SINGLE CHANNEL FULL-DUPLEX WIRELESS RADIOS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Mayank Jain
August 2011

This dissertation is online at: http://purl.stanford.edu/zh047jt6489

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Philip Levis, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Sachin Katti

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Nick McKeown

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Daniel O'Neill

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

Wireless networking is fast becoming the primary method for people to connect to the Internet and with each other. The available wireless spectrum is increasingly loaded, with users demanding higher performance and reliability from their wireless connections. This dissertation proposes single channel full-duplex as a new paradigm for wireless system design that can mitigate some of the throughput and reliability problems of today's wireless systems.

With full-duplex wireless radios, a node can receive and transmit data at the same time, without using multiple wireless channels. At the physical layer, this capability can double the available throughput at a node. Further, sending and receiving at the same time allows a wireless node to exchange control messages while receiving data, making real-time feedback schemes possible. Such feedback, assumed to be impossible till now, allows network researchers to rethink the way wireless networks are designed. This dissertation shows that with the extra feedback channel available through full-duplexing, reliability in existing wireless networks can be significantly improved. Motivated by the promise that full-duplex wireless holds, this dissertation explores the challenges in implementing such radios and how these radios could influence the design of wireless networking stacks.

The primary challenge in implementing a full-duplex wireless system is removing the self-interference from a node's transmit antenna to its receive antenna. This dissertation discusses various analog and digital techniques to cancel this self-interference. It presents an adaptive full-duplex wireless system that combines analog and digital self-interference cancellation to remove up to 73dB of self-interference. Exploiting full-duplexing to the fullest extent requires a redesign of higher layers, especially the

MAC and network layers. This dissertation presents a full-duplex MAC layer implementation to show how full-duplexing can be leveraged to improve wireless reliability and performance. This implementation reduces hidden terminal losses by up to 88% and significantly improves network fairness and throughput.

The main contribution of this dissertation is to motivate full-duplex radios as a direction for research on future wireless systems. It shows that designing full-duplex systems, while challenging, is feasible. It also presents many ideas on how full-duplexing could be used to improve a variety of wireless systems, including multi-hop data networks, cellular systems, and wireless LANs.

To my parents and my brother

Acknowledgements

The endeavor of doctoral research entails many trials and tribulations. My time at Stanford, through ups and downs, was enriched in a variety of ways with the care and contributions of many people.

Philip Levis, my advisor, has been very supportive of me throughout this journey. His attention to detail and focus on getting practical results have been formative in all research that I have done since joining his group. His ability to conduct research in diverse areas like networking, wireless communications, distributed systems and operating systems continues to amaze me. Phil saw the possible impact of full-duplex radios before any of us and kept pushing us through numerous discussions to improve the system and realize its full potential. I would also like to thank him for instilling in me the importance of good writing and presenting research in a human understandable form. His enthusiasm for both technical and non-technical aspects of life is infectious.

I would like to thank Sachin Katti for many technically deep discussions. It was in one of his classes that the seed of this dissertation was sown. He provided valuable guidance in the transformation of full-duplex radios from a small class project idea to a full PhD dissertation. He is also full of new ideas for where this technology could be applied. Working on research with him has been a pleasure and a great learning opportunity.

Nick McKeown is amazing in how he handles big research projects with utmost ease. I thank him for all his technical insights during the early part of my Ph.D., while working on a joint project with his group. His questions about my research always made me think harder, dig deeper and often come up with better solutions.

Daniel O'Neill has been a pleasure to know both as a research guide and a friend.

My discussions with him covering subjects as varied as research, technology, politics, business, and good restaurants made for very enjoyable and educational lunch meetings. I was fortunate to be able to tap into his wisdom and knowledge and will continue to value his advice.

The work presented in this research would not have been possible without significant contributions from very motivated individuals. I'd like to thank Jung Il, whose contributions to this research probably exceed mine. Kannan was deeply involved with and excited about this research from the first day and I hope he will successfully carry it forward in his new capacity as a professor. Taemin and Dinesh spent many sleepless nights trying to get the full-duplex system working days before the Mobicom'11 paper deadline. I would also like to thank Prasun Sinha and Siddharth Seth for valuable contributions in the form of technical discussions and suggestions. Patrick Murphy from Mango Communications was very helpful, providing technical guidance for using WARP boards and loaning us a few testboards while we were getting up to speed on the hardware platform.

My stay at Stanford benefited tremendously from interactions with all members of the SING group, past and present. Jung Il has been a co-author on all the papers that I have written since joining this group. I do not think I have ever worked as efficiently in a team of two as I have with Jung Il. Kannan was the one responsible for pulling me in to the group by introducing me to Phil and has since been an awesome collaborator. He constantly keeps coming up with ideas to apply theoretical concepts to practical problems which led to some very interesting research. Maria is the lab's "Chief Fun Officer" and takes her responsibilities very very seriously. I would miss our early morning working sessions and her various baked goodies. She also serves as an aesthetics advisor for presentations and loves being a grammar Nazi. Ewen is the most knowledgeable computer programmer around, which made programming help in almost any language literally available next door. Even though I do not drink beer, Ewen's beer brewing experiments made for very interesting conversation. Behram's affinity to all things food and his awesome maple ice cream made it easy to talk about and consume good food at the same time. I also thank him for being my soldering tutor. Tahir was the measure of how good a joke was; making him laugh

confirmed the quality of a joke. Jung Woo, Kevin, Brano, Wanja, Om and Martin all provided very interesting conversations, both technical and non-technical. Richard, Juan and Eddie were all very bright and extremely hard working interns whom I had the pleasure of working with. My work also benefited through interactions with members outside SING, including Brandon, Masa and KK from the McKeown group, Jeff, Manu, Aditya and Steven from SNSG, and several others.

The daily grind in the office became easier because our awesome admins took such good care of us. Alexis Wing, Mary Jane Swenson, Marianne Siroker and all the admins in the Gates 2B wing made bureaucratic stuff completely transparent to us. Alexis in particular worked untiringly with the Stanford procurement department and various equipment vendors to make sure that we received test boards and other essential hardware on-time for meeting various deadlines. Charlie Orgish is the best network admin one could ever hope for. With him around, setting up new networking nodes through the whole building was made as painless as possible, and troubleshooting any networking problem was as simple as knocking on his door.

My time at Stanford was enriched by many interactions outside the scope of research as well. I made some of my best friends after I came to Stanford while further strengthening several existing friendships.

I hold Sunny responsible for me applying to Stanford. I remember him telling me if he could get into Stanford, so could I. I still consider that to be among the biggest compliments I have received. Riti and Sunny have been a constant source of encouragement since the day I arrived at Stanford. From helping me setting up a bank account to getting me familiarized to the bus routes and providing a place on campus where I could crash whenever I had a very late night, they were always there for me. Riti's great cooking and Sunny's late night movie plans kept things fun at Stanford. Rajan decided that since Sunny and I were here, he should come to Stanford as well. Him being here was both a reminder of good old times and an opportunity to develop many new experiences.

Sachin Adlakha was probably my first Stanford friend. He has been immensely helpful, providing research ideas when I was stuck, providing company during lunch and post lunch walks, cooking great food and being a great roommate for a brief period

of time. Vineet was the willing target of constant teasing. His last few weekends in the Bay Area before going to UIUC was the most driving and hiking I have ever done. Kannan, Kadambari and their son Abhay are great friends and provided wonderful company while at Stanford. I'll fondly remember the long homework sessions with Kannan and the delicious idlis, dosas and rasam chawal that I had for dinner at their place. Forum is probably the happiest person I know. Her constant enthusiasm for fun activities and positive approach to life were as enjoyable as eating bowlfuls of her mother's undhiyu. Mridul is a strong believer of work hard and play hard. When he wasn't slogging away in office, he was looking for places to go rafting, hiking, driving or some combination thereof. I would also like to thank Dinkar for being a great conversationist and an awesome cook. Manasi, Rohit, Ashutosh, Shivani, Sandeep and Urmila all helped ease the stress of research with weekend outings, movie nights and several home cooked dinners. My lunch group at Stanford including Vinay, Samar, Sara, Michelle, Hattie, Tom and Rebecca made me look forward to noon everyday.

I had a lot of support from my extended family in the Bay Area and India throughout my Ph.D. I thank Ranu didi, Hemu Jijaji, Karnika Didi, Nidhi, Deepti, Ashish and Sudhanshu for their constant encouragement and words of wisdom. My nieces and nephews, Dhruv, Devika, Naman, Namya and Chandrika, kept me entertained with their antics and also made me a very patient person. I especially thank Hem Mamiji and Ramesh Mamaji for supporting me and taking care of me during one the toughest times in my life.

Finally, I would like to thank my parents and my brother Saurabh. Saurabh was my roommate, friend and local guardian for the last six years. He supported me when I decided to come to Stanford without any guaranteed funding and never let me worry about anything other than my work during my Ph.D. He also arranged all our vacations and weekend hiking trips and I just had to show up. He is responsible for me never having to experience the life of a poor grad student. My parents have been a constant source of support throughout my life. My father got me interested in engineering early on with all our projects at home trying to fix broken electronics, while my mother drilled in me the virtue of working hard. My parents encouraged me

in all my decisions, including quitting a well paying job to go to grad school. They took care of me and helped me get through some very tough times in the middle of my Ph.D. My family constantly inspires me to attain greater heights. This dissertation is dedicated to them.

Contents

List of Tables

List of Figures

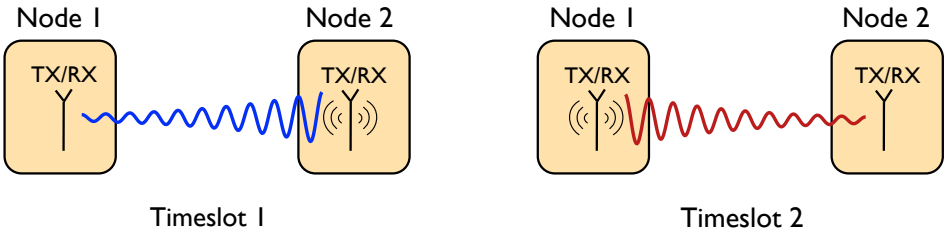
Chapter 1

Introduction

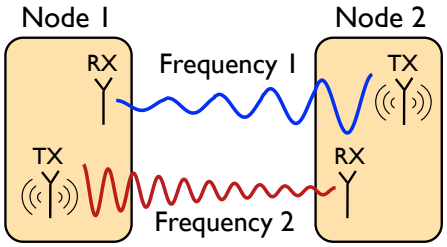
Wireless radios are increasingly pervasive in everyday life. With laptops connecting to WiFi hotspots and cellphones streaming videos through cellular basestations, users are demanding higher speeds and higher availability from their wireless networks. Even with the significant advancements made in wireless network designs over the last couple of decades, wireless networks are plagued with problems such as intermittent connection losses and unexplained loss of performance. An example that many readers would identify with is seeing a good signal level on the WiFi connection of a laptop and still experiencing page load errors in a web browser. Such problems are much more noticeable in highly crowded wireless environments, such as enterprise buildings or conferences with hundreds of users using the wireless medium for data transfer at the same time.

The challenges in wireless originate from the shared, broadcast nature of the wireless medium. A shared medium implies that communication devices need to contend amongst themselves, requiring specific sharing mechanisms to use the medium efficiently. The wireless medium also exhibits rapid attenuation of signals. With such attenuation, different devices in a network can have very different and inconsistent views of the wireless channel.

Before we consider a whole network of wireless nodes trying to use the wireless channel efficiently, we should consider the simple case of two nodes trying to send data to each other. For example, in an access point based WiFi network, a laptop



(a) Time-Division Duplexing



(b) Frequency-Division Duplexing

Figure 1.1: Current wireless systems create a bi-directional communication channel using either TDD or FDD

connects to the Internet by sending uplink packets to the access point and receiving downlink packets from the access point. Thus, conceptually, most networks require a bi-directional communication channel between communicating nodes. The notion of a communication channel allowing bi-directional data transfer is called duplexing. Interference between transmissions and receptions has to be avoided for duplex communication to be possible.

1.1 Wireless Today

Current wireless systems achieve the isolation required between the two directions of communication using independence in either time or frequency. Accordingly, these duplexing techniques are called Time-Division Duplexing, or Frequency-Division Duplexing.

- **Time-Division Duplexing (TDD)** is when nodes divide access in time as shown in Figure 1.1(a). When Nodes 1 and 2 want to send data to each other,

they break time into slots. Then, the two nodes can take turns in sending data to each other. Dividing channel time in this way prevents the transmissions from the two nodes from interfering with each other. Many data networks use TDD as it is simple to implement, especially in ad-hoc networks where frequency use is not tightly controlled. Time-division duplexing is also commonly known as half-duplexing. Examples include wireless LANs (802.11) and Bluetooth [1, 17].

- **Frequency-Division Duplexing** uses two different carrier frequencies for carrying transmissions, as shown in Figure 1.1(b). In this case, Nodes 1 and 2 can send data to each other at the same time, albeit using two different frequencies. The use of different frequencies prevents the two signals from interfering with each other, even though the two transmissions occur at the same time. Many cell networks use FDD to enable simultaneous uplink and downlink transmissions in the network. Examples of FDD use include GSM and CDMA2000 cellular systems and DSL internet connections [2, 3, 4, 5]. Since FDD uses separate frequencies for sending and receiving, it uses double the bandwidth used by TDD for achieving the same physical layer data rate.

Using time-division duplexing exacerbates the inconsistency in the channel views across nodes. Since only one node among a pair of communicating nodes can transmit at a given time, the wireless channel around the transmitting node may look occupied, while the wireless channel around the receiving node may look unoccupied. Such inconsistencies are the root cause of many of the problems with TDD wireless networks, such as packet losses due to hidden terminal effects.

On the other hand, frequency-division duplexing requires a wireless node to use twice the frequency bandwidth for sending and receiving signals of a given bandwidth. In some cases, this is expensive. For example, using FDD in cellular networks requires a carrier to license paired spectrum, i.e. buy the license for the use of two frequencies instead of one, to support FDD. In other cases, FDD may simply be infeasible. For example, 802.11b networks operate in the 70MHz wide 2.4GHz band with a signal bandwidth of 20MHz. This allows for up to three non-overlapping channels of operation for independent 802.11 networks co-existing in close proximity. With FDD,

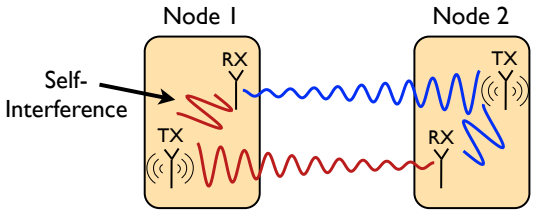


Figure 1.2: Self-interference is the main challenge in implementing single channel full-duplex wireless.

each 802.11 network would have to use twice the frequency bandwidth, i.e. 40MHz, thus not allowing more than a single 802.11 network to co-exist without interference between networks.

One way to mitigate the problem of inconsistent channel views in TDD without the excess bandwidth required with FDD would be to design a radio that does not use frequency division or time division for duplexing; a radio that can send and receive data at the same time using a single carrier frequency. This leads us to the question: ***Why are wireless radios not single channel full-duplex?***

1.2 Single Channel Full-Duplex and Self-Interference

> *"It is generally not possible for radios to receive and transmit on the same frequency band because of the interference that results. Thus, bidirectional systems must separate the uplink and downlink channels into orthogonal signaling dimensions, typically using time or frequency dimensions."*
>
> - Andrea Goldsmith, "Wireless Communications," Cambridge Press.

A basic precept of wireless communication is that a radio cannot transmit and receive on the same frequency at the same time, i.e. operate in a full-duplex fashion. As wireless signals attenuate quickly over distance, the signal from a local transmitting antenna is hundreds of thousands of times stronger than transmissions from other nodes. Figure 1.2 shows an example where Nodes 1 and 2 are trying to send data to each other simultaneously using the same frequency. Node 1's own transmission is

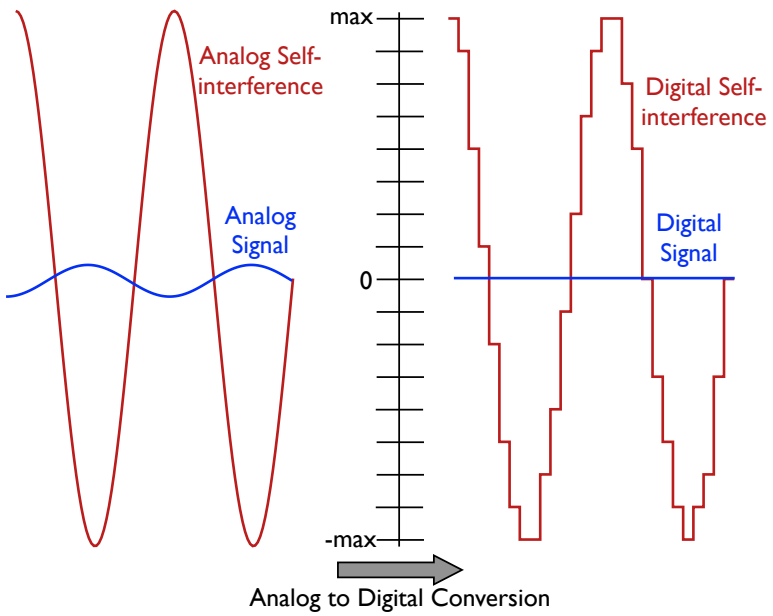


Figure 1.3: With very strong self-interference, the Analog to Digital Converter (ADC) saturates. The digital samples have little or no information of the intended digital signal.

much stronger at its receive antenna, compared to the signal it receives from Node 2. With such strong self-interference, the receiver of Node 1 is unable to decode any signals that Node 2 is trying to send to Node 1.

This example shows that the biggest challenge in designing a single channel full-duplex wireless radio is eliminating the self-interference signal from the receiver of the wireless node. In theory, this problem should be easy to solve. For a system with an antenna each for transmit and receive, since the system knows the transmit antenna's signal, it can subtract it from the receive antenna's signal and decode the remainder. For example, a typical 802.11 system uses 20dBm transmit power. The power of the transmit antenna's signal at a receive antenna placed 6 inches away is ∼-10dBm, while the noise floor is ∼-85dBm. Thus, if a system can remove ∼75dB of self-interference by cancellation, it can decode the receive antenna's signal.

The idea of interference cancellation is not new. Wireless researchers have used interference cancellation techniques to either exploit collisions [40] or recover from collision losses [32, 34]. Full-duplexing using some cancellation techniques has also

been explored in the literature. Analog cancellation techniques using noise canceling chips have been proposed to subtract the self-interference signal (the "noise") from the received signal [52]. Digital cancellation, used in CSMA/CN [55], optical networks [30], and proposals for full-duplex operation [16], subtracts self-interference in the digital domain, after receiver has converted the baseband signal to digital samples. Based on existing work, it should be possible to construct a system that subtracts self-interference from the received signal to enable full-duplex operation.

Unfortunately, solving this problem in practice is much harder. Strong self-interference saturates the receive circuitry, specifically the analog to digital converter (ADC), thus making it impossible for the receiver to decode a packet after removing the self-interference signal. This precludes implementing a full-duplex system using only digital cancellation. Figure 1.3 shows how self-interference can saturate a receiver's ADC. A typical wireless receiver takes an analog radio frequency signal from the receive antenna, and converts it to a series of digital samples using the ADC. These digital samples are used to process data packets. Digital conversion involves defining quantization levels for discretizing the continuous analog input. The ADC adapts the quantization levels based on the strongest signal being received to avoid clipping effects.

As Figure 1.3 shows, when a receiver gets a very strong self-interference signal, the ADC scales its quantization levels to match the level of the self-interference. Since the ADC has a finite resolution (typically 8-12 bits,) it has a finite number of quantization levels. If the intended receive signal is much weaker than the self-interference, the signal at the receiver after the ADC may contain no information from the intended signal. In such a case, even if all the self-interference is removed from the received digital samples, the receiver will not be able to process the intended packet. Similarly, strong self-interference can also cause saturation in other components of the receiver, especially, analog amplifiers. To solve the self-interference problem, a receiver has to at least reduce, if not remove, the self-interference before it causes saturation to any component in the receive chain.

Digital cancellation cannot be used on its own to cancel self-interference because of ADC saturation. We need to develop self-interference cancellation techniques for

analog signals as well. Canceling self-interference using analog signals can remove or reduce the self-interference signal before the ADC, avoiding saturation. Existing work has used off-the-shelf noise cancellation chips to implement analog cancellation [52]. This cancellation technique does not work with a fairly high power, wideband wireless signal, such as that used for 802.11 (WiFi).

Wired networks, such as telephone networks, have used echo-cancellation schemes for combining voice data going in opposite directions onto a single twisted-pair, effectively establishing a full-duplex channel. Although the wired voice channel is much more static and narrowband than wireless data channels, echo-cancellation schemes provide a good reference for techniques to reduce self-interference.

1.3 Contributions

This dissertation proposes and explores single channel wireless full-duplexing as a new paradigm for designing wireless networks. Specifically, it makes three contributions.

- It presents novel designs and improvements to existing techniques for canceling self-interference [22, 39]. It also evaluates these techniques through analysis and implementation using hardware components, and the USRP and WARP software radio platforms [9, 8].

- It presents the design of a working wireless full-duplex prototype that brings analog and digital cancellation techniques together and proposes algorithms to adapt the design to wireless channel changes. The prototype can cancel up to 73dB of self-interference making full-duplex feasible for WiFi like systems [39].

- It explores the larger implications of full-duplexing on network performance. It also evaluates a subset of these implications with an implementation of a full-duplex MAC on a 5-node software radio testbed. The full-duplex setup can achieve a network throughput of 11Mbps vs 7.5Mbps for half-duplex while significantly improving fairness and can reduce hidden terminal losses by up to 88%.

1.3.1 Self-interference Cancellation Techniques

The first contribution of this dissertation is to implement and compare many existing and novel cancellation techniques that can be used to mitigate the self-interference problem. We identify different analog radio frequency and digital baseband techniques that have been used in current systems, or can be used for implementing interference cancellation mechanisms. These techniques provide varying levels of performance and introduce different constraints on the design of the system, which makes certain techniques less feasible for current wireless systems. For example, this work shows that a phase-offset approach, a seemingly simple and elegant way to implement analog cancellation, restricts the maximum bandwidth of the wireless signal and may not be useful for systems such as 802.11n. A signal inversion approach, on the other hand addresses the problems with the phase-offset approach to give much better cancellation performance, even with very wideband signals.

We use a tightly controlled channel sounder[1] to compare the performance of phase-offset cancellation and signal inversion cancellation, and the limits of the existing circuitry at the physical layer. We find that well-tuned signal inversion cancellation circuit built from commodity components can cancel over 45dB across a 40MHz bandwidth.

1.3.2 Adaptive Wireless Full-Duplex Prototype

The second contribution of this dissertation is to combine analog and digital cancellation techniques for improving the overall performance of self-interference cancellation. Combining an analog signal inversion cancellation technique with a digital cancellation technique designed for OFDM systems allows a full-duplex radio to cancel up to 73dB of self-interference: consequently, full-duplex 802.11n devices are possible with a separation of 20cms between TX and RX antennas.

Furthermore, this dissertation presents algorithms that can adaptively tune both cancellation mechanisms to quickly, accurately, and automatically adapt to changes

[1]These channel sounders are wideband (∼240MHz) radios used for RF profiling, programmed to generate a single wideband pilot pattern for measuring the channel.

in the wireless channel. A simple gradient descent over received interference power is used to adapt the analog cancellation mechanism, while the digital cancellation scheme uses OFDM based channel estimation to adaptively estimate and cancel the self-interference in the digital domain. The tuning algorithms are fast, typically requiring less than one millisecond to retune the cancellation mechanism from scratch.

We also explore the engineering challenges that arise in making full-duplex practical. While signal inversion can, in theory, provide perfect cancellation, this assumes that the signal inversion circuit has a flat frequency response: if not flat, the inverted signal might differ slightly from the transmitted signal. The cancellation performance depends on the resolution and range of the devices used and the extent of non-linearities introduced by those devices. These results indicate possible future challenges in large scale full-duplex radio production.

1.3.3 Full-Duplexing Networking: A Real-time MAC Implementation

Although wireless full-duplexing is a physical layer mechanism, its implications go beyond physical layer throughput. With new media access control (MAC) layer designs that support full-duplex, some of the most challenging problems in wireless networks can be mitigated, including hidden terminals, fairness in wireless LANs, and end-to-end delay in multihop networks. This dissertation presents ideas from different wireless domains, including wireless LANs, cellular networks, and multi-hop wireless networks to show how full-duplexing could be used to improve network performance.

As an attempt to understand some of the gains with full-duplexing, this dissertation presents the design and implementation of a real-time full-duplexing MAC layer. This design is largely based on a WiFi (802.11) like MAC layer, modified to use the full-duplexing physical layer. A 5-node testbed evaluation shows that the full-duplexing MAC can reduce hidden-terminal losses in WiFi networks by up to 88% and significantly improve the fairness between downstream and upstream flows in these networks. Full-duplexing in the testbed gives a combined network throughput of 11Mbps vs 7.5Mbps for half-duplex.

This dissertation only scratches the surface in looking at some implications that a full-duplexing physical layer would have on the performance of wireless networks. Understanding the overall impact of full-duplexing on different wireless networks is a rich area of future work.

1.4 Outline

This dissertation proposes a design for an adaptive single channel full-duplex wireless system. The design is applicable across multiple wireless domains and allows a rethinking of the way wireless networks are designed today. By properly exploiting the full-duplex capability, many problems with existing wireless networks can be mitigated. The rest of this dissertation digs deeper into the design and technical challenges of wireless full-duplexing, and discusses applications of wireless duplexing.

Chapter 2 explores the design space of cancellation schemes applicable for implementing self-interference cancellation. It discusses both analog and digital cancellation schemes, and their strengths and limitations. Chapter 3 looks at practical hardware concerns affecting the implementation of a full-duplex system. Based on the performance and practical constraints of different cancellation schemes, Chapter 4 describes the design of a full-duplex wireless system. This system combines an analog and a digital cancellation scheme to get an overall 73dB reduction in self-interference. This chapter also describes auto-tuning mechanisms for the design to adapt to changes in the wireless channel. Using the full-duplex physical layer system, Chapter 5 exemplifies some of the higher layer gains possible using a WiFi like MAC design modified for full-duplex operation. This real-time MAC implementation shows full-duplexing mitigating hidden terminal losses and improving fairness in a wireless LAN based setup. Chapter 6 further discusses other possible applications of full-duplexing in domains ranging from cellular networks to multi-hop access networks to secure networks for medical applications. Chapter 7 discusses challenges and directions for future research in making wireless full-duplex transition from a research prototype to a commercial product.

Chapter 2

Self-Interference Cancellation

Interference cancellation is a well researched topic in both wired and wireless systems. Wireline systems in telephone networks used echo-cancellation techniques, enabling voice duplexing on a single twisted pair connection [29]. Since then, cancellation techniques have been used in commercial systems and in research.

Techniques for interference cancellation in research can be categorized into digital and analog cancellation techniques. Digital cancellation operates on digital samples. If a full-duplex radio has a good estimate of the phase and amplitude of its transmitted signal at the receive antenna, it can generate the digital samples for its transmitted signal and subtract them from its received samples. Digital cancellation in current literature shows cancellation performance of up to 20-25dB, which is insufficient for a full-duplex system [34, 32]. Analog cancellation uses knowledge of the transmission to cancel self-interference in the RF signal, before it is digitized. One approach to analog cancellation uses a second transmit chain to create an analog cancellation signal from a digital estimate of the self-interference [28, 33], canceling ≈33dB of self-interference over narrow band signals.

Another approach taps the transmit signal to create an interference sample, modifies it and adds it to the receive signal to cancel interference. Some work suggests modifying the interference sample using phase-shifting techniques [21, 20], while other work uses techniques similar to noise-canceling headphones [52]. The self-interference signal is the "noise" which a circuit subtracts from the received signal. Many active

filter based techniques have also been suggested to cancel adjacent channel interference [60, 41]. On similar lines, various feed-forward architectures have been proposed to estimate and cancel adjacent channel interference using either analog estimation or analog to digital conversion and digital estimation of the interference [36, 11, 18]. Most such cancellation techniques in literature are targeted towards canceling either a narrow bandwidth, low power interference signal, or an out-of-band self-interference signal which may leak some power in the signal being received. This dissertation targets canceling a strong in-band self-interference signal with bandwidth as high as 40MHz.

Radio designs also often use duplexing circuits, such as circulators, to share the same antenna for transmit and receive [41, 21, 20, 47, 48]. The design presented in this dissertation does not preclude the use of duplexers for antenna sharing, these techniques are complementary. Finally, one design suggests using optical components for obtaining high bandwidth, high power interference cancellation [58]. While showing promising results, the integration of optical components in production wireless designs does not seem feasible.

Despite the relatively wide use of cancellation techniques, both for interference mitigation and for duplexing, single channel full-duplex operation does not exist in wireless data networks such as WiFi. The challenge in making single channel full-duplex work comes from two factors: the significant power difference between the self-interference and the desired receive signal and the variability of the wireless channel, requiring fast adaptation to a changing channel.

The self-interference can be millions to billions of times stronger (60-90 dB) than a received signal. For example, a radio with a transmit power of 20dBm and a noise floor of approximately -90dBm needs to cancel nearly 110dB of self-interference to ensure that its own transmissions do not disrupt reception. wireless channels also introduce other complexities such as frequency selective channel response and multipath fading. These variations can occur within the period of a few packet times, even for static WiFi networks. Self-interference cancellation schemes have to adapt at the same rate as the rate of channel changes.

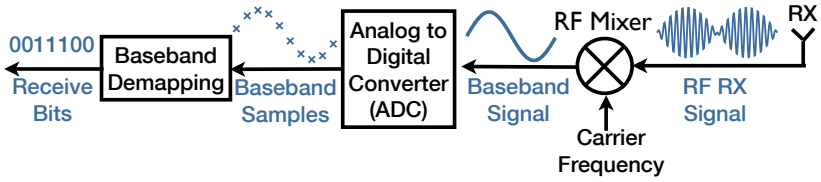


Figure 2.1: Simplified block diagram of an RF Receiver

This chapter presents three different techniques for analog self-interference cancellation: phase offset cancellation, vector modulation cancellation and signal inversion cancellation. It also presents the basics of digital cancellation. The chapter compares the various cancellation techniques through analysis and implementation using hardware components to identify theoretical and practical limits to the performance of the different techniques. It shows how phase offset cancellation and vector modulation cancellation are both limited in the signal power and bandwidth that they can handle and how signal inversion cancellation does not have the same constraints.

2.1 Radio Design

A wireless signal is processed through a series of stages in a wireless receiver and different cancellation techniques try to remove interference at different stages in the receive chain. This section provides background on the basics of radio design to facilitate the explanation of cancellation techniques in later sections.

Figure 2.1 shows the basic design of a modern radio receiver. We walk through these details because the underlying data representations determine how and when a full-duplex radio can cancel signals. We use channel 1 of 802.11b as a running example to ground the concepts in concrete numbers.

A wireless signal occupies a bandwidth, a range of frequencies. 802.11b channels, for example, are 22MHz wide. Channel 1 of 802.11b is centered at 2.412GHz: it spans 2.401GHz to 2.423GHz. The signal transmitted and received at this frequency range is called the *RF (Radio Frequency) signal*. Because digitally sampling a 2.4GHz signal would require very high speed sampling at the Nyquist frequency of 4.8GHz, radios downconvert a RF signal to a *baseband signal* centered around 0Hz. The baseband

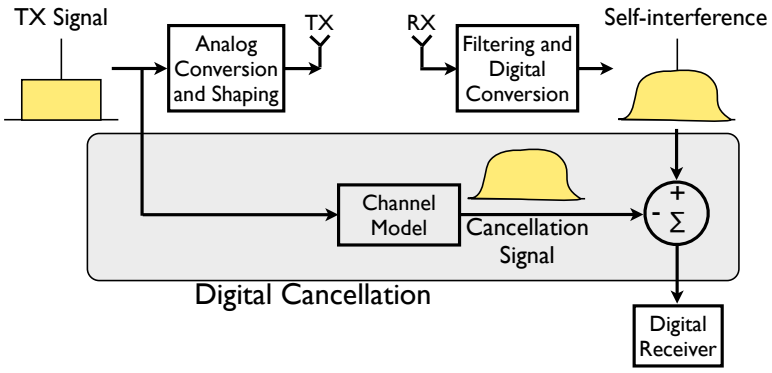


Figure 2.2: Basic block diagram for implementing digital cancellation. The transmitted digital samples are passed through a self-interference channel model to create a digital cancellation signal, which is subtracted from received digital samples.

signal of 802.11b channel 1 occupies -11 to 11 MHz.

The baseband signal is still an analog waveform, which is then converted to digital samples for further processing using an analog-to-digital converter (ADC). Downconverting allows the radio to use a much lower speed ADC: the 22MHz baseband signal needs an analog to digital converter (ADC) operating at or slightly above the Nyquist rate of 22 MHz. Commodity WiFi cards typically use 8-bit samples, though some software radios can provide 12 bit resolution.

To transmit a packet, a radio generates digital samples for the desired waveform, converts them to a baseband signal with a digital to analog converter (DAC) and upconverts the baseband to RF before transmitting.

Cancellation of self-interference can be implemented at different stages in the receiver chain. Specifically, cancellation may take place on the analog RF signal, before digitization of the signal through the ADC; or on digital samples after the ADC. Accordingly, there are various digital and analog techniques to cancel interference. The following sections go in some detail describing different cancellation schemes and their advantages or limitations.

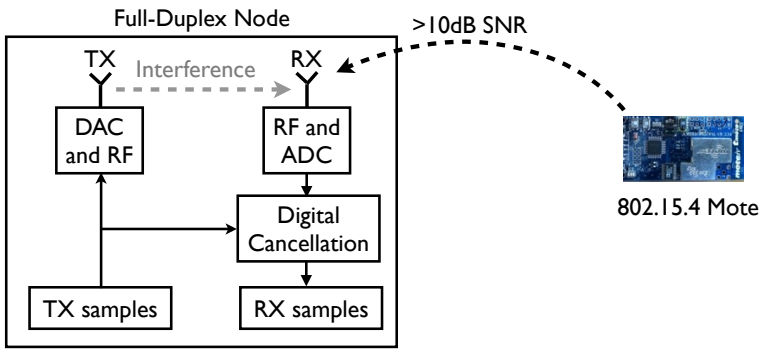


Figure 2.3: Setup for evaluating the efficacy of using only digital cancellation.

2.2 Digital Cancellation

Digital cancellation operates on digital samples. If a full-duplex radio has a good estimate of the phase and amplitude of its transmitted signal at the receive antenna, it can generate the digital samples for its transmitted signal and subtract them from its received samples. Figure 2.2 shows the basic digital cancellation operation.

Digital cancellation, while helpful, is insufficient: current systems in the research literature cancel up to 20-25 dB [34, 32]. The limitation is that ADCs have a limited dynamic range: since self-interference is extremely strong, an ADC can quantize away the received signal, making it unrecoverable after digital sampling.

A small experiment shows the inefficacy of using only interference cancellation on digital samples to implement a full-duplex node. The full-duplex node used for this test has a receive RF board trying to decode packets from a 802.15.4 transmitter placed a few meters away. The 802.15.4 node transmits packets at 0dBm power. The receiver has a perfect link with an SNR of >10dB to the 802.15.4 transmitter. A second RF board on the full-duplex node continuously transmits packets causing interference at the receiver. A simple digital cancellation technique is used to try and cancel the nodes self-interference. Figure 2.3 shows the test setup used for this experiment.

Figure 2.4 shows the resulting throughput for different transmit powers of the self-interference signal. Even with digital cancellation, the self-interference signal transmit power needs to be ∼36dB lower than the transmit power of the intended transmitter

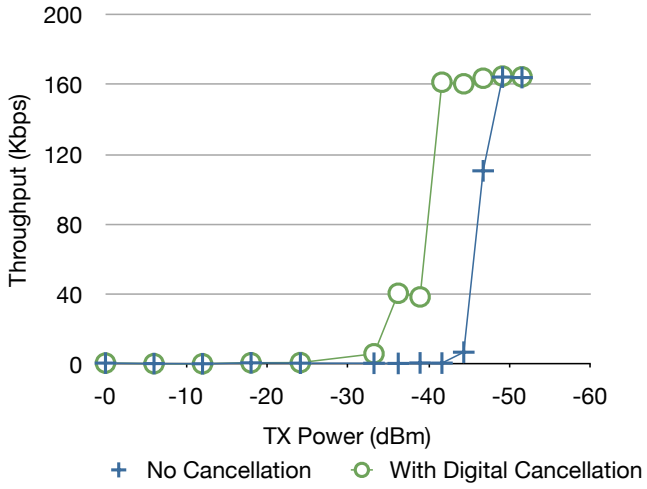


Figure 2.4: Receive throughput using digital interference cancellation with varying self-interference signal power. Digital interference cancellation gives an SNR gain of only about 10dB, while full-duplexing in this setup requires ∼46dB.

for the receiver to receive any intended packets. As a comparison, the figure also shows that the receiver can receive intended packets, without any digital cancellation, only if the transmit power of the (self-)interferer is ∼46dB lower than the intended transmitter. Thus, in this case, digital cancellation gives an SNR gain of 10dB. For a true full-duplex operation, we want the transmit powers of the intended and interfering transmitters to be equal. With better digital cancellation techniques, the 36dB gap can be reduced, but eliminating the gap completely with digital cancellation is unlikely. Current state-of-the-art digital cancellation schemes report up to 25dB cancellation [34], which would still leave a gap of 21dB between the power of the self-interferer and the intended transmitter. The gap becomes even more pronounced with higher power systems like 802.11.

This shows the limitation of using only digital interference cancellation techniques for achieving full-duplex. A node's transmit signal completely overwhelms its receive analog-to-digital converter (ADC) such that the digital samples do not retain any information of the weaker signal that a node is trying to receive.

2.3 Analog Cancellation Using Phase Offset

Analog cancellation techniques use knowledge of the transmission to cancel self-interference in the RF signal, before it is digitized. Analog cancellation can be implemented using a variety of techniques such as phase offset, vector modulation or signal inversion.

Phase offset cancellation uses the insight that sending the same signal from two or more paths results in constructive and destructive interference patterns where those signals add. For example, if the transmission signal from a node is split among two paths, both of which cause self-interference at the receiver, the second path can be offset from the first by an odd multiple of half the carrier wavelength. This causes the two signals to add destructively at the receiver, thus significantly reducing self-interference.

One way to implement phase offset cancellation is using multiple antennas with controlled placement for achieving the correct phase offset. We call this implementation *antenna cancellation*. We explore antenna cancellation in some depth to better understand the performance and limitations of phase-offset cancellation schemes

2.3.1 Antenna Cancellation Overview

Figure 2.5 shows a basic implementation of antenna cancellation. The transmission signal from a node is split across two transmit antennas. A separate receive antenna is placed such that its distance from the two transmit antennas differs by an odd multiple of half the wavelength of the center frequency of transmission.

For example, if the wavelength of transmission is λ, and the distance of the receive antenna is d from one transmit antenna, then the other transmit antenna is placed at $d + \lambda/2$ away from the receive antenna. This causes the signal from the two transmit antennas to add destructively, causing significant attenuation of the signal at the receive antenna.

Destructive interference is most effective when the signal amplitudes at the receiver from the two transmit antennas match. The input signal to the closer transmit antenna is attenuated to get the received amplitude to match the signal from the

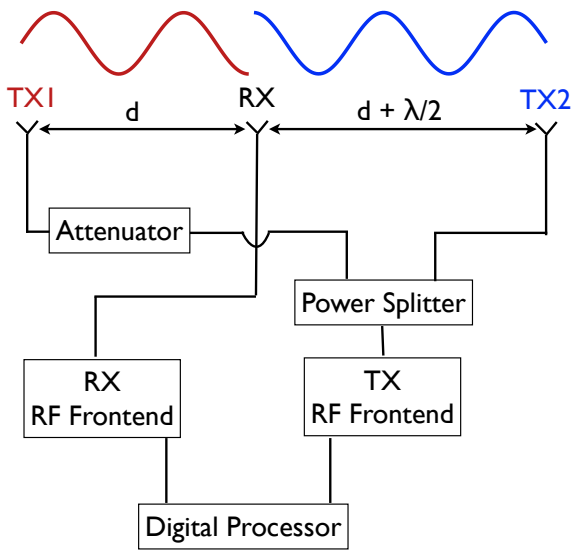


Figure 2.5: Basic block diagram showing phase offset cancellation implemented using antenna cancellation. Cancellation is achieved through the destructive addition of the signals coming from two transmit antennas at the receive antenna.

second transmit antenna, thus achieving better cancellation. A general implementation could use differently placed or more than three antennas to achieve better cancellation.

With the antenna cancellation scheme presented, antennas are optimally placed only for line-of-sight (LOS) components between the two transmit and one receive antennas. If such a node is placed in a corner, for example, the reflected signals from the transmit antennas will not necessarily cancel. While this puts a limitation on the performance of the antenna cancellation, signal strength of the reflected signals is typically much weaker than LOS due to longer signal path and attenuation when reflected. It is possible to bring this signal into the dynamic range of the ADC by using RF interference cancellation after the antenna cancellation stage.

2.3.2 Performance of Antenna Cancellation

In an ideal antenna cancellation scenario, the amplitudes from the two transmit antennas would be perfectly matched at the receiver and the phase of the two signals

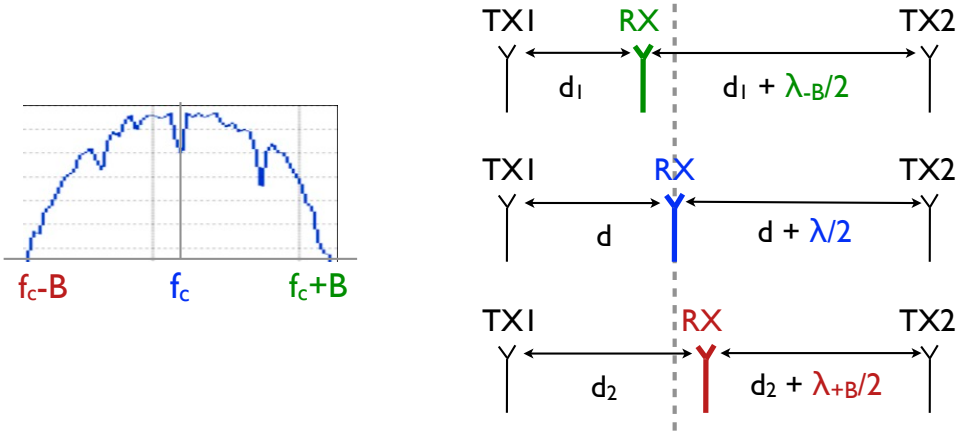


Figure 2.6: Effect of bandwidth on antenna cancellation accuracy. Even with perfect placement for the center frequency at distance d from antenna TX 1, there is a placement error for frequencies $f_c - B$ and $f_c + B$.

would differ by exactly π. However, we find that the bandwidth of the transmitted signal places a fundamental bound on the performance of antenna cancellation. Furthermore, real world systems are prone to engineering errors which limit system performance. The sensitivity of the antenna cancellation to amplitude mismatch at the receive antenna and to the error in receive antenna placement is an important consideration.

To analyze the reduction in interference using antenna cancellation, we look at the self-interference signal power at the receive antenna after antenna cancellation. The received self-interference power is derived in Appendix A.1 as:

$$2A_{ant}\left(A_{ant} + \epsilon^{A}_{ant}\right)|x[t]|^2\left(1 - cos\left(\frac{2\pi\epsilon^{d}_{ant}}{\lambda}\right)\right) + \left(\epsilon^{A}_{ant}\right)^2|x[t]|^2$$

where A_{ant} is the amplitude of the baseband signal, $x[t]$, at the receive antenna received from a single transmit antenna. ϵ^{A}_{ant} is the amplitude difference between the received signals from the two transmit antennas at the receive antenna. ϵ^{d}_{ant} represents the error in receiver antenna placement compared to the ideal case where the signals from the two antennas arrive π out of phase of each other. This equation lets us evaluate the sensitivity of antenna cancellation to receive antenna placement, change

of transmit frequency, and amplitude matching at the receive antenna.

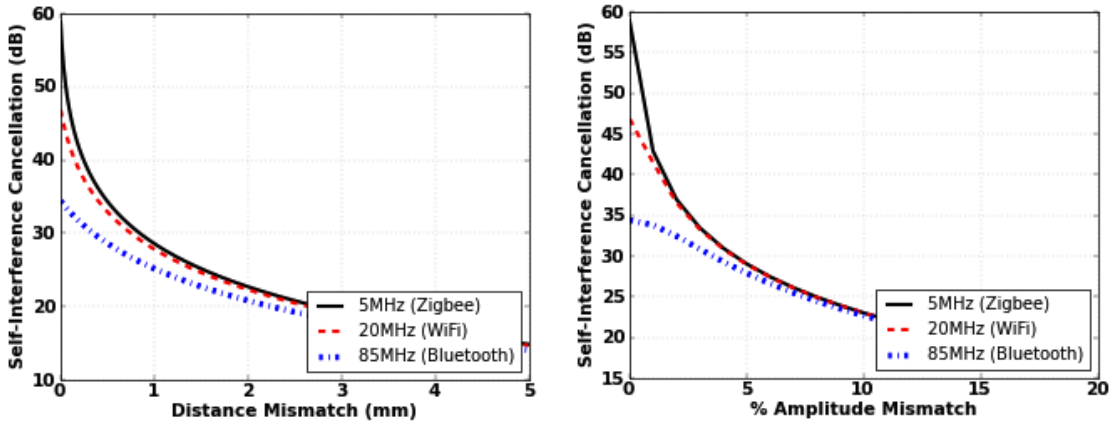
ϵ^d_{ant} also captures the effect of bandwidth on antenna cancellation. Consider a 5MHz signal centered at 2.48GHz. The signal has frequency components between 2.4775GHz and 2.4825GHz. If the receive antenna is placed perfectly for the center frequency, there is a small error in placement for the other frequencies within that bandwidth as Figure 2.6 shows. The "half-wavelength" offset required is different for different frequencies within the bandwidth.

We can map the difference in wavelength to the error in receiver placement. For example, a δ difference in wavelength is similar to a $\delta/4$ error in receiver placement. Thus, ϵ^d_{ant} for 2.4775GHz in this case would be $\sim \frac{1}{4}\left(\frac{c}{2.4775*10^6} - \frac{c}{2.48*10^6}\right)$, where c is the speed of light. This gives $\epsilon^d_{ant} \sim 0.025mm$, corresponding to 60.7dB antenna cancellation for the 2.48GHz center frequency. Thus, 60.7dB is the best antenna cancellation performance possible for a 5MHz signal in the 2.4GHz band using the three-antenna scheme described in this section. Similarly, using 20MHz and 85MHz bandwidths gives best case reduction of 46.9dB and 34.3dB respectively. The constraint on bandwidth for a given cancellation using antenna cancellation, or any other phase-offset technique, is a fundamental one. Thus, with antenna cancellation it would be impossible to implement full-duplexing for very wideband systems.

As can be seen from the effect of bandwidth, antenna cancellation does not provide a frequency flat channel at the receiver even if there is perfect amplitude matching. This distortion in the received signal can be a problem for the decoder of the received signal. It can also complicate the incorporation of other cancellation schemes after antenna cancellation.

Any error in receive antenna placement adds to ϵ^d_{ant}. To see the effect of receive antenna placement error, suppose the receive antenna is 1mm off from the optimal position, i.e. $\epsilon^d_{ant} = 1mm$. With perfect amplitude matching and with a λ of 12.1cm (for a center frequency of 2.48GHz), we see a 28.7dB reduction in power compared to no antenna cancellation. Figure 2.7(a) shows the theoretical performance of antenna cancellation with error in receiver placement, for different bandwidths.

Figure 2.7(b) shows the theoretical performance of antenna cancellation with error in amplitude matching, assuming perfect center frequency receiver placement, for



(a) RX power with distance mismatch

(b) RX power with amplitude mismatch

Figure 2.7: Performance of antenna cancellation with distance and amplitude mismatch for signals with different bandwidth. A 1mm mismatch can restrict the receive power reduction to ∼29dB. An amplitude mismatch of 10%, corresponding to 1dB variation, can restrict the receive power reduction to ∼20dB.

different bandwidths. For example, say the amplitude of one signal is 10% higher than the other, i.e. $\epsilon^A_{ant} = 0.1 * A_{ant}$. In this case, the powers of the two signals differ by $\sim 1dB$. With this ϵ^A_{ant}, the reduction in received power due to antenna cancellation is 23dB, if we ignore the effect of bandwidth. For a 5MHz bandwidth, the same ϵ^A_{ant} gives a 22.994dB reduction. Thus, a small amplitude mismatch tends to dominate the performance restrictions on antenna cancellation.

2.3.3 Antenna Cancellation in Practice

Figure 2.8 shows the effect of antenna cancellation with transmitter TX1 attenuated by 6dB compared to TX2. Experiments show that the received power from the two TX antennas differs by about 5.1dB when the receiver is placed at the null point. Thus, this setup has an amplitude mismatch of ∼1dB causing the cancellation to be restricted to ∼20dB as shown in the previous analysis. The above analysis did not consider the multipath effect. However, results from the measurements show that the multipath effect is not a dominant component in this experimental setup.

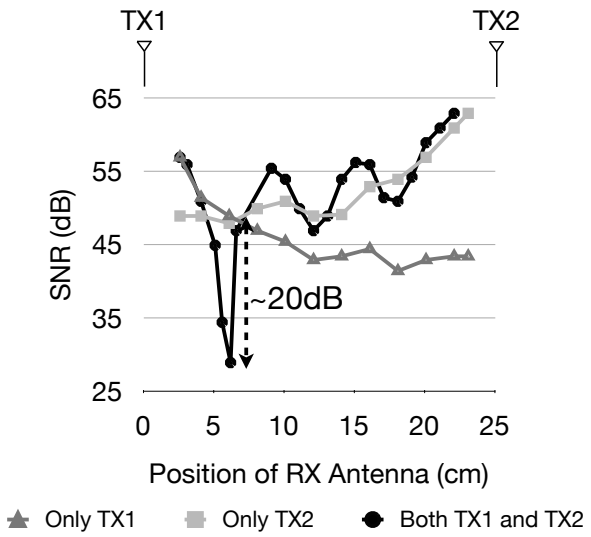
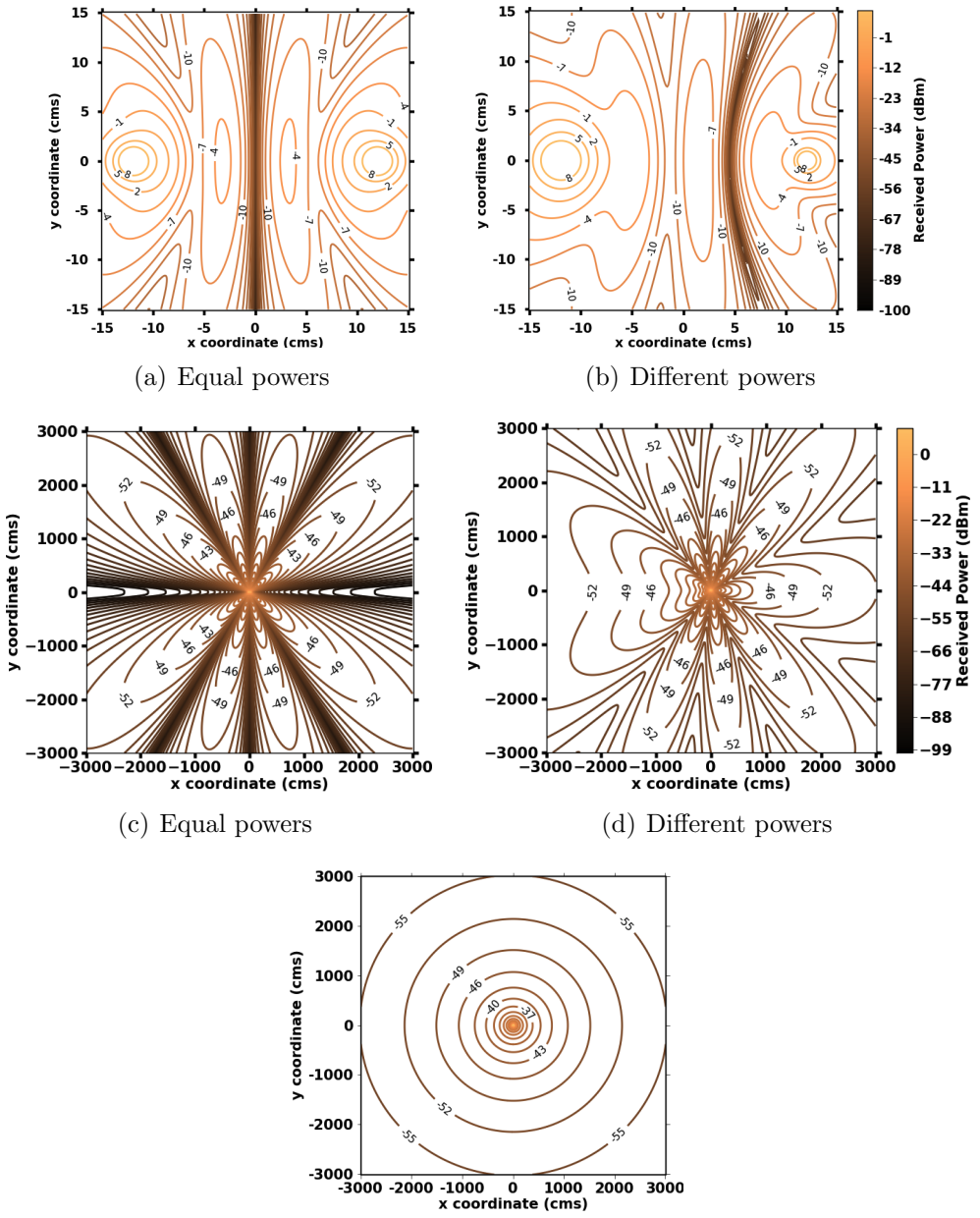


Figure 2.8: Received SNR for different receive antenna placements. The received SNR is fairly monotonic with distance when any one transmit antenna is active. With both transmit antennas active, there is a sharp reduction in receive power at the null point.

2.3.4 Effect of Antenna Cancellation on Intended Receivers

While antenna cancellation can reduce self-interference from a node's own transmitter, an important question is how this affects the received signal at nodes other than the transmitter. Another question is how does introducing the phase offset on air compare to having the phase offset introduced in the wires leading to the two transmit antennas by differing the two transmit antenna wire lengths by $\lambda/2$. Unlike the on-air phase offset, the wired phase offset approach does not require an attenuator and gives a null point exactly at the center of the two transmit antennas.

The contour map in Figure 2.9(a) shows received power at different points in space with both transmit antennas transmitting a single frequency tone at the same power with a phase difference of π using a simple simulation with a freespace propagation model. Each contour line corresponds to a specific received power. Figure 2.9(b) shows the received signal strength with different transmit powers from the transmit antennas such that amplitudes match at the null point without any phase shift in antenna signals. The null points achieved in the two cases are at different locations, but both schemes are equally good in terms of signal reduction at the null point.



(a) Equal powers

(b) Different powers

(c) Equal powers

(d) Different powers

(e) Single transmitter

Figure 2.9: Freespace signal strength profiles for equal transmit powers and different transmit powers on two transmit antennas. This simulation uses a pathloss exponent of 2. Figures (a) and (b) correspond to a short-range study. When transmit powers are equal, the minimum received signal is in the middle and when the transmit powers are different, the minimum is closer to the lower transmit power antenna. Figures (c), (d) and (e) correspond to a long-range study. When transmit powers are equal, receivers equidistant from the transmit antenna pair can see huge differences in the received signal strength. When transmit powers are different, however, such differences are much smaller.

The difference between these two cases becomes clearer by looking at the received signal at larger distances. Figure 2.9(e) shows the received signal strength profile, over space, for a single transmit antenna over a distance of 30m from the transmitter. This is the baseline for comparison of the two schemes with antenna cancellation. Figure 2.9(c) shows the contours over larger distances for the same setup as Figure 2.9(a). It is apparent that even in normal communication range, there are locations with very low received power due to the destructive interference.

Figure 2.9(d) shows the contours of received power when one transmit signal is attenuated by 6dB compared to the other and there is no phase shift between the two transmitted signals. The effect of destructive interference is much lower in this case.

With two transmit antennas, the signals from the two antennas get added constructively or destructively at the receiver. At distances much larger than the spacing between the transmit antennas, the signals from both antennas undergo almost equal attenuation. With equal receive power from both antennas, a perfectly destructive combining of the two signals causes the received signal to be zero power. In case of unequal transmit powers, the received power at these distances is different from the two transmit antennas. Even when the signals combine perfectly out of phase, the resulting signal is not zero power. Comparing with the single antenna case, using the antenna cancellation scheme leads to a maximum degradation of 6dB at any receiver location. In a real network setting, diversity gains due to two transmit antennas may offset this degradation.

Antenna cancellation can give significant reduction in self-interference, which shows the promise of using the phase-offset cancellation scheme. That said, the bandwidth constraint in phase-offset cancellation shows that this type of cancellation can only be effective when used for very narrowband (<5MHz) signals, while many current wireless systems are going towards larger bandwidths. Further, the destructive interference used in antenna cancellation may also adversely affect receptions at intended receivers, though this effect may be less pronounced in an indoor setting.

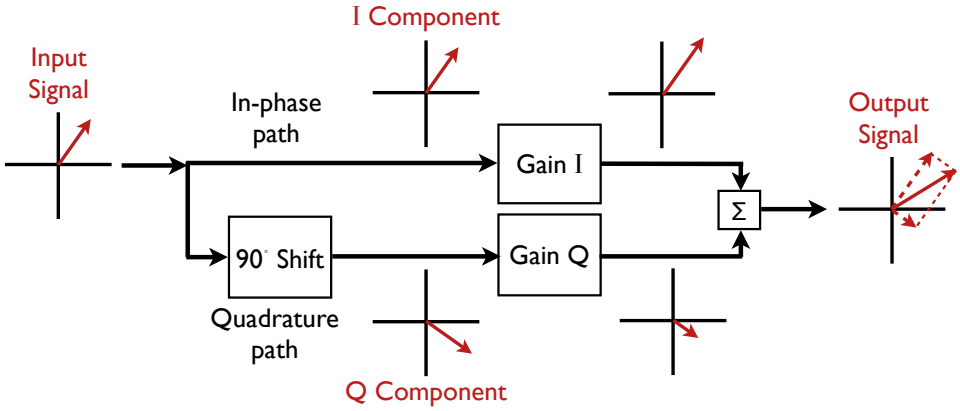


Figure 2.10: Basic block diagram of a linear vector modulator. Any scaling and phase shift can be applied to the input signal by appropriately adjusting the in-phase (I) and quadrature (Q) gains. The 90° shift is typically implemented by delaying the signal by a quarter wavelength ($\lambda/4$ delay).

2.3.5 Phase-offset Cancellation in Other Forms

Phase-offset cancellation can also be implemented using specialized hardware such as hybrid rings [61]. Hybrid rings are 4-port devices which provide coupling and isolation between its ports by exploiting constructive and destructive interference between two paths (clockwise and counterclockwise) taken by all waves entering any port. This allows implementation of a single antenna full-duplex solution with hybrid rings. Explaining the operation of the hybrid ring is beyond the scope of this work, but it suffices to say that hybrid rings also suffer from the same bandwidth limitation as antenna cancellation due to their wavelength specific design.

2.4 Analog Cancellation using Vector Modulation

A vector modulation allows a modulator to control the angle and amplitude of an input signal. A vector modulator can scale and rotate an input signal by a desired value. Figure 2.10 shows the basic block diagram of a vector modulator. An input signal is split into in-phase (I) and quadrature phase (Q) components, with the Q

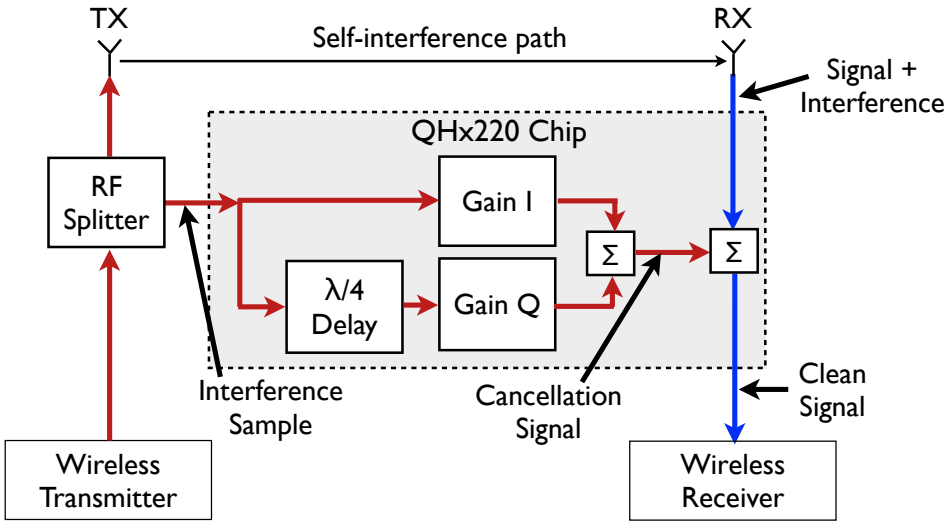


Figure 2.11: Block diagram of the QHx220 chip using vector modulation for removing interference from RF signals. The chip can be used to self-interference cancellation for full-duplexing by feeding a sample of the transmit signal as the interference sample.

component 90° out of phase of the I component. The I and Q components are individually scaled using variable amplifiers, and then combined to achieve any angular and scaling shift in the input signal. The 90° phase shift is typically implemented by delaying the input signal by a quarter wavelength ($\lambda/4$ delay).

Some existing noise cancellation chips, such as the QHx220, use vector modulation to create a cancellation signal for removing interference or noise from a received signal [51]. Figure 2.11 shows a block diagram of the QHx220 chip. The chip operates on the RF signal and takes two inputs, an input signal and an interference sample. The input signal typically contains a desired signal with some form of ambient interference. The interference sample input tries to independently measure only the ambient interference. The QHx220 chip feeds the interference sample as an input to a vector modulator to adjust the phase and gain of the sample, creating a cancellation signal that can remove the ambient interference from the input signal. It then outputs the sum of the input and the cancellation signal.

As Figure 2.11 shows, the QHx220 chip can be used to implement self-interference cancellation. The transmit signal is sampled through an RF signal splitter and fed to

the interference sample input. The signal at the receive antenna is fed to the signal input of the chip. The I and Q gains are adjusted to reflect the on-air attenuation and phase change of the transmit signal, creating a cancellation signal that matches the self-interference at the receive antenna.

Cancellation based on vector modulation is limited by the frequency dependence of the $\lambda/4$ delay. The quadrature component has a fixed delay with respect to the in-phase component. For a single frequency, this approach can correctly emulate a 90° phase shift. However, for signals with bandwidth, the fixed delay only matches a $\lambda/4$ delay for one frequency. Thus, this technique suffers from a similar bandwidth constraint as antenna cancellation. Prior work also discusses the limitation of the cancellation model used by QHx220 [42]. Based on expressions derived in this work, we can estimate the best cancellation achievable with vector modulation to be ≈28dB for a 100MHz signal and ≈36dB for a 40MHz signal.

The performance of cancellation using QHx220 also depends on the linearity of the active amplifiers used in the vector modulator. Just as saturation of the receive amplifiers makes cancellation less effective when it is implemented purely in the digital samples, saturation in the vector modulator amplifiers can also limit self-interference cancellation performance. This effect is explored in more detail in Section 3.2.

2.5 Analog Cancellation using Signal Inversion

The motivation of using signal inversion for self-interference cancellation comes from a simple observation: any radio that creates a cancellation signal through adjusting phase will always encounter a bandwidth constraint that bounds its maximum cancellation. This limits the performance of both phase offset cancellation and cancellation using vector modulation. To cancel beyond this bound, a radio needs to obtain the perfect *inverse* of a signal, that is, a signal which is the perfect negative of the transmitted signal at all instants. Combining this inverse with the transmitted signal can, in theory, completely cancel self-interference.

All a radio needs is to invert a signal without adjusting its phase. Luckily, there is a component that does exactly that: a balanced/unbalanced (balun) transformer.

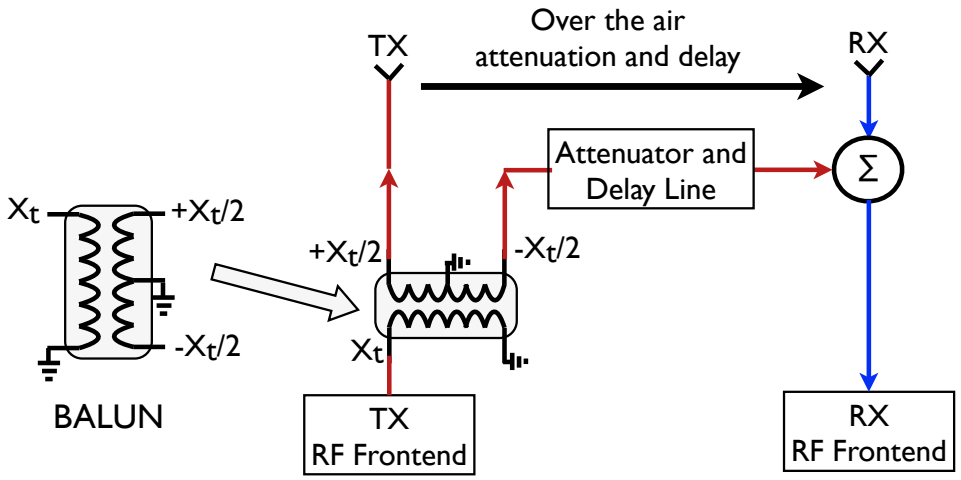
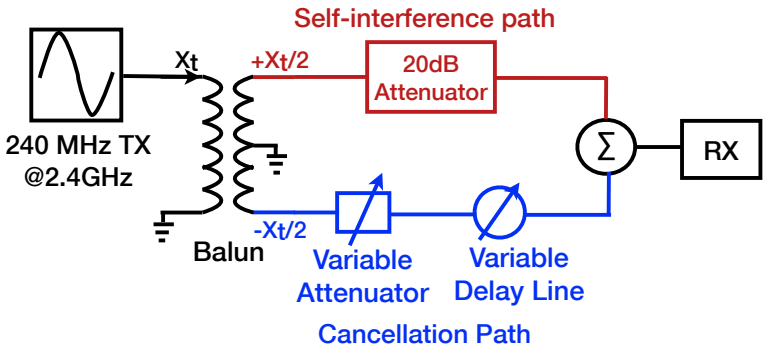


Figure 2.12: Block diagram of self-interference cancellation using signal inversion. The inverse of the self-interference signal is generated using a Balun. An adjustable attenuator and delay is needed to compensate for on-air delay and attenuation.

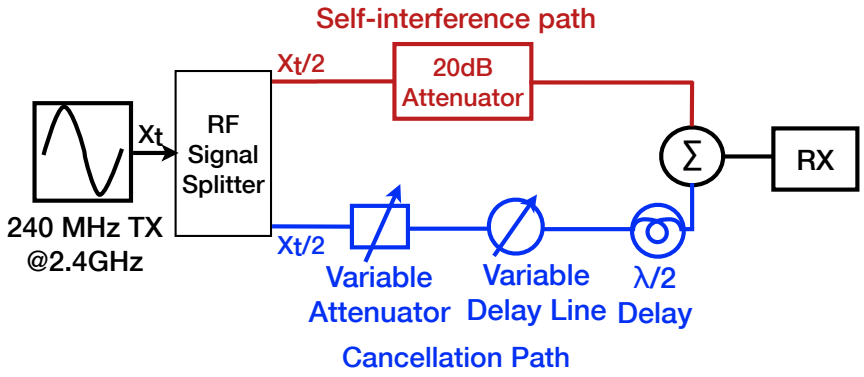
Baluns are a common component in RF, audio and video circuits for converting back and forth between single-ended signals – single-wire signals with a common ground – and differential signals – two-wire signals with opposite polarity. For example, converting a single-ended signal on a co-axial cable to a differential signal for transmission on a twisted pair cable (such as Ethernet), or vice-versa, uses a balun to take the signal as input and output the signal and its inverse.

Baluns can be used to obtain the inverse of a self-interference signal and the inverted signal can then be used to cancel the interference. Figure 2.12 shows a 2-antenna design that uses a balun to cancel self-interference. The transmit antenna transmits the positive signal. The negative signal goes over wire to generate an interference cancellation signal. A passive variable delay and attenuator is used to match the cancellation signal to the self-interference at the receive antenna. The receiver then combines the received signal with the cancellation signal to significantly reduce the residual self-interference.

This technique ideally uses high precision passive components to realize the variable attenuation and delay in the cancellation path. While signal inversion cancellation can theoretically cancel perfectly, there are practical limitations. For example,



(a) Signal inversion based cancellation test setup



(b) Phase offset based cancellation test setup

Figure 2.13: Wired setup to measure the cancellation performance of signal inversion vs phase offset. The phase offset experiment uses an RF splitter instead of a balun to split the signal.

the transmitted signal on the air experiences attenuation and delay. To obtain perfect cancellation the radio must apply identical attenuation and delay to the inverted signal before combining it, which may be hard to achieve in practice. Moreover, the balun may have engineering imperfections, such as leakage or a non-flat frequency response.

2.5.1 Canceling Larger Bandwidths with Signal Inversion

Generating a cancellation signal using signal inversion and passive components does not have any theoretical constraints on the maximum cancellation achievable for a

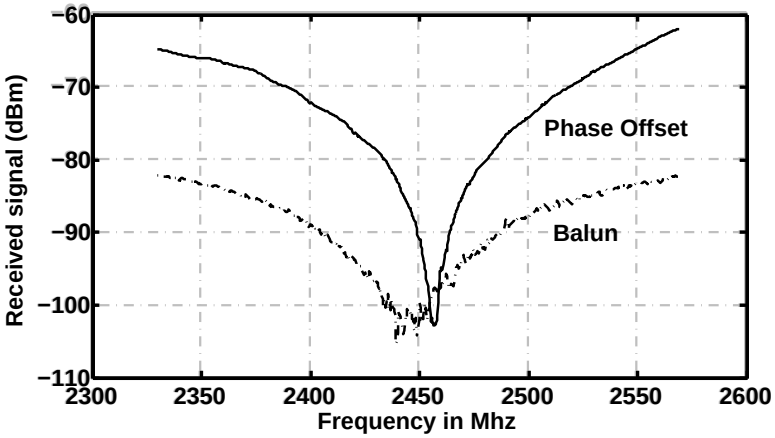


Figure 2.14: Cancellation of the self-interference signal with the balun vs with phase offset. The received signal is -49dBm without any cancellation. Using a balun gives a flatter cancellation response.

signal with a given bandwidth or given power level. With this technique, we should be able to cancel signals with wider bandwidths much better than if the cancellation signal was generated using phase offset.

To understand the practical benefits and limitations of inverting a signal with a balun, compared to using phase offset for generating a cancellation signal, we conduct a tightly controlled RF experiment. Figure 2.13 shows the two experimental setups. We program a signal generator to generate a wideband 240MHz chirp with a center frequency of 2.45GHz. This signal goes over two wires. The first wire is an ideal self-interference path and has a 20dB attenuator representing the over-the-air loss from the transmit to the receive antenna. The second wire goes through a cancellation path, consisting of a variable attenuator and variable delay element that can be controlled to modify the cancellation path signal to match the self-interference. The combination of the two signals feeds into a signal receiver. The variable delay line and attenuator in this experiment are manually tunable passive devices which allow for a high degree of precision in tuning.

We implement phase offset cancellation using an RF splitter to split the transmit signal into the self-interference and cancellation signals; and by making the cancellation path one half of a wavelength longer than the self-interference path. An RF

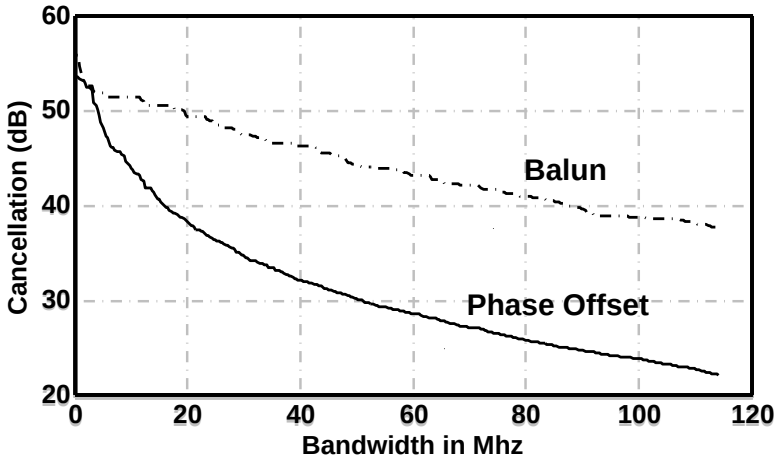


Figure 2.15: Cancellation performance with increasing signal bandwidth when using the balun method vs using phase offset cancellation.

combiner adds the two signals on the received side to measure the canceled signal. The balun setup, on the other hand, uses a balun to split the transmit signal, and uses wires of the same length for the self-interference and cancellation paths. In both cases, the passive delay line and attenuator provide fine-grained control to match phase and amplitude for the interference and cancellation paths to maximize cancellation.

Figure 2.14 compares the frequency response of the residual self-interference at the receiver using signal inversion with a balun versus using phase-offset cancellation. Using a phase-offset signal cancels well over a narrow bandwidth, but is very limited in canceling wideband signals. Phase offset cancellation can cancel 50dB for a 5MHz signal, but only provide 25dB of cancellation for a 100MHz signal. In comparison, signal inversion through the balun provides a good degree of cancellation over a much wider bandwidth. For example, balun based cancellation would provide 52dB of cancellation for a 5MHz signal and 40dB of cancellation for a 100MHz signal.

Balun cancellation is not perfect across the entire band. The key reason is that the balun circuit is not frequency flat, i.e., different parts of the band are inverted with different amplitudes. Consequently applying a single attenuation and delay factor to the inverted signal will not cancel the transmitted signal perfectly: this is a simple instance of real-world engineering tolerances limiting theory. Based on Figure 2.14, we can obtain the best possible cancellation with balun and phase-offset cancellation

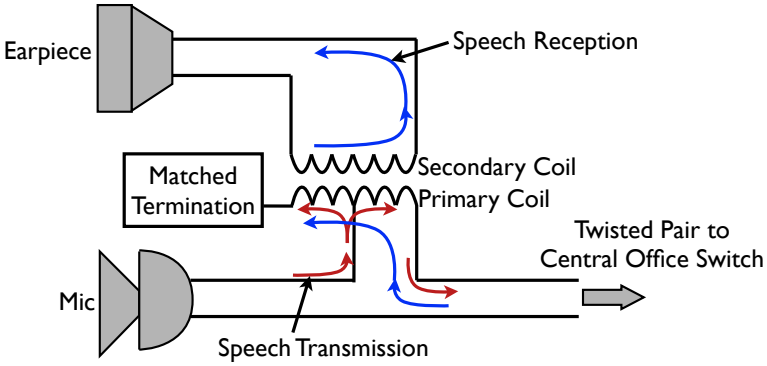


Figure 2.16: A telephone instrument uses a hybrid coil to duplex its speech transmission and reception on a single twisted pair connection to the central office.

for a given signal bandwidth. Figure 2.15 shows the best cancellation achieved using each method for signals of varying bandwidths.

2.5.2 Signal Inversion Cancellation in Telephones: Hybrid Coils

The idea of using transformers to create the inverse of a signal, thereby canceling the self-interference, has been used in telephone lines since the 1960s [29]. Telephone lines operate using a twisted wire pair to connect a telephone with the central switching station. The single twisted pair carries both the speech signal from the mic of the telephone to the central office switch and the speech signal from the central office to the earpiece of the telephone. Since the same twisted pair wire carries speech data in both directions, the telephone instrument needs to cancel or isolate its speech transmission at its mic from the speech reception at its earpiece, a problem identical to the wireless full-duplex problem in concept. Telephone instruments use transformer structures called hybrid coils that allow this full-duplexing operation.

Figure 2.16 shows the connection of a phone to the twisted pair going to the central office switch through a hybrid coil. The splitting of the transmit signal across two branches in the primary coil of the transformer creates opposing fields on the secondary, which cancel out isolating the receiver (earpiece) from the transmitter

(mic). The wired self-interference channel between the mic and the earpiece is fairly static: a well tuned hybrid coil can give sufficient self-interference cancellation in this scenario, without requiring adjustments. In contrast, the self-interference channel in wireless devices tends to be much more dynamic, requiring tuning components in the cancellation path to adjust to channel changes. The self-interference cancellation design presented in this dissertation uses real-time tuning algorithms to adapt quickly to such changes.

2.6 Summary

This chapter explored the design space of cancellation techniques for removing self-interference from an RF signal. It outlined four main methods of self-interference cancellation:

- Digital Cancellation
- Analog Cancellation using Phase-Offset
- Analog Cancellation using Vector Modulation
- Analog Cancellation using Signal Inversion

Further, this chapter presented some performance results to compare the different cancellation mechanisms and discuss their relative merits and constraints. Digital cancellation is effective in capturing many channel effects including multipath, but is constrained by the fact that the self-interference may saturate some of the receive circuitry, making cancellation less effective. Both phase-offset and vector modulation cancellation can present a bandwidth constraint on the maximum bandwidth signal that can be canceled based on the amount of cancellation required. Vector modulation typically also needs to use active components, which are susceptible to saturation. Signal inversion solves the bandwidth concern in theory, but its performance is still limited by the accuracy and the flatness of frequency response of the passive components used. None of the analog cancellation techniques presented here can cancel all the self-interference if any multipath components exist, thus emphasizing the importance of digital cancellation.

Chapter 3

Hardware Concerns

Chapter 2 discussed the basic techniques used for cancellation of interference, including self-interference. The chapter also looked in detail into the theoretical and practical performance limits of cancellation techniques. This chapter discusses how hardware constraints affect the performance of cancellation techniques.

Interference cancellation poses two hardware challenges. The first challenge comes from the requirements of range and resolution of parameter values for any interference cancellation mechanism. The second challenge is preventing saturation in the cancellation circuitry from the high powered self-interference signal.

3.1 Resolution and Range of Components

Each type of cancellation typically involves setting some parameters in the cancellation mechanism. In case of signal inversion cancellation as described in Section 2.5, the parameters are the delay and attenuation values of the passive components used in the cancellation path. Similarly, in case of phase-offset cancellation implemented using the antenna cancellation mechanism, the parameters include the exact location of the receive antenna and the value of the attenuator placed in one of the transmit antenna paths. For making any cancellation scheme practical, the hardware implementation should provide flexibility in setting the parameters associated with the specific cancellation mechanism.

The range and resolution for setting cancellation parameters is an important and challenging requirement for hardware used to implement cancellation. We use the example of signal inversion cancellation to illustrate this point.

3.1.1 Resolution

The adjustable parameters for signal inversion cancellation are the values of the delay line and of the attenuation in the cancellation signal path. Ideally, the two values have to exactly match the delay and attenuation of the over-the-air interference signal. A minimum resolution in the hardware means that we cannot always exactly match the delay and attenuation, but rather achieve the closest values possible with the given hardware. This introduces small errors in matching the delay and attenuation of the cancellation signal, resulting in imperfect cancellation. The analysis of cancellation with these errors follows very easily from the analysis of cancellation using antenna cancellation with imperfect antenna placement or amplitude matching discussed in Section 2.3.2.

In signal inversion cancellation, if the hardware used has a minimum attenuation resolution of att_{res} in dB, the error in amplitude of the cancellation signal can be up to $10^{att_{res}/2*1/20} = 10^{att_{res}/40}$ times the amplitude of the self-interference signal in linear scale. Similarly, if the resolution of the delay element is del_{res}, the maximum error in delay matching would be $del_{res}/2$. Given these errors, the self-interference power at the receiver after cancellation would be:

$$A_{si}^2|x[t]|^2 \left(1 + \left(10^{att_{res}/40}\right)^2 - 2 * \left(10^{att_{res}/40}\right) * cos\left(2\pi f_c del_{res}/2\right)\right),$$

where A_{si} is the amplitude of the baseband self-interference, $x[t]$, at the receive antenna.

Based on this equation, we can evaluate the sensitivity of cancellation performance to the resolution of delay and attenuation. Figure 3.1 shows the best possible cancellation for hardware with different resolution. The plot uses four different resolutions for the variable attenuator, 0.0dB, 0.01dB, 0.1dB and 1dB, and shows the cancellation performance in each case for varying resolution of the delay line. It is

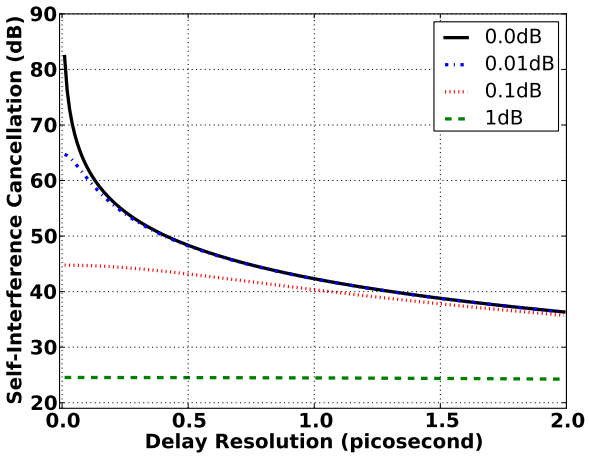


Figure 3.1: Effect of hardware resolution on cancellation performance. The four plots correspond to different resolutions of the variable attenuator and show the amount of cancellation for each with varying resolution of the delay line. Good resolution is needed in both attenuation and delay for achieving good cancellation performance.

interesting to note that for different combinations of hardware resolution, the resolution of one parameter may restrict the overall cancellation performance, such that there is little use in improving the resolution of the other parameter. For example, if the attenuator has a resolution of 0.1dB we get a cancellation performance of about 40dB for a delay line resolution of 1ps. Improving the delay line resolution to 0.5ps in this case only marginally improves cancellation performance to 42dB. On the other hand, if the same resolution improvement was possible for the delay line, with the attenuator having a resolution of 0.01dB, the performance would improve significantly from 42dB to 49dB. Resolution improvements in attenuation and delay lines should go hand-in-hand to get the maximum benefit out of those improvements. Recent work in developing new structures for implementing on-chip variable delay lines and attenuators provide promising results [46, 19, 23, 62, 38, 26].

In our experience, a resolution of 0.1dB attenuation and 0.3 picosecond delay can feasibly be achieved in today's hardware, thus giving a maximum cancellation of $\sim$45dB. The same absolute resolution in delay gives different performance for different carrier frequencies. Higher carrier frequencies require finer delay line resolution for getting the same cancellation performance, and vice-versa.

3.1.2 Range

Range of hardware is another concern for implementing cancellation schemes. This is especially true for applications that require the design to fit in a small area on a circuit board, or even within a chip. As an example, a delay line that can incorporate a whole wavelength's delay would have to be as much as 12cm long for the 2.4GHz band. Such lengths are not achievable on a chip and are hard to put in a small area on a circuit board. One option is to implement larger delays using variable capacitor elements, which can now be put on chips.

Another option is to carefully engineer the duplex system and operate it in an environment where the delay variation is not significant. In this case, we do not need the variable delay to cover a whole wavelength's delay, rather it needs to cover a small fraction of the wavelength. Our experimental setup works well with a delay line that covers about a 15th of a wavelength for the 2.4GHz band. The same may not be true for practical systems that are subject to many extreme changes in the self-interference channel, thus requiring a wider delay adjustment range.

The attenuation of the self-interference signal tends to be less fickle than delay; achieving a reasonable range in the variable attenuator is easier than achieving enough range in the delay line. Typical variable attenuators can accommodate up to 8 bits of range with a resolution of 0.1dB, thus giving an attenuation range of 0 to 25dB. Depending on the average over-the-air loss of the self-interference signal, the cancellation signal may be attenuated by a fixed amount before passing it through the variable attenuator.

For example, if the over-the-air attenuation varies between 15dB and 25dB, which is a fairly typical range, we can attach a 15dB fixed attenuator in the cancellation path before passing the signal through the variable attenuator. This would give a cancellation signal attenuation range from 15-40dB, covering the range of attenuation for the self-interference signal.

3.2 Non-linearity in Hardware

Typical digital systems are designed assuming that the digital signal from the transmitter undergoes only linear transformations in reaching the digital receiver. In a non-linear system, different frequencies are not independent, i.e. a non-zero value on one frequency at the input of the system causes a non-zero value at the output of the system on several frequencies. This breaks the operating assumption for both the frequency based channel estimation and frequency based signaling used by modern OFDM systems. Typical radio systems have non-linearities, but they are weak enough such that they can be treated as noise.

The main causes of non-linearities are responses of transmitter amplifiers and saturation in receiver circuits. Saturation in receiver circuits leads to effects like clipping which can severely distort the received signal's frequency response. Wireless receivers use an Automatic Gain Control (AGC) circuit to prevent clipping, while maintaining a reasonable dynamic range at the same time. That said, if the input signal is very strong, even the AGC may not be able to prevent clipping.

The main source of non-linearities in both the transmitter and receiver circuits tend to be the portions with active components like amplifiers, and transistor based filters. Passive components have a higher dynamic range and can handle much higher input powers before giving a non-linear response. Implementing certain analog cancellation schemes can also introduce non-linearities in the signal response. For example, using cancellation with vector modulation, as shown in Section 2.4, could involve using a QHx220 noise cancellation chip, that uses active amplifiers for setting gains for both I and Q signal components. The high power associated with the self-interference and cancellation signals causes the amplifiers in QHx220 to saturate thus giving a very non-linear output.

Non-linearities also complicate digital cancellation. As mentioned in Section 2.2, digital cancellation involves creating a channel model and passing digital samples through that model to create a cancellation signal that is subtracted from the received samples. Typically, this channel model is created assuming a linear channel. But strong non-linearities violate this assumption and stop digital cancellation from being

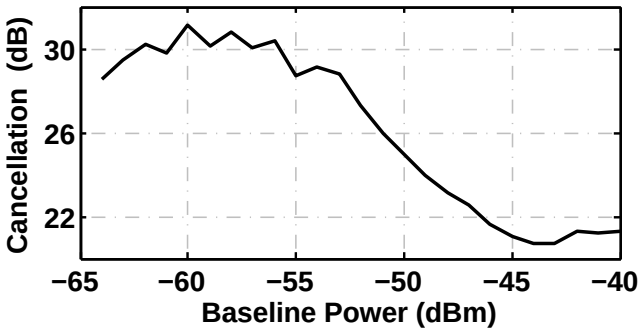
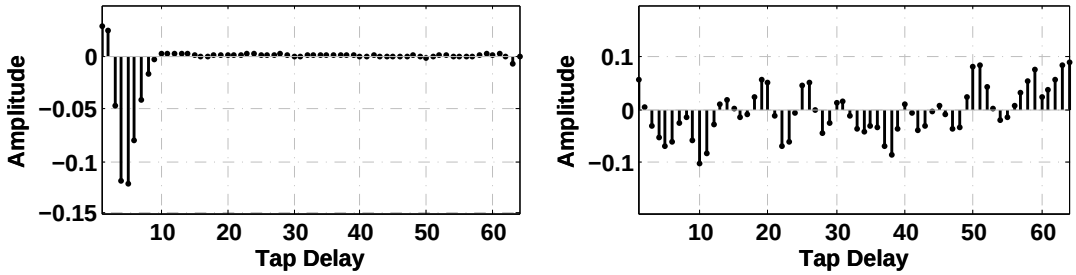


Figure 3.2: Performance of cancellation using the active QHx220 chip with increasing received power. The QHx220 limits the cancellation to 30 dB at lower powers, and 20 dB at higher powers indicating the effect of saturation and non-linearity on cancellation performance.



(a) Passive Components (b) Active Components

Figure 3.3: Real part of digital channel estimates with signal inversion cancellation using passive components and vector modulation cancellation using the active QHx220 chip. The active components in the QHx220 chip introduce non-linearities leading to invalid estimation.

effective.

An experiment shows the effect of non-linearities caused by QHx220. The experimental setup combines vector modulation cancellation using QHx220 described in Section 2.4 with a digital cancellation scheme. The digital cancellation scheme models the self-interference channel as a linear time-invariant system. If there are strong non-linearities in the channel, they would lead to inaccurate channel estimates and digital cancellation performance would suffer.

Figure 3.2 shows the cancellation achieved with the combination of vector modulation cancellation and digital cancellation as the transmit power of the self-interfering

signal varies. At an input power of -60 dBm, the circuit can cancel approximately 30 dB. As input power increases, however, cancellation deteriorates due to saturation in QHx220's active components. QHx220 cannot handle a high input power, and beyond a threshold clips and introduces non-linearities. These non-linearities reduce the efficacy of digital cancellation.

Figure 3.3(a) and 3.3(b) show the issues the non-linearities introduce in greater detail. They show the channel impulse response estimate after cancellation using signal inversion with passive components as described in Section 2.5 compared to the impulse response for vector modulation cancellation using QHx220 which contains active components. When using passive circuits, the channel estimate contains only a few strong taps with small delays. However, with the active QHx220 circuit, the channel estimate has a non-trivial value until the 64th tap, which represents a 1 km long strong multipath component. Such multipath cannot exist: it is an outcome of linear estimation of a non-linear channel.

Maintaining good linearity of response while implementing analog cancellation schemes is an important consideration both for implementing digital cancellation on top of analog cancellation and ensuring good performance of the digital receiver since OFDM receivers require a linear channel response to properly decode a received signal.

3.3 Summary

Although different cancellation schemes have different benefits and drawbacks from a conceptual point of view, as discussed in Chapter 2, their are many practical concerns that also come into play when trying to design a feasible full-duplex system. This chapter discussed some of the hardware considerations for implementing these systems. The resolution of the devices used for implementing the cancellation scheme dictates how much cancellation can possibly be achieved with a given system. The total range of those devices restricts the magnitude of changes in the wireless environment that the design can accommodate. Finally, linearity of the devices used for cancellation is important for the digital processing that takes place on the received signal.

Chapter 4

Full-Duplex Radio Design

Chapters 2 and 3 discussed conceptual and practical constraints in implementing cancellation techniques for designing a wireless full-duplex system. This chapter details the design of a practical, real-time full-duplex wireless radio based on the principles discussed in the previous chapters. It starts by outlining some of the requirements for the system based on static IEEE 802.11 (Wi-Fi) like networks. Cellular and mobile networks pose more stringent requirements, some of which are discussed later in Section 7.3.

Requirements of the design

- High Bandwidth: Data networks are increasingly using higher bandwidths. The current WiFi standard uses 20MHz or 40MHz signals, and other technologies, such as whitespace networking may use even higher bandwidths [12, 13]. Current full-duplexing implementations in research [28, 52] and in industry [29] focus on narrowband signals and may not be directly applicable to wider bandwidth data signals. *The first requirement of the design is that it should work for fairly wide bandwidth signals, at least up to 40MHz.*

- Adaptive: Full-duplexing has been used in wired networks for some time. One such example is use of hybrid coils in telephone networks [29]. Wired networks tend to have fairly stable characteristics, thus not requiring constant retuning of the cancellation mechanism. The wireless channel, on the other hand, is

constantly changing. Since cancellation mechanisms require very fine tuning to perform well, the design has to adapt to channel changes. *The second requirement is that the design adapt cancellation settings to changes in the wireless channel.*

- Relatively High Power: For full-duplexing to work well, the cancellation techniques need to cancel enough self-interference to bring it close to the noise floor of the receiver. As the transmit power increases, so does the self-interference. Existing methods of full-duplexing can achieve reasonable performance for applications with very low power signals, requiring 30dB cancellation of self-interference. With WiFi like devices, the self-interference is much stronger, requiring 75dB cancellation of self-interference. *The third requirement is that the design should achieve 75dB cancellation of self-interference to work with relatively high transmission powers.*

4.1 Design Overview

The requirements posed for the full-duplex system guide the design decisions for the system. The system uses a combination of analog and digital cancellation techniques to meet the 75dB cancellation requirement. The analog cancellation technique used is signal inversion since it does not have any inherent bandwidth constraint, and does not use active components that may get saturated with higher self-interference power.

Figure 4.1 shows the block diagram of the adaptive full-duplex design combining signal inversion based analog cancellation and digital cancellation. As shown in Section 2.5, an industry-grade balun in an analog cancellation circuit can cancel up to 45dB of a 40MHz signal. But this cancellation only handles the dominant self-interference component between the receive and transmit antennas. A node's self-interference may have other multipath components, which, although much weaker than the dominant one, are strong enough to interfere with reception. Furthermore, the balun circuit may distort the cancellation signal slightly, such that it introduces

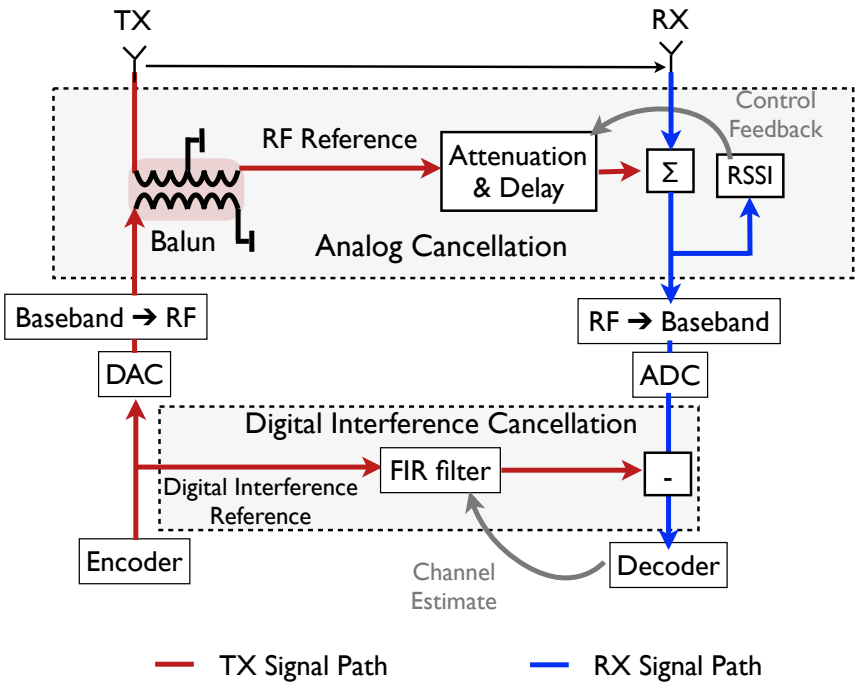


Figure 4.1: Block diagram of full-duplex system. The ideal cancellation setup uses passive, high precision components for attenuation and delay adjustment.

some interference leakage. Digital cancellation can create a channel model that cancels multipath components. The 45dB cancellation from signal inversion reduces the self-interference level such that the received digital samples aren't saturated, thus making digital cancellation feasible.

The digital cancellation portion of the design uses a finite impulse response (FIR) filter to store a model of the self-interference channel after analog cancellation. Transmit digital samples are passed through the FIR filter to create digital cancellation samples which are subtracted from the received samples to further clean interference from the received signal.

This system achieves both the high power and the wide bandwidth requirements. Now we are left with coming up with a mechanism for making the cancellation setup adapt to changes in the wireless channel. Section 4.2 explains how the analog cancellation mechanism can be made adaptive using RSSI as a minimization objective for a control system. Section 4.3 describes an adaptive digital cancellation design for an

OFDM receiver.

4.2 Adaptive Analog Cancellation

The results evaluating the performance of the signal inversion technique in Figure 2.15 show that, if the phase and amplitude of the inverted signal are set correctly, signal inversion cancellation can have impressive results across a wide bandwidth. This raises a simple follow-on question. Is it possible to automatically adjust the phase and amplitude, thereby self-tuning cancellation in response to channel changes? In this section we describe an algorithm that can accurately and quickly self-tune a cancellation circuit.

The basic approach is to estimate the attenuation and delay of the self-interference signal and match the inverse signal appropriately. Ideally, the auto-tuning algorithm would adjust the attenuation and delay to minimize the residual energy after analog cancellation (assuming no other signal is being received on the RX antenna). Let G_c and τ_c be the variable attenuation and delay factors respectively, and $s(t)$ be the signal received at the input of the programmable delay and attenuation circuit. The delay over the air (wireless channel) relative to the programmable delay is τ_a. The attenuation over the wireless channel is G_a. The energy of the residual signal after analog cancellation is derived in Appendix A.2 to be:

$$E = \int_{T_o} (G_a s(t - \tau_a) - G_c s(t - \tau_c))^2 \mathrm{d}t \tag{4.1}$$

where T_o is the baseband symbol duration. The goal of the algorithm is to adjust the parameters G_c and τ_c so that energy of the residual signal is minimized.

Our key insight is that the residual energy function in Equation 4.1 has a pseudo-convex relationship with G_c and τ_c for WiFi style OFDM signals for the case when $|\tau_a - \tau_c| < 1/4fc$. We refer the reader to Appendix A.2 for the mathematical details. Figure 4.2 shows the theoretical RSSI output for different values of G_c and τ_c around the optimal for a specific setting where the self-interference signal has 20dB attenuation and zero delay. This plot clearly shows the pseudo-convex nature of the RSSI function with attenuation and delay. We can exploit this structure to design a simple gradient descent algorithm to converge to the optimal setting of delay and

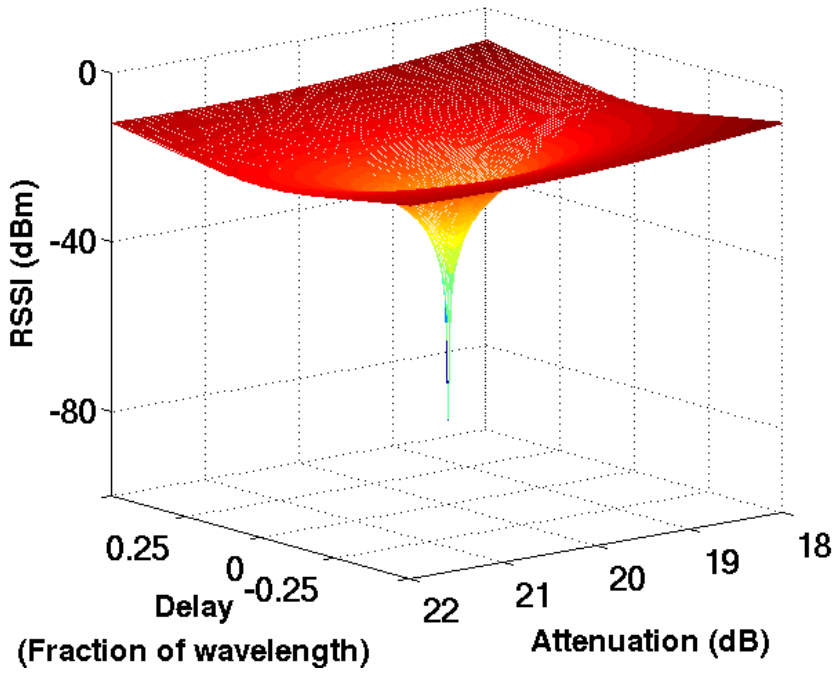


Figure 4.2: Theoretical RSSI of the residual signal after signal inversion cancellation with varying delay and attenuation. Note the deep null at the optimal point and the pseudo-convex shape of the RSSI function.

attenuation.

While a gradient descent algorithm would work well, fine-grained programmable analog attenuation and delay lines are unfortunately not typical commodity components and so are expensive [46]. So we do not provide results from an implementation of adaptive analog cancellation using signal inversion with passive components. Recent research on implementing these components on-chip [46, 19, 23, 62, 38, 26], combined with an increased demand for such components for full-duplex radios should help commoditize these components.

4.2.1 Practical Algorithm with QHx220

Our current implementation of adaptive analog cancellation uses the QHx220 noise cancellation chip as an approximation for the delay line and attenuator needed for

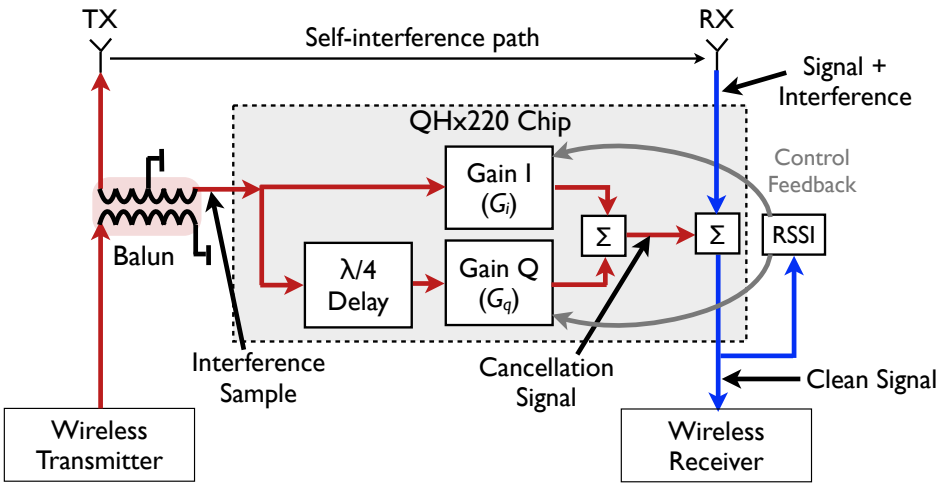


Figure 4.3: Block diagram of analog cancellation with signal inversion using the QHx220 chip as an approximation for delay and attenuation. The RSSI values represent the energy remaining after cancellation. The auto-tuning algorithm adapts gain parameters G_i and G_q to minimize this energy.

implementing analog signal inversion cancellation [51]. QHx220 uses linear quadrature modulation with active components to create a cancellation signal, and is susceptible to all the problems mentioned in Section 2.4, such as saturation and bandwidth limitations. The saturation in the chip also causes non-linear distortion as discussed in Section 3.2, especially for input powers beyond -40dBm. Hence, cancellation will not be perfect for typical wireless input powers (0-30dBm). Non-linear distortion also impacts digital cancellation, as we have seen in Section 3.2.

Our implementation uses the QHx220 despite its imperfections because it is inexpensive and easily available. However, we believe that it is feasible to build a full-duplex radio using an electronically tunable delay and attenuation chipset, since they are commercially available (but not yet widely and inexpensively.) Furthermore, including them as small parts of a full-duplex radio hardware design would not be particularly complex or expensive.

Figure 4.3 shows the block diagram of signal inversion cancellation using the QHx220 chip with the auto-tuning circuit. The RSSI value provides the residual signal energy after analog cancellation has subtracted self interference from the received

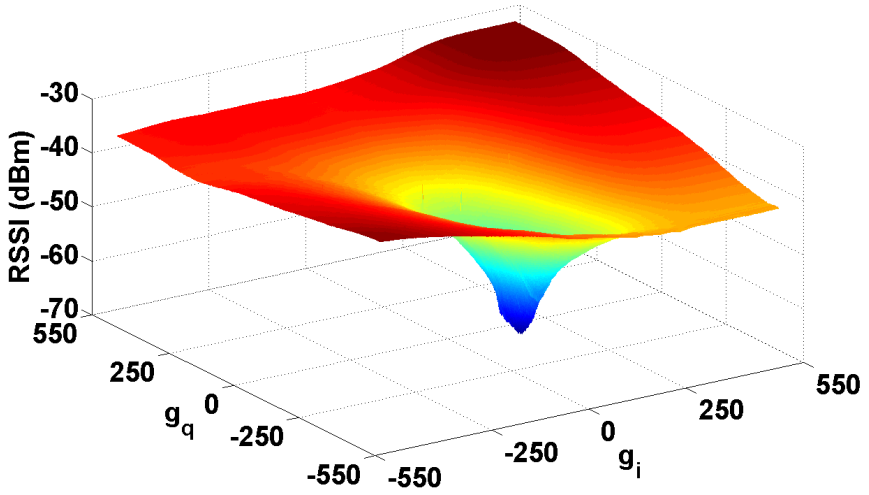
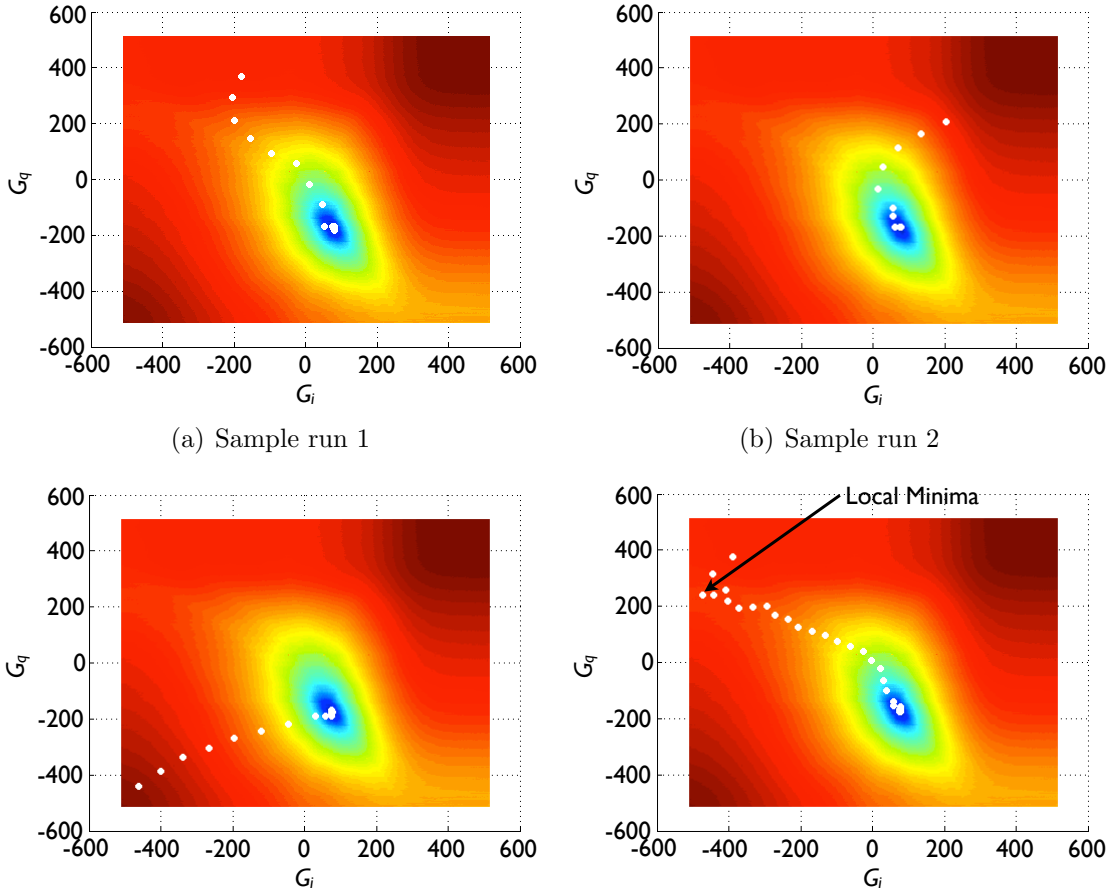


Figure 4.4: RSSI of the residual signal after analog cancellation as we vary G_i and G_q in the QHx220. G_i and G_q can each be varied from a value of -512 to +512. Note the deep null at the optimal point.

signal.

The goal of the auto-tuning algorithm in this case is to find the gains G_i and G_q such that the QHx220 chip output is the best approximation of the self-interference we need to cancel from the received signal. Fortunately, we can show that even with this approximate version, we still retain a pseudo-convex structure as shown in Appendix A.2.2. To see this empirically, we conduct an experiment where the TX antenna transmits a 10MHz OFDM signal, and we vary the two gains in QHx220. We plot the RSSI output in Figure 4.4, where a deep null exists at the optimal point. Hence we can use the same gradient descent algorithm for tuning the two attenuation factors in QHx220, G_i and G_q. G_i and G_q can be varied between the values -512 and +512. A value of 0 corresponds to no output and a value of 512 corresponds to a gain of 20dB. We do not know exactly how much gain each gain value corresponds to, but the functioning of the algorithm does not depend on this knowledge.

The gradient descent algorithm works in steps, and at each step it computes the slope of the residual RSSI curve by changing G_i and G_q by a fixed step size. A step of the algorithm requires five measurements, each involving a small change in G_i and G_q followed by RSSI sampling. The combination of these readings is used to determine



(a) Sample run 1

(b) Sample run 2

(c) Sample run 3

(d) Sample run 4

Figure 4.5: Sample runs of the adaptive analog cancellation mechanism with random starting points on the mesh shown in Figure 4.4. Each white dot represents one iteration.

the slope of the RSSI function at the current G_i and G_q value and the new G_i and G_q values to be used for the next step. If the residual RSSI is lower at the new G_i and G_q values than before, then the algorithm moves to the new settings for the gains, and repeats the process. If at any point it finds that the residual RSSI increases, it knows that it is close to the optimal point. It then reverses direction, reduces the step size and attempts to converge to the optimal point. The algorithm also checks for false positives, caused due to noisy minimas and saddle points. Appendix A.3 provides a pseudo-code describing the gradient descent algorithm used for implementing adaptive

Figure 4.6: CDF of Algorithm convergence on hardware. About 30% of the runs have to recover from noisy minimas, but do so quickly.

cancellation with the QHx220 chip.

While the cancellation the system can obtain with QHx220 is limited, we can evaluate how well the autotuning algorithm works in terms of convergence time. Each step in the auto-tuning algorithm measures RSSI for five points along the residual RSSI curve. Each data point takes $26\mu s$, $10\mu s$ to change the QHx220 settings and $16\mu s$ to measure RSSI over two OFDM symbols. Each iteration therefore takes $130\mu s$.

Figure 4.5 shows a few runs of the algorithm with random starting points. The algorithm is fast; it typically converges to the minimum in $8 - 15$ iterations, depending on the choice of the starting point. Figure 4.5(d) shows a run where the algorithm initially converges to a saddle point, and then recovers to converge to the true minima within 20 iterations. Figure 4.6 shows the cumulative distribution of how many iterations the algorithm takes to converge. The median convergence for the algorithm is 8 steps: the average execution time of the algorithm is approximately 1 ms. Currently we transmit one short ($64\mu s$) packet for each RSSI measurement, but the whole process can be fit in one full packet of $1ms$ duration.

4.3 Adaptive Digital Cancellation

The full-duplex radio design uses digital cancellation to cancel any residual interference that persists after analog signal inversion cancellation. Implementing digital cancellation for a full-duplex radio, however, is more challenging than other uses of digital cancellation, such as successive interference cancellation (SIC) and ZigZag

decoding [34, 32]. Unlike SIC or ZigZag, which uses digital cancellation to recover packets which would have otherwise been lost, a full-duplex radio uses digital cancellation to prevent the loss of packets which a half-duplex radio could receive. While an SIC implementation that recovers 80% of otherwise lost packets is a tremendous success, a full-duplex radio that drops 20% of packets is barely usable.

To the best of our knowledge, our digital cancellation system has three advancements compared to existing software radio implementations in the literature. First, it is the first real-time cancellation implementation that runs in hardware: this is necessary for designing and testing a real-time full-duplex MAC described in Chapter 5. Second, it is the first cancellation implementation that can operate on 10MHz signals. Finally, it is the first digital cancellation that operates on OFDM signals.

Section 2.2 discussed the basics of digital cancellation. Adaptive digital cancellation has two components: estimating the self-interference channel; and using the channel estimate on the known transmit signal to generate digital samples to subtract from the received signal. Although seemingly simple, implementing digital cancellation on a real OFDM system is complicated. Typical OFDM systems use received digital samples both in the time domain, for operations like packet detection and carrier offset correction, and the frequency domain, for operations like channel estimation, equalization and decoding. A fast fourier transform (FFT) circuit is used to convert the time domain received samples to frequency domain. Channel estimation in OFDM systems is done in the frequency domain since estimating a channel requires a simple division operation in the frequency domain, while requiring complex deconvolution in the time domain.

The channel estimation block present in an OFDM system can be re-used for estimating the self-interference channel for digital cancellation. Using this can give the self-interference channel estimate in the frequency domain. But digital cancellation has to be applied before any other digital processing happens on the signal to prevent operations like packet detection being affected by self-interference. Thus, the frequency domain estimate has to be converted to time domain using inverse fast fourier transform (IFFT). This time domain estimate can be programmed into a digital finite impulse response (FIR) filter to create the self-interference channel model.

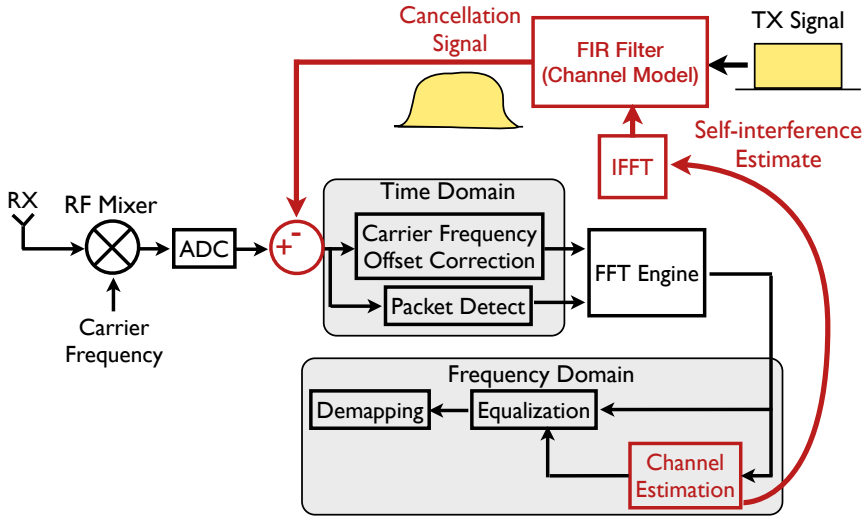


Figure 4.7: Simplified block diagram of an OFDM receiver with digital cancellation. The cancellation uses frequency domain channel estimation but cancels self-interference in the time domain samples at the input of the digital receiver chain.

Then digital cancellation can be implemented by passing transmit digital samples through this FIR filter to create a cancellation signal which is subtracted from the received digital samples at the beginning of the digital receive chain. Figure 4.7 shows digital cancellation implemented in an OFDM system with frequency domain channel estimation and time domain cancellation.

We now describe the steps involved in digital cancellation in detail.

Channel Estimation: To estimate the channel, the radio uses known training symbols at the start of a transmitted OFDM packet. It models the combination of the wireless channel and analog cancellation circuitry effects together as a single self-interference channel, estimating its response. The estimation uses the least square algorithm [57] due to its low complexity. Since the training symbols are defined in the frequency domain – each OFDM subband is narrow enough to have a flat frequency response – the radio estimates the frequency response of the self-interference channel as a complex scalar value at each subcarrier. Specifically, let $\mathbf{X} = (X[0], \cdots, X[N-1])$ be the vector of the training symbols used across the N subcarriers for a single OFDM symbol, and M be the number of such OFDM training

symbols. Let $\mathbf{Y}^{(m)}$, $m = 1, \cdots, M$, be the corresponding values at the receiver after going through the self interference channel. The least squares algorithm estimates the channel frequency response of each subcarrier k, $\hat{H}_s[k]$, as follows:

$$\hat{H}_s[k] = \frac{1}{M}\left[\frac{1}{X[k]}\left(\sum_{m=1}^{M} Y^{(m)}[k]\right)\right]$$

Next, the radio applies the inverse fast Fourier transform (IFFT) to the frequency response to obtain the time domain response of the channel. Upon transmission, it generates digital samples from the time domain response and subtracts them from the observed signal. The time domain response of the self interference channel can be emulated using a standard finite impulse response (FIR) filter in the digital domain. Standard FPGA implementations of FIR filters are widely available end efficient.

By estimating the frequency response in this way, the least squares algorithm finds the best fit that minimizes overall residual error. The algorithm is more robust to noise in samples than prior approaches, such as simple preamble correlation. Furthermore, unlike more complex algorithms such as minimum mean squared error (MMSE) estimation, which requires a matrix inversion, least squares is simple enough to implement in existing software radio hardware for real-time packet processing. FIR based implementations have been used previously for echo cancellation in wired networks [25]. One approach combines decision feedback equalizers with digital echo cancelers to adaptively tune a receiver for both echo cancellation and inter-symbol interference reduction [50]. Although this work presents results from using a linear equalizer for implementing frequency domain estimation and time-domain cancellation, using frequency domain decision feedback equalizers could lead to improved digital cancellation performance [14].

Applying Digital Cancellation: Next, the radio applies the estimated time domain channel response to the known transmitted baseband signal and subtracts it from the received digital samples. To generate these digital samples, the hardware convolves the known signal with the FIR filter representing the channel. Let $s[n]$ be the known transmitted digital sample at time n fed into the FIR filter. The output $i[n]$ of the

filter is the linear convolution of $h_s[n]$ and $s[n]$:

$$i[n] = \sum_{k=0}^{N-1} \hat{h}_s[k]s[n-k]$$

The receive antenna of the node gets both the self-interference signal from its transmit antenna and a signal from an intended transmitter. The signal output from the receiver analog-to-digital converter (ADC) is given by:

$$r[n] = \sum_{k=0}^{N-1} h_d[k]d[n-k] + \sum_{k=0}^{N-1} h_s[k]s[n-k] + z[n]$$

where $d[n]$ and $h_d[n]$ are the transmitted signal and channel impulse response from the intended transmitter, and $z[n]$ is additive white Gaussian noise.

The radio subtracts the estimates of the transmit signal from the received samples $r[n]$. The received signal after digital cancellation, $\hat{r}[n]$, is given by:

$$\begin{aligned} \hat{r}[n] &= r[n] - i[n] \\ &= \sum_{k=0}^{N-1} h_d[k]d[n-k] + \sum_{k=0}^{N-1} \Big(h_s[k] - \hat{h}_s[k] \Big) s[n-k] + z[n], \end{aligned}$$

The quality of digital cancellation depends on how well the digital channel estimate $\hat{h}_s$ models the self-interference channel h_s.

Efficacy of Digital Cancellation

Channel estimation accuracy significantly affects digital cancellation's performance. Poor channel estimates can cause the system to generate digital samples different from what the node hears, such that their subtraction corrupts a received waveform. Accuracy suffers if there is another interfering transmitter present during the channel estimation phase. Hence, the MAC protocol must provide an interference-free period for channel estimation via carrier sense.

The second factor is the coherence time of the self-interference channel, i.e. the duration over which the channel's state is stable. A node needs to re-estimate its channel state at a period below the channel coherence time. When the coherence time is much longer than a packet time, periodic channel estimates can be sufficient. However, when the coherence time is comparable to or less than a single packet transmission, the MAC layer may need to introduce a "silence period" within each

packet so that the full-duplex node can recalibrate its digital cancellation block in the middle of a packet transmission.

The third and final factor is possible non-linearities in the self interference channel. Digital cancellation as presented above assumes that the self-interference can be modeled as the output of a linear time-invariant system. However, in practice, the analog cancellation step may introduce a non-linear distortion of the transmitted signal which cannot be modeled using an LTI system. As we have shown in Section 3.2, active components can introduce non-linearities, thus constraining the performance of Digital Cancellation.

4.4 Cancellation Performance

We evaluate the complete cancellation design, described in Figure 4.1, by measuring its attenuation of the self-interference signal. Unlike the cancellation results in Section 2.5.1, which evaluated the performance of analog cancellation using signal inversion in isolation, these experiments evaluate the entire radio design.

We focus on self-interference cancellation obtainable with tunable passive attenuation and delay components. These components are not electronically programmable: they can only be controlled manually. Consequently we cannot use the passive components when analog cancellation needs frequent tuning, such as wireless channels. We therefore evaluate its performance using a wired setup by connecting the TX and RX antennas using a wire to simulate a controlled wireless channel. We measure self-interference cancellation obtained with a combination of analog signal-inversion cancellation and digital cancellation. As the passive components are tuned manually, human imprecision makes it unlikely that the system is at the optimal point: it may be possible to obtain even stronger cancellation with a programmable passive component.

Figure 4.8 plots the total cancellation obtained for a 10MHz WiFi signal with increasing self-interference power. Analog signal-inversion cancellation, in agreement with the results in Figure 2.14, cancels around 45 dB of self interference. Digital cancellation can cancel as much as 30 dB. This is 10 dB more cancellation over a

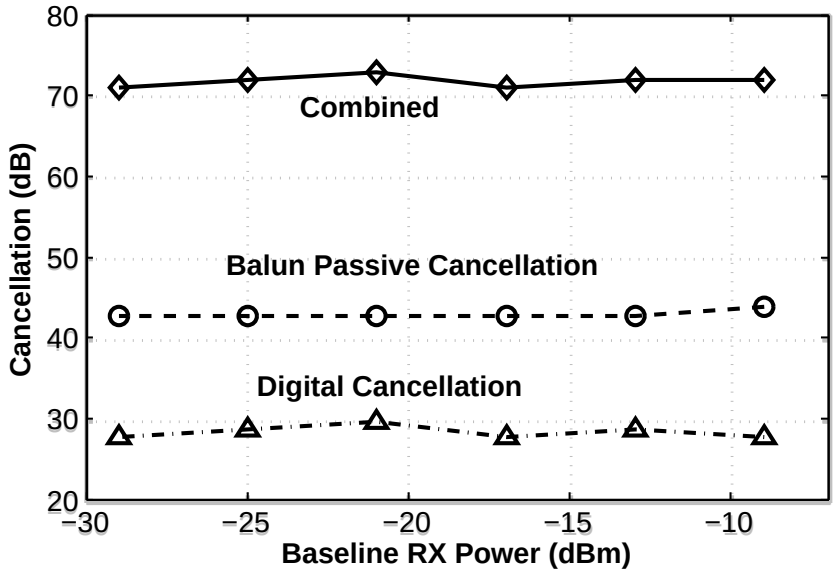


Figure 4.8: Cancellation performance of analog signal-inversion cancellation combined with digital cancellation in a controlled wired setting, where phase and amplitude are controlled by manually tuned, precision passive components. Together they cancel 70-73 dB of self-interference.

sixteen times wider bandwidth (10 MHz vs 612 kHz) compared to the full-duplex approach reported by Duarte et al. [28], which uses two separate TX chains. The 25-30dB of digital cancellation adds on top of the 40-45 dB of analog cancellation to provide a total cancellation of 70-73 dB.

We omit plots for digital cancellation alone for brevity, but report some observations here. Typically, digital cancellation can reduce self-interference by up to 30dB for lower power signals. The performance of digital cancellation degrades by up to 9 dB at higher received powers, for example when not using analog cancellation, due to receiver saturation.

The above setup does not benefit from the self-interference reduction from antenna separation between the TX and RX antennas, because they are connected directly by a low loss wire. In practice with wireless channels, we observe around 40dB of attenuation (20 cm, 8 inches) from antenna separation with 3dBi antennas. Combined with the 73dB reduction from analog signal-inversion cancellation, the proposed full-duplex design could bring down the self-interference by up 113 dB. Such cancellation would

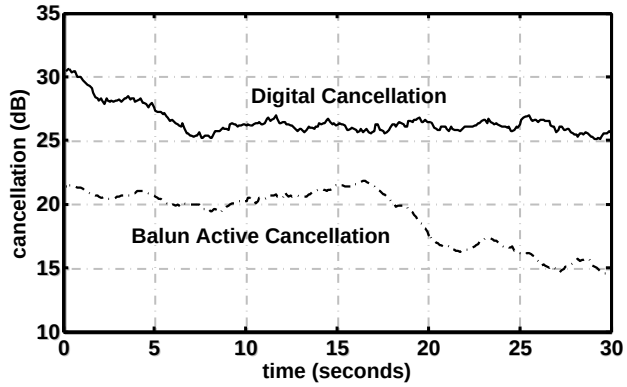


Figure 4.9: Performance of adaptive analog cancellation and of digital cancellation over time. Before cancellation, the received power is -45dBm for the analog cancellation experiment, and -58dBm for digital cancellation. Once tuned, the QHx220 settings are stable for over 10 seconds. The 20 dB maximum is caused by the nonlinearities of the QHx220. Digital cancellation performance, on the other hand deteriorates within a span of 3-4 seconds and needs more frequent tuning.

bring a 20 dBm transmit signal to -93 dBm, close to the noise floor on commodity hardware.

In summary, analog signal-inversion cancellation in conjunction with digital cancellation can cancel enough self-interference from wideband signals such as WiFi to enable full-duplex operation.

4.5 Self-Interference Coherence Time

An important factor to consider is the need to recalibrate the system periodically. The recalibration rate depends on how fast the wireless channel for the self-interference signal changes, or the self-interference coherence time.

Figure 4.9 shows a plot of how auto-tuned analog signal inversion cancellation (using the QHx220 chip) and digital cancellation decay over time in a daytime office wireless environment. While it takes approximately 1 ms to tune the analog cancellation circuit, this tuning is typically stable for over 10 seconds. Also, the auto-tuning algorithm can run on-demand: a node should recalibrate when it finds that its noise floor has increased when it starts transmitting a beacon or a data packet.

Looking at digital cancellation, immediately after estimation, there is a 32 dB reduction in self-interference. Performance quickly degrades over the first two seconds as the channel changes, and after 7 seconds settles at approximately 25 dB. This result makes sense: unlike calibrating analog signal-inversion cancellation, which is mainly for the line-of-sight component between the two antennas, digital cancellation is handling varying multipath components. A full-duplex radio needs to recalibrate its digital cancellation much more frequently than its analog cancellation. Each recalibration for digital cancellation takes only a few OFDM symbols.

4.6 Summary

This chapter described the design of a full-duplex wireless system. It used the insights into the performance and practical constraints associated with different cancellation schemes developed in earlier chapters to develop a design that works for wideband and fairly high power wireless nodes. The design combines analog signal-inversion cancellation and digital cancellation to get an overall 73dB reduction in self-interference. A practical full-duplex system also needs to adaptively re-tune itself to changes in the wireless channel. This chapter describes auto-tuning mechanisms for the design to adapt to changes in the wireless channel. Specifically, analog cancellation uses a gradient descent algorithm to received power by adjusting the delay and attenuation of the cancellation path, thus optimizing cancellation. Digital cancellation uses OFDM channel estimation to estimate the self-interference channel and a reconfigurable FIR filter to implement a channel model. Each of these techniques show very fast adaptation rates, <1ms for analog cancellation adaptation and a few OFDM symbols for digital cancellation, thus showing the feasibility of implementing a full-duplex Wi-Fi transceiver.

Chapter 5

Full-Duplex MAC

Earlier sections showed that a wireless full-duplex system that can nearly double the throughput of a single hop link is practically implementable. On the other hand, the implementation uses additional resources that could otherwise be used to implement a 2x2 MIMO system, that may provide similar physical layer gains. It is unclear if only the physical layer gains of full-duplex would justify the engineering and cost needed to implement these systems. Section 7.1 provides a theoretical comparison of full-duplexing vs 2x2 MIMO with half-duplexing.

However, the most interesting possible benefits of full-duplex occur above the physical layer. Research into applications of full-duplex wireless has suggested that a full-duplex system may mitigate many of the problems with wireless networks today [22, 53, 33]. For example, full-duplexing can help address the hidden terminals problem, improve fairness in wireless systems, reduce congestion due to MAC scheduling, and reduce end-to-end delays in multihop wireless networks.

While the expected gains from full-duplex MACs look promising, the lack of a real-time full-duplex MAC layer implementation has prevented experimentally evaluating these claims. In this chapter, we take the case of an access point based wireless LAN and describe how a standard half-duplex MAC for such networks can be modified to incorporate full-duplexing capabilities. We then implement a real-time full-duplex MAC on a software radio platform and evaluate its performance on a network of 5 full-duplex nodes with one of the nodes acting as the access point. This evaluation

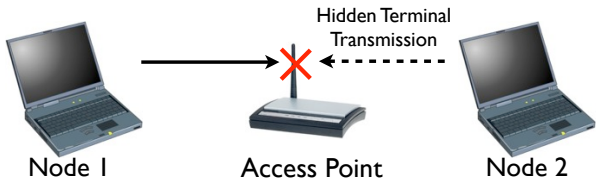


Figure 5.1: An infrastructure Wi-Fi setup. A hidden terminal occurs at the AP when Node 1 and Node 2 cannot hear each other's transmissions leading to collisions.

shows that full-duplex reduces packet losses due to hidden terminals by up to 88%. Full-duplex also mitigates unfair channel allocation in AP-based networks, increasing fairness from 0.85 to 0.98 while improving downlink throughput by 110% and uplink throughput by 15%. To the best of our knowledge, this is the first implementation of a full-duplex MAC that works in real-time on real hardware.

5.1 MAC Gains with Full-Duplex

As alluded to earlier, full-duplexing can help improve the performance of existing wireless systems by mitigating problems such as hidden terminal losses and unfairness in access point based networks. This section shows how a full-duplex MAC can solve such problems.

5.1.1 Reducing Hidden Terminals

Figure 5.1 shows a typical home or office Wi-Fi setup with the hidden terminal problem. End nodes connect to the backbone network through an access point. The hidden terminal problem occurs when Node 2 is unable to hear Node 1's transmissions to the access point and starts sending data to the access point at the same time, thus causing a collision at the access point.

This problem can be solved using full-duplex nodes. Suppose all nodes always have data to send to and receive from the access point. Then, as soon as Node 1 starts transmitting data to the access point, the access point starts transmitting data back to Node 1 simultaneously. Node 2 hears the transmission from the access point and

(a) Half-Duplex

(b) Full-Duplex

Figure 5.2: An access point based network with 1 AP connected to 3 nodes. MAC scheduling results in unfairly low channel allocation for downlink traffic for half-duplex. Full-duplex solves the problem balancing uplink and downlink channel access.

delays its transmission, thereby avoiding a collision. If the access point does not have any packets to send back to Node 1, it can repeat whatever it hears. This repetition serves as an implicit ACK for Node 1 and prevents Node 2 from transmitting. This scheme for mitigating hidden terminals also applies to multihop wireless networks.

Full-duplexing does not completely prevent the hidden terminal problem. In order for the receiver to respond, it needs to receive the destination address of the link layer header. However, typically the destination address is preceded by the preamble, PHY header, and part of the MAC layer header, where collisions can still occur. For example, for 802.11g, the receiver needs to receive 15 bytes before it can decode the receiver address, which leaves the initial ∼2.5% of the packet time to be vulnerable for 6Mbps and ∼10% for 54Mbps, for a 1500 byte packet. This vulnerability is inevitable, but can be reduced by changing the packet format such that the destination address is placed earlier in the packet.

5.1.2 Improved Fairness in Access Point Networks

Existing work [15, 44] has studied the problem of fairness between upstream and downstream flows in access point (AP) based networks. Since 802.11 CSMA provides

the same transmit opportunities to all clients and the AP, the AP only gets $1/(N+1)$ of the channel when there exist N clients. Thus, the downstream flows only get an aggregate throughput $1/(N+1)$ of the channel capacity, while upstream flows get $N/(N+1)$.

Figure 5.2(a) shows this fairness problem for an access point (AP) connected to 3 clients. Suppose all client stations have uplink data to send to the access point and the access point has downlink data to be sent to each client. If all the uplink and downlink flows have saturated traffic, the clients constantly contend with each other and with the access point for channel access. With half-duplex nodes, typical MAC scheduling gives the access point $1/4^{th}$ the total transmission opportunities. This restricts the aggregate downlink throughput of the network to $1/4^{th}$ the capacity of one link. On the other hand, each client also gets $1/4^{th}$ the total transmission opportunities to serve only its own uplink flows, thus resulting in an aggregate uplink throughput of $3/4^{th}$ the capacity of one link. This leads to a network with very low fairness. Some suggested solutions to this problem include controlling the channel access priority or incorporating rate control mechanisms above the MAC layer [15, 44].

However, with full-duplexing, the access point can transmit a downlink packet whenever it receives an uplink packet from a client that it has traffic for as shown in Figure 5.2(b). This balances the aggregate uplink and downlink throughputs, thus elegantly solving this fairness problem.

5.2 Design

Fully exploiting full-duplex wireless requires redesigning the MAC. For example, existing MACs based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) preclude a wireless node from transmitting while it is receiving a packet irrespective of whether the packet is addressed to the node or not. With full-duplex a node should in fact try to transmit a packet whenever it is receiving a packet to maximize the use of the wireless channel. In a sense, a full-duplex PHY creates a parallel channel that a transmitter can simultaneously receive on. However, the parallel channel exists only for transmitters for full-duplex: signals add up on other nodes.

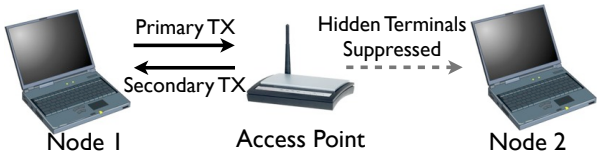
(a) Symmetric dual link (b) Asymmetric dual link

Figure 5.3: Symmetric and asymmetric dual links in the Contraflow full-duplex MAC framework.

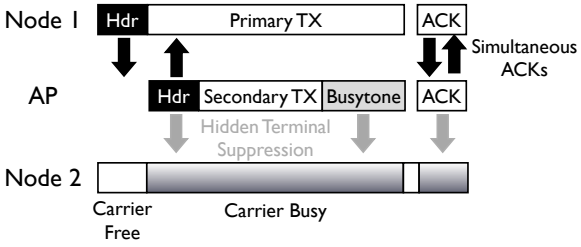
A full-duplex MAC must utilize this rather limited parallel channel in full. Thus, full-duplexing requires a redesign of the MAC layer. We leverage existing research on a new MAC design for full-duplex wireless systems, named Contraflow, to guide our MAC design [53].

Contraflow is a general framework for a single channel full-duplex MAC protocol. In Contraflow, *a primary transmitter* initiates a transmission via standard CSMA/CA. Once the *primary receiver* decodes the header of the primary transmission, it can initiate a *secondary transmission*. The secondary transmission can be destined to any node where the primary and the secondary transmitters do not collide, either the transmitter (*symmetric dual-link*) or nodes outside the interference range of the primary transmission (*asymmetric dual-link*). To support asymmetric dual-link, each Contraflow node must learn its interference range and exchange this information. Figure 5.3 shows examples of symmetric and asymmetric dual-links in contraflow.

We have designed and implemented a MAC based on the above framework. Limiting the scope of this MAC to WiFi-like networks where multiple clients connect to an access point, we focus on the symmetric dual-links. In this MAC design, the primary transmitter is always the secondary receiver. Figure 5.4(a) illustrates the behavior of a symmetric dual-link in an AP-based network. The primary (Node 1) sends a packet to the AP after carrier sense. As soon as the AP receives the header, it starts a secondary transmission back to Node 1. Even if Node 2 is hidden to Node 1, it senses the secondary transmission and keeps quiet. Thus, in theory, full-duplex can prevent hidden terminal collisions.



(a) Full-duplex with hidden terminals



(b) Full-duplex packet exchange

Figure 5.4: The full-duplex MAC protects primary and secondary transmissions from losses due to hidden terminals. A *busytone* is used to protect periods of single-ended data transfer

Since the primary and secondary packets are offset in time and may have different lengths, relying solely on data packets does not completely protect from hidden terminals. If a node finishes its transmission but has not finished receiving its duplex packet, it may still experience a collision. Figure 5.4(b) shows an example of such a packet exchange. The AP's secondary transmission to Node 1 may finish before Node 1's primary transmission to the AP. Our MAC implementation uses *busytones* as a way to mitigate this problem. Whenever a node finishes transmitting a packet and sees that it still has not finished receiving, it transmits a predefined signal until its reception ends. If a node receives a primary transmission and does not have a corresponding secondary packet to send, it sends the busytone immediately after decoding the header of the primary packet.

Full-duplex mitigates hidden terminals, but does not completely eliminate them. A primary receiver is susceptible to collisions until it has finished receiving the primary transmission's packet header. In 802.11a, for example, this period is $\approx 56\mu$s, much shorter than a typical data packet.

5.3 Real-time MAC Implementation

The previous section has described the design of the full-duplex MAC protocol. Implementing the MAC protocol, however, is not trivial: it imposes certain requirements on the radio hardware. This section focuses on identifying the minimal features that a full-duplex MAC requires, and shows how it is actually implemented.

5.3.1 Challenges

Maximizing the overlap of two transmissions increases throughput and improves collision avoidance. Transmission overlap depends on solving two technical challenges, minimizing secondary response latency and having transmission flexibility in preloaded packets.

Secondary response latency: A primary transmission is a "cue" for the secondary transmission to start. A faster response to this cue enables a longer overlap of the two transmissions. The earlier the destination address is in the packet and the faster the hardware can initiate a secondary transmission, the lower the secondary response latency.

Flexibility in preloaded packets: Starting a secondary transmission immediately after primary header reception requires having a packet destined for the primary transmitter already loaded in the hardware. Having multiple packets loaded increases the probability of having a packet ready for secondary transmission. This calls for the radio driver to have per-destination transmission queues, with the head of each queue preloaded in hardware.

5.3.2 Platform

The challenges for realizing a full-duplex MAC layer place requirements on the system hardware. Specifically, the hardware must notify the start of reception once the destination address has been decoded, must be able to start a secondary transmission quickly, and needs enough memory to store multiple transmit packets.

Typical software-defined radios such as USRP [9] cannot meet these requirements, due to latency between hardware and the host PC as well as their need to store packets as digital samples rather than bits. Off-the-shelf WiFi cards also do not meet the requirements due to the lack of header reception indicators and general difficulty in programming low-level mechanisms such as backoff and clear channel assessments.

We use the WARP V2 platform [8] from Rice University. WARP handles physical layer packet processing and latency-sensitive MAC operations in a powerful on-board FPGA. As the FPGA can convert packet bits to digital baseband samples on-the-fly, it can preload multiple packets stored concisely as bits. Furthermore, it has flexible "autoresponder" hardware triggers that can be used to minimize secondary response latency. Finally, the FPGA is powerful enough to incorporate digital cancellation. An on-chip embedded processor implements the MAC and the auto-tuning algorithm in real-time.

5.3.3 Implementation Details

Our full-duplex MAC uses the OFDM Reference Design v15 from Rice University [8]. The design uses a WiFi-like packet format and 64-subcarrier OFDM physical layer signaling using a 10MHz bandwidth. Each OFDM symbol is $8\mu s$ long, thus QPSK modulation achieves 12Mbps bitrate without any channel coding (there is currently no library support for channel coding in the WARP release). The PHY frame has a $32\mu s$ preamble, followed by a $16\mu s$ training sequence, and 24-byte MAC header that is $16\mu s$ long. A 1484-byte long packet takes $1056\mu s$ to transmit.

The implementation is based on the half-duplex MAC implementation from the same reference design which mimics 802.11 behavior. Our full-duplex MAC uses the same CSMA/CA behavior for primary transmissions. The MAC design uses autoresponder logic, which is designed for low-latency link-layer ACK transmission, to automatically trigger a secondary transmission if the node detects a primary transmission addressed to it. A measurement shows that the latency of this logic is $11\mu s$, which results in the total secondary response latency of $75\mu s$ including header reception latency.

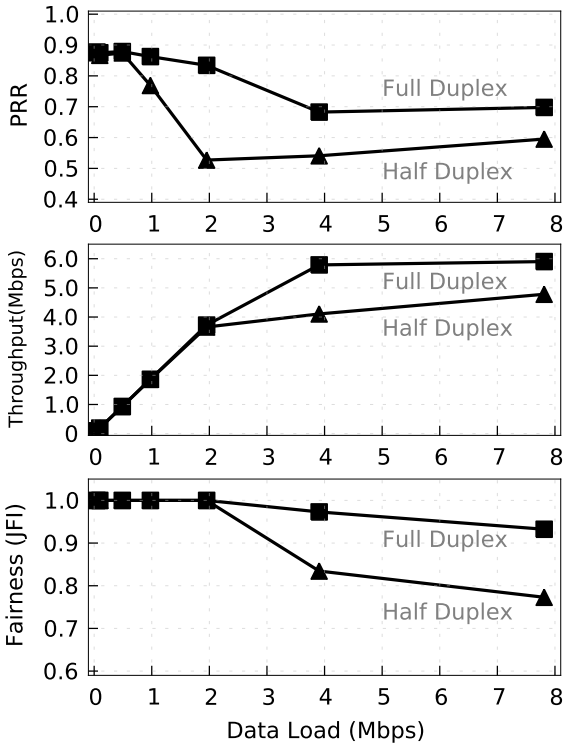


Figure 5.5: Two upstream UDP flows from two hidden terminals to an AP. Full-duplexing mitigates collisions due to hidden terminals.

The hardware auto responder logic automatically picks the correct transmission queue to send a secondary packet from based on the header of the primary reception. For primary transmissions, the software maintains a record of the order of arrival of packets from the host computer for different destinations, and sends packets from different queues in the same order.

5.4 MAC Evaluation

This section evaluates the benefits of a full-duplexing MAC layer with experiments on a testbed of WARP based full-duplex wireless nodes. These experimental results show that full-duplex radios can help solve long standing problems in wireless LAN MACs. Specifically we focus on the hidden terminal and fairness problems discussed in Section 5.1.

	Throughput (Mbps)		Fairness (JFI)
	Up	Down	
Half Duplex	5.18	2.36	0.845
Full-Duplex	5.97	4.99	0.977

Table 5.1: Throughput and fairness for four bi-directional UDP flows between an AP and four clients without hidden terminals. Fairness is measured using Jain's fairness index (JFI). Full-duplexing helps improve the fairness in Wi-Fi like networks.

5.4.1 Hidden Terminals

We setup the following hidden terminal experiment. An AP node is in the middle of 2 nodes which are hidden to each other. Both nodes constantly try to send UDP data to the AP. There is no downstream traffic from the AP to the nodes. The hidden terminal effect causes packets to collide at the AP, thus causing link layer failures. Since all traffic is unidirectional, full-duplex does not increase the physical layer capacity in this scenario.

Figure 5.5 shows the effect of using a full-duplex AP in preventing hidden terminals. Both flows maintain a fair throughput until the data load becomes 2Mbps for each flow. At the load of 2Mbps, the Packet Reception Ratio (PRR) for half-duplex drops to 52.7%, but full-duplex maintains a PRR of 83.4%. Excluding the effect of inherent link losses, full-duplex prevents 88% of collision losses because the busy tone of full-duplex prevents hidden terminals.

As the data load reaches 4Mbps, the total load exceeds the link capacity. In this case, half-duplex cannot maintain both flows due to heavy collisions. The effect can be seen in fairness, which starts to collapse for half-duplex. Because there is only one dominant flow active, PRR and throughput for half-duplex start to increase.

Full-duplex does not perfectly prevent hidden terminal collisions because secondary transmissions start only after header reception. We can see this effect at the data load of 4Mbps, where the PRR of full-duplex decreases to 68.3%. Header reception in our reference design takes $64\mu s$. This is longer than the typical 802.11 physical layer which takes $\approx 24\mu s$ for header reception. Thus, we expect the collision avoidance performance of full-duplex to be better with 802.11.

5.4.2 Fairness

Table 5.1 shows experimental results when an AP is connected to four clients. All nodes are within the carrier sense range of each other, thus removing hidden terminal effects. Each node makes a bidirectional UDP flow to the AP, making 8 active UDP flows, each with a 3Mbps load. The fairness index is computed over the individual throughput of 8 flows.

With half-duplex, when traffic is saturated, the AP gets the same share of the channel as all other nodes. However, the AP potentially has four times the traffic as any other node, since it is sending traffic to all four nodes. Consequently downstream flows may get an unfairly low share of the channel if the network is fully congested. In our experiment however, we see that the downstream flows do manage to get higher throughput than what we theoretically expect. The reason is that the data load of each flow is 3Mbps, which is lower than the link capacity of around 8Mbps. Consequently, we sometimes have nodes with empty transmit queues that do not contend for the channel, thereby leaving a larger share of the channel for downstream traffic compared to the theoretical throughput limit of $C/(n+1)$, where n denotes the number of clients and C the network capacity.

Since the traffic load is bidirectional, it is trivial that full-duplex gets higher throughput than half-duplex. However, what is interesting is how full-duplex distributes the additional throughput. With full-duplex, whenever the AP gets an upstream packet from any node, it is able to send a downstream packet to the same node, thus achieving fairness between upstream and downstream flows. Therefore, full-duplex improves the downstream throughput 111%, while the upstream throughput increases only by 15%.

In theory, full-duplex should increase the overall throughput by a factor of two, while the results show only a 45% overall increase. The reason is the limited queue sizes at the AP to send to the wireless clients. Each node can queue 16 packets that come from a host via Ethernet. Due to bursty traffic, sometimes the queue at the AP does not have packets for all clients. If the AP receives a primary transmission from a client and the AP has no packets to respond, the AP loses an opportunity for secondary TX, decreasing throughput. Looking at one of the logs verifies this, where

it shows that the AP is able to exploit secondary transmissions for only 52% of the primary receptions.

5.5 Summary

This chapter exemplified some of the higher layer gains possible with a full-duplex capable wireless physical layer. It described the design and implementation of a WiFi like MAC design modified for full-duplex operation. The implementation addresses certain challenges such as latency requirements for secondary transmissions and maximizing the opportunities for secondary transmissions by having per-destination queues. On the other hand, the implementation is restricted in only supporting symmetric full-duplex operation. This real-time MAC implementation shows full-duplexing reducing hidden terminal losses by up to 88% and improving fairness in an access point based network from 0.85 to 0.98.

Chapter 6

Redesigning Wireless with Full-Duplex

Full-duplexing provides a fundamental shift in the way wireless radios are designed and used. Chapter 5 concentrates on applying a full-duplex wireless radio to a wireless LAN based network, but this capability can be applied to and improve the performance of wireless devices across many application domains such as cellular systems, whitespace networks and multihop wireless data networks. This chapter looks at how different wireless domains could use full-duplexing in the future.

This chapter discusses the use of full-duplexing with three different abstractions. The first abstraction is using the additional channel available with full-duplex as a low bandwidth real-time control channel, the second abstraction is to use the second channel as an additional data channel for forwarding received data, and the third is to use full-duplexing as a means to introduce a gatekeeper for securing wireless networks.

6.1 Control Backchannel

When full-duplex links are used in a symmetric setup, i.e. for two nodes sending data to each other at the same time, it provides a real-time, inband backchannel for the primary receiver to send data back to the primary transmitter. The backchannel can

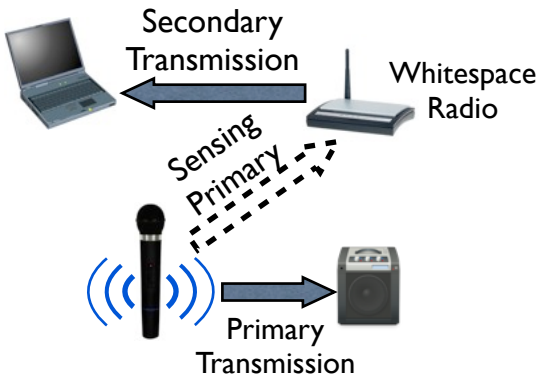


Figure 6.1: Whitespace radios need to co-exist with incumbent primary transmitters. The whitespace radio senses a wireless channel before using it to avoid interfering with primary transmissions.

be used for sending data or control traffic. Depending on the usage there are several applications possible with full-duplex systems. If the backchannel is used for sending control traffic then such a system can be used in whitespaces, for immediate collision notification and for sending in-band channel status. The following subsections discuss these applications in detail.

6.1.1 Opportunistic Spectrum Use (White Spaces)

Much of the licensed spectrum is under-utilized: only 5.2% occupancy between 30MHz to 3GHz [54, 13]. For this reason, in 2008 the FCC issued a ruling to allow for unlicensed (secondary) users to use licensed frequency bands as long as the licensed (primary) users do not experience perceivable interference [54]. The FCC requires that a secondary user be able to detect a primary signal that is as low as -114dBm. This requirement implies that current sophisticated solutions for a secondary user system cannot detect a primary user's presence while it's using a spectrum [13]. Although this requirement has since been removed due to the technical challenges in implementing it, the FCC, in its ruling encouraged the further development of spectrum sensing capability for future development of this space [10]

Figure 6.1 shows a secondary whitespace radio co-existing with a primary wireless device, such as a wireless mic. Without a full-duplex antenna, secondary transmitters

need to be very conservative in choosing transmission slots [37]. It is not necessarily safe for them to transmit even when the channel is sensed as vacant because they must account for the possibility that the primary might begin transmitting in the middle of their transmissions. This limits the utility that can be extracted from the vacant spectrum. By inferring the statistical properties of primary occupancy, smarter secondary strategies can be devised, but the basic problem remains [37].

A full-duplex system can fundamentally alter this balance because the secondary transmitters can sense primary activity even while they are transmitting and quickly vacate the spectrum. This ability will allow for significantly more efficient use of the vacant spectrum.

Research on opportunistic spectrum use has also shown the effectiveness of cooperation among secondary nodes for more accurate sensing of primary activity [49]. This ability is also easier to engineer using a full-duplex system. A secondary receiver can use the full-duplex backchannel to periodically report the state it observes on all the channels including the one that is currently used. This in-band shared information can be used by the secondaries to select channels with very low probability of being used by the primaries.

6.1.2 Packet Error Notification

The full-duplex backchannel can be used for notifying the transmitter about packet errors. This notification can be either explicit or implicit. For explicit notification, the receiver can send an abort packet back to the transmitter as soon as it encounters erroneous bits. This scheme works for errors both due to signal variation and due to collision from another node. CSMA/CN, an existing error notification technique, can reliably send a notification back to the transmitter only if the received notification power is within 36dB of the transmit power level [55]. With a full-duplex system, this notification can be received even when it is 80dB lower than the transmit power level.

An implicit way to inform the transmitter of packet errors is for the receiver to simply transmit whatever it is receiving, back to the transmitter. This “mirroring”

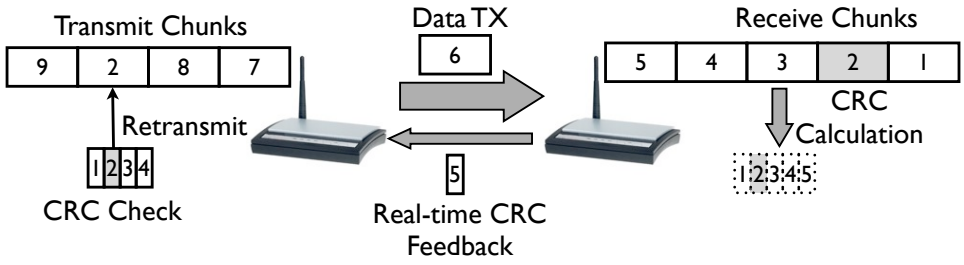


Figure 6.2: Real-time error notification using CRC feedback over small blocks of data. The transmitter checks the CRC feedback for each block and retransmits blocks that have the wrong CRC. Erroneous blocks are marked grey.

allows the transmitter to identify, if any, portions of the packet likely in error. This knowledge may be used by the transmitter to retransmit only the portions that it deems to be in error, implementing a real-time partial packet recovery scheme.

Maranello, an existing partial packet recovery scheme, splits a packet into blocks and computes a CRC on every block before sending the packet [35]. The receiver, after receiving the entire packet, sends the CRCs for all the blocks. This allows the transmitter to determine which blocks are in error and then send only the erroneous blocks. Figure 6.2 shows a similar technique implemented with full-duplex where the receiver sends the block CRCs as it's receiving data blocks. The transmitter receives these CRCs and can interleave retransmits in the middle of other data blocks. This saves the time equivalent to one packet transmission from the receiver and reduces the latency for getting retransmitted blocks.

6.1.3 In-Band Channel Status

In current wireless systems, a transmitter uses feedback from receivers for past transmissions to form a best guess of what the current wireless channel state may be. As wireless channels tend to be highly variable in nature, systems either use conservative guesses to ensure a high packet success rate, or use higher layer mechanisms such as retries. Essentially, not having real-time channel state information at the transmitter leads to sub-optimal wireless channel use.

The full-duplex backchannel may be used for sending real-time channel state as

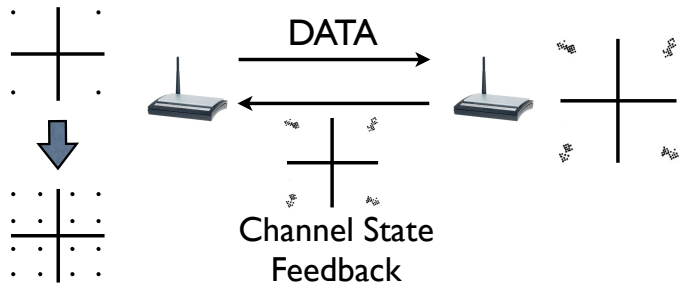


Figure 6.3: Real-time feedback for rate adaptation. Receiver sends perceived constellation. Transmitter uses this feedback to adapt constellation real-time.

observed by the receiver. This real-time knowledge of the receiver's channel state is known in information theory as Channel Side Information at the Transmitter (CSIT) [31]. CSIT has been assumed to be unrealistic and used in many theoretic algorithms to show achievability of channel capacity.

With a full-duplex system, CSIT is now practical. Therefore, many capacity-achieving theoretic schemes such as waterfilling are now practical as well [31]. Waterfilling schemes provide a framework for a transmitter to change its transmit power, modulation and datarate, according to the channel state, to maximize link throughput. As an example, Figure 6.3 shows how feedback from a receiver allows its transmitter to adapt modulation in real-time. The receiver simply can send the received constellation periodically, while still receiving packets from the transmitter. This knowledge is useful for the transmitter to decide whether to use denser (sparser) constellation when the channel is good (bad). Current techniques allow a transmitter to change this modulation for every packet [56]. With the real-time feedback, a full-duplex transmitter can do this adaptation during a packet transmission. Specifically, this real-time adaptation can be used by wireless video streaming devices that, when ON, continuously send video streams to their receivers such as a TV set.

As an example, we can consider OFDM signaling, which uses multiple sub-channels to encode and send data. With waterfilling based on real-time channel state feedback, a transmitter may decide to use different modulation densities for different frequency sub-channels, based on the state of each sub-channel. Such mechanisms have been successfully used in wired networks, such as DSL, where the channel tends to be fairly

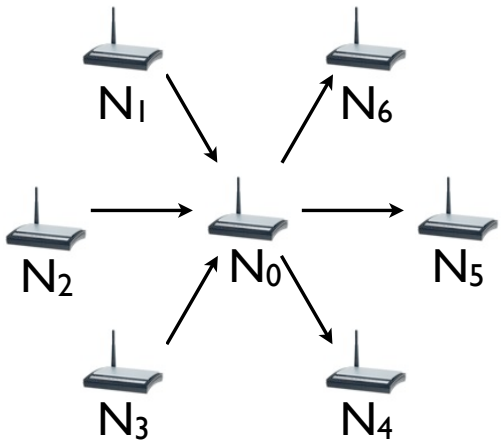


Figure 6.4: A star topology multihop network. Node N0 becomes a congested node. The network throughput in regular MAC operation is 1/n for 2n+1 nodes.

stable, but have been difficult to use in wireless due to its dynamic nature and lack of real-time channel state information. The real-time feedback channel realized with full-duplexing solves this problem.

The channel state feedback being available at the transmitter is even more beneficial for MIMO systems. MIMO systems use channel state information to pick an optimal operating point between using multiple antennas for sending multiple streams, or for sending fewer streams more robustly, or with a higher data rate. A real-time feedback mechanism can thus increase the gains achieved with MIMO systems.

6.2 Data Forwarding in Multihop Networks

Full-duplexing wireless nodes can provide significant performance gains in multihop networks. In a multihop wireless setup, a full-duplex forwarding node can forward data to the next hop, while simultaneously receiving data from the previous hop. This mechanism can reduce congestion in the network and significantly decrease the end-to-end latency of the network by using wireless cut-through routing.

6.2.1 Reducing Congestion due to MAC Scheduling

Figure 6.4 shows a network in star topology. Nodes N1, N2, and N3 have data to send to nodes N4, N5, and N6 respectively. All data has to be routed through node N0, and N0-N3 are in the interference range of each other. If all three source nodes have saturated flows to be sent to their respective destinations, nodes N0-N3 constantly contend with each other for channel access. Assuming typical MAC scheduling, N0 gets $1/4^{th}$ the total transmission opportunities. This restricts the aggregate network throughput to $1/4^{th}$ the capacity of one link.

In a general star topology with 2n+1 nodes with n nodes trying to route data via node N0, the aggregate network throughput is 1/n.

With full-duplexing, N0 can transmit and receive at the same time. For each transmission from either node N1, N2, or N3, N0 can forward a packet to a destination. Thus, the aggregate network throughput is equal to the single link capacity. Full-duplex eliminates the loss of network throughput due to congestion and MAC scheduling by allowing congested nodes to forward out packets and receive packets at the same time.

6.2.2 Cut-through Routing in Multihop Networks

Multihop networks suffer from long end-to-end delays causing loss in performance for delay sensitive protocols like TCP. Further, multihop networks have a $1/3^{rd}$ throughput scaling compared to single hop networks due to interference between forwarding hops.

The idea of receiving and forwarding at the same time can be extended to solve these problems. The insight is that as a full-duplex node is starting to receive a packet it can simultaneously start to forward it. Thus, instead of the default store-and-forward architecture, full-duplex nodes can forward a packet while receiving it. This idea is similar to cut-through switching [24] used for multihop wired communication networks. The technique can theoretically reduce the end-to-end delay for packet delivery through a multihop network from a packet time multiplied by number of hops to a little more than a single packet time.

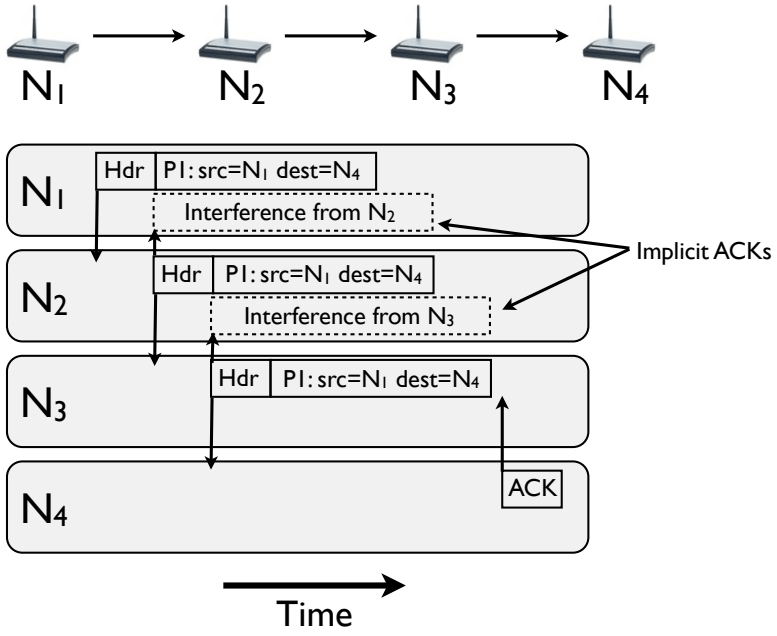
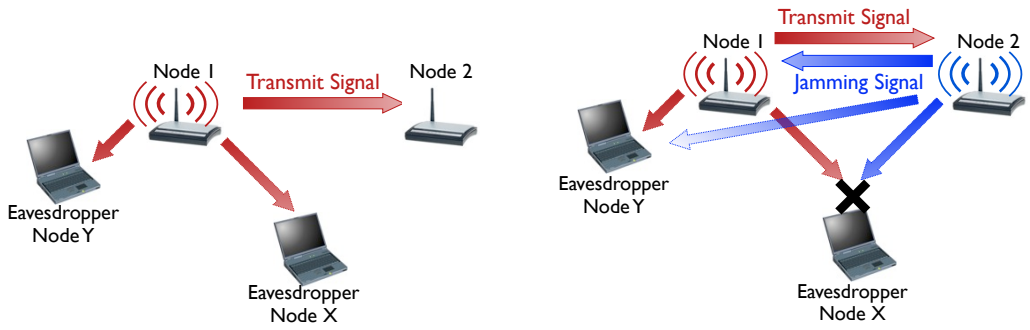


Figure 6.5: Wormhole switching in a multihop network. Interference from forwarding hops can be canceled using digital cancellation and can also serve as implicit ACKs.

Figure 6.5 shows the way cut-through switching can work on full-duplex wireless links. N2 starts receiving a packet from N1. As soon as N2 has processed the packet header, it knows where to forward the packet and starts transmitting the packet to N3. Similarly, N3 starts forwarding the packet to N4. At this time, N3's transmission also interferes with the reception at N2. Since N2 knows the part of the packet N3 would be transmitting at this time, it can use digital cancellation techniques to cancel N3's transmission. Further, once N2 has finished receiving the packet from N1, it can again apply digital cancellation to previously received samples from N1 and N3 to cancel the samples received from N1. This allows N2 to check the packet transmission from N3. This can act as an implicit ACK mechanism, removing the need of an explicit ARQ scheme. The last node in the route sends an explicit ACK to the second-to-last node in the route. Existing work has suggested a similar implicit ARQ scheme for a multi-channel wireless network used as an interconnect backbone for chip multi-processors [43].



(a) Half-Duplex (b) Full-Duplex

Figure 6.6: Full-duplexing can prevent eavesdropping of wireless data. Eavesdropper Node X cannot decode Node 1's data when Node 2 sends a jamming signal at the same time. A well placed eavesdropper (Node Y) may still successfully eavesdrop.

6.3 Security with Full-Duplex

Full-duplex wireless can also help make wireless networks inherently more secure to eavesdropping. Typically, the broadcast nature of wireless enables eavesdroppers to easily listen to any user traffic and higher layer encryption mechanisms are needed to prevent such eavesdropping. Figure 6.6(a) shows such a situation. Node 1 is transmitting data to Node 2 and eavesdroppers Node X and Node Y can easily listen to this transmission. With full-duplex, such eavesdropping becomes much more difficult. An eavesdropper trying to listen to a full-duplex communication hears the superposition of the signals transmitted from the two communicating nodes. As Figure 6.6(b) shows, with full-duplex Node X hears the sum of the signals coming from Node 1 and Node 2, and thus is unable to decode either of the two signals. This provides an extra layer of security in the wireless communication stack.

Even when Node 2 does not have data to send back to Node 1, it can send some form of a jamming signal which precludes almost all other nodes in the network from decoding Node 1's transmission. Recent work has applied full-duplexing to improving security in implanted medical devices [33]. The same principles can also be applied to other domains, including enterprise data networks and military communications.

Although full-duplexing enhances the security of the network by making it harder for eavesdroppers to decode overheard data, this method is not foolproof and should

not be considered a complete security solution. Specifically, if the eavesdropper is able to place itself fairly close to the transmitting node, while being away from the jamming node, it may be able decode the transmitted signal in spite of the jamming signal. For example, going back to Figure 6.6(b), even though Node X cannot eavesdrop on Node 1's transmission, Node Y which is much closer to Node 1 may be able to decode Node 1's transmissions even with Node 2 transmitting a jamming packet at the same time.

Chapter 7

Discussion

This dissertation motivates a new paradigm in wireless network design: *full-duplex wireless networking*. The most important challenge in implementing a full-duplex wireless system is canceling self-interference. This dissertation discusses various implementation techniques for self-interference cancellation, ultimately leading to a prototype achieving ≈73dBm reduction in self-interference. This prototype uses a combination of signal inversion RF cancellation using a balun circuit and adaptive digital cancellation.

Further, this dissertation discusses the many gains possible with full-duplex wireless and evaluates a small subset of those gains by implementing a real-time MAC for a full-duplex wireless LAN network. While this work provides a first look and an important step towards changing wireless networks with full-duplexing, it leaves several open questions and possibilities for future research. This chapter briefly discusses some of them. Specifically, this chapter compares how full-duplex would fare against 2x2 MIMO purely from a physical layer throughput standpoint, discusses some of the RF engineering challenges to be addressed in making full-duplexing radio feasible and suggests some improvements in the current hardware implementation of full-duplex radios to extend their usefulness.

7.1 Comparison with MIMO

The performance evaluation in Chapter 5 shows the MAC benefits of full-duplex. Can full-duplex also provide any gain over half-duplex in terms of physical layer throughput? In particular, balun cancellation requires two antennas and can double throughput. However, with the same resources, one can build a half-duplex multi-input multi-output (MIMO) system which achieves the same gain. This raises a natural question: under what conditions might a wireless system benefit more from one technique than the other?

If all communication is in one direction, then MIMO performs better, as it doubles the throughput of a single direction. If communication is bidirectional, however, the tradeoff is less clear. This section aims to provide some insight into the tradeoffs between half-duplex MIMO and full-duplex in terms of information theoretic channel capacity under different conditions.

The capacity analysis in this section, for both a 2x2 MIMO channel and a full-duplex channel, uses the simplest case of two nodes constantly trying to send data to each other. Both nodes have two antennas that can be used for transmit or receive. The analysis compares performance when the nodes use the two antennas to implement a 2x2 MIMO system vs implement a full-duplex system.

This analysis makes two assumptions:

- A wireless node has the *same power constraint during transmission* regardless of its duplex mode. This means that a half-duplex node cannot double its transmit power even though it remains silent half of the time (compared to the full-duplex node which always transmits). This is a practical constraint since FCC regulations and circuit limitations on transmit power restrict the maximum power coming out of a single device regardless of the duplex mode, making it infeasible to perform such power pulling across time for half-duplex nodes. Under this assumption, a 2x2 MIMO system is able to use the maximum power of one node at any time, since only one of two communicating nodes can transmit at a given time. On the other hand, with full-duplexing, the system can use the maximum power of both communicating nodes at the same time,

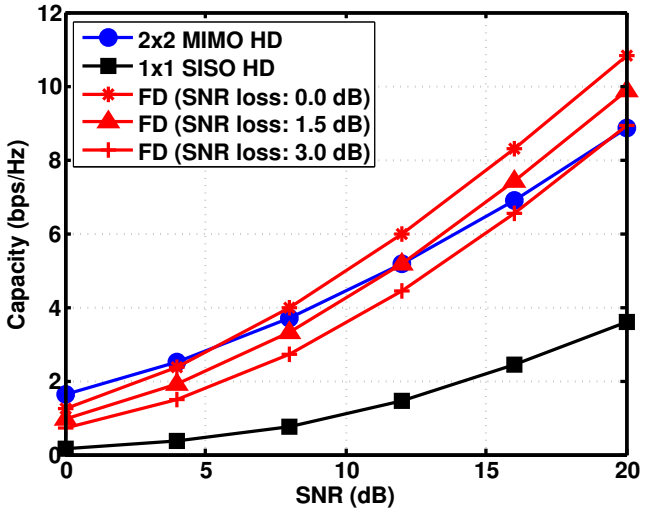


Figure 7.1: Capacity comparison of the proposed full-duplex system and the 2×2 MIMO half-duplex system

thus allowing the use of more total power in the system.

- The transmitter knows the state of the wireless channel from itself to the receiver perfectly. For a MIMO system, this increases capacity through an additional transmitter processing technique, called MIMO pre-coding. In case of full-duplexing, if both the nodes know the channels between all the antenna pairs, they can agree on the best transmit-receive antenna pairs to maximize the sum capacity in both directions. This assumption is valid for most new wireless systems, like 802.11n and LTE [6, 7], which use periodic feedback to inform the transmitter about the channel state.

Figure 7.1 shows the 10%-outage capacity of the wireless link for 2x2 MIMO half-duplex vs full-duplex for different levels of self-interference cancellation performance. Cancellation performance is modeled in terms of the SNR loss due to residual self-interference compared to half-duplex. The details of channel capacity analysis for full-duplex and 2x2 MIMO half-duplex, and the formal definition of the outage capacity are presented in Appendix A.4.

The figure shows that at low SNR, MIMO half-duplex outperforms full-duplex.

This result is expected with the diversity gains in MIMO helping its performance for lower SNR values. However, at higher SNRs, full-duplex achieves higher average capacity as long as the SNR loss remains below 1.5 dB. While surprising at a first glance, we can see that such gain actually comes from the power constraint per device allowing full-duplex to use twice the energy per unit time as compared to MIMO half-duplex.

Knowing channel state information in practice means different things for a full-duplex system and a MIMO system. Channel state information in the full-duplex setup allows the system to adaptively pick one of its antennas for transmit, rather than always using the same transmit antenna. This means that a full-duplex system only requires a one bit feedback from receiver to transmitter for the full-duplex system to exploit some gains of channel state knowledge at the transmitter. MIMO systems, on the other hand, require at least a few bits of feedback to program transmit precoders for achieving near-optimal performance [45].

Although not discussed in this section, we have also analyzed the case when the transmitter does not have any information about the state of the wireless channel. Without channel state, the transmitter cannot change its rate in response to channel changes and should fix its rate in advance. In this case, MIMO half-duplex outperforms full-duplex over the entire SNR range even with ideal cancellation. The spatial diversity of the MIMO system improves the reliability of the wireless channel, leading to better performance. However, having no channel state information is not a practical scenario for current multi-antenna systems, which always use some form of state feedback for rate adaptation.

This section shows that for different channel conditions the performance of full-duplex can exceed or lag behind the performance of a similarly resourced 2x2 MIMO system. Interestingly, on top of its superior MAC performance, full-duplexing can also provide better physical layer performance at high SNRs.

More importantly, this evaluation shows that MIMO half-duplex and full-duplex offer their respective advantages under different scenarios: robustness in low SNR scenarios using MIMO, and higher efficiency with full-duplex under high SNR. Thus, high performing systems can adopt a hybrid of the two modes depending on instantaneous

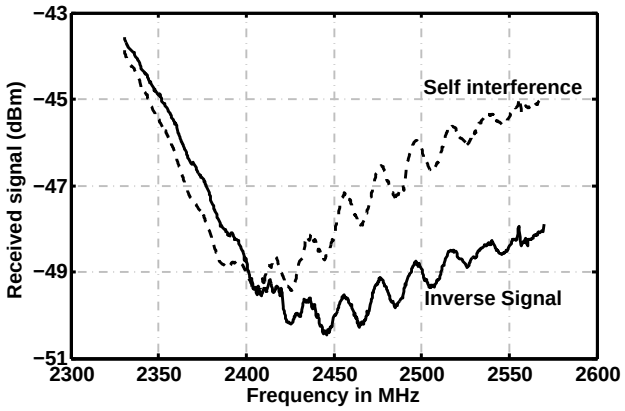


Figure 7.2: Frequency response of a previous version of our balun circuit. The frequency selective mismatch, caused by poor layout, prevented balun cancellation beyond 25 dB.

channel conditions and availability of channel state at the transmitter. A significant question going forward is whether full-duplex can be combined with MIMO systems. To make a MIMO system full-duplex, each receive antenna would have to cancel the self-interference introduced by all transmit antennas, requiring multiple cancellation circuits per node. This requirement seems challenging right now, but with miniaturization of components, full-duplex MIMO systems can be made possible. If so, this introduces an interesting degree of freedom.

7.2 RF Engineering

One issue that comes up numerous times in the full-duplex design is the sensitivity and precision of the cancellation components. For example, one early design of the balun circuit had a much less even frequency response, shown in Figure 7.2. This uneven response was partially from RF echoes in the balun board and placed an upper bound on the maximum cancellation possible. Consider the precision involved: canceling 50 dB of self-interference requires that the inverted signal be within 10^{-5} of the reference signal, or 99.999% accurate. These accuracies are clearly possible – a small group of researchers were able to achieve them with commodity components.

The practical limitations of balun cancellation remain an open question: if engineered as carefully as mobile phones, for example, much greater cancellation may be possible. The need for in-line high-precision attenuation and delay circuits may introduce additional, practical challenges: the QHx220 is a poor substitute.

One question left unanswered is if high-precision RF components can be mass manufactured on a chip small enough to fit in a mobile phone. Fitting to such small form factors bring in challenges from a size point of view as discussed in Chapter 3: a 2.4GHz delay line would require an effective range of 12cms length to allow delay adjustments up to 1 wavelength. This is clearly impractical and needs a rethinking in terms of the design of programmable attenuation and delay. One option to consider, for example, is MEMS based varactor devices for delay lines [27]. These devices provide voltage controlled capacitance which can provide the delay range required in form factors that can easily fit a small chip.

Ultimately, the performance of self-interference cancellation and the success of full-duplex systems would depend heavily on the size and accuracy of the devices used to implement the system. Research on new structures for implementing on-chip variable delay lines and attenuators for a variety of range and accuracy requirements can lead to an improved cancellation design in the future [46, 19, 23, 62, 38, 26].

7.3 Protocol Implementation Improvements

The current implementation of a full-duplex MAC described in Chapter 5 gives evidence that a full-duplex MAC can significantly mitigate some of the problems that plague wireless networks today, such as hidden terminals and unfair channel allocation between uplink and downlink flows. On the other hand, Chapter 6 describes many other improvements to wireless networks possible with full-duplexing used in various contexts to improve control feedback or data delivery. Implementing and evaluating these applications of full-duplex would require significant improvements in our current design. This section discusses some of those improvements.

Asymmetric Traffic Handling in MAC Our current MAC design only sends one secondary packet for a primary packet. Further, the secondary receiver in the

current design is always the same as the primary transmitter. However, for many scenarios, such as multihop networks, having a different secondary receiver is more beneficial as discussed in Section 6.2. Further, for asymmetrically sized traffic like TCP, multiple short TCP acks can be transmitted while receiving one long TCP packet. More generally, efficiently utilizing the additional capacity of the secondary channel, rather than wasting it with busy tones, is an open question.

In-Packet Channel Estimation The current full-duplex prototype uses periodic sounding packets for tuning the cancellation mechanisms. This method would suffice for a network with static or slowly moving nodes. For more dynamic environments, such as cellular and mobile networks, the channel state can change very quickly. This requires very frequent updates to the self-interference channel estimate. Using in-packet techniques to update channel estimates on a per-packet basis can address this challenge.

Half-Duplex Compatibility One interesting question going forward is how full-duplex systems can be incrementally deployed. For example, while the full-duplex system presented does not preclude coexistence with existing half-duplex systems, secondary transmissions need to know whether the primary is full-duplex capable, otherwise there may be poor interactions with link layer retry counts. Or perhaps secondary transmissions should be considered as simple opportunistic receptions.

7.4 Conclusion

Full-duplex wireless breaks a fundamental assumption that has dictated the design of wireless systems. The aim of this dissertation is to motivate full-duplexing as a direction for research into what future wireless systems would look like. To address the underlying challenge of self-interference cancellation, this dissertation presents the design and evaluates various analog and digital signal processing techniques which show promising results. The prototype presented in this work combines signal inversion analog cancellation with digital cancellation to achieve 73dB reduction in self-interference cancellation. The prototype can also adapt to channel dynamics with adaptive digital and analog schemes. Further improvements in the design and

faster adaptation techniques can enable full-duplexing in very challenging wireless environments such as cellphone handsets and basestations.

Exploiting full-duplex requires redesigning higher layers in the networking stack. Chapter 5 provides promising results for full-duplexing applied to a wireless LAN setting with simple MAC layer changes and Chapter 6 discusses how full-duplex could significantly improve many other types of wireless networks.

This initial work shows how different areas of research can come together to improve wireless systems in a very fundamental way. Future research opportunities in full-duplex wireless include core RF engineering to push the limits of cancellation performance and innovations in physical layer design, such as incorporating MIMO based full-duplexing. Full-duplexing also leads to a rethinking of how networking researchers design routing algorithms for multi-hop wireless networks. This dissertation provides many ideas on how full-duplexing could affect various aspects of wireless research. However, it is impossible to gauge the full impact of this technology on the design of systems of the future. Full-duplexing holds the promise to replace wireless everywhere.

Appendix A

Mathematical Derivations and Psuedo Code

A.1 Received Power with Phase Offset Cancellation

Let the unit power baseband signal be $x[t]$. The signal is scaled by different transmission amplitudes A_1 and A_2 at the two transmit antennas. The transmitted signals undergo attenuations Att_1 and Att_2 and phase shifts ϕ_1 and ϕ_2 in the wireless channel before reaching the receive antenna. The received signal is then given by:

$$\frac{A_1}{Att_1}x[t]e^{j(2\pi f_c t+\phi_1)} + \frac{A_2}{Att_2}x[t]e^{j(2\pi f_c t+\phi_2)}$$

Ideally, $\frac{A_1}{Att_1} = \frac{A_2}{Att_2}$, but in practical systems, it would be impossible to get the amplitudes from the two transmit signals to match exactly at the receive antenna.

We let $\frac{A_1}{Att_1} = A_{ant}$ and represent the amplitude mismatch by ϵ^A_{ant}, thus giving $\frac{A_2}{Att_2} = A_{ant} + \epsilon^A_{ant}$. Further, the two transmit symbols ideally are exactly π out of phase from each other when they are received at the receive antenna ($\phi_2 = \phi_1 + \pi$). Since the signal transmitted is not a single frequency, but rather a band of frequencies, and due to the constraints of practical systems, we take $\phi_2 = \phi_1 + \pi + \epsilon^\phi_{ant}$. This gives

the received signal as:

$$\begin{aligned}
&A_{ant}x[t]e^{j(2\pi f_c t+\phi_1)} + \left(A_{ant}+\epsilon^A_{ant}\right)x[t]e^{j\left(2\pi f_c t+\phi_1+\pi+\epsilon^\phi_{ant}\right)}\\
=&A_{ant}x[t]e^{j2\pi f_c t}e^{j\phi_1}\left(1-e^{j\epsilon^\phi_{ant}}\right) - \epsilon^A_{ant}x[t]e^{j\left(2\pi f_c t+\phi_1+\epsilon^\phi_{ant}\right)}
\end{aligned}$$

The instantaneous power of any complex signal $r[t]$ is given by $r[t]\overline{r[t]}$ where $\overline{r[t]}$ is the complex conjugate of the signal. Thus, the received signal power is:

$$\begin{aligned}
&\left\{A_{ant}x[t]e^{j2\pi f_c t}e^{j\phi_1}\left(1-e^{j\epsilon^\phi_{ant}}\right) - \epsilon^A_{ant}x[t]e^{j\left(2\pi f_c t+\phi_1+\epsilon^\phi_{ant}\right)}\right\} *\\
&\left\{A_{ant}\overline{x[t]}e^{-j2\pi f_c t}e^{-j\phi_1}\left(1-e^{-j\epsilon^\phi_{ant}}\right) - \epsilon^A_{ant}\overline{x[t]}e^{-j\left(2\pi f_c t+\phi_1+\epsilon^\phi_{ant}\right)}\right\}\\
=&A^2_{ant}x[t]^2\left(2-e^{j\epsilon^\phi_{ant}}-e^{-j\epsilon^\phi_{ant}}\right)+\\
&A_{ant}\epsilon^A_{ant}x[t]^2\left(2-e^{j\epsilon^\phi_{ant}}-e^{-j\epsilon^\phi_{ant}}\right)+\left(\epsilon^A_{ant}\right)^2|x[t]|^2\\
=&2A_{ant}\left(A_{ant}+\epsilon^A_{ant}\right)|x[t]|^2\left(1-cos\left(\epsilon^\phi_{ant}\right)\right)+\left(\epsilon^A_{ant}\right)^2|x[t]|^2
\end{aligned}$$

The phase error occurs due to a small deviation in the receiver antenna placement. The phase shift ϕ depends on the distance d between the transmit and receive antennas and is given by $\frac{2\pi d}{\lambda}$, where λ is the transmission wavelength. Thus, the phase error ϵ^ϕ_{ant} can be represented as $\frac{2\pi\epsilon^d_{ant}}{\lambda}$, where ϵ^d_{ant} is the error in receiver antenna placement. The received power thus becomes:

$$2A_{ant}\left(A_{ant}+\epsilon^A_{ant}\right)|x[t]|^2\left(1-cos\left(\frac{2\pi\epsilon^d_{ant}}{\lambda}\right)\right)+\left(\epsilon^A_{ant}\right)^2|x[t]|^2.$$

A.2 Received Power Convexity With Analog Cancellation

The adaptive analog cancellation uses received energy as a measure for cancellation performance and tries to minimize the energy to maximize performance. Here we analyze mathematically the expression for energy for analog cancellation implemented on a system using OFDM signalling. We model both the ideal setup involving signal inversion and passive adjustable delay and attenuation, and the practically implemented setup that uses the QHx220 noise cancellation chip as a substitute for the delay and attenuator.

We first model the OFDM signal. Assume the signal uses M subcarriers and $g(t)$ is the rectangular pulse of length T_o used for the pulse shaping of the OFDM signal. Assume M to be even. We can write $x(t)$, the baseband OFDM pulse as follows,

$$x(t) = \sum_{k=-\frac{M}{2}}^{k=\frac{M}{2}-1} x_k g(t) \exp(jk2\pi f_o t)$$

where $f_o = 1/T_o$. The up converted OFDM signal is given by:

$$\begin{aligned} s(t) =& \Re\{x(t) \exp(j2\pi f_c t)\} \\ s(t) =& \sum_{k=-\frac{M}{2}}^{k=\frac{M}{2}-1} r\ g(t) \cos(2\pi(f_c + kf_o)t) \end{aligned} \tag{A.1}$$

$s(t)$ is the transmitted OFDM signal.

A.2.1 Modeling For an Ideal Delay and Attenuator

We assume that there is no multipath. Let τ_a be the delay of the self-interference signal over the air, relative to the cancellation signal received via wire. We also assume that the self-interference signal is delayed more than the cancellation signal i.e. $\tau_a > 0$. Let G_a be the attenuation of the signal via air, and assume the signal over wire does not attenuate. We have two control variables to modify the cancellation signal, delay and attenuation. Let the control delay and attenuation be τ_c and G_c respectively. These two variables have to be adjusted for optimal cancellation. The

receiver gets the sum of two signals, the self-interference $r_a(t) = G_a * s(t - \tau_a)$ and the cancellation signal $r_c(t) = -G_c * s(t - \tau_c)$.

The phase and attenuation match when $G_a = G_c$ and $\tau_a = \tau_c$. We assume that the system is engineered to keep the required control delay range to less than half a wavelength, i.e. $\tau_a \leq 1/2f_c$. We also assume that the delay control line has a range of one full wavelength's delay, i.e. $0 \leq \tau_c \leq 1/fc$

For a T_o length OFDM symbol, the received energy E is given by the square of the received signal integrated over time T_o. Thus:

$$E = \int_{T_o} (r_a(t) + r_c(t))^2 \, \mathrm{d}t.$$

$$E = \int_{T_o} (G_a s(t - \tau_a) - G_c s(t - \tau_c))^2 \, \mathrm{d}t.$$

Define $\tau = \tau_c - \tau_a$. Note that$-1/2f_c \leq \tau \leq 1/2f_c$ since $\tau_a \leq 1/2f_c$ and control delay is always positive, i.e. $\tau_c > 0$.

Changing the limits $\hat{t} = t - \tau_a$

$$E = \int_{T_o} (G_a s(\hat{t}) - G_c s(\hat{t} - \tau))^2 \, \mathrm{d}\hat{t}.$$

$$E = \int_{T_o} r^2 \Big\{ \sum_k \big(\, G_a g(\hat{t}) \cos(2\pi(f_c + kf_o)\hat{t}) - G_c g(\hat{t} - \tau) \cos(2\pi(f_c + kf_o)(\hat{t} - \tau)) \, \big) \Big\}^2 \, \mathrm{d}\hat{t}.$$

The above expression is the product of two identical summations. We consider the l_{th} term of the first sum and the m_{th} term of the second sum. The product of those two terms is given by:

$$\begin{aligned} E_{l,m} = \int_{T_o} r^2 \Big\{ \big(\, G_a g(\hat{t}) \cos(2\pi(f_c + lf_o)\hat{t}) - G_c g(\hat{t} - \tau) \cos(2\pi(f_c + lf_o)(\hat{t} - \tau)) \, \big) \Big\} \cdot \\ \Big\{ \big(\, G_a g(\hat{t}) \cos(2\pi(f_c + mf_o)\hat{t}) - G_c g(\hat{t} - \tau) \cos(2\pi(f_c + mf_o)(\hat{t} - \tau)) \, \big) \Big\} \, \mathrm{d}\hat{t} \\ = r^2 \{ E_{11} + E_{12} + E_{21} + E_{22} \} \end{aligned} \tag{A.2}$$

where

$$E_{11} = \int_{T_o} G_a^2 g^2(\hat{t}) \cos(2\pi(f_c + lf_o)\hat{t}) \cos(2\pi(f_c + mf_o)\hat{t})\, \mathrm{d}\hat{t}$$
$$E_{12} = \int_{T_o} -G_a G_c g(\hat{t}) g(\hat{t}-\tau) \cos(2\pi(f_c + lf_o)\hat{t}) \cos(2\pi(f_c + mf_o)(\hat{t}-\tau))\, \mathrm{d}\hat{t}$$
$$E_{21} = \int_{T_o} -G_a G_c g(\hat{t}-\tau) g(\hat{t}) \cos(2\pi(f_c + lf_o)(\hat{t}-\tau)) \cos(2\pi(f_c + mf_o)\hat{t})\, \mathrm{d}\hat{t}$$
$$E_{22} = \int_{T_o} G_c^2 g^2(\hat{t}-\tau) \cos(2\pi(f_c + lf_o)(\hat{t}-\tau)) \cos(2\pi(f_c + mf_o)(\hat{t}-\tau))\, \mathrm{d}\hat{t}.$$

We analyze each of these terms individually.

$$\begin{aligned} E_{11} &= \int_{T_o} G_a^2 g^2(\hat{t}) \cos(2\pi(f_c + lf_o)\hat{t}) \cos(2\pi(f_c + mf_o)\hat{t})\, \mathrm{d}\hat{t}. \\ &= \int_{T_o} \frac{1}{2} G_a^2 g^2(\hat{t}) \{\cos(2\pi(2f_c + (l+m)f_o)\hat{t}) + \cos(2\pi(l-m)f_o\hat{t})\}\, \mathrm{d}\hat{t}. \\ &= \int_{T_o} \frac{1}{2} G_a^2 g^2(\hat{t}) \{\cos(2\pi(l-m)f_o\hat{t})\}\, \mathrm{d}\hat{t}. \end{aligned}$$

The last equality follows since f_c is large frequency integrated over multiple cycles resulting into zero. Further for $l \neq m$, E_{11} goes to zero for a rectangular pulse shape $g(t)$. For $l = m$, expression of E_{11} is,

$$E_{11} = T_o \frac{G_a^2}{2}. \tag{A.3}$$

Similarly, we can show that E_{22} is given by:

$$E_{22} = T_o \frac{G_c^2}{2}. \tag{A.4}$$

From this point on, we would always ignore the high frequency term involving the carrier frequency as it integrates to zero over T_0 We define autocorrelation of $g(t)$

$$R_g(\tau) = \int_{T_o} g(\hat{t}) g(\hat{t}-\tau) \mathrm{d}\hat{t}$$

Using the autocorrelation function, we can write E_{12} as:

$$E_{12} = \int_{T_o} -G_a G_c g(\hat{t}) g(\hat{t}-\tau) \cos(2\pi(f_c + lf_o)\hat{t}) \cos(2\pi(f_c + mf_o)(\hat{t}-\tau))\, \mathrm{d}\hat{t}$$

Ignoring the high frequency term,

$$E_{12} = \int_{T_o} -\frac{1}{2} G_a G_c g(\hat{t}) g(\hat{t}-\tau) \cos(2\pi((l-m)f_o\hat{t} + (f_c + mf_o)\tau))\, \mathrm{d}\hat{t}.$$

Thus for $l \neq m$, we have $E_{12} = 0$ when $g(t)$ is rectangular. For $l = m$, E_{12} is given

by:

$$E_{12} = -\frac{G_a G_c}{2} R_g(\tau)\cos(2\pi(f_c + mf_o)\tau)$$

Similarly, we can show $E_{21} = 0$ for $l \neq m$ and for $l = m$, $E_{21} = E_{12}$, i.e.

$$E_{21} = -\frac{G_a G_c}{2} R_g(\tau)\cos(2\pi(f_c + lf_o)\tau)$$

For a rectangular pulse shape $g(t) = rect(\frac{t}{T_o})$,

$$R_g(\tau) = T_o - |\tau|$$

Note that range of values τ can take is small, $-1/2f_c \leq \tau \leq 1/2f_c$. For these small values of τ, $R_g(\tau)$ is almost constant. So E_{12} and E_{21} can be written as:

$$E_{12} = E_{21} \approx -\frac{G_a G_c}{2} T_o \cos(2\pi(f_c + mf_o)\tau) \tag{A.5}$$

Collecting all terms and substituting in Equation A.2,

$$E = T_o\ r^2 \sum_k \{\frac{G_a^2 + G_c^2}{2} - G_a G_c \cos(2\pi(f_c + kf_o)\tau)\}$$

$$E = T_o\ r^2 \sum_k \{\frac{G_a^2 + G_c^2}{2} - G_a G_c \cos(2\pi(f_c + kf_o)(\tau_a - \tau_c))\}$$

Taking the second derivative w.r.t. G_c:

$$\frac{\partial^2 E}{\partial G_c^2} = T_o r^2 k$$

which is always positive. Taking the second derivative w.r.t. τ_c:

$$\frac{\partial^2 E}{\partial \tau^2} = T_o r^2 G_a G_c \sum_k (2\pi(f_c + kf_o))^2 \cos(2\pi(f_c + kf_o)(\tau_a - \tau_c))$$

For delay mismatch such that $-1/4f_c \leq \tau_a - \tau_c \leq 1/4f_c$, the the second derivative is positive, but it goes negative beyond these values. So the function is not technically convex in the entire operating range of $-1/2f_c \leq \tau_a - \tau_c \leq 1/2f_c$, so some traditional convex optimization algorithms based on second derivatives, such as Newton method, may not be used. On the other hand, with the optimal point at $\tau_a - \tau_c = 0$, the energy function E is monotonically decreasing for $\tau_a - \tau_c < 0$ and monotonically increasing for $\tau_a - \tau_c > 0$ within the operating range of $\tau_a - \tau_c$. Thus, first derivative methods, such as gradient descent can still be used effectively in this context.

A.2.2 Modelling for QHX220

With the noise canceler chip QHx220, we can tune the gain parameters for I-phase and Q-phase. The chip introduces a fixed delay multipath for generating the Q-phase component. Let the fixed delay introduced by the chip be τ_q, where $\tau_q = 1/4f_c$ for carrier frequency f_c. Using the OFDM signal described in Equation A.1:

$$s(t) = \sum_{k=1}^{M} x_k g(t) \cos(2\pi(f_c + kf_o)t),$$

we can write the input to the receive terminal after analog noise cancellation as:

$$r(t) = G_a s(t - \tau_a) - G_i s(t) - G_q s(t - \tau_q)$$

where G_a and τ_a are the over the air gain and delay for the self-interference signal, and G_i and G_q are the I and Q phase gain settings in the QHx220 chip respectively. The corresponding signal energy is:

$$E = \int_{T_o} (r(t))^2 \mathrm{d}t$$

$$E = \int_{T_o} (G_a s(t - \tau_a) - G_i s(t) - G_q s(t - \tau_q))^2 \mathrm{d}t$$

Differentiating with respect to G_i twice gives:

$$\frac{\partial^2 E}{\partial G_i^2} = \int_{T_o} (s(t))^2 \mathrm{d}t$$

Similarly, for G_q:

$$\frac{\partial^2 E}{\partial G_q^2} = \int_{T_o} (s(t - \tau_q))^2 \mathrm{d}t$$

Both the second derivatives are clearly positive, thus showing the convexity of the received energy with the two gain parameters.

A.3 Pseudocode for Adaptive Analog Cancellation Using QHx220

Data: Use previous G_i^{intial}, G_q^{intial}
$G_i \leftarrow G_i^{intial}$, $G_q \leftarrow G_q^{intial}$;
Initialize $R_{cur} = read\ rssi(G_i,\ G_q)$;
$d \leftarrow derivative\ step\ size$, $\Delta \leftarrow step\ size$;
$start \leftarrow 1$;
while *start* **do**
  // Sample four points $G_i \pm d$ and $G_q \pm d$ for RSSI ;
  $R_1 \leftarrow read\ rssi(G_i + d, G_q)$;
  $R_2 \leftarrow read\ rssi(G_i, G_q + d)$;
  $R_3 \leftarrow read\ rssi(G_i - d, G_q)$;
  $R_4 \leftarrow read\ rssi(G_i, G_q - d)$;
  calculate slopes, $s_i \leftarrow \frac{R1-R3}{2d}$, $s_q \leftarrow \frac{R2-R4}{2d}$;
  Update G_i, G_q in the $(s_i\ , s_q)$ direction with Δ units as radius ;
  $R_{new} = read\ rssi(G_i, G_q)$;
  if $R_{new} >= min(R_{cur}, R_1, R_2, R_3, R_4)$ **then**
    close to the minimum, so decrease both Δ and d ;
    check false alarm, if step size decreases and stuck in the noisy minimum ;
    if $R_{cur} = min(R_{cur}, R_1, R_2, R_3, R_4)$ **then**
      if this occur for few times consecutively, minimum is reached;
      Store G_i^{intial} , G_q^{intial} and $start \leftarrow 0$;
    end
    $R_{cur} \leftarrow min(R_{cur}, R_1, R_2, R_3, R_4)$;
    update $G_i^{intial}, G_q^{intial}$ to minimum point;
  else
    $R_{cur} \leftarrow R_{new}$;
    update $G_i^{intial}, G_q^{intial}$ to new point;
    check for false alarm, if the step size has decreased and global minimum is far away
  end
end

A.4 Capacity Analysis

We investigate the information theoretic channel capacity of the full-duplex with balun cancellation and the 2x2 MIMO half-duplex. A simple frequency flat block fading model is considered to highlight the underlying tradeoff between each duplex mode.

A.4.1 System Model

Consider a point-to-point link between two nodes A and B. Each node has two antennas. The channel between the nodes is specified by a 2×2 matrix $\mathbf{H}$ where (i, j)-th element of $\mathbf{H}$, denoted by h_{ij}, corresponds to the channel gain from the j-th antenna in node A to the i-th antenna in node B. For example, $h_{1,2}$ is the channel between antenna 1 of A and antenna 2 of B. Each h_{ij} is assumed to be an i.i.d. complex Gaussian random variable with zero mean and unit variance, i.e. Rayleigh fading, and remain the same for a given communication interval of T.

For full-duplex, we model the effect of the residual self-interference after all cancellation as an relative SNR loss $\beta \geq 1$. $\beta = 1$ implies an ideal full-duplex system that cancels self-interference perfectly, while $\beta = 2$ raises the noise floor by 3 dB.

Given a communication interval of T, each MIMO half-duplex node exclusively occupies $\frac{T}{2}$, while full-duplex nodes can transmit simultaneously over entire T. We do not consider inter-packet intervals or MAC backoff. This benefits MIMO because these "dead" periods do not overlap, unlike in a full-duplex system.

We assume the same average transmit power P for both duplex modes, but the average is taken over their respective duration when the node is actively transmitting, i.e. $\frac{T}{2}$ for half-duplex and T for full-duplex. This means that we do not allow a half-duplex node to double its power during transmission even though it remains silent the other half of the time. In practice, this can be more realistic assumption in the sense that such power pulling across time may not be feasible for power amplifiers with fixed dynamic range. Note that under this assumption, each half-duplex node spends $P \cdot (\frac{T}{2}) + 0 \cdot (\frac{T}{2}) = P \cdot (\frac{T}{2})$ amount of energy over the communication interval of T while each full-duplex node can use twice more energy, $P \cdot T$, to increase its data

rate.

A.4.2 Capacity Analytical Formulation

We investigate the capacity of the two modes in two different settings depending on the availability of the channel state information at the transmitter (CSIT). Let $\rho = \frac{P}{\sigma^2}$ denote the average received SNR at the receiver for a given average transmit power P and the noise variance at the receiver σ^2. From [59], given a MIMO channel $\mathbf{H}$, the capacity in bps/Hz without CSIT is:

$$\mathcal{C}_{\text{hd}}^{(\text{w/o CSIT})}(\mathbf{H}, \rho) = \sum_{i=1}^{2} C\left(\frac{\rho}{2}\lambda_i\right)$$

while capacity with CSIT is

$$\mathcal{C}_{\text{hd}}^{(\text{CSIT})}(\mathbf{H}, \rho) = \max_{P_i : \sum_i P_i \leq P} \sum_{i=1}^{2} C(\rho \frac{P_i}{P}\lambda_i),$$

where $C(x) := \log_2(1+x)$ and λ_i is the i-th largest eigenvalue of $\mathbf{H}\mathbf{H}^H$. For the half-duplex, the sum capacity of the two nodes over T is $\mathcal{C}_{\text{hd-sum}}^{(.)}(\mathbf{H}, \rho) = \mathcal{C}_{\text{hd}}^{(.)}(\mathbf{H}, \rho)$ since each node uses the channel half of the time, and the channel capacity in both direction is the same from the reciprocity of the channel.

For a full-duplex system without CSIT, we assume without loss of generality that each node uses the first antenna as its transmit antenna. Then, the sum capacity is the sum of the capacity of two independent SISO channels:

$$\mathcal{C}_{\text{fd-sum}}^{(\text{w/o CSIT})}(\mathbf{H}, \rho) = \sum_{i=1}^{2} C\left(\frac{\rho}{\beta}|h_{i\bar{i}}|^2\right),$$

where $\bar{i} = \{1, 2\} - i$. When CSIT is available, the sum capacity of the full-duplex system is

$$\mathcal{C}_{\text{fd-sum}}^{(\text{CSIT})}(\mathbf{H}, \rho) = \max\left(\mathcal{C}_{\text{fd-sum}}^{(\text{w/o CSIT})}(\mathbf{H}, \rho), \sum_{i=1}^{2} C\left(\frac{\rho}{\beta}|h_{ii}|^2\right)\right),$$

where the sum capacity gain from the $\max(\cdot, \cdot)$ operation is due to the adaptive transmit antenna selection based on CSIT.

Finally, the performance metric used for measuring the capacity of the wireless link is the 10% outage capacity. In practice, this measure gives the capacity of the

link such that it achieves a PRR of 90%. The ϵ-outage capacity is defined as

$$\mathcal{C}_{\text{outage}}(\epsilon)(\rho) = \max\{r \geq 0 | \Pr\{\mathcal{C}(\mathbf{H}, \rho) \leq r\} \leq \epsilon\}$$

Bibliography

[1] ANSI/IEEE Std 802.11 2003 Edition.

[2] 3GPP TS 45.001 version 10.0.0 specification.

[3] 3GPP2 A.S0014-D v3.0: Interoperability Specification (IOS) for cdma2000 Access Network Interfaces - Part 4.

[4] ADSL2+M: ITU G.992.5 Annex M Standard Document.

[5] VDSL2: ITU-T G.993.2 Standard Document.

[6] ANSI/IEEE Std 802.11n-2009 Amendment 5: Enhancements for Higher Throughput.

[7] `http://www.3gpp.org/article/lte`.

[8] Rice university warp project. http://warp.rice.edu.

[9] Universal software radio peripheral, ettus research llc. http://www.ettus.com.

[10] Second memorandum and opinion: Unlicensed operation in the tv broadcast bands, additional spectrum for unlicensed devices below 900 mhz and in the 3 ghz band. `http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-10-174A1.pdf`, 2010.

[11] S. Ayazian and R. Gharpurey. Feedforward interference cancellation in radio receiver front-ends. *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 54(10):902 –906, oct. 2007.

[12] Paramvir Bahl, Atul Adya, Jitendra Padhye, and Alec Walman. Reconsidering wireless systems with multiple radios. *SIGCOMM Comput. Commun. Rev.*, 34(5):39–46, 2004.

[13] Paramvir Bahl, Ranveer Chandra, Thomas Moscibroda, Rohan Murty, and Matt Welsh. White space networking with wi-fi like connectivity. *SIGCOMM Comput. Commun. Rev.*, 39(4):27–38, 2009.

[14] K. Berberidis and J. Palicot. A frequency-domain decision feedback equalizer for multipath echo cancellation. In *Global Telecommunications Conference, 1995. GLOBECOM '95., IEEE*, volume 1, pages 98 –102 vol.1, nov 1995.

[15] N. Blefari-Melazzi, A. Detti, I. Habib, A. Ordine, and S. Salsano. TCP Fairness Issues in IEEE 802.11 Networks: Problem Analysis and Solutions Based on Rate Control. *IEEE Transactions on Wireless Communications*, 6(4):1346–1355, 2007.

[16] D. W. Bliss, P. A. Parker, and A. R. Margetts. Simultaneous transmission and reception for improved wireless network performance. In *Proceeings of the 2007 IEEE Workshop on Statistical Signal Processing*, 2007.

[17] Bluetooth SIG, Inc. `http://www.bluetooth.org`.

[18] Danijela Cabric, Mike S. W. Chen, David A. Sobel, Stanley Wang, Jing Yang, and Robert W. Brodersen. Novel radio architectures for uwb, 60 ghz, and cognitive wireless systems. *EURASIP Journal on Wireless Communications and Networking*, 2006(17957):1–18, January 2006.

[19] Peiyan Cao. Optical or electrical—novel approaches to design true time delay phased array antennas. *International Journal of Advances in Optical Communication and Networks*, 1(1):1–18, December 2010.

[20] Y. K. Chan, V. C. Koo, B.-K. Chung, and H.-T. Chuah. A cancellation network for full-duplex front end circuit. *Progress In Electromagnetics Research Letters*, 7:139–148, 2009.

[21] Wei-Han Cheng. Cancellation circuit for transmit-receive isolation. Master's thesis, Naval Postgraduate School, 2010.

[22] Jung Il Choi, Mayank Jain, Kannan Srinivasan, Phil Levis, and Sachin Katti. Achieving single channel, full duplex wireless communication. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, MobiCom '10, pages 1–12, New York, NY, USA, 2010. ACM.

[23] Ta-Shun Chu, J. Roderick, and H. Hashemi. An integrated ultra-wideband timed array receiver in 0.13*mu*m cmos using a path-sharing true time delay architecture. *Solid-State Circuits, IEEE Journal of*, 42(12):2834 –2850, dec. 2007.

[24] William J. Dally and Charles L. Seitz. The torus routing chip. *Distributed Computing*, 1(4):187–196, 1986.

[25] Paul Defraeye, Dirk Rabaey, Wim Roggeman, Johan Yde, and Layos Kiss. A 3*mu*m; cmos digital codec with programmable echo cancellation and gain setting. In *Solid-State Circuits Conference, 1984. ESSCIRC '84. Tenth European*, pages 239 –243, sept. 1984.

[26] H. Dogan, R.G. Meyer, and A.M. Niknejad. Analysis and design of rf cmos attenuators. *Solid-State Circuits, IEEE Journal of*, 43(10):2269 –2283, oct. 2008.

[27] C. W. Domier and Jr. N. C. Luhmann. Rf mems extended tuning range varactor and varactor based true time delay line design. *PIERS Online*, 4(4):433–436, 2008.

[28] Melissa Duarte and Ashutosh Sabharwal. Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results. In *Forty-Fourth Asilomar Conference on Signals, Systems, and Components*, 2010.

[29] E. M. T. Jones George L. Matthaei, Leo Young. *Microwave filters, impedance-matching networks, and coupling structures*. Artech House Books, 1980.

[30] Ioan L. Gheorma and Ganesh K. Gopalakrishnan. Rf photonic techniques for same frequency simultaneous duplex antenna operation. *IEEE Photonics Letters*, 19(13), July 2007.

[31] Andrea Goldsmith. *Wireless Communications*. Cambridge Press, 2005.

[32] Shyamnath Gollakota and Dina Katabi. ZigZag decoding: combating hidden terminals in wireless networks. In *SIGCOMM '08: Proceedings of the ACM SIGCOMM 2008 conference on Data communication*, pages 159–170, New York, NY, USA, 2008. ACM.

[33] Shyamnath Gollakota and Dina Katabi. They can hear your heartbeats: Non-invasive security for implanted medical devices. In *SIGCOMM '11: Proceedings of the ACM SIGCOMM 2011 conference on Data communication*, New York, NY, USA, 2011. ACM.

[34] Daniel Halperin, Thomas Anderson, and David Wetherall. Taking the sting out of carrier sense: interference cancellation for wireless lans. In *MobiCom '08: Proceedings of the 14th ACM international conference on Mobile computing and networking*, pages 339–350, New York, NY, USA, 2008. ACM.

[35] Bo Han, Aaron Schulman, Francesco Gringoli, Neil Spring, Bobby Bhattacharjee, Lorenzo Nava, Lusheng Ji, Seungjoon Lee, and Robert Miller. Maranello: Practical partial packet recovery for 802.11. In *NSDI*, 2010.

[36] S. Hori and B. Murmann. Feedforward interference cancellation architecture for short-range wireless communication. *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 58(1):16 –20, jan. 2011.

[37] Senhua Huang, Xin Liu, and Zhi Ding. Optimal transmission strategies for dynamic spectrum access in cognitive radio networks. *IEEE Transactions on Mobile Computing*, 8, 2009.

[38] Yan-Yu Huang, Wangmyong Woo, Youngchang Yoon, and Chang-Ho Lee. Highly linear rf cmos variable attenuators with adaptive body biasing. *Solid-State Circuits, IEEE Journal of*, 46(5):1023 –1033, may 2011.

[39] Mayank Jain, Jung Il Choi, Taemin Kim, Dinesh Bharadia, Siddharth Seth, Kannan Srinivasan, Philip Levis, Sachin Katti, and Prasun Sinha. Practical real-time full-duplex wireless. In *Proceedings of the seventeenth annual international conference on Mobile computing and networking*, MobiCom '11, New York, NY, USA, 2011. ACM.

[40] Sachin Katti, Shyamnath Gollakota, and Dina Katabi. Embracing wireless interference: analog network coding. In *SIGCOMM '07: Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 397–408, New York, NY, USA, 2007. ACM.

[41] H. Khatri, P.S. Gudem, and L.E. Larson. A saw-less cmos cdma receiver with active tx filtering. In *Custom Integrated Circuits Conference, 2009. CICC '09. IEEE*, pages 379 –382, sept. 2009.

[42] Sang Wu Kim, Young Jin Chun, and Sangmun Kim. Co-channel interference cancellation using single radio frequency and baseband chain. *Communications, IEEE Transactions on*, 58(7):2169 –2175, 2010.

[43] Suk-Bok Lee, Sai-Wang Tam, Ioannis Pefkianakis, Songwu Lu, M. Frank Chang, Chuanxiong Guo, Glenn Reinman, Chunyi Peng, Mishali Naik, Lixia Zhang, and Jason Cong. A scalable micro wireless interconnect structure for cmps. In *MobiCom '09: Proceedings of the 15th annual international conference on Mobile computing and networking*, pages 217–228, New York, NY, USA, 2009. ACM.

[44] D. J. Leith, P. Clifford, D. Malone, and A. Ng. TCP Fairness in 802.11e WLANs. *IEEE Communications Letters*, 9(12), 2005.

[45] D.J. Love, Jr. Heath, R.W., and T. Strohmer. Grassmannian beamforming for multiple-input multiple-output wireless systems. In *Communications, 2003. ICC '03. IEEE International Conference on*, volume 4, pages 2618 – 2622 vol.4, may 2003.

[46] M. Meghdadi, M. Azizi, M. Kiani, A. Medi, and M. Atarodi. A 6-bit cmos phase shifter for *s*-band. *Microwave Theory and Techniques, IEEE Transactions on*, 58(12):3519 –3526, dec. 2010.

[47] M. Mikhemar, H. Darabi, and A. Abidi. A tunable integrated duplexer with 50db isolation in 40nm cmos. In *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, pages 386 –387,387a, feb. 2009.

[48] M. Mikhemar, H. Darabi, and A. Abidi. An on-chip wideband and low-loss duplexer for 3g/4g cmos radios. In *VLSI Circuits (VLSIC), 2010 IEEE Symposium on*, pages 129 –130, june 2010.

[49] Shridhar Mubaraq Mishra, Anant Sahai, and Robert W. Brodersen. Cooperative sensing among cognitive radios. In *In Proc. of the IEEE International Conference on Communications (ICC)*, 2006.

[50] K. H. Mueller. Combining echo cancellation and decision feedback equalization. In *The Bell System Technical Journal*, volume 58, pages 491–500, February 78.

[51] Quellan Inc. Qhx220 narrowband noise canceller ic. `http://www.quellan.com/products/qhx220_ic.php`.

[52] Bozidar Radunovic, Dinan Gunawardena, Peter Key, Alexandre Proutiere, Nikhil Singh, Vlad Balan, and Gerald Dejean. Rethinking indoor wireless mesh design: Low power, low frequency, full-duplex. In *Wireless Mesh Networks (WIMESH 2010), 2010 Fifth IEEE Workshop on*, pages 1 –6, june 2010.

[53] Bozidar Radunovic, Dinan Gunawardena, Alexandre Proutiere, Nikhil Singh, Vlad Balan, and Peter Key. Efficiency and fairness in distributed wireless networks through self-interference cancellation and scheduling. Technical Report MSR-TR-2009-27, Microsoft Research, 2009.

[54] Anant Sahai, Shridhar Mubaraq Mishra, and Rahul T. Spectrum sensing: Fundamental limits. *draft chapter for a Springer Book: Cognitive Radios: System Design Perspective*, June 2009.

[55] Souvik Sen, Romit Roy Choudhury, and Srihari Nelakuditi. CSMA/CN: carrier sense multiple access with collision notification. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, MobiCom '10, pages 25–36, New York, NY, USA, 2010. ACM.

[56] Souvik Sen, Naveen Santhapuri, Romit Roy Choudhury, and Srihari Nelakuditi. Accurate: Constellation based rate estimation in wireless networks. In *NSDI*, 2010.

[57] Yushi Shen and Ed Martinez. Channel Estimation in OFDM Systems. *Application Note, Freescale Semiconductor*, (AN3059), 2006.

[58] J. Suarez and P.R. Prucnal. System-level performance and characterization of counter-phase optical interference cancellation. *Lightwave Technology, Journal of*, 28(12):1821 –1831, june15, 2010.

[59] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecom.*, 10:585–595, November 1999.

[60] T.D. Werth, C. Schmits, R. Wunderlich, and S. Heinen. An active feedback interference cancellation technique for blocker filtering in rf receiver front-ends. *Solid-State Circuits, IEEE Journal of*, 45(5):989 –997, may 2010.

[61] Joseph F. White. *High frequency techniques: an introduction to RF and microwave engineering*. John Wiley and Sons, 2004.

[62] Tang Xinyi. *Broadband phase shifter design for phased array radar systems*. PhD thesis, National University of Singapore, August 2010.