensional vector space, it is always possible to find a set of basis vectors \mathbf{v}_i which yield independent random variables $T(\mathbf{v}_i)$ for which (7.3) holds. Therefore, we expect the minimum interference power to decrease by at least a factor of $1/N$. This means that the SIR should increase at least quadratically with N .

7.3.2 Constructive algorithms for improving interference suppression

While a naive approach based on scalar quantization does not achieve the optimum interference suppression, the space of all possible quantized vectors grows exponentially with NB , and thus a vector quantizer that exhaustively searches through the set of reconstruction levels is impractical. Instead, in this section, we focus on a set of simple and computationally efficient algorithms based on greedy coordinate descent optimization. This approach substantially improves over the naive method, yet has computational cost that is linear in the number of antennas. We will present a detailed analysis of three variants of this approach: greedy coordinate descent (GCD), iterative greedy coordinate descent (*i*GCD), and least-significant-bit greedy coordinate descent (LSB-GCD).

Greedy coordinate descent (GCD)

The greedy coordinate descent algorithm (GCD) quantizes the weight vector by searching through the set of discrete quantization levels for each element individually, instead of jointly as in the exhaustive search algorithm, and chooses the value that maximizes the SIR. This process is repeated for each coefficient until algorithm converges. More precisely, the algorithm, proceeds as follows. We first start by choosing a vector from the set of valid reconstruction levels. Next, we search through all 2^{2B} valid reconstruction levels of the first coefficient⁸, while keeping the other N coefficients fixed. The value which maximizes the output SIR (SINR) is chosen as the quantized value of the first coefficient. We then proceed

⁷The mean of the exponential distribution is equal to the variance of the Gaussian random variable.

⁸Recall that each coefficient is represented by $2B$ bits: B bits for the real part and B bits for the imaginary part.

to quantize the second coefficient, keeping the first coefficient fixed at its quantized value and coefficients 3 through N fixed at their initial values, by a similar search. This method is applied to each of the N coefficients of the weight vector \mathbf{w} . This process is repeated for several rounds until the weights stabilize (converge). Our simulation results indicate that the number of rounds required for convergence is constant and independent of N , K , and B .

We ran several simulations to evaluate the performance of GCD as a function of N , K , and B . The results are shown in Figures 7.37/7.38, 7.39/7.40, and 7.41/7.42. The plots in Figures 7.37/7.38 are for the average SIR under GCD. In Figure 7.37a, we fixed the number of interferers $K = 15$ and plot the average SIR for different values of B as we vary the number of antennas N from 20-245 on the x-axis (on the y-axis is the SIR normalized by N^2 in dB scale). The graph suggests that the increase in SIR versus N is quadratic (compared with a linear growth in the case of scalar quantization) since SIR/N^2 remains relatively flat as we increase N . In Figure 7.37b, we fix the number of antennas N while varying the number of interferers K on the x-axis (on the y-axis is the average SIR multiplied by K^2 in dB scale). Since $K^2 \text{SIR}$ remains relatively flat as we increase K , we can conclude that the average SIR decays as $1/K^2$. However, similar to the scalar quantized zero-forcing beamformer, the functional dependence of the SIR under GCD on N and K individually is a little more complicated and depends on the ratio N/K . When N/K is large, then the SIR is approximately quadratic in N and $1/K$. However, more generally, the SIR is quadratic in the ratio N/K . The functional dependence of the SIR on N and K through the ratio N/K is corroborated by Figure 7.38a, where vary the number of interferers K on the x-axis while keeping the ratio $N/K = 3$ fixed. The graph shows that the average SIR stays flat as we increase K as long as the ratio N/K remains constant. The dependence of GCD on the number of quantization bits is shown in Figures 7.37ab/7.38a (different curves correspond to different number of quantization bits) and further corroborated by Figure 7.38b (number of bits on the x-axis). Just as in scalar quantization, the average SIR grows exponentially in the bit rate, with a gain of approximately 6dB per bit. Therefore, the functional dependence of the SIR under GCD on N , K , and B is approximately:

$$\overline{\text{SIR}}(\mathbf{w}_{\text{gcd}}) = \frac{E[|\mathbf{w}_{\text{gcd}}^H \mathbf{h}_d|^2]}{\sum_{i=1}^K E[|\mathbf{w}_{\text{gcd}}^H \mathbf{h}_i|^2]} \sim \left(\frac{N}{K}\right)^2 2^{2B} \quad (7.4)$$

We also analyzed the desired signal and interference components separately to see how each contributes to the overall average SIR. The average desired signal power is shown in Figures 7.39/7.40. In Figure 7.39a, the x-axis is the number of antennas N (the number of interferers $K = 15$ is fixed), and the y-axis is the average desired signal power normalized by N (the different curves represent different quantization levels). In Figure 7.39b, the x-axis is the number of interferers K (the number of antennas N is fixed), and the y-axis is the average desired signal power (the different curves represent different quantization levels). Taken together, these two figures show an approximate linear growth of the desired signal power with N , little dependence on B , and a small decay versus K . The slight non-linear increase versus and the small decay versus K has to do with desired signal component lost in the interference subspace. However, when the ratio N/K is kept constant, the desired signal power grows linearly with N (Figure 7.40a). However, the desired signal power under GCD is slightly below the quantized matched filter and ZF beamformer (Figure 7.40b).

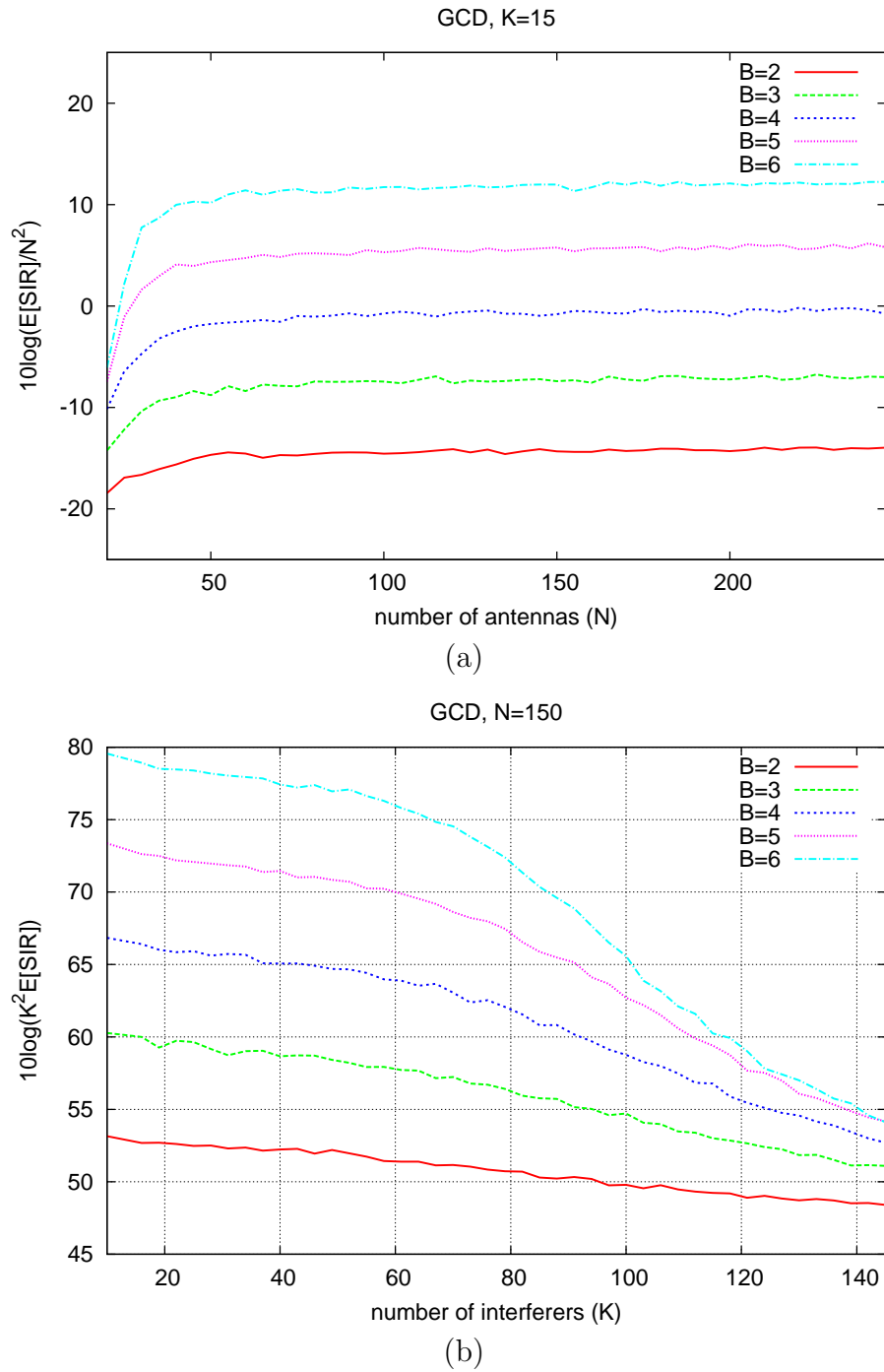
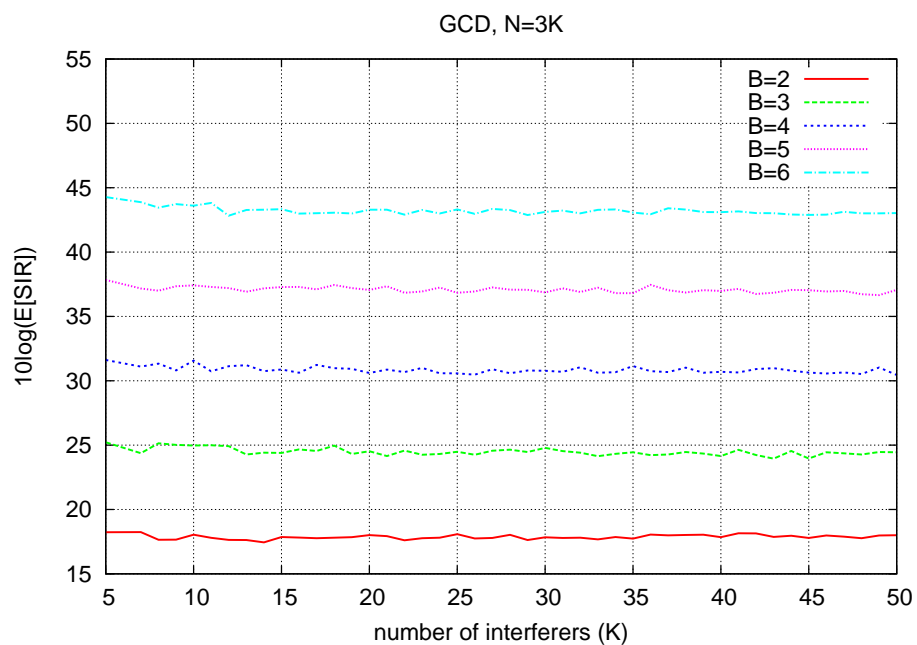
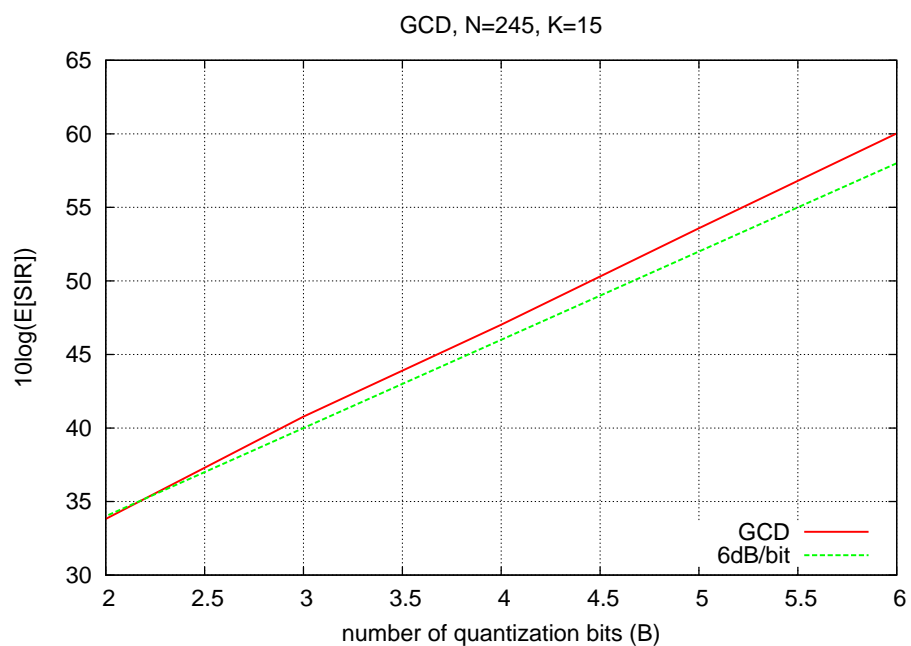


Figure 7.37: (a) Simulated (averaged) SIR (normalized by N^2 in dB scale) under GCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) SIR (multiplied by K^2 in dB scale) under GCD as a function of the number of interferers K for different quantization levels B ($N = 150$). The simulations in both Figures were under the Rayleigh channel model.



(a)



(b)

Figure 7.38: (a) Simulated (averaged) SIR (in dB scale) in under GCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Simulated (averaged) SIR (in dB scale) under GCD as a function of the number of quantization bits B ($N = 245, K = 10$). The simulations in both Figures were under the Rayleigh channel model.

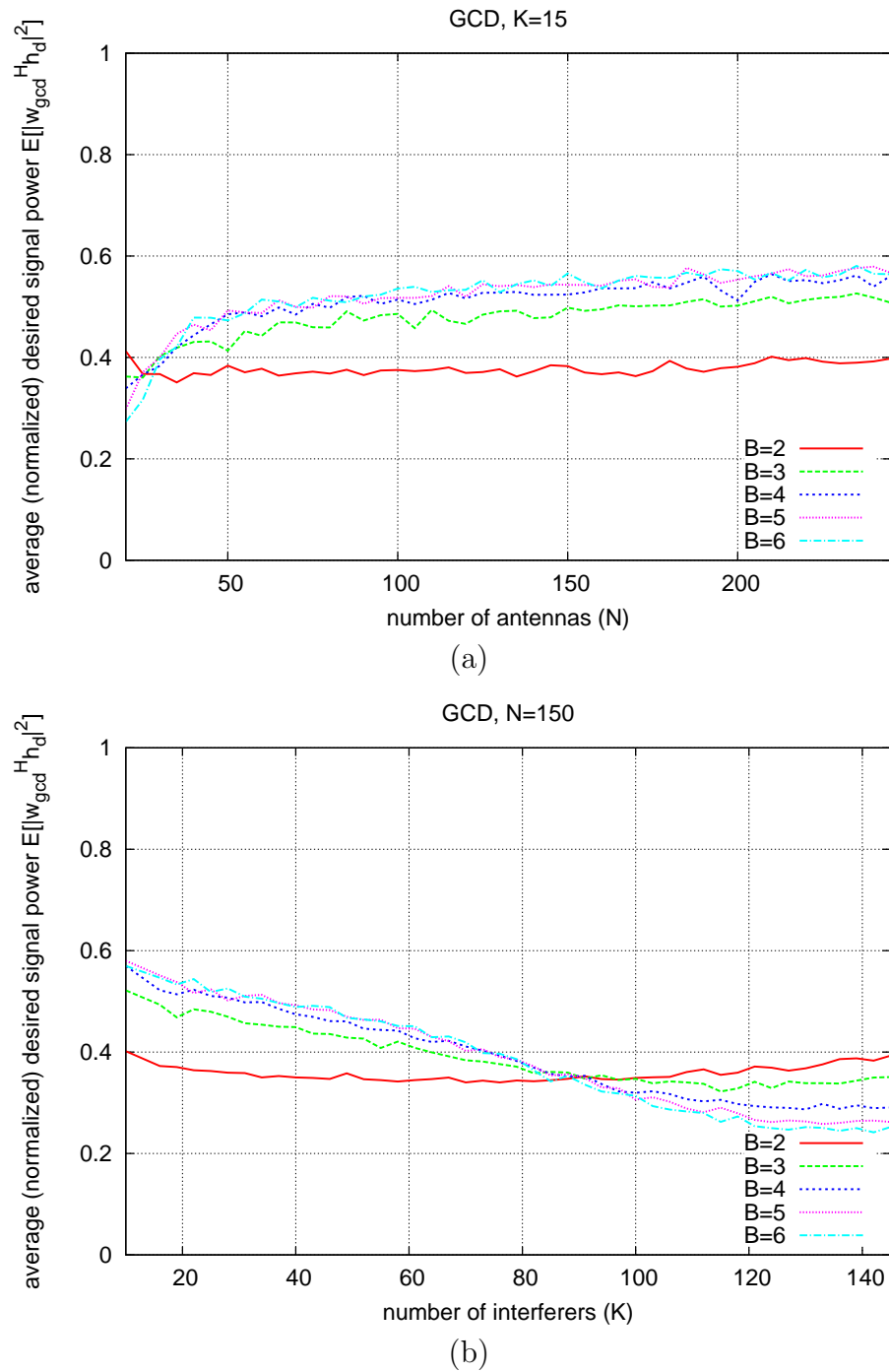


Figure 7.39: (a) Simulated (averaged) desired signal power (normalized by N) under GCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) desired signal power (normalized by N) under GCD as a function of the number of interferers K for different quantization levels B ($N = 150$). The simulations in both Figures were under the Rayleigh channel model.

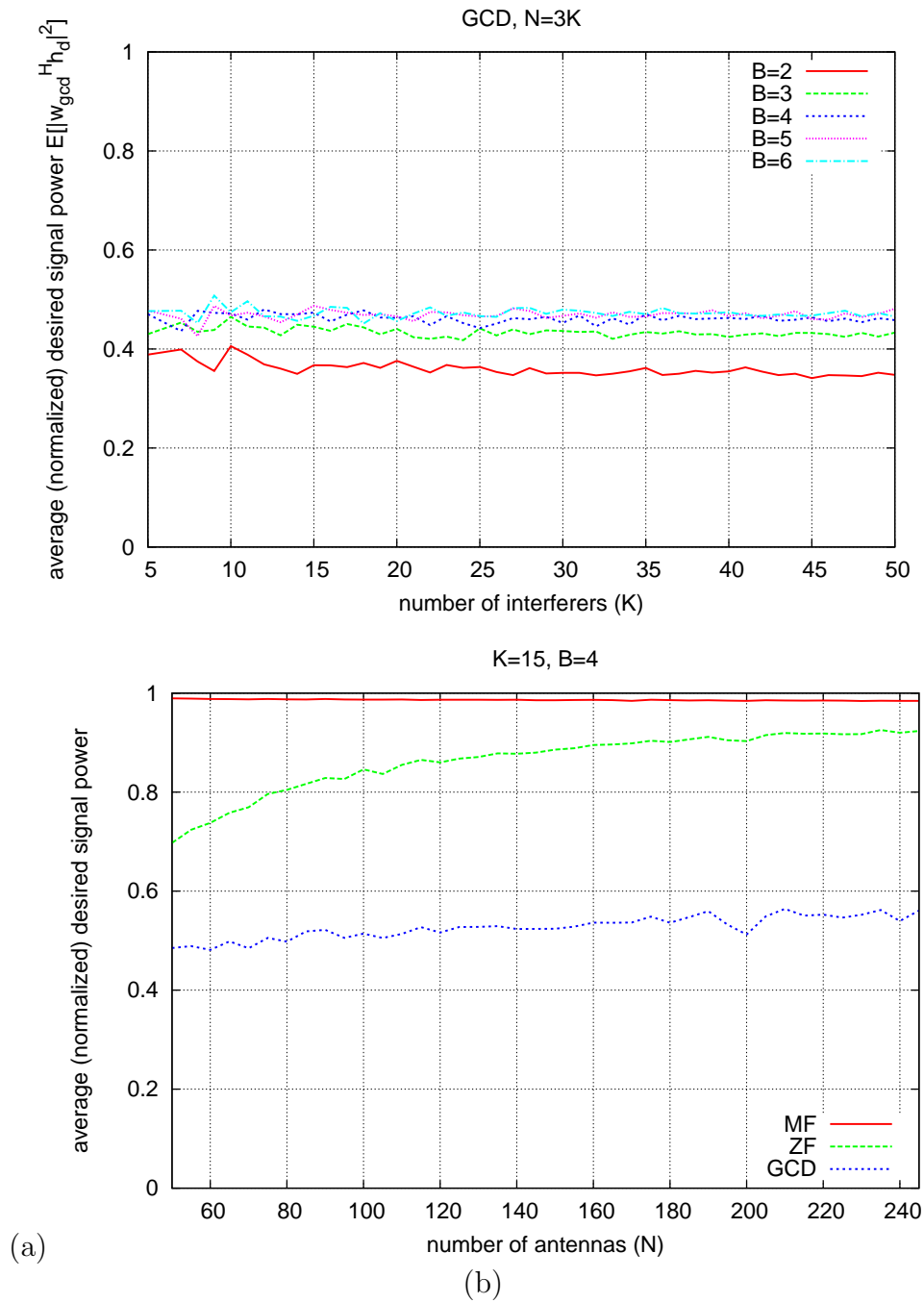


Figure 7.40: (a) Simulated (averaged) desired signal power (normalized by N) under GCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Comparing the desired signal power under GCD to the desired signal power under the scalar quantization schemes described in Section 7.2 (quantized spatial matched filtering and quantized zero-forcing) as a function of the number of antennas (N) ($K = 15$, $B = 4$). The simulations in both Figures were under the Rayleigh channel model.

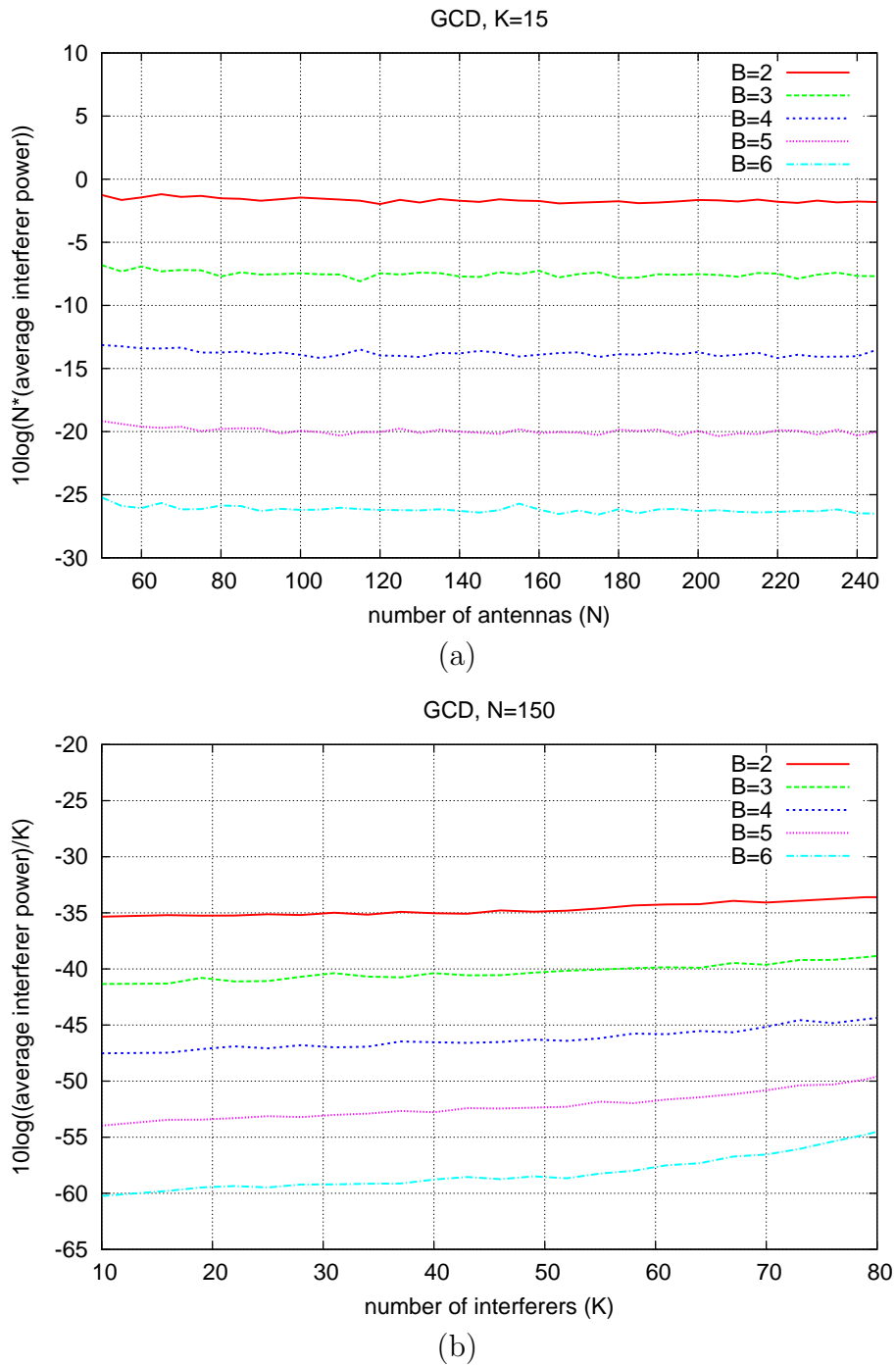
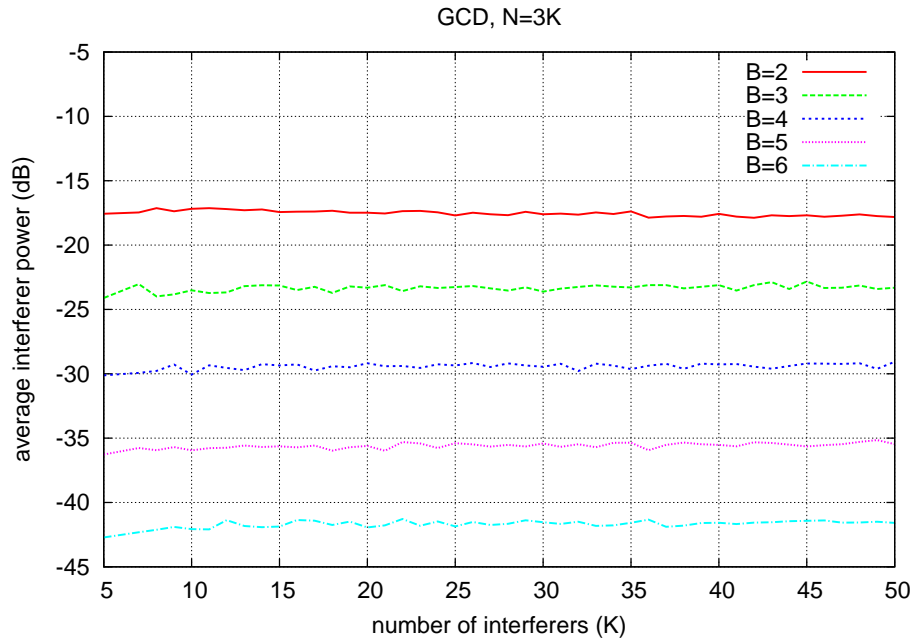
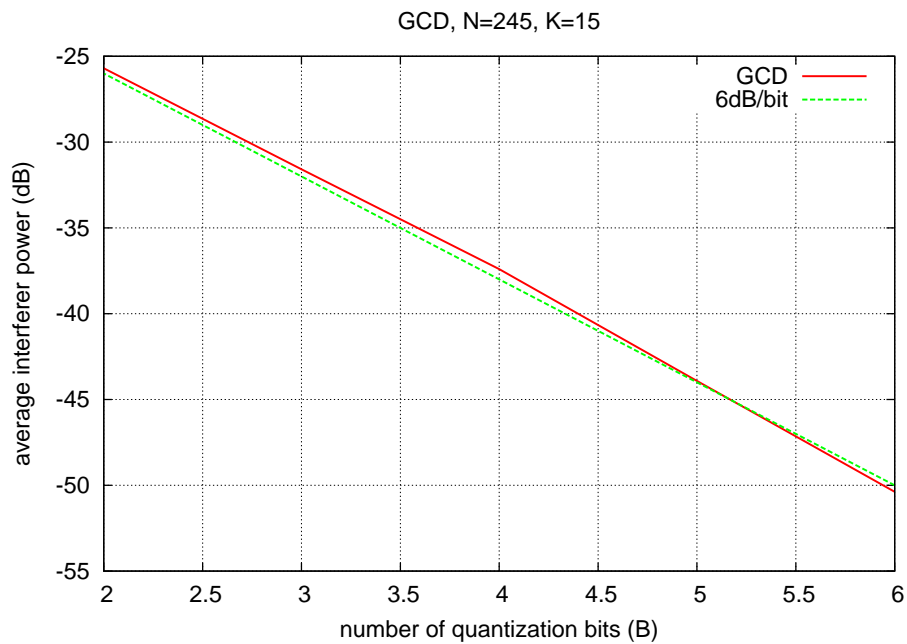


Figure 7.41: (a) Simulated (averaged) power of a single interferer (multiplied by N in dB scale) under GCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) power of a single interferer (divided by K in dB scale) under GCD as a function of the number of interferers K for different quantization levels B ($N = 150$). The simulations in both Figures were under the Rayleigh channel model.



(a)



(b)

Figure 7.42: (a) Simulated (averaged) power of a single interferer (in dB scale) in under GCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Simulated (averaged) power of a single interferer (in dB scale) under GCD as a function of the number of quantization bits B ($N = 245, K = 10$). The simulations in both Figures were under the Rayleigh channel model.

However, the difference is a constant factor (compared with the factor of N in interference reduction). Therefore, a factor of N in Equation 7.4 comes from the desired signal and the rest ($2^{2B}N/K^2$) comes from the interference rejection. This is validated by the results in Figures 7.41/7.42, which plots the average interferer power under GCD versus N , K , and B . Figure 7.41a shows that the average interferer power scales as $1/N$ when K is fixed, while Figure 7.41b shows that the average interferer power increases linearly with K when N is fixed. Therefore, the average interferer power scales with the ratio K/N (Figure 7.42a). Furthermore, the gap between the different curves corresponding to the different quantization levels is approximately 6dB indicating an exponential dependence on the number of quantization bits B . Therefore, the total interference power scales as $\frac{K^2}{N}2^{-2B}$ (recall that we have a total K interferers).

Iterative greedy coordinate descent (*i*GCD)

The GCD algorithm improves the overall SIR by an extra factor of N/K , and achieves this with a computational complexity that grows linearly with N (as opposed to exponentially in the case of exhaustive search). However, GCD still searches through all valid reconstruction levels for each coefficient, and thus the running time grows exponentially with $2B$. In this Section, we present a variant of the conventional greedy coordinate descent algorithm called the iterative greedy coordinate descent algorithm or *i*GCD that eliminates the exponential dependence on B , while retaining the extra factor of N/K in SIR improvement. *i*GCD computes the bits of each antenna weight separately, instead of searching over all 2^{2B} possible weights as in GCD. We start by running GCD on the most significant bit of each weight. After converging on a single bit solution, the next iteration runs a single bit GCD for the next significant bit. In this iteration, the most significant is fixed at the optimal solution found in the previous iteration, and the search is only performed over the values of the second bit. This process is repeated B times (for each bit, while it is being modified, all previous more significant bits are fixed). Therefore, the overall running time will be linear in B instead of exponential as in GCD.

Figures 7.43/7.44, 7.45/7.46, and 7.47/7.48 show the same results as Figures 7.37/7.38, 7.39/7.40 and 7.41/7.42 for *i*GCD. These plots show that the behavior of *i*GCD is almost identical to that of GCD. In Figure 7.55a, we plot both GCD and *i*GCD on the same graph. This shows that we incur no penalty by optimizing the quantized beamforming vector one bit at a time instead of searching all reconstruction levels.

Least significant bit-greedy coordinate descent (LSB-GCD)

A straight forward algorithm that can be derived directly from the proof in Section 7.3.1 is least significant bit-greedy coordinate descent or LSB-GCD. LSB-GCD, as the name suggests, is a variant of the general coordinate descent algorithm that optimizes only the least significant bit or bits. In other words, we start with the quantized least squares (ZF) solution, then we optimize the least significant bit using GCD. Figures 7.49/7.50, 7.51/7.52, and 7.53/7.54 show the simulation results for LSB-GCD (similar to Figures 7.37/7.38, 7.39/7.40, and 7.41/7.42). The plots show that LSB-GCD exhibits similar behavior to that of GCD and *i*GCD. A direct comparison between the three algorithms in Figure 7.55a shows that

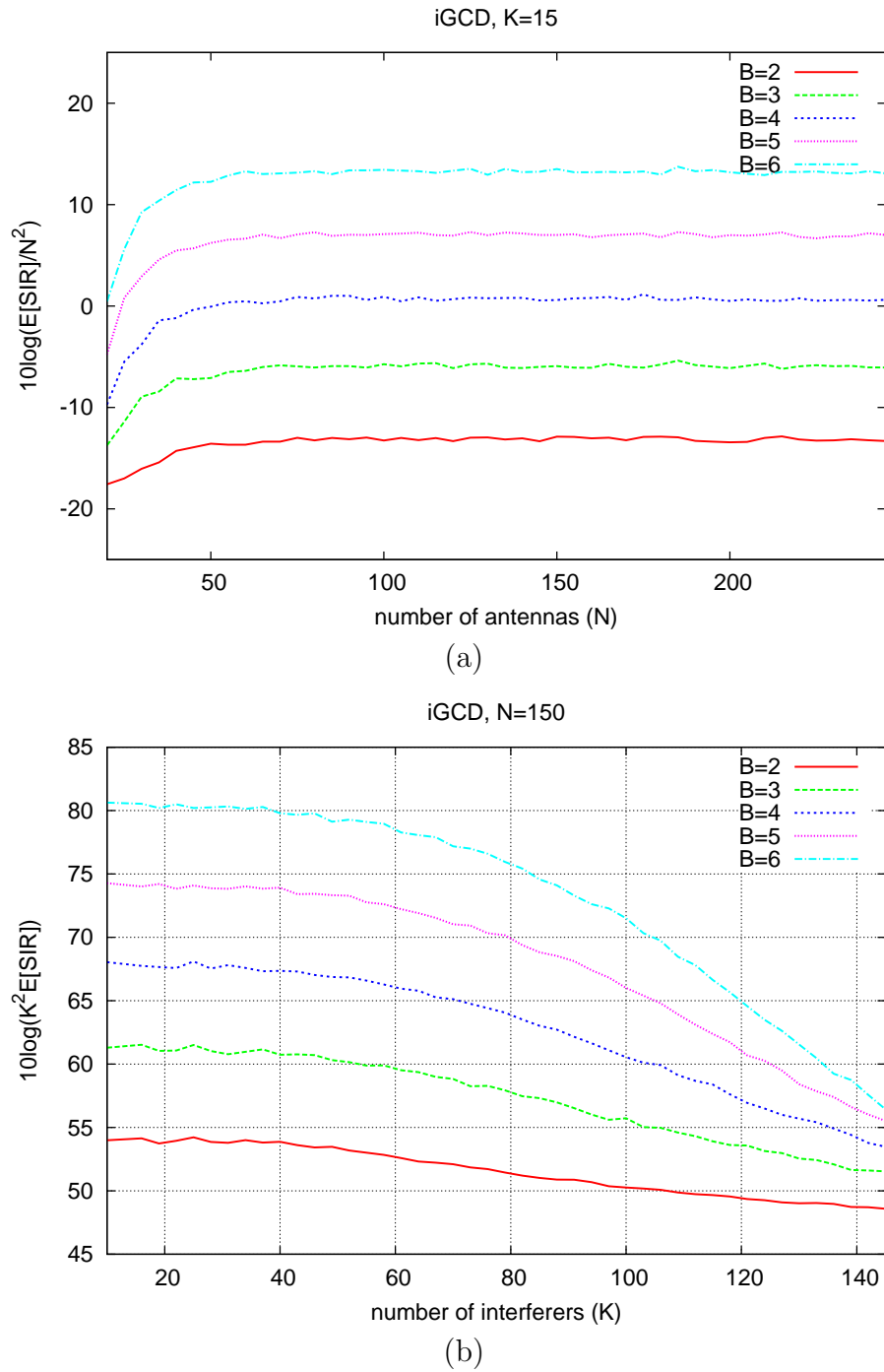
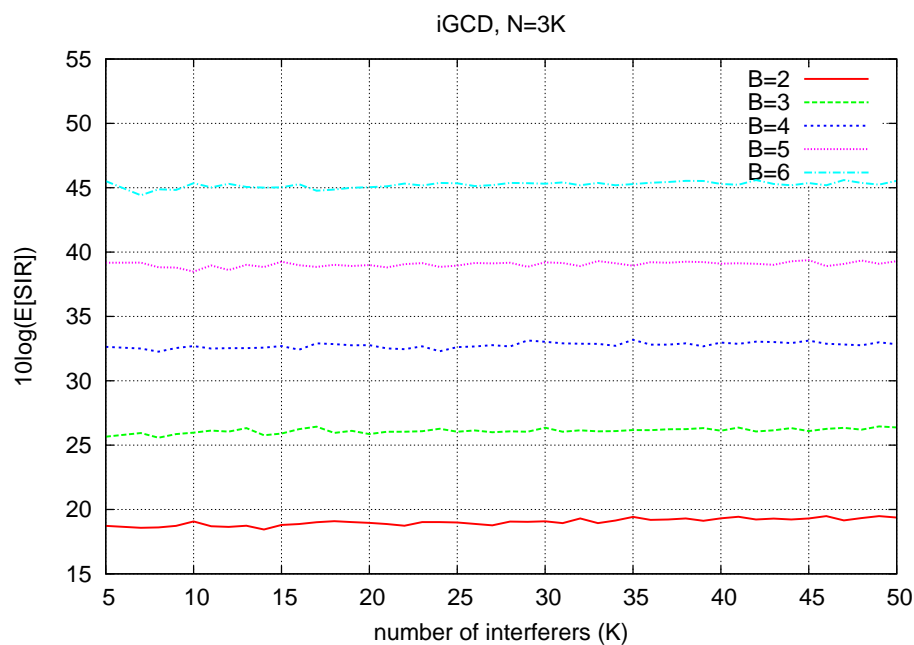
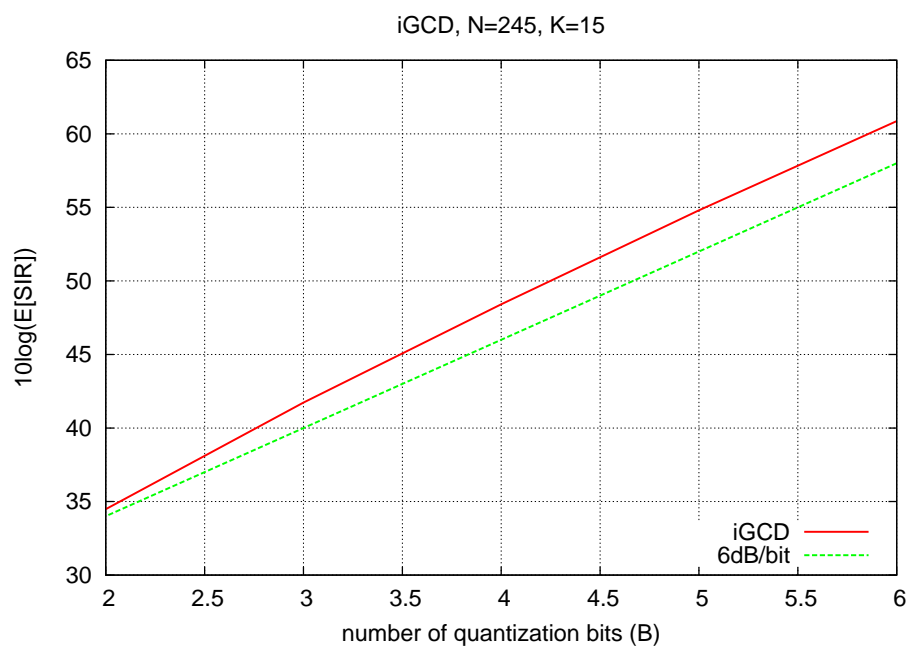


Figure 7.43: (a) Simulated (averaged) SIR (normalized by N^2 in dB scale) under iGCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) SIR (multiplied by K^2 in dB scale) under iGCD as a function of the number of interferers K for different quantization levels B ($N = 150$). The simulations in both Figures were under the Rayleigh channel model.



(a)



(b)

Figure 7.44: (a) Simulated (averaged) SIR (in dB scale) in under iGCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Simulated (averaged) SIR (in dB scale) under iGCD as a function of the number of quantization bits B ($N = 245, K = 10$). The simulations in both Figures were under the Rayleigh channel model.

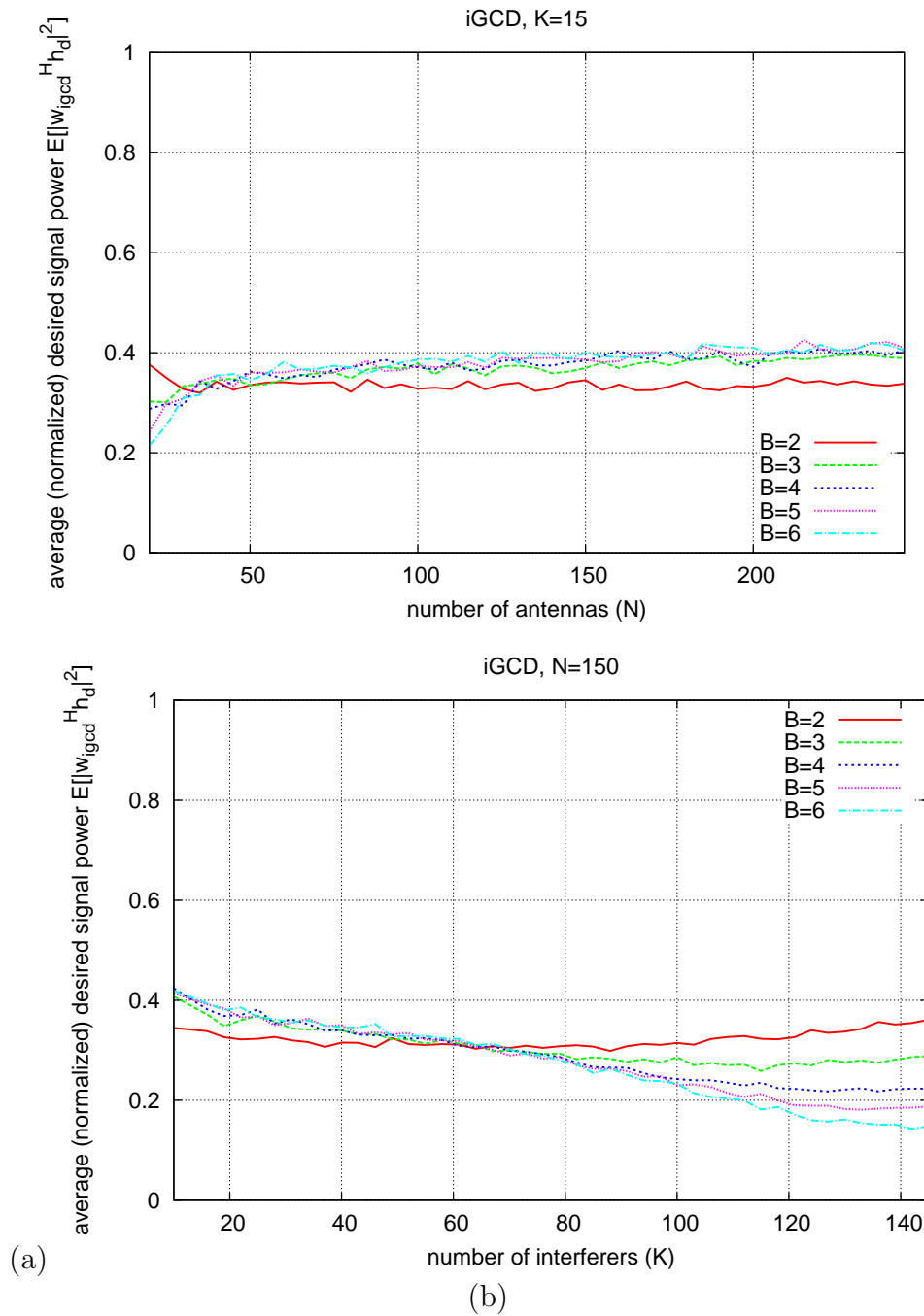


Figure 7.45: (a) Simulated (averaged) desired signal power (normalized by N) under iGCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) desired signal power (normalized by N) under iGCD as a function of the number of interferers K for different quantization levels B ($N = 150$). The simulations in both Figures were under the Rayleigh channel model.

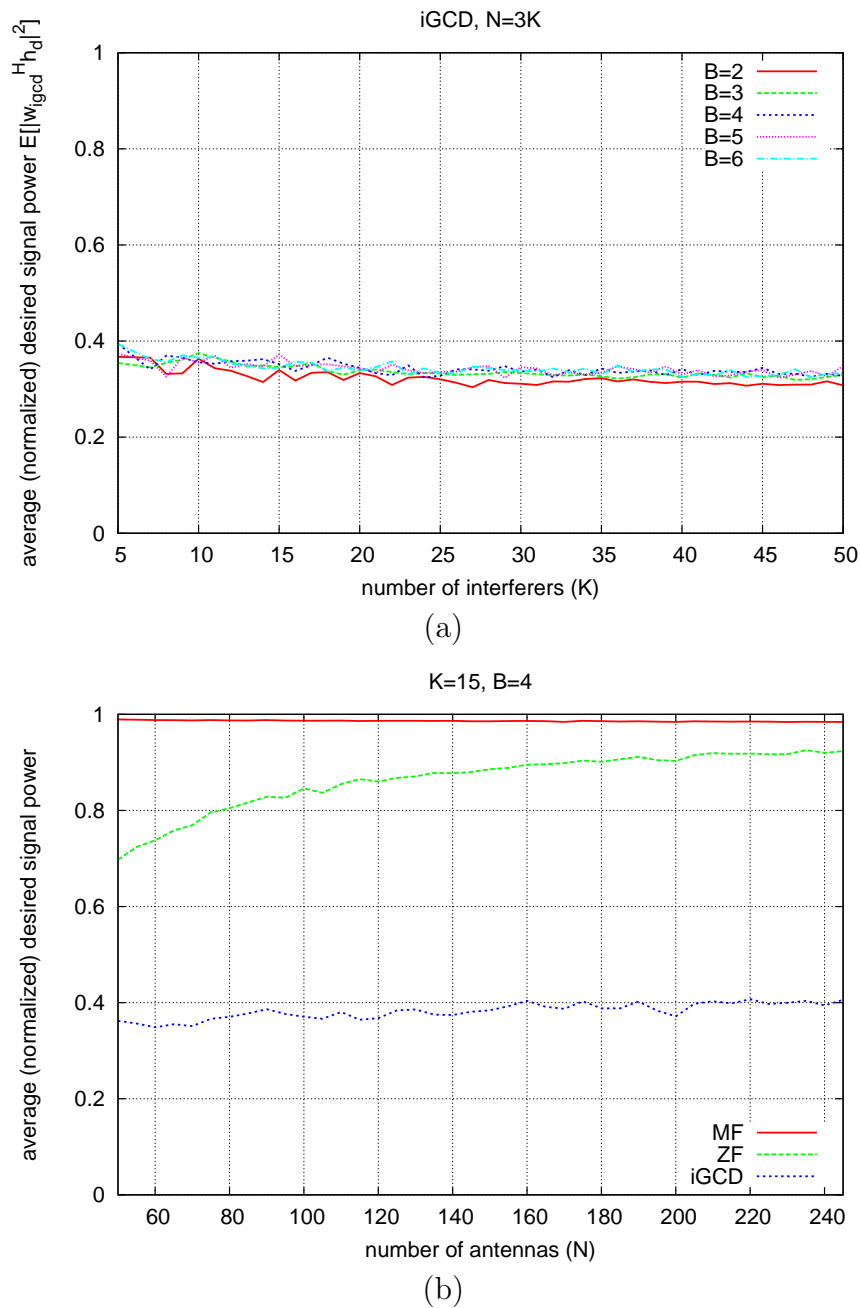


Figure 7.46: (a) Simulated (averaged) desired signal power (normalized by N) under iGCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Comparing the desired signal power under iGCD to the desired signal power under the scalar quantization schemes described in Section 7.2 (quantized spatial matched filtering and quantized zero-forcing) as a function of the number of antennas (N) ($K = 15$, $B = 4$). The simulations in both Figures were under the Rayleigh channel model.

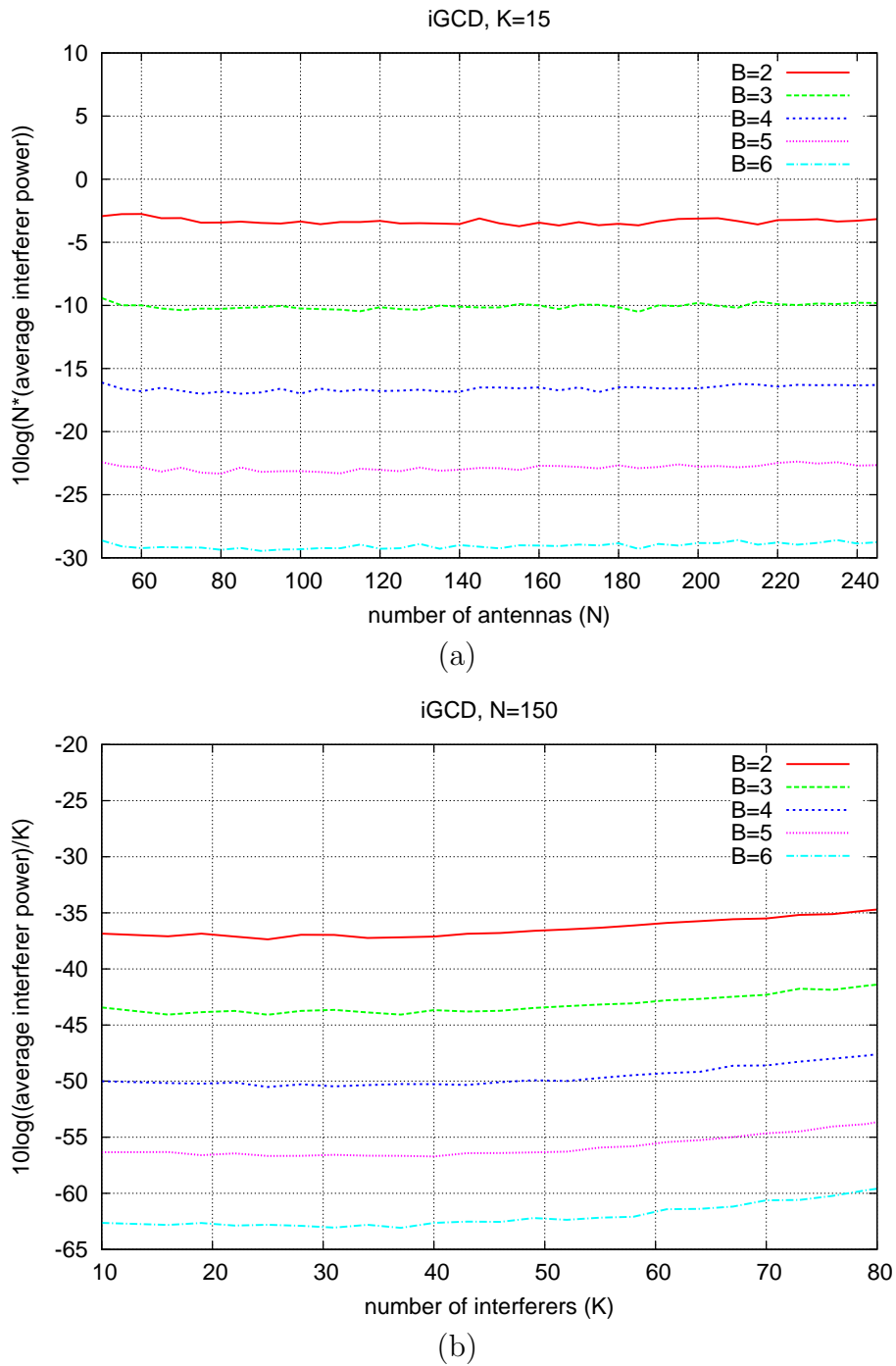


Figure 7.47: (a) Simulated (averaged) power of a single interferer (multiplied by N in dB scale) under iGCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) power of a single interferer (divided by K in dB scale) under iGCD as a function of the number of interferers K for different quantization levels B ($N = 150$). The simulations in both Figures were under the Rayleigh channel model.

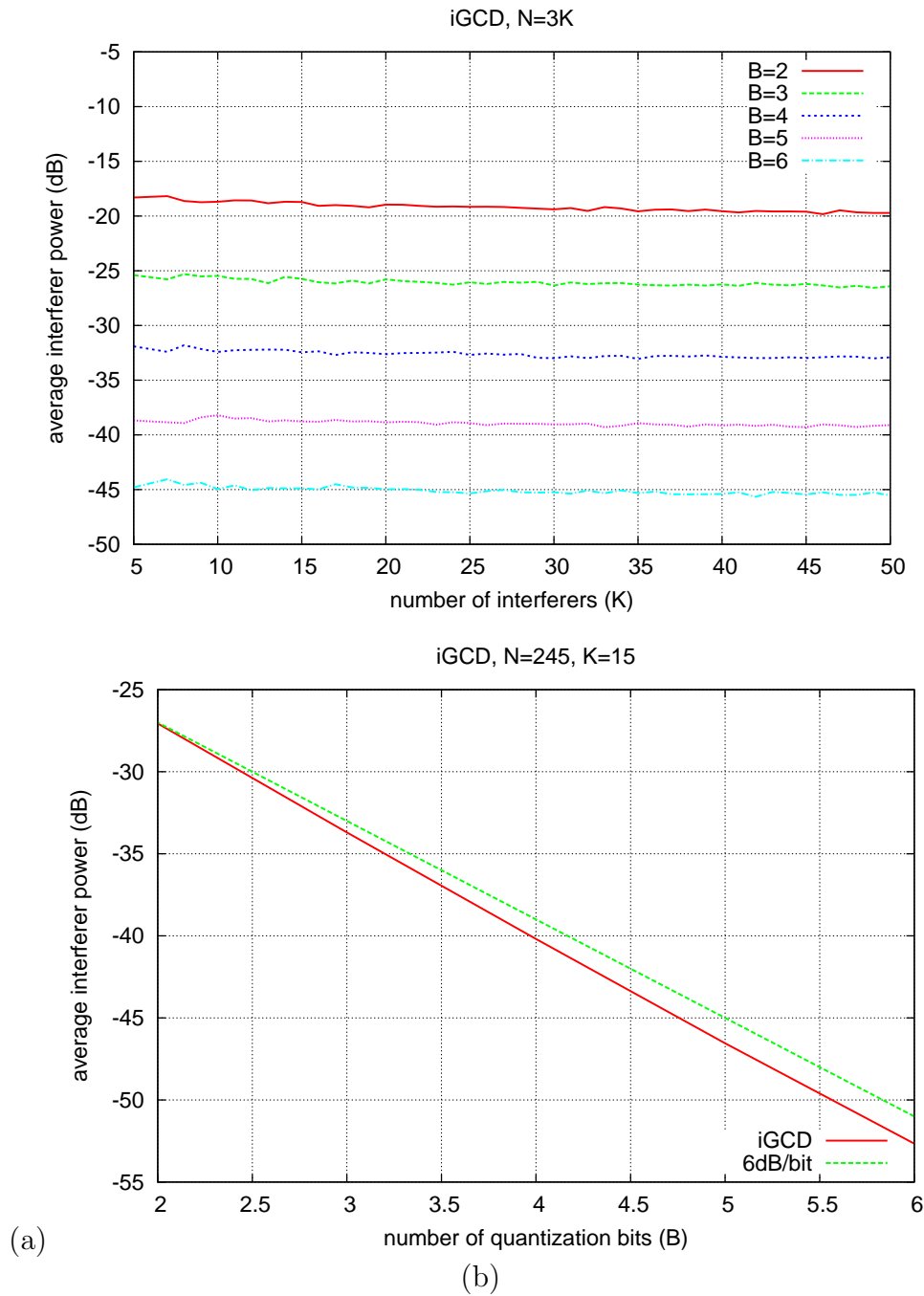


Figure 7.48: (a) Simulated (averaged) power of a single interferer (in dB scale) in under iGCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Simulated (averaged) power of a single interferer (in dB scale) under iGCD as a function of the number of quantization bits B ($N = 245$, $K = 10$). The simulations in both Figures were under the Rayleigh channel model.

their performance is almost identical. While the desired signal power under LSB-GCD is slightly higher than both GCD and i GCD (Figure 7.55b), the interference power is also slightly higher (Figure 7.55c).

Impact of channel model

The proofs and simulation results presented in this section assumed a Rayleigh fading model (i.e. the complex channel weights are *i.i.d* complex Gaussian random variables) for its simplicity. However, these results extend to other channel models as well. In Figure 7.56, we compare the performance of i GCD under two extreme channel conditions: Rayleigh fading and line of sight (LOS). For the LOS channel, we used both uniform linear and square arrays with $\lambda/2$ spacing. For square arrays, we randomized the direction of arrival, whereas for linear arrays, we only randomized the azimuth angle of arrival. In the figure, we plot the SIR (normalized by N^2) on the y-axis versus the number of antennas N on the x-axis. The results show that the SIR maintains a quadratic growth as a function of N for both Rayleigh and LOS channels.

Optimality of GCD based algorithms

Our simulations show that GCD-based algorithms converge after $O(N)$ steps. However, a guaranteed convergence does not necessarily mean that it will converge to the optimum result. In order to verify this, let us consider the impact of the initial state of the algorithm on the final output. If the algorithm is optimal, then it should converge to the same result regardless of how the algorithm is initialized. In Sections 7.3.2 and 7.3.2, we did not tackle the question of initializing the GCD and i GCD algorithms, and assumed that both are initialized to the quantized matched filter. Here, we consider three different initializations: quantized matched filter, quantized ZF beamformer, and a random vector. Figure 7.57 shows the average SIR under i GCD for the three different initializations. To get a fair and accurate comparison, we used the same values of $\mathbf{h}_d, \mathbf{h}_1, \dots, \mathbf{h}_K$ for each initialization. The plots show that even though the channel responses are the same, the average SIR changes when we change the initial state of the algorithm. These results suggest that these algorithms converge to local instead of global optima. However, this still does not tell us how far from the optimum is the GCD solution. To answer this question we simulated the average SIR using a full exhaustive search of all possible quantized weight vectors. The results are shown in Figures 7.58ab. while it is difficult to draw conclusions from these graphs since we can only simulate a full exhaustive search for small values of N and B , Figures 7.58ab suggest an exponential growth in average SIR as a function of N (K is fixed in Figure 7.58a) and an exponential decay as a function of K (N is fixed in Figure 7.58b). In [7], we derived an upper bound on the achievable SIR and showed that it is indeed exponential, and the results in Figure 7.58a would suggest that this bound is tight. Therefore, while GCD-based algorithms provide substantial improvements over scalar quantization, they are still far from optimum.

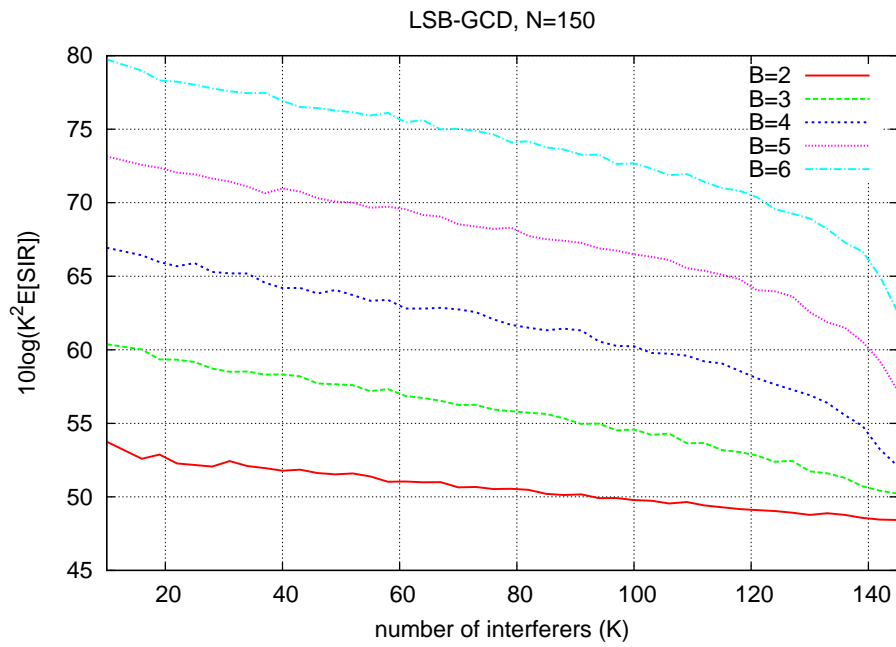
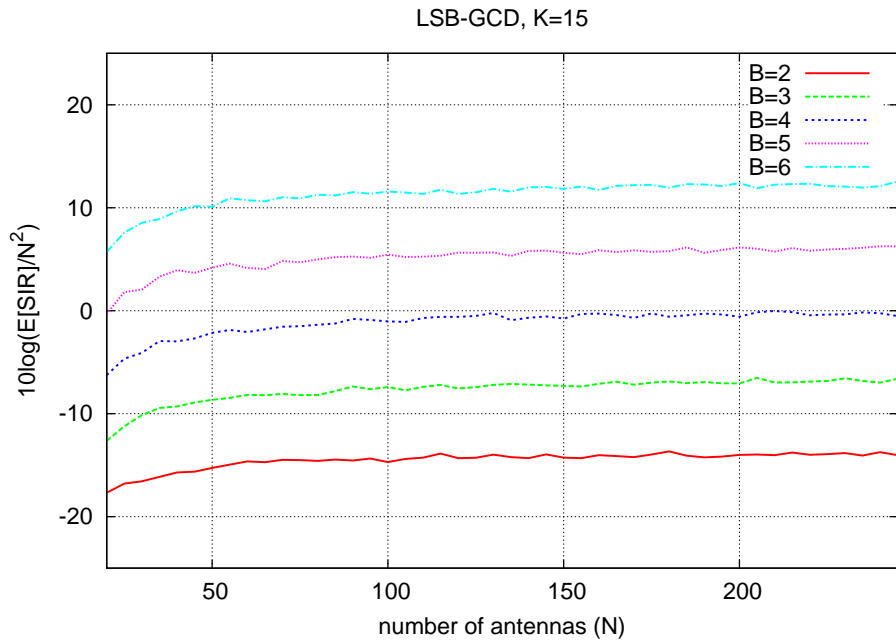


Figure 7.49: (a) Simulated (averaged) SIR (normalized by N^2 in dB scale) under LSB-GCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) SIR (multiplied by K^2 in dB scale) under LSB-GCD as a function of the number of interferers K for different quantization levels B ($N = 150$). The simulations in both Figures were under the Rayleigh channel model.

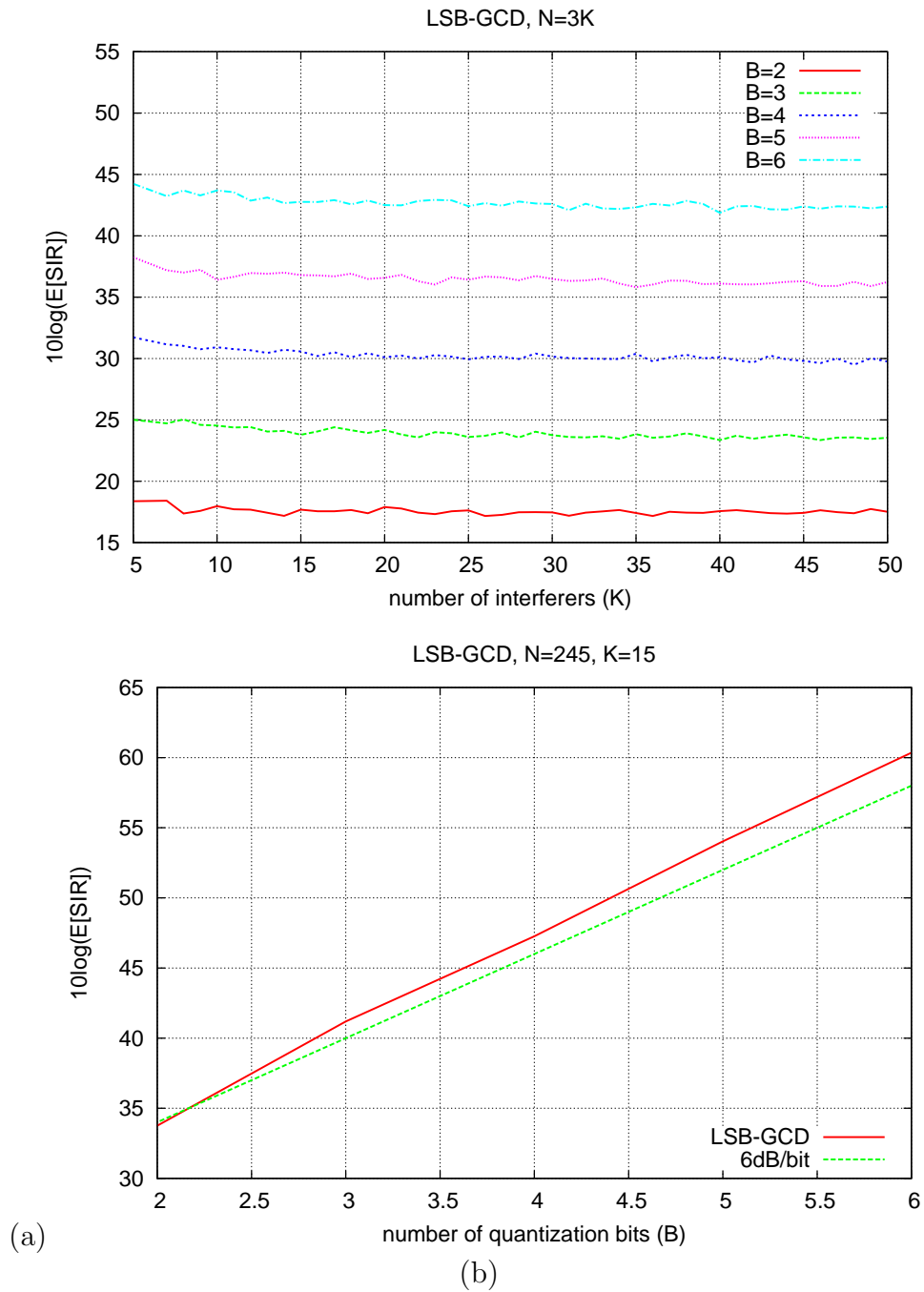


Figure 7.50: (a) Simulated (averaged) SIR (in dB scale) in under LSB-GCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Simulated (averaged) SIR (in dB scale) under LSB-GCD as a function of the number of quantization bits B ($N = 245$, $K = 10$). The simulations in both Figures were under the Rayleigh channel model.

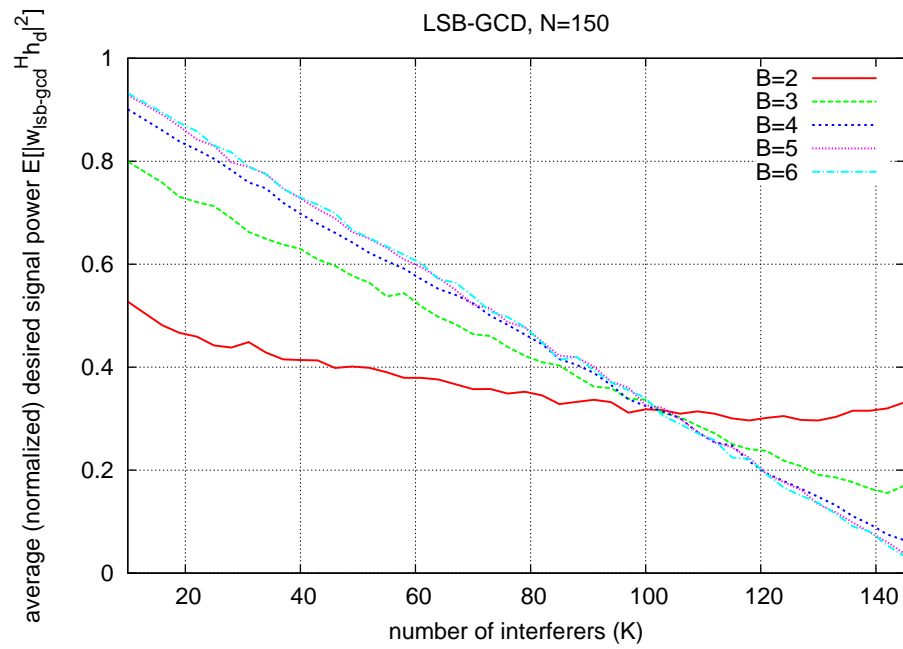
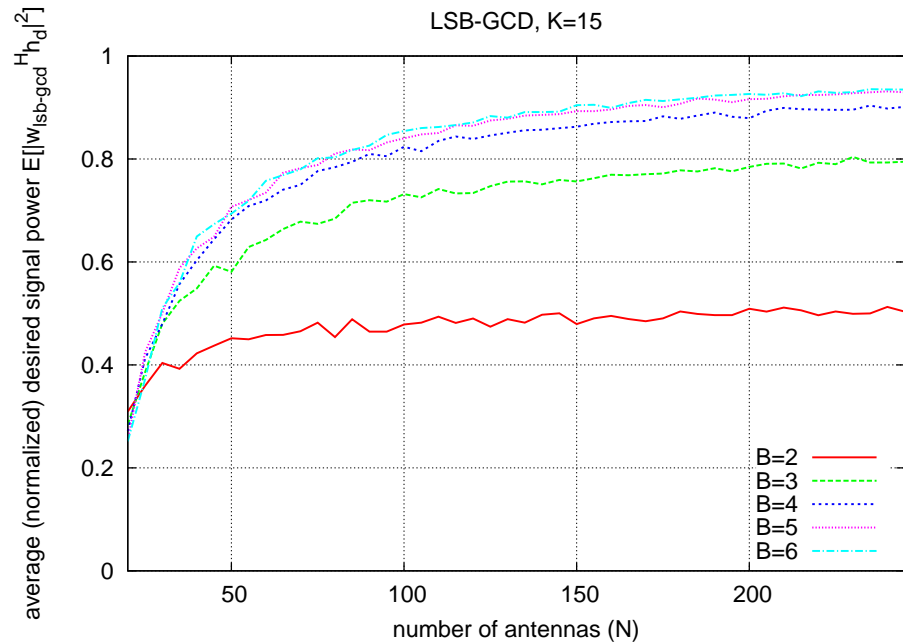


Figure 7.51: (a) Simulated (averaged) desired signal power (normalized by N) under LSB-GCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) desired signal power (normalized by N) under LSB-GCD as a function of the number of interferers K for different quantization levels B ($N = 150$). Both simulations were under the Rayleigh channel model.

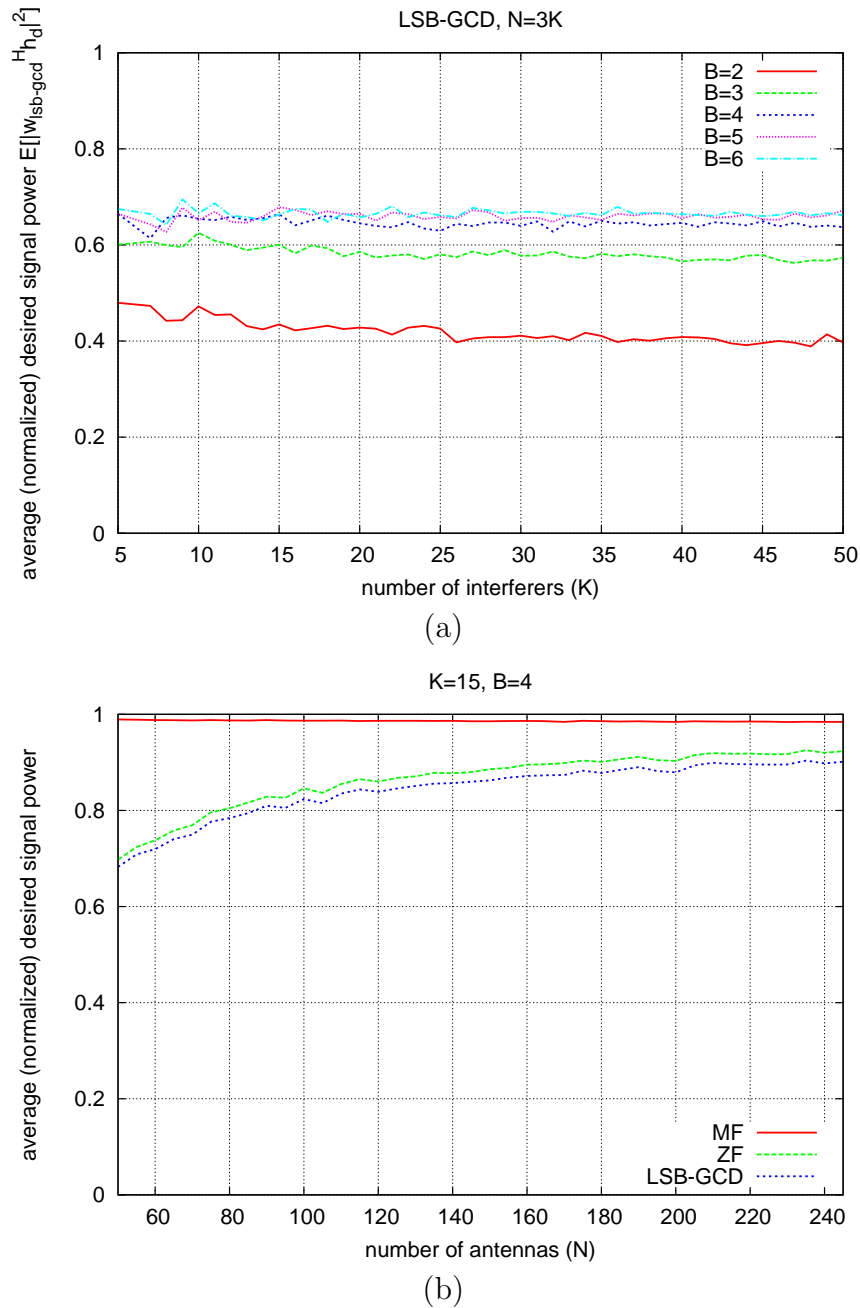
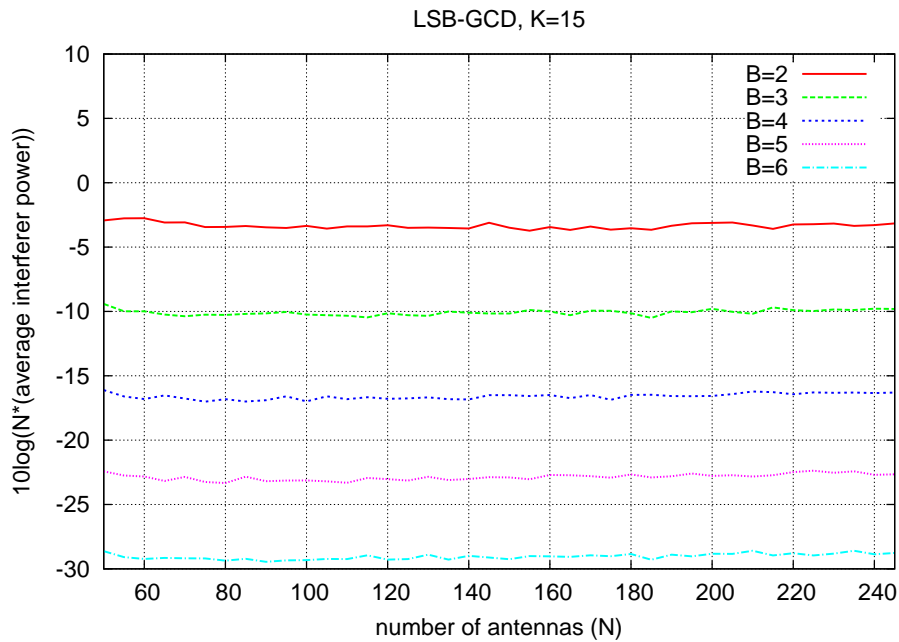
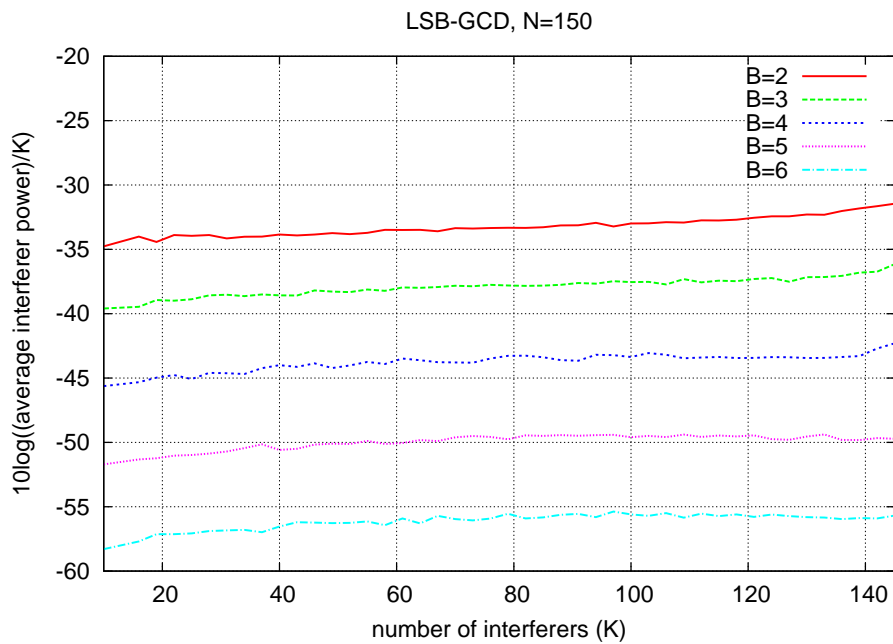


Figure 7.52: (a) Simulated (averaged) desired signal power (normalized by N) under LSB-GCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Comparing the desired signal power under LSB-GCD to the desired signal power under the scalar quantization schemes described in Section 7.2 (quantized spatial matched filtering and quantized zero-forcing) as a function of the number of antennas (N) ($K = 15$, $B = 4$). Both simulations were under the Rayleigh channel model.

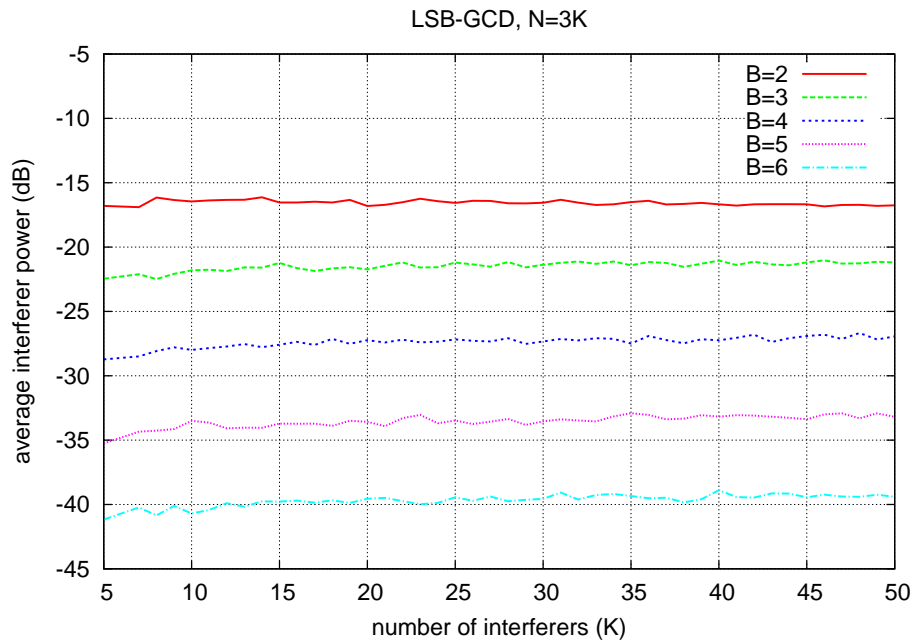


(a)

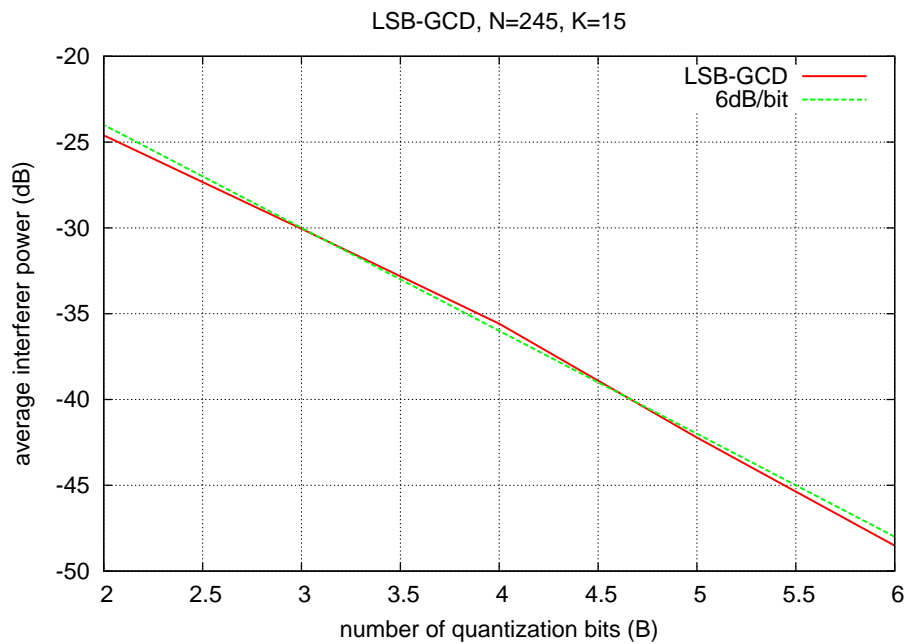


(b)

Figure 7.53: (a) Simulated (averaged) power of a single interferer (multiplied by N in dB scale) under LSB-GCD as a function of the number of antennas N for different quantization levels B ($K = 15$). (b) Simulated (averaged) power of a single interferer (divided by K in dB scale) under LSB-GCD as a function of the number of interferers K for different quantization levels B ($N = 150$). Both simulations were under the Rayleigh channel model.



(a)



(b)

Figure 7.54: (a) Simulated (averaged) power of a single interferer (in dB scale) in under LSB-GCD as a function of the number of interferers K for different quantization levels B when the ratio $N/K = 3$. (b) Simulated (averaged) power of a single interferer (in dB scale) under LSB-GCD as a function of the number of quantization bits B ($N = 245, K = 10$). Both simulations were under the Rayleigh channel model.

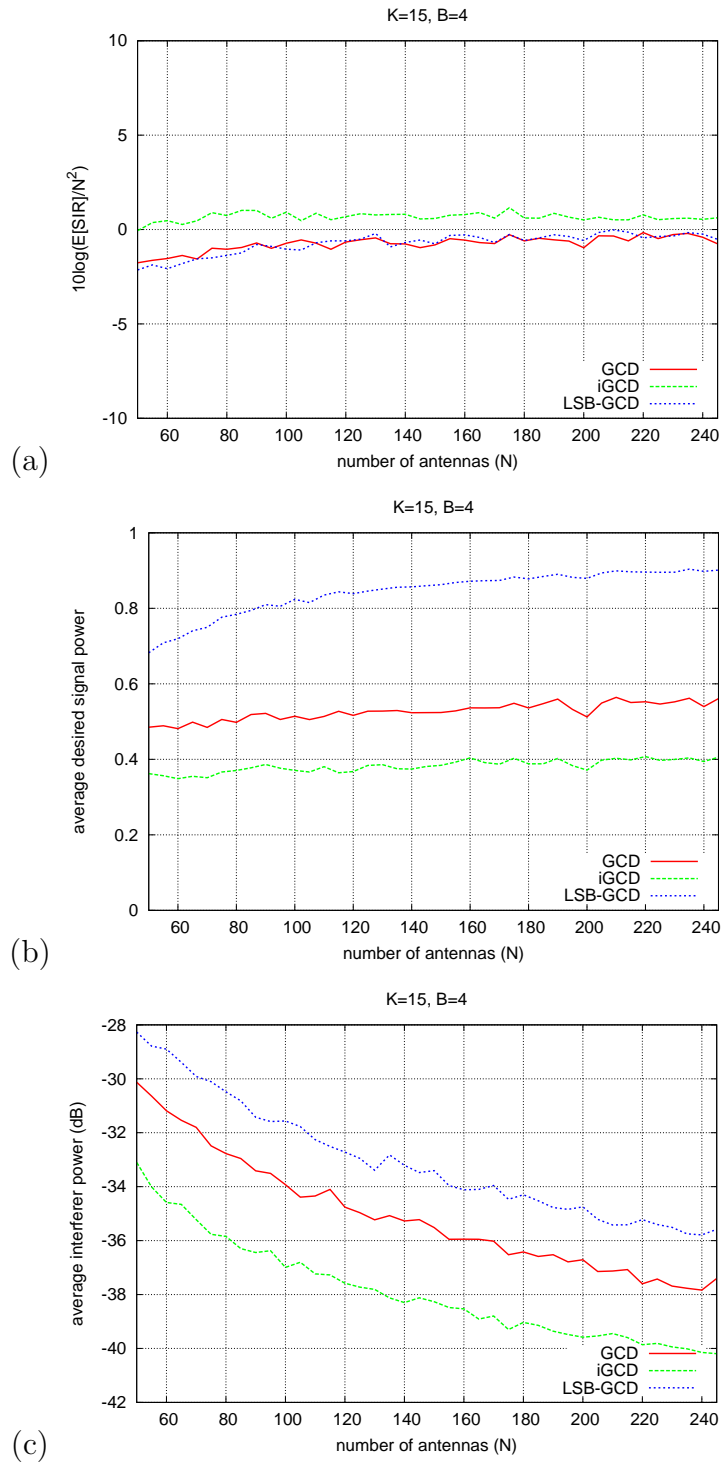


Figure 7.55: Comparing the performance of three vector quantization techniques (GCD, i GCD, LSB-GCD) versus the number of antennas (N) under a Rayleigh channel model ($K = 15$ and $B = 4$). (a) Average SIR normalized by N^2 in dB scale. (b) Desired signal power (normalized by N). (c) Average interferer power in dB scale.

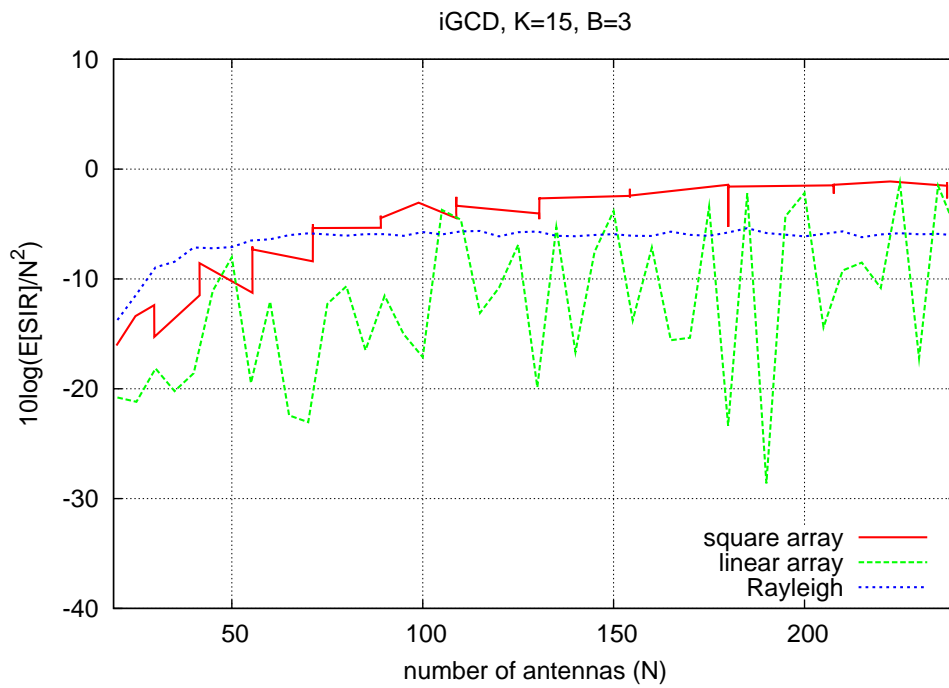


Figure 7.56: Performance of *iGCD* under different channel models (Rayleigh/LOS). On the x-axis, the number of antennas (N). On the y-axis, the average SIR (normalized by N^2) in dB scale. The number of interferers and bits are fixed at $K = 15$ and $B = 3$ respectively.

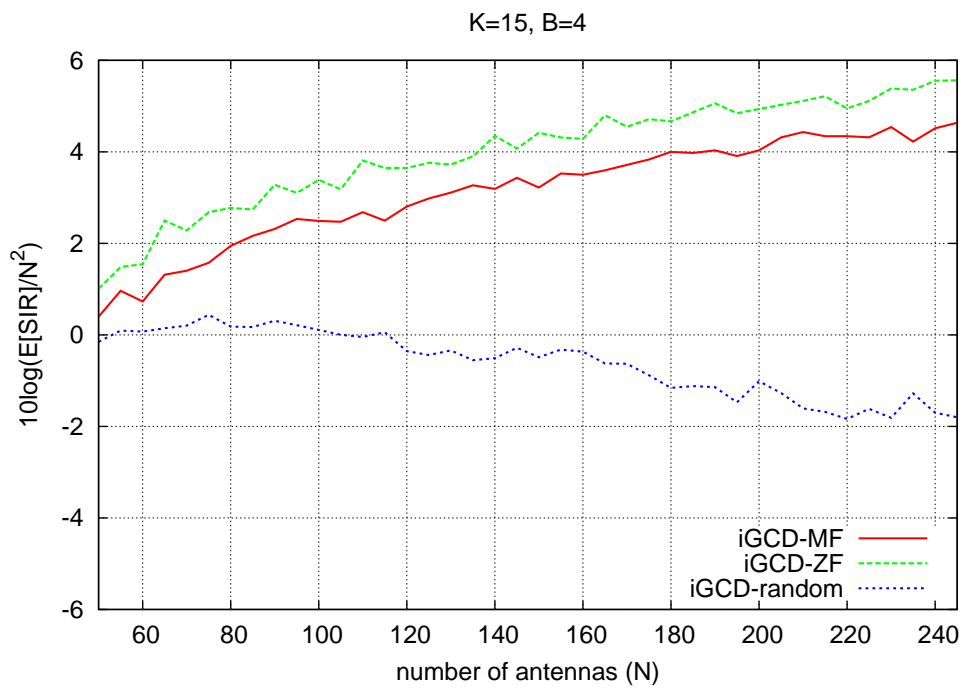
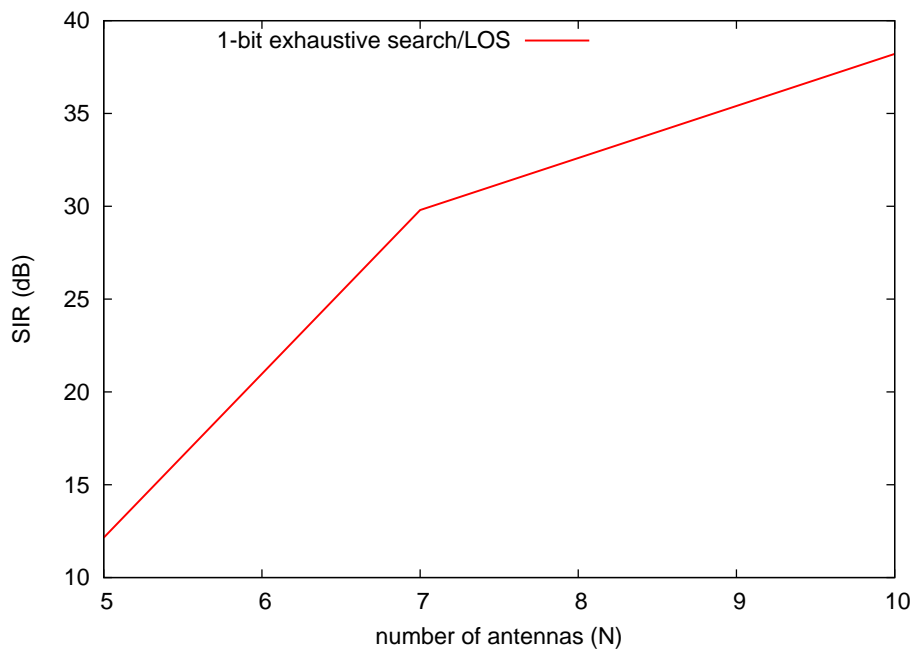
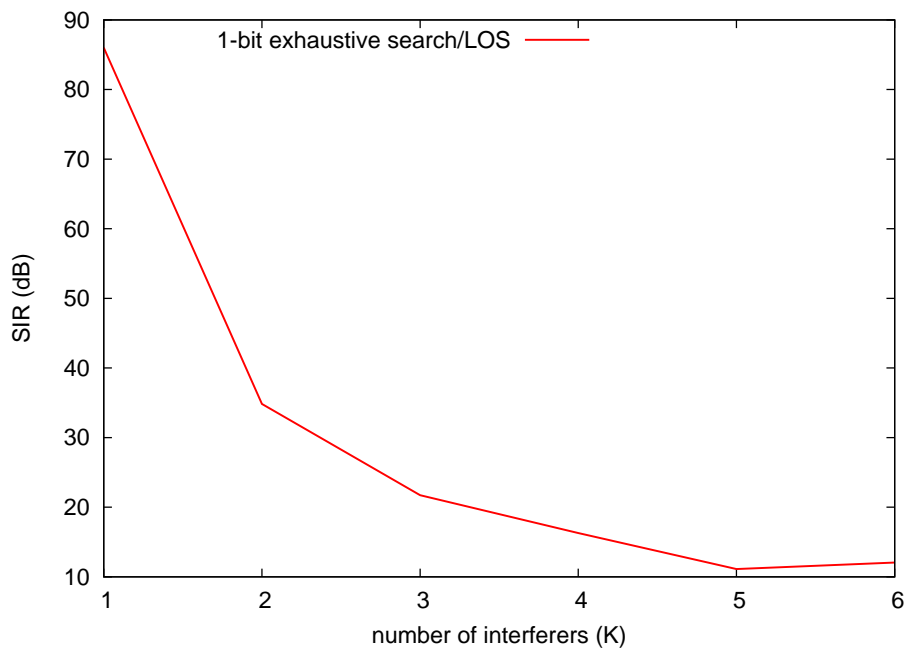


Figure 7.57: Average SIR normalized by N^2 under *iGCD*, for three different initializations (quantized matched filter, quantized zero-forcer, and random quantized vector), versus the number of antennas N (the x-axis), with a fixed number of interferers $K = 15$, and quantization bits $B = 4$. The simulation was under the Rayleigh channel model.



(a)



(b)

Figure 7.58: (a) Average SIR under exhaustive (optimal) searching versus the number of antennas N (the x-axis), with a fixed number of interferers $K = 2$, and quantization bits $B = 1$. (b) Average SIR under exhaustive (optimal) searching versus the number of interferers K (the x-axis), with a fixed number of antennas $N = 10$, and quantization bits $B = 1$. The simulation was under an LOS channel model with a uniform linear array with $\lambda/2$ spacing.

Computation complexity

Simulation results show that the number of steps (iterations) required for convergence is proportional N (for all three algorithms). At each step, the SIR needs to be re-evaluated. However, since we are only changing one coefficient at a time, the number of complex operations (multiplications/additions) needed to compute (update) the new SIR is proportional to K . Since both GCD and i GCD can be initialized using the scalar quantized matched filter (requires only $O(N)$ operations), the total number operations required to compute the GCD and i GCD solutions is $O(NK2^{2B})$ and $O(NKB)$ respectively⁹. Since LSB-GCD only works with the least significant bit(s), it eliminates the dependence on B . However, LSB-GCD requires computing the least-squares (ZF) solution as a starting point. This requires $O(NK^2)$ operations. Therefore, the total running time of LSB-GCD is proportional to $O(NK^2)$. Therefore, if $K \gg B$, then the i GCD can even outperform the scalar quantized ZF beamformer from a computational complexity standpoint.

Switched antenna arrays

In Chapter 2, we listed several alternative architectures for implementing dynamic beamforming in addition to phased antenna arrays, including switched antenna arrays and switched parasitic antennas. While the underlying theory and physical properties behind these different architectures may differ, they all share several key properties. First, they all require an array of antennas¹⁰. Second, the overall gain (if well designed) is proportional to the size of the array. Finally, in all the architectures, some aspect (weight) of each antenna element is dynamically controlled. Therefore, from a mathematical point view, all these different antenna systems can be represented by a size N complex column vector of beamforming weights (channel responses). The universe from which these weights can be drawn will be different for each architecture. For switched antenna arrays and switched parasitic antennas, this universe is usually $\{0, 1\}^N$ (i.e. antennas are either turn on or off). This very low resolution makes vector quantization algorithms like GCD a natural fit for these systems.

The performance is similar to GCD with conventional I/Q quantization. The SIR grows almost quadratically with the ratio N/K (Figure 7.59a shows quadratic increase with N while Figure 7.59b shows a fixed SIR when N/K is fixed). The desired signal grows linearly with N and is largely independent of the number of interferers K (Figures 7.61ab), while the average interferer power shrinks with N/K (Figures 7.60ab).

7.3.3 Improving the beamforming gain using vector quantization

Our focus in this chapter has been almost exclusively on improving the interference suppression since the beam nulls tend to be a lot more sensitive to weight errors than peaks. However, even in the absence of interference, the signal to noise ratio (SNR) can still be improved by using a vector quantization technique like GCD or i GCD although the extra gain from using these techniques will be very small since the loss from scalar quantization is

⁹GCD is exponentially dependent on B while i GCD is only linearly dependent on B .

¹⁰Not all these antennas are necessarily driven by a signal, and may act only as passive reflectors (e.g. switched parasitics).

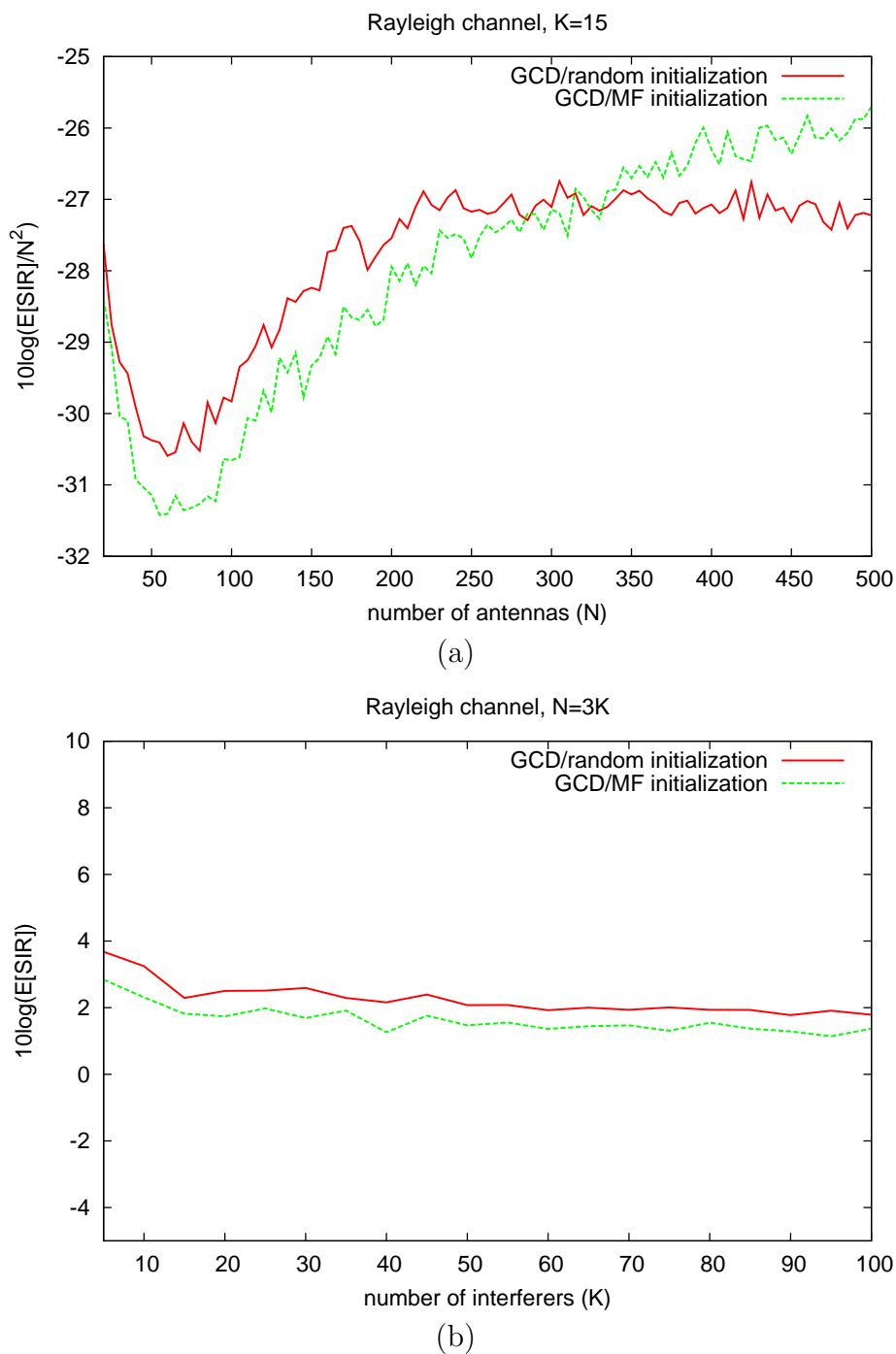


Figure 7.59: (a) Average SIR normalized by N^2 under GCD with two quantization levels $\{0, 1\}$ versus the number of antennas N (the x-axis), with a fixed number of interferers $K = 15$. (b) Average SIR under GCD with two quantization levels $\{0, 1\}$ versus the number of interferers K (the x-axis), when the ratio N/K is fixed. The plots in this figure used the Rayleigh channel model.

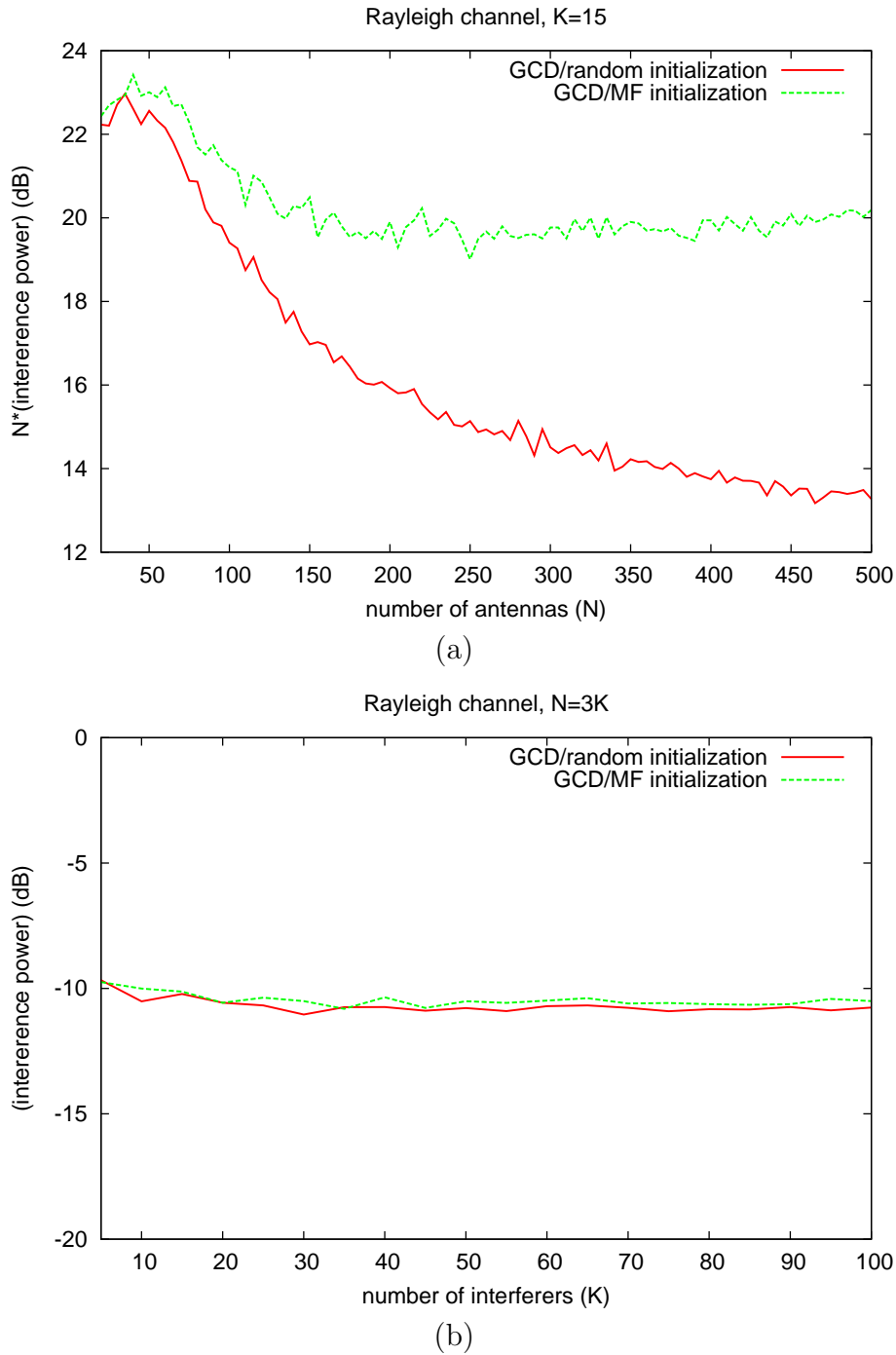


Figure 7.60: (a) Average interferer power multiplied by N (in dB scale) under GCD with two quantization levels $\{0, 1\}$ versus the number of antennas N (the x-axis), with a fixed number of interferers $K = 15$. (b) Average interferer power (in dB scale) under GCD with two quantization levels $\{0, 1\}$ versus the number of interferers K (the x-axis), when the ratio N/K is fixed. The plots in this figure used the Rayleigh channel model.

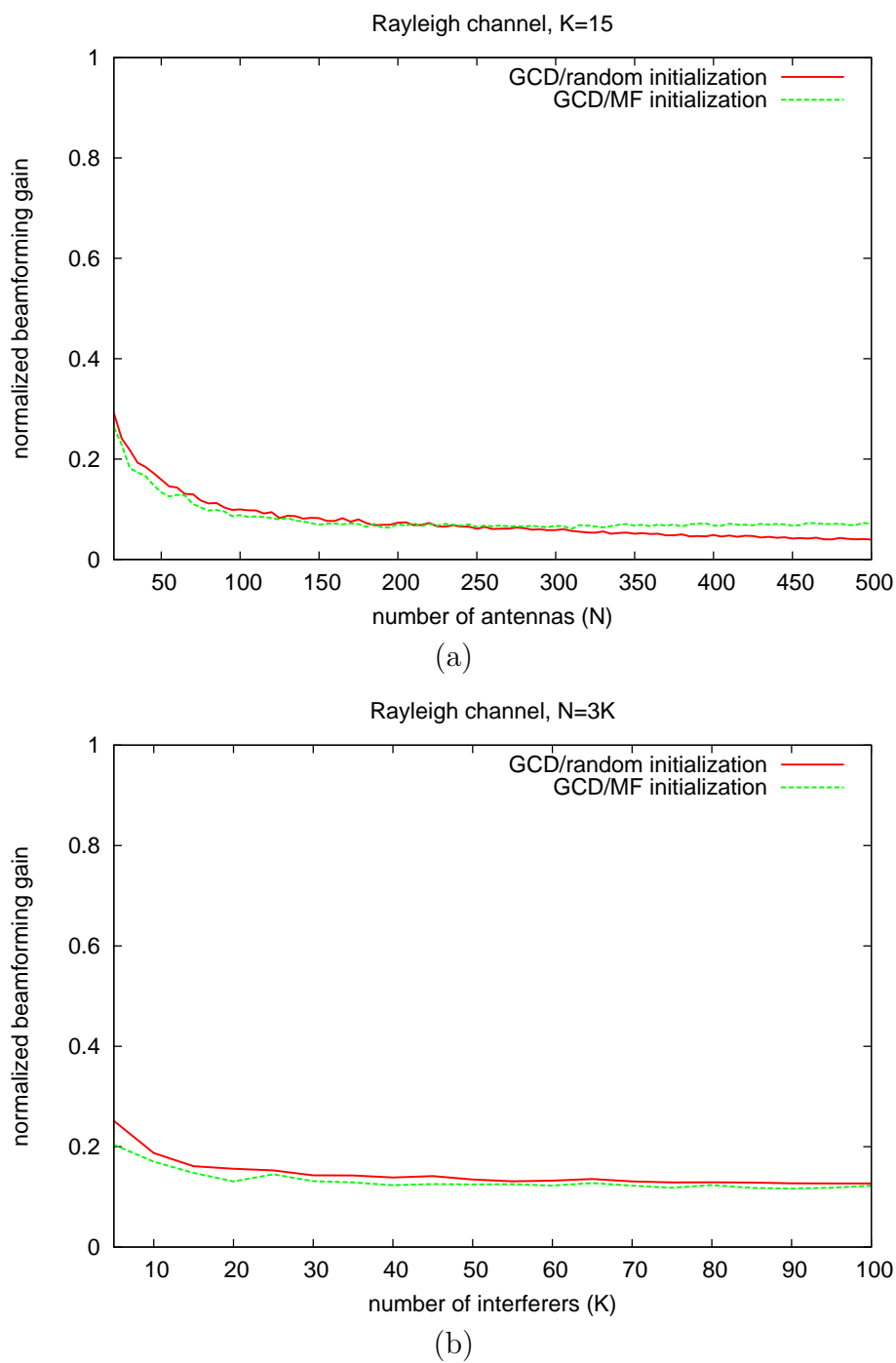


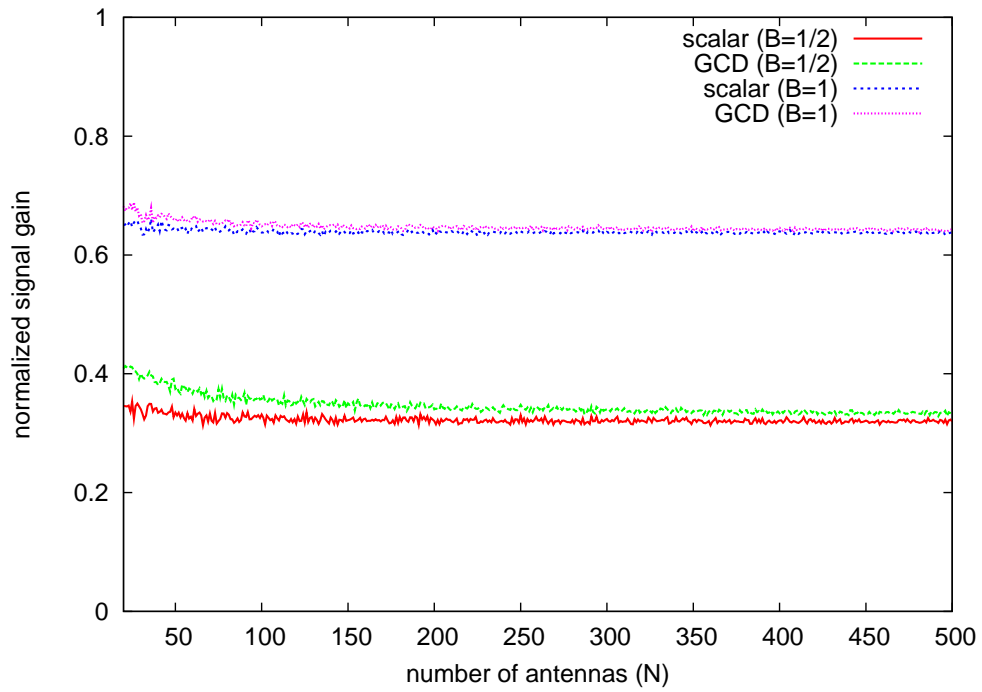
Figure 7.61: (a) Normalized desired signal gain under GCD with two quantization levels $\{0, 1\}$ versus the number of antennas N (the x-axis), with a fixed number of interferers $K = 15$. (b) Normalized desired signal gain under GCD with two quantization levels $\{0, 1\}$ versus the number of interferers K (the x-axis), when the ratio N/K is fixed. The plots in this figure used the Rayleigh channel model.

already small to begin with as we showed in Chapter 6. Figure 7.62a shows the normalized beamforming gain of the scalar quantized matched filter and the GCD quantizer for two quantization levels: $B = 1$ and $B = 1/2$ (for a Rayleigh channel)¹¹. The graph shows that GCD provides a slight improvement over a scalar quantizer and the gap gets smaller as we increase N and B . Furthermore, GCD is still below the optimum though the gap is extremely small (Figure 7.63a shows that the normalized beamforming gain is slightly below optimum solution obtained by exhaustive search). Figure 7.64a shows the normalized beamforming gain when GCD is initialized to two different values (the quantized matched filter and a random vector). The graph shows that the final result is slightly different depending on how we initialize the GCD algorithm, which shows that GCD does not converge to a global minimum. However, the difference is almost negligible. Figures 7.62b, 7.63b, and 7.64b show that the same results hold for an LOS channel (a uniform linear array ($\lambda/2$ spaced) with the azimuth angle chosen randomly and the elevation angle fixed at 90° from the z-axis) as well.

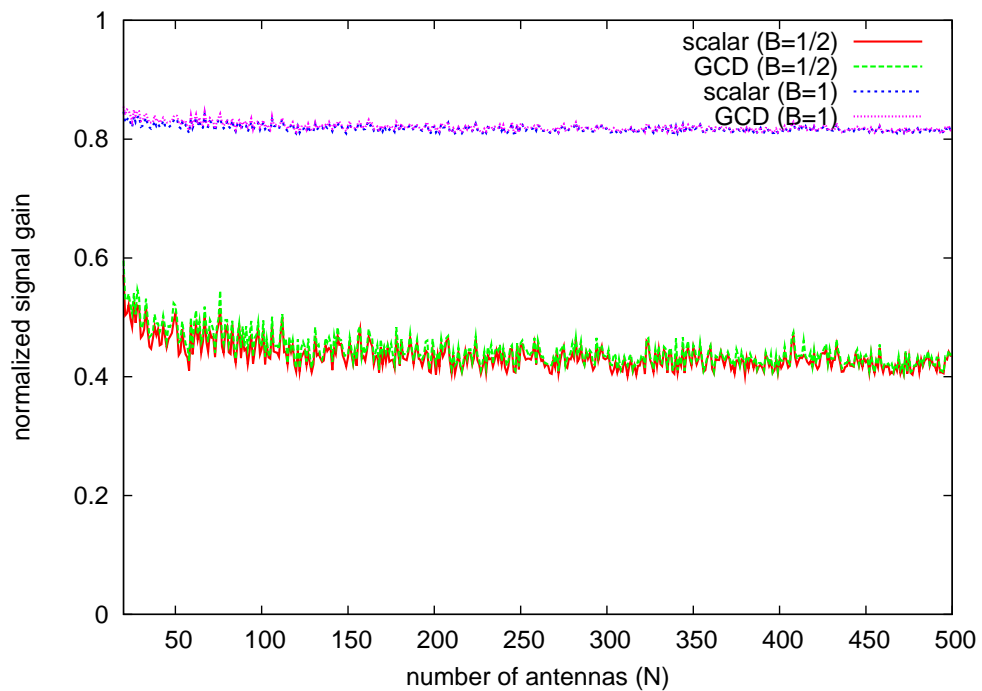
7.4 Conclusion

In this chapter, we have shown that coefficient quantization in digital receiver implementations can have a profound effect on the performance of interference-limited multi-antenna systems. In particular, quantization of beamforming weights critically affects interference suppression performance, and must be performed carefully in order to exploit the degrees of freedom gains from using an increased number of antennas. We have analyzed the performance of both scalar and vector quantization techniques, and presented computationally efficient greedy sequential quantization algorithms that provide substantial gains over naive scalar quantization. However, we have also shown that the interference rejection achieved by those algorithms (polynomial in the number of antennas) is still far from optimum (exponential in the number of antennas). Important open problems include refinement of the analysis to provide tight upper and lower bounds on SIR scaling, and devising efficient algorithms for finding the quantized weights that approach the performance of the optimal quantized weights.

¹¹ $B = 1$ corresponds to four quantization levels: $\{1 + j, 1 - j, -1 + j, -1 - j\}$. $B = 1/2$ corresponds to two quantization levels: $\{1, -1\}$

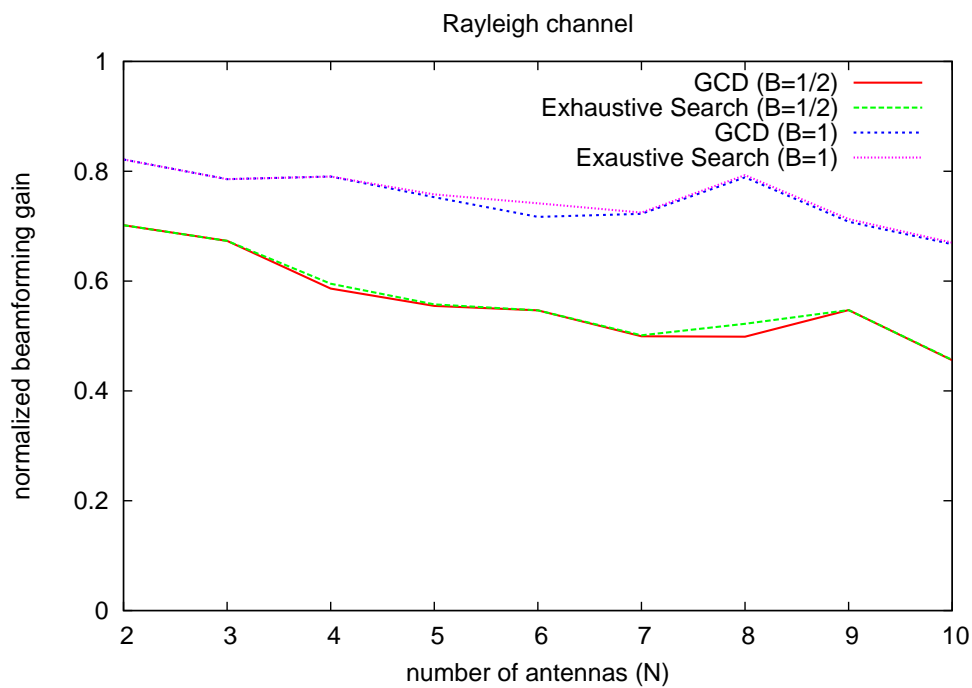


(a)

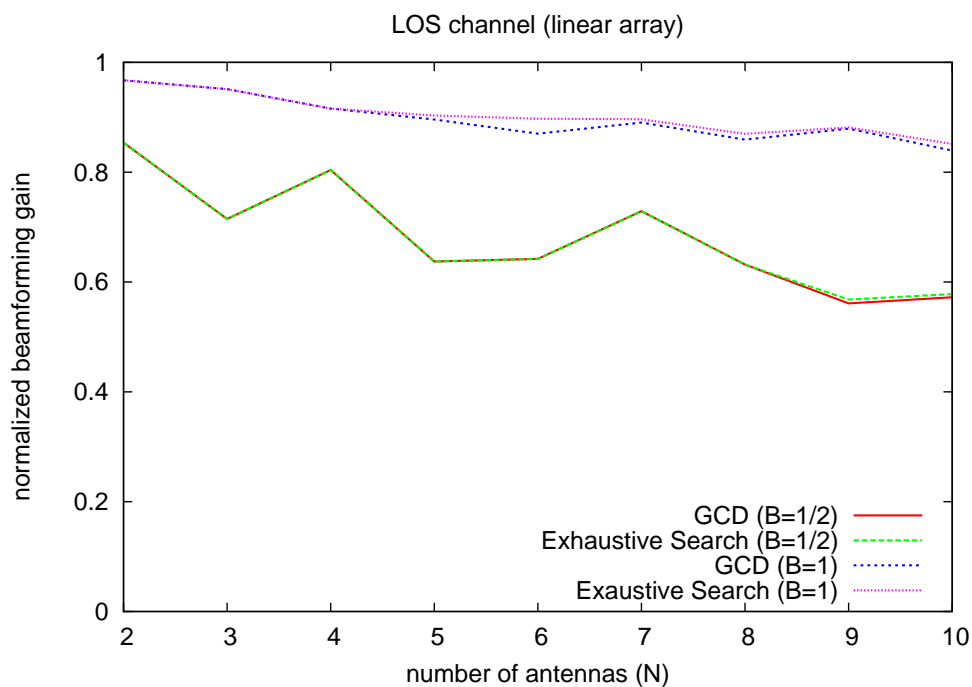


(b)

Figure 7.62: Normalized beamforming gain under GCD and scalar quantization (quantized matched filtering) versus the number of antennas N for two different quantization levels ($B = 1, \frac{1}{2}$): (a) Rayleigh channel. (b) LOS channel (a uniform linear array with a $\lambda/2$ spacing).



(a)



(b)

Figure 7.63: Normalized beamforming gain under GCD and exhaustive search versus the number of antennas N for two different quantization levels ($B = 1, \frac{1}{2}$): (a) Rayleigh channel. (b) LOS channel (a uniform linear array with a $\lambda/2$ spacing).

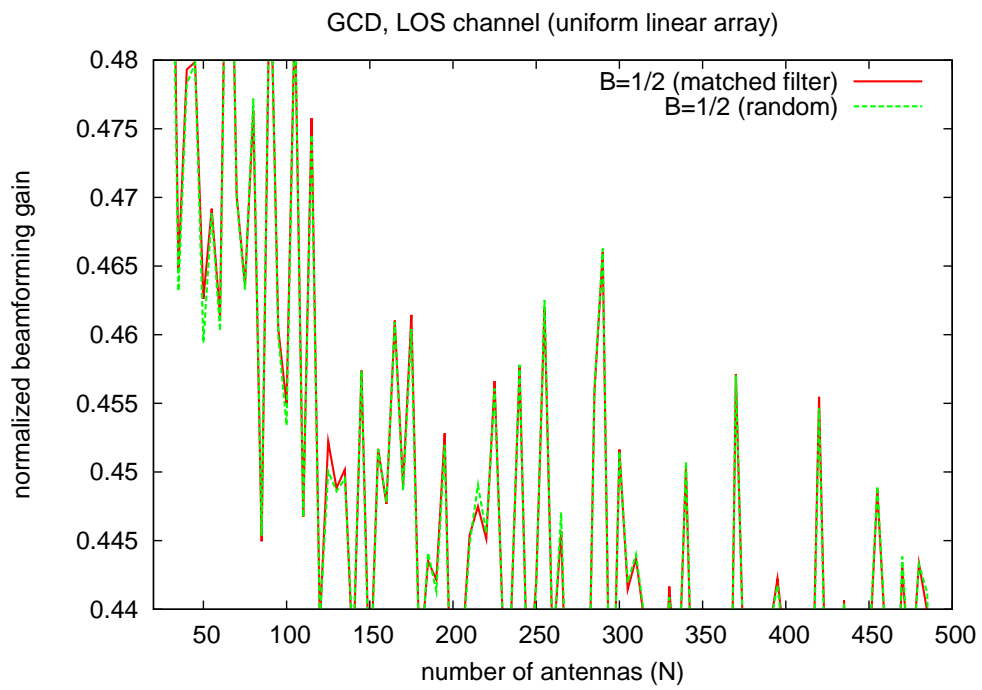
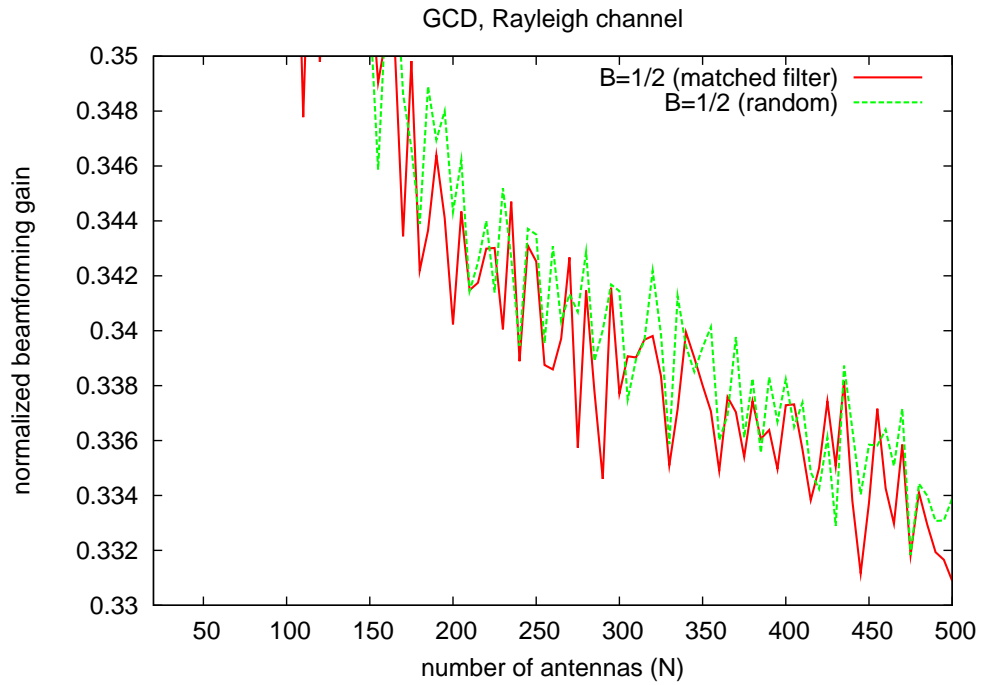


Figure 7.64: Normalized beamforming gain under GCD with two different initializations (quantized matched filter and a random initialization) versus the number of antennas N for two different quantization levels ($B = 1, \frac{1}{2}$): (a) Rayleigh channel. (b) LOS channel (a uniform linear array with a $\lambda/2$ spacing).

Chapter 8

Conclusion and future work

In this work, we have presented an architecture for building high gain adaptive antenna array systems for long range and high capacity systems. The hardware component of the proposed architecture covers both the radio circuit design as well as the antenna design. We found that a hybrid RF/digital beamforming radio architecture provides the best trade off between performance (speed/precision/bandwidth) on the one hand, and cost (complexity/power consumption) on the other hand. We also found that we can utilize 3-dimensional array structures to improve the signal to noise ratio and the beam scan angle. These structures are more compact than traditional planar antennas. In both cases, we proposed different techniques for implementing the radio and antenna architecture with low cost technologies like integrated circuits and printed circuit boards.

The software component of the framework is composed of signal processing and adaptive algorithms for optimizing the performance of arrays that utilize our proposed hardware architecture. We introduced signal processing techniques that improve the performance of conventional adaptive filters used for beamforming in the presence of adverse channel conditions (e.g. noise, interference, fast fading and frequency offsets). We also proposed computationally efficient vector quantization techniques for improving the interference suppression under low resolution and error prone RF and analog beamformers in large arrays.

Up to this point, we have validated our approach mostly with mathematical proofs and computer simulations. The next logical step would be to implement those techniques in hardware and test them in real world settings. Some of this work is already underway. We have already built several prototypes of phased array antennas with 16 and 32 elements operating at frequency bands in the 2.4GHz and 5GHz range (WiFi bands). The prototypes were built with discrete circuit components mounted on PCBs. So far we have had success with initial lab tests and some limited outdoor tests. However, more rigorous testing is required before these prototypes can be deployed in a large scale outdoor testbed.

The next major step in fulfilling our ultimate goal is the realization of a fully integrated beamforming solution. Although our project has been mainly focused on frequencies near cellular bands (e.g. ISM bands), There are ongoing parallel efforts at the Berkeley Wireless Research Center for building and testing a fully integrated phased array system at 60GHz [4, 52, 34, 56]. As of now (December 2010), The project is at the stage where a fully integrated 4-channel phased array transceiver has been built and successfully tested (in 65nm CMOS process) [56].

There also remains a lot of work to be done on the algorithms side. We showed that even though our proposed vector quantization algorithms improve the performance of the array by a factor proportional to the number of antennas compared with scalar techniques, there is still a lot of room for improvement. The performance of the techniques we have proposed is quadratic in the number of antennas, whereas an exhaustive search of all reconstruction levels can be exponential. Therefore, designing a tractable algorithm that approaches this upper bound remains an open problem.

Another direction we would like to pursue in the future is to consider the impact of incorporating beamforming and adaptive antenna techniques on the design of higher layers in the network stack. In this work, we have restricted our attention to the physical layer. However, existing higher layers in the protocol stack (e.g. MAC, network, transport) have not taken into account recent advances in wireless technologies when originally designed. In fact, some of these protocols, like TCP, have not been designed with wireless networks in mind. Therefore, care must be taken when introducing new technologies to ensure that the performance of upper layer protocols is not adversely affected. At the same time, we need to consider how to design future networking protocols in order to take full advantage of state of the art technologies in the wireless space, and how much improvement in the system performance can be achieved when the different components of the system are designed jointly compared when they are designed independently. There has already been some work in this area [51, 16, 17, 19, 18]. However, we believe we barely scratched the surface.

Bibliography

- [1] www.ansoft.com.
- [2] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2009-2014. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html.
- [3] E. Adabi and A. Niknejad. Broadband variable passive delay elements based on an inductance multiplication technique. In *Proc. of IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, June 2008.
- [4] S. Alalusi. *A 60GHz Adaptive Antenna Array in CMOS*. PhD thesis, University of California at Berkeley, Berkeley, CA, USA, 2005.
- [5] S. Alalusi and R. Broderon. A 60GHz Phased Array in CMOS. In *Proc. of IEEE Custom Integrated Circuits Conference*, September 2006.
- [6] O. Bakr and M. Johnson. Impact of phase and amplitude errors on array performance. Technical Report UCB/EECS-2009-1, EECS Department, University of California, Berkeley, January 2009.
- [7] O. Bakr, M. Johnson, R. Mudumbai, and U. Madhow. Interference suppression in the presence of quantization errors. In *Allerton Conference on Communication, Control, and Computing*, September 2009.
- [8] O. Bakr, M. Johnson, R. Mudumbai, and K. Ramchandran. Multi-antenna Interference Cancellation Techniques for Cognitive Radio Applications. In *Proc. of IEEE Wireless Communications and Networking Conference*, April 2009.
- [9] O. Bakr, M. Johnson, J. Park, E. Adabi, K. Jones, and A. Niknejad. A Scalable, Low Cost Architecture for High Gain Beamforming Antennas. In *Proc. of IEEE International Symposium on Phased Array Systems and Technology*, October 2010.
- [10] O. Bakr, M. Johnson, B. Wild, and K. Ramchandran. A multi-antenna framework for spectrum reuse based on primary-secondary cooperation. In *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, October 2008.
- [11] O. Bakr, M. Johnson, B. Wild, and K. Ramchandran. Performance analysis of a multi-antenna framework for spectrum reuse. Technical Report UCB/EECS-2008-101, EECS Department, University of California, Berkeley, August 2008.

- [12] C. Balanis. *Antenna Theory*. John Wiley & Sons, 2005.
- [13] F. Behbahani, Y. Kishigami, J. Leete, and A.A. Abidi. CMOS mixers and polyphase filters for large image rejection. *IEEE Journal of Solid-State Circuits*, 36(6):873–887, June 2001.
- [14] J. Blass. Multidirectional Antenna: A New Approach to Stacked Beams. *IRE International Conference Record*, 8(1), 1960.
- [15] J. Butler and R. Lowe. Beam-Forming Matrix Simplifies Design of Electronically Scanned Antennas. *Electronic Design*, pages 170–173, 1961.
- [16] R. R. Choudhury and Nitin Vaidya. Performance of Ad Hoc Routing using Directional Antennas. *Elsevier Ad Hoc Networks Journal*, November 2004.
- [17] R. R. Choudhury and Nitin Vaidya. On Designing MAC Protocols for Wireless Networks using Beamforming Antennas. *IEEE Transactions of Mobile Computing (TMC)*, May 2006.
- [18] R. R. Choudhury and N. Vaidya. Deafness: A MAC Problem in Ad Hoc Networks when using Directional Antennas. In *IEEE ICNP*, September 2004.
- [19] R. R. Choudhury, X. Yang, R. Ramanathan, and N. Vaidya. Using Directional Antennas for Medium Access Control in Ad Hoc Networks. In *ACM MOBICOM*, September 2002.
- [20] D. Chowdhury, C. Hull, O. Degani, Y. Wang, and A. Niknejad. A Fully Integrated Dual-Mode Highly Linear 2.4 GHz CMOS Power Amplifier for 4G WiMax Applications. *IEEE Journal of Solid-State Circuits*, 44(12):3393–2402, December 2009.
- [21] C. B. Dietrich. *Adaptive Arrays and Diversity Antenna Configurations for Handheld Wireless Communication Terminals*. PhD thesis, Virginia Polytechnic Institute, Blacksburg, VA, USA, 2000.
- [22] H. Dogan, R. Meyer, and A. Niknejad. A DC-10GHz Linear-in-dB Attenuator in 0.13 μ m CMOS Technology. In *Proc. of IEEE Custom Integrated Circuits Conference*, November 2004.
- [23] D. Ehyaie and A. Mortazawi. A 24 GHz Modular Phased Array Based on Extended Resonance Technique. In *Proc. of IEEE International Symposium on Phased Array Systems and Technology*, October 2010.
- [24] P. Haldi, D. Chowdhury, P. Reynaert, G. Liu, and A. Niknejad. A 5.8 GHz 1 V Linear Power Amplifier Using a Novel On-Chip Transmormer Power Combiner in Standard 90 nm CMOS. *IEEE Journal of Solid-State Circuits*, 45(5):1054–1063, May 2008.
- [25] H. Hashemi, X. Guan, A. Komijani, and A. Hajimiri. A 24-ghz sige phased-array receiver-lo phase-shifting approach. *IEEE Transactions on Microwave Theory and Techniques*, 53(2):614–26, February 2005.

- [26] M. Hayes. *Statistical Digital Signal Processing and Modeling*. Wiley, 1996.
- [27] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 2002.
- [28] T. Lee. *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2004.
- [29] J. C. Liberti and T. Rappaport. *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*. Prentice Hall, 1999.
- [30] X. Liu, A. Sheth, M. Kaminsky, K. Papagiannaki, S. Seshan, and P. Steenkiste. DIRC: increasing indoor wireless capacity using directional antennas. In *ACM SIGCOMM*, 2009.
- [31] U. Madhow, K. Bruvold, and L. J. Zhu. Differential MMSE: A framework for robust adaptive interference suppression for DS-CDMA over fading channels. *IEEE Transactions on Communications*, 53(8):1377–1390, Aug. 2005.
- [32] A. Mainwaring, A. L. Umnov, V. A. Filimonov, I. S. Shishalov, and M. Shuralev. Theoretical and Experimental Investigations of Adaptive Reactance Parasitic Antenna Array. In *Mobile and Wireless Communications Summit*, 2007.
- [33] D. Manolakis, V. Ingle, and S. Kogon. *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. McGraw-Hill, 2000.
- [34] C. Marcu, D. Chowdhury, C. Thakkar, J. Park, L. Kong, M. Tabesh, Y. Wang, B. Afshar, A. Gupta, A. Arbabian, S. Gambini, R. M. Zamani, E. Alon, and A. Niknejad. A 90nm cmos low-power 60 ghz transceiver with integrated baseband circuitry. *IEEE Journal of Solid-State Circuits*, 44(12), December 2009.
- [35] S. M. Mishra, J. Hwang, D. Filippini, T. Du, R. Moazzami, and L. Subramanian. Economic Analysis of Networking Technologies for Rural Developing Regions. In *WINE*, December 2005.
- [36] R. Mudumbai, J. Hespanha, and U. Madhow. Scalable feedback control for distributed beamforming in sensor networks. In *IEEE International Symposium on Information Theory (ISIT)*, September 2005.
- [37] R. Mudumbai, B. Wild, U. Madhow, and K. Ramchandran. Distributed beamforming using 1 bit feedback: from concept to realization. In *Allerton Conference on Communication, Control, and Computing*, September 2006.
- [38] A. Natarajan, A. Komijani, and A. Hajimiri. A Fully Integrated 24-GHz Phased-Array Transmitter in CMOS. *IEEE Journal of Solid-State Circuits*, 40(12):2502–14, December 2005.
- [39] S. Nedeveschi, R. Patra, S. Surana, S. Ratnasamy, L. Subramanian, and E. Brewer. An Adaptive, High Performance MAC for Long-Distance Multihop Wireless Networks. In *ACM MOBICOM*, 2008.

- [40] A. Niknejad, M. Bohsali, E. Adabi, and B. Heydari. Integrated circuit transmission-line transformer power combiner for millimetre-wave applications. *Electronic Letters*, 43(5):290–291, March 2007.
- [41] A. Niknejad and H. Hashemi. *mm-Wave Silicon Technology: 60 GHz and Beyond*. Springer, 2008.
- [42] A. Oppenheim and R. Schaffer. *Discrete-Time Signal Processing*. Prentice Hall, 1999.
- [43] J. Paramesh, R. Bishop, K. Soumyanath, and D.J. Allstot. A 1.4V 5GHz four-antenna Cartesian-combining receiver in 90nm CMOS for beamforming and spatial diversity applications. In *Proc. of IEEE International Solid-State Circuits Conference*, February 2005.
- [44] R. Patra. *A Multi-Tier Network Architecture for Long Distance Rural Wireless Networks in Developing Regions*. PhD thesis, University of California at Berkeley, Berkeley, CA, USA, 2009.
- [45] D. Pozar. *Microwave Engineering*. John Wiley & Sons, 2005.
- [46] J. G. Proakis and D. G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*. Prentice Hall, 1996.
- [47] J. G. Proakis and M. Salehi. *Communication Systems Engineering*. Prentice Hall, 2002.
- [48] RNCOS. Emerging Mobile Wireless Technologies – Road to a Wireless World, 2005.
- [49] K. Rosen. *Discrete Mathematics and its Applications*. McGraw Hill, 2007.
- [50] A. Sayed. *Adaptive Filter*. Wiley & Sons, 2008.
- [51] S. Sen, X. Jie, R. Ghosh, and R. R. Choudhury. Link Layer Multicast using Smart Antennas: No Client Left Behind. In *IEEE ICNP*, 2008.
- [52] D. Sobel. *A Baseband Mixed-Signal Receiver Front-End for 1Gbps Wireless Communications at 60GHz*. PhD thesis, University of California at Berkeley, Berkeley, CA, USA, 2008.
- [53] G. Strang. *Linear Algebra and Its Applications*. Thomson Learning, 1988.
- [54] W. Stutzman and G. Thiele. *Antenna Theory and Design*. John Wiley & Sons, 1998.
- [55] S. Surana, R. Patra, S. Nedeveschi, M. Ramos, L. Subramanian, Y. Ben-David, and E. Brewer. Beyond Pilots: Keeping Rural Wireless Networks Alive. In *NSDI*, 2008.
- [56] M. Tabesh, J. Chen, C. Marcu, L. Kong, S. Kang, E. Alon, and A. Niknejad. A 65nm CMOS 4-Element Sub-34mW/Element 60GHz Phased-Array Transceiver. In *Proc. of IEEE International Solid-State Circuits Conference*, To appear in February 2011.
- [57] D. V. Thiel and S. Smith. *Switched Paracitic Antennas for Cellular Communications*. Artech House, 2002.

- [58] H.L. Van Trees. *Optimum Array Processing*. Wiley-Interscience, 2002.
- [59] D. Tse and P. Viswanath. *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.
- [60] F. Ulaby. *Fundamentals of Applied Electromagnetics*. Prentice Hall, 2004.
- [61] H. Wang and A. Hajimiri. A Wideband CMOS Linear Digital Phase Rotator. In *Proc. of IEEE Custom Integrated Circuits Conference*, September 2007.
- [62] M. R. Williamson, G. E. Athanasiadou, and A. R. Nix. Investigating the effects of antenna directivity on wireless indoor communication at 60GHz. In *Proceedings of the PIMRC Conference*, 1997.
- [63] H. Zarei, C.T. Charles, and D.J. Allstot. Reflective-type phase shifters for multiple-antenna transceivers. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 54(8):1647–56, August 2007.