

Full-Duplex Transceivers for Next-Generation Wireless Communication Systems

Ian P. Roberts and Himal A. Suraweera

Abstract Wireless communication systems can be enhanced at the link level, in medium access, and at the network level when transceivers are equipped with full-duplex capability: the transformative ability to simultaneously transmit and receive over the same frequency spectrum. Effective methods to cancel self-interference are required to facilitate full-duplex operation, which we overview herein in the context of traditional radios, along with those in next-generation wireless networks. We highlight advances in self-interference cancellation that leverage machine learning, and we summarize key considerations and recent progress in full-duplex millimeter-wave systems and their application in integrated access and backhaul. We present example design problems and noteworthy findings from recent experimental research to introduce and motivate the advancement of full-duplex millimeter-wave systems. We conclude this chapter by forecasting the future of full-duplex and outlining important research directions that warrant further study.

1 Introduction

For more than a century, wireless communication systems have almost exclusively operated in a half-duplex fashion, where transmission and reception of radio waves have typically been separated—or orthogonalized—in the time domain, frequency domain, or both. Put simply, signals transmitted or received by a traditional half-duplex system exist in different frequency bands or at different times, referred to as frequency-division duplexing (FDD) and time-division duplexing (TDD), respec-

Ian P. Roberts

Department of Electrical and Computer Engineering, University of California, Los Angeles, California 90095, USA, e-mail: ianroberts@ucla.edu

Himal A. Suraweera

Department of Electrical and Electronic Engineering, University of Peradeniya, Peradeniya 20400, Sri Lanka, e-mail: himal@eng.pdn.ac.lk

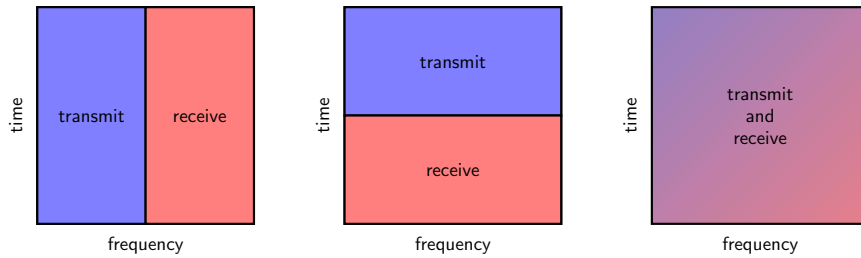


Fig. 1 The time-frequency resources consumed by transmission and reception with (left) FDD; (middle) TDD; and (right) full-duplex. In practice, guard bands are usually necessary for both FDD and TDD, which consumes additional radio resources.

tively. Half-duplex operation is necessitated by the manifestation of *self-interference* (SI) when a transceiver attempts to receive signals while simultaneously transmitting in the same spectrum. In most cases, SI is many orders of magnitude stronger than a relatively weak signal-of-interest (or *desired receive signal*), which has presumably propagated tens or hundreds of meters. This makes it virtually impossible to accurately recover the desired receive signal from their combination without taking additional measures to mitigate the effects of SI [1, 2]. By receiving in neighboring frequency spectrum or on a separate time slot as its transmissions, a half-duplex transceiver can avoid inflicting SI onto a desired receive signal, hence the usage of FDD and TDD.

By their nature, FDD and TDD both consume radio resources by dedicating time-frequency resources to either transmission or reception. Of course, this would not be an issue if practical systems were not resource-constrained. In reality, all practical wireless communication systems operate on limited time-frequency resources. At the very least, most systems are confined to certain frequency spectrum by regulatory bodies, such as the Federal Communications Commission (FCC) in the United States. The consumption of radio resources by half-duplex operation has motivated researchers to explore in-band *full-duplex* operation¹ [1, 3–7]. Starting in the late 2000s, researchers began heavily investigating and developing means to mitigate SI and bring full-duplex to life. Since then, full-duplex has matured and has recently found new life in millimeter-wave (mmWave) networks [8, 9], in joint communication and sensing [10], and through advancements in machine learning [11].

In this chapter, we introduce full-duplex operation and highlight its enhancements. Then, we overview full-duplex solutions for traditional radios and those for modern and next-generation wireless systems. We conclude with a look ahead at the future of full-duplex to motivate and steer its research, development, and deployment.

¹ We use the term “full-duplex” to refer to in-band full-duplex operation, in particular, as opposed to out-of-band full-duplex, which has been used to describe systems capable of simultaneously transmitting and receiving via FDD.

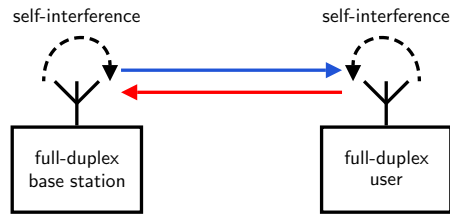


Fig. 2 A full-duplex base station transmitting and receiving in-band with a full-duplex user. SI manifests at both devices, requiring each to take measures to sufficiently cancel it.

1.1 What is Full-Duplex?

Full-duplex capability allows a transceiver to concurrently transmit and receive over the same frequency spectrum. In other words, transmission and reception can both make use of the full available frequency spectrum all the time. As mentioned, when operating in this full-duplex fashion, SI is inflicted onto a desired receive signal. To equip communication systems with full-duplex capability, engineers have developed a variety of ways to mitigate SI using radio frequency (RF) components, analog and digital filters, and other creative means. We will outline a variety of these strategies in this chapter. If SI can be sufficiently mitigated, a full-duplex transceiver can reliably receive while transmitting in-band, unlocking a number of enhancements, which we overview in the next section.

As depicted in Fig. 2, a full-duplex base station may transmit and receive concurrently with a neighboring user that also has full-duplex capability. This makes better use of radio resources, and as intuition may suggest, leads to a potential doubling of spectral efficiency, as compared to half-duplex techniques. In other words, radio resources are being used twice as efficiently with full-duplex, since they are used for both transmission and reception, rather than divided as with FDD and TDD, as illustrated in Fig. 1.

Fig. 3 depicts another full-duplex operating mode, where a full-duplex base station transmits to a half-duplex user while receiving from another half-duplex user. This mode of operation can also potentially double spectral efficiency. It is important to note that cross-link interference is inflicted on the user receiving by the user transmitting, the level of which depends on a number of factors, chiefly the users' locations and the environment.

Takeaways. Full-duplex operation is an exciting alternative to half-duplexing with FDD and TDD since it makes better use of radio resources. To enable full-duplex operation, however, self-interference must be mitigated sufficiently in order to reliably recover desired receive signals while transmitting in-band.

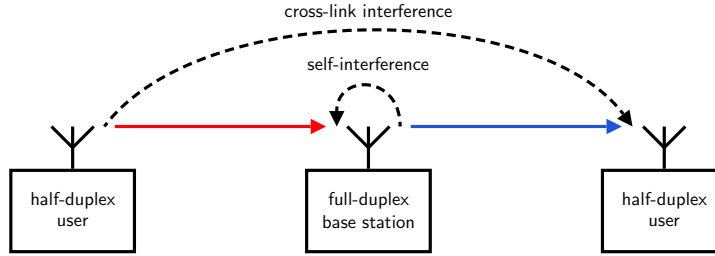


Fig. 3 A full-duplex base station transmitting to one half-duplex user while receiving from another half-duplex user in-band. SI manifests at the base station, whereas cross-link interference is inflicted onto the user receiving by the user transmitting.

2 Enhancements Introduced by Full-Duplex

Compared to half-duplex operation, full-duplex can introduce enhancements to communication systems on a link level [1, 2, 5, 6], in medium access and spectrum sharing [12–17], and at the network level [18, 19]—which we overview in this section.

2.1 Link-Level Analysis

The impacts of full-duplex can be readily observed by examining familiar link-level expressions foundational to wireless communication systems. To do so, consider the full-duplex operating mode depicted in Fig. 3, where a full-duplex base station communicates with two half-duplex users. Taking the perspective of the full-duplex base station, we refer to its *transmit link* and *receive link*. The full-duplex capacity of the system, in the absence of interference, can be written as

$$C_{\text{fd}} = \log_2(1 + \text{SNR}_{\text{tx}}) + \log_2(1 + \text{SNR}_{\text{rx}}), \quad (1)$$

where SNR_{tx} and SNR_{rx} are the maximum signal-to-noise ratios (SNRs) of the transmit and receive links, respectively. If employing TDD to duplex transmission and reception instead of full-duplex, the achievable sum spectral efficiency is

$$R_{\text{TDD}} = \alpha \cdot \log_2(1 + \text{SNR}_{\text{tx}}) + (1 - \alpha) \cdot \log_2(1 + \text{SNR}_{\text{rx}}), \quad (2)$$

where α is the fraction of time allocated to transmission, with the remainder allocated to reception. If employing FDD, this achievable sum spectral efficiency becomes

$$R_{\text{FDD}} = \alpha \cdot \log_2\left(1 + \frac{\text{SNR}_{\text{tx}}}{\alpha}\right) + (1 - \alpha) \cdot \log_2\left(1 + \frac{\text{SNR}_{\text{rx}}}{1 - \alpha}\right), \quad (3)$$

where α is the fraction of the total bandwidth allocated to transmission, with the remainder allocated to reception. Here, since the integrated noise power is proportional

to bandwidth, FDD enjoys SNR increases inversely proportional to bandwidth. The expression of R_{TDD} as presented implicitly assumes an instantaneous transmit power constraint. Under an average transmit power constraint, which is less practical, R_{TDD} and R_{FDD} coincide [20].

Under full-duplex operation, the signal-to-interference-plus-noise ratio (SINR) of a desired receive signal (i.e., a signal-of-interest) on a given link is

$$\text{SINR} = \frac{P_{\text{des}}}{P_{\text{noise}} + P_{\text{int}}} = \frac{\text{SNR}}{1 + \text{INR}}, \quad (4)$$

where P_{des} is the power of the desired receive signal; P_{noise} is the noise power; and P_{int} is the power of interference (i.e., SI on the receive link, cross-link interference on the transmit link). The right-hand side of (4) can be obtained by dividing the numerator and denominator by the noise power P_{noise} , where SNR is the SNR of the desired receive signal and INR is the interference-to-noise ratio (INR). The power of SI—and hence INR on the receive link—depends on the quality of SI mitigation employed at the full-duplex transceiver. For now, we can consider the degree of (residual) SI P_{SI} as being some level below the transmit power P_{tx} at the full-duplex transceiver. For instance, suppose our full-duplex base station is capable of reducing SI to a power level of

$$P_{\text{SI}} = \frac{P_{\text{tx}}}{L}, \quad (5)$$

where L is the total amount of mitigation (i.e., cancellation) achieved by its full-duplex solution. The total amount of SI mitigation L may capture a variety of efforts, including antenna isolation and/or additional SI cancellation filtering, which we will discuss further in the next section. Since the power of a desired receive signal P_{des} can be quite close to the noise floor P_{noise} in practical systems, SI power P_{SI} must be near or below the noise floor to ensure it does not prohibitively erode full-duplex resource gains. This means is not uncommon for L to be on the order of 100 dB for practical systems. Consider a transmit power of $P_{\text{tx}} = 20$ dBm and a noise floor of $P_{\text{noise}} = -90$ dBm: $L = 110$ dB is required for $P_{\text{SI}} = P_{\text{noise}}$ (i.e., INR = 0 dB). Achieving this degree of mitigation is precisely what has hindered the adoption of full-duplex since the advent of wireless communications and what has made successful demonstrations of full-duplex so impressive [7, 21–24].

A summary of key power ratios of the full-duplex system in Fig. 3.

SNR_{tx}	Strength of the desired signal on the transmit link.
SNR_{rx}	Strength of the desired signal on the receive link.
INR_{tx}	Strength of cross-link interference on the transmit link.
INR_{rx}	Strength of SI on the receive link.
SINR_{tx}	Effective quality of the desired signal on the transmit link.
SINR_{rx}	Effective quality of the desired signal on the receive link.

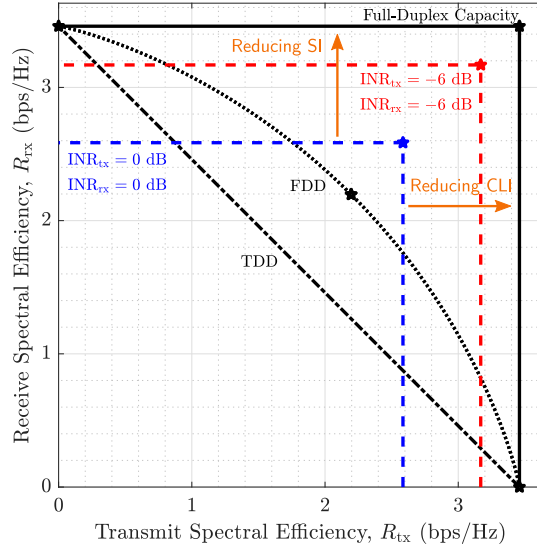


Fig. 4 The rate region boundaries for various multiplexing strategies when $\text{SNR}_{\text{tx}} = \text{SNR}_{\text{rx}} = 10$ dB. Full-duplex, shown as dashed blue and red lines, can outperform TDD and FDD with SI and cross-link interference (shown as CLI) sufficiently mitigated. Stars (★) indicate points that maximize the sum spectral efficiency.

When full-duplexing transmission and reception, the maximum achievable spectral efficiency when treating residual interference as noise can be expressed as

$$R_{\text{fd}} = \log_2(1 + \text{SINR}_{\text{tx}}) + \log_2(1 + \text{SINR}_{\text{rx}}) \leq C_{\text{fd}}. \quad (6)$$

When $\text{SINR}_{\text{tx}} \rightarrow \text{SNR}_{\text{tx}}$ and $\text{SINR}_{\text{rx}} \rightarrow \text{SNR}_{\text{rx}}$, then $R_{\text{fd}} \rightarrow C_{\text{fd}}$. These expressions illustrate the potential resource gains offered by full-duplex compared to FDD and TDD, since there are no pre-log fractions; the effectiveness of such depends heavily on the presence of low SI and low cross-link interference, however.

In Fig. 4, we compare full-duplex operation against half-duplex operation with FDD and TDD by plotting their *rate regions* for various levels of SI INR_{rx} and cross-link interference, where $\text{SNR}_{\text{tx}} = \text{SNR}_{\text{rx}} = 10$ dB. Here, each line depicts the boundary of its rate region, encompassing all feasible transmit-receive spectral efficiency pairs $(R_{\text{tx}}, R_{\text{rx}})$, and each star (★) indicates the point that maximizes the sum spectral efficiency. The simple time-sharing nature of TDD is shown as the diagonal line connecting the two points of greedy time-sharing. FDD offers improvements over TDD, courtesy of its SNR gains when shrinking bandwidth, as mentioned. When SI power is mitigated to a level equal to the noise power (i.e., $\text{INR}_{\text{rx}} = 0$ dB) and cross-link interference is also at the noise floor, the sum spectral efficiency can clearly exceed FDD and TDD, but only marginally. As SI and cross-link interference are reduced to below the noise floor, however, the achievable rate region of full-duplex expands, approaching that of its capacity region.

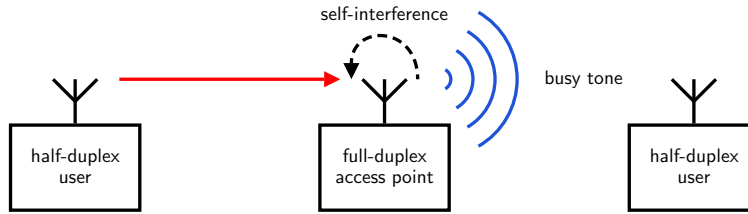


Fig. 5 While receiving from one user, a full-duplex access point broadcasts a busy tone instructing all other users not to transmit, preventing collisions and overcoming the hidden node problem.

At lower SNRs, lower INRs are required for appreciable full-duplex operation, since the effects of interference magnify due to the steep nature of $\log_2(1+x)$ at low x , tightening the requirements SI and cross-link interference mitigation. In addition, the SNR gains introduced by FDD are magnified since doubling x nearly doubles $\log_2(1+x) \approx x$ at low x , reducing the gap between FDD and the full-duplex capacity.

At higher SNRs, the gap between full-duplex and half-duplex grows, and the effects of interference diminish due to the saturating nature of $\log_2(1+x)$ at high x . Higher INRs can be tolerated at high SNRs, relaxing the requirements on mitigating SI and cross-link interference.

2.2 In Medium Access and Spectrum Sharing

Half-duplex transceivers have been ubiquitous in wireless networks, and for good reason, communication standards and practices have been built on this half-duplex assumption. The ability to transmit and receive simultaneously and in-band is a transformative technology that can unlock new approaches to medium access and spectrum sharing that are more efficient than those built on a half-duplex assumption. In turn, full-duplex can facilitate wireless networks that deliver higher throughput, inflict lower interference, and make better use of spectrum.

Overcoming the Hidden Node Problem. To illustrate this, consider the famous hidden node problem in wireless networks: if two users are distant from one another but both within earshot of a nearby access point, the two users may be unaware when the other is transmitting to the access point. This can lead to collisions at the access point—and hence a waste of radio resources—if not dealt with. Conventional approaches to overcome this use handshaking between users and the access point to grant a user permission to transmit (e.g., request-to-send (RTS) and clear-to-send

(CTS) mechanisms) along with random backoffs. By upgrading the access point with full-duplex capability, more efficient approaches to medium access can be employed [12, 15, 16]. For instance, as illustrated in Fig. 5, the full-duplex access point can broadcast a busy tone anytime it is receiving from a user. This busy tone can be heard by all users in the network, informing them to not transmit. Compared to conventional approaches to medium access, this approach consumes no additional radio resources and avoids the overhead associated with handshaking between users and the access point and reduces latency.

Dynamic Spectrum Access and Cognitive Radios. The number of devices with wireless connectivity has grown profoundly over the past two decades and will seemingly continue for years to come. The amount of available spectrum has not grown at a comparable pace, however. In light of this, researchers have proposed spectrum sharing and cognitive radios to dynamically and opportunistically make better use of frequency spectrum when it is free [25]. For instance, a cognitive radio may sense a frequency band to see if it is being used by nearby devices. If there appears to be no activity, the cognitive radio may begin transmitting information. Periodically, the cognitive radio may halt transmission to sense the channel to ensure that it does not inflict interference on incumbents that have rights to the band—a waste of resources if none are detected. Full-duplex can make this process more efficient by empowering the cognitive radio to continuously sense the channel while transmitting [13, 14, 16]. This allows it to more efficiently transmit information, since it does not have to halt transmission to sense the channel, and enables the cognitive radio to react more quickly to the presence of incumbents. Without full-duplex capability, the cognitive radio would presumably be unaware of the presence of incumbents until the end of its transmission, potentially causing interference that leads to collisions. Techniques discussed for overcoming the hidden node problem can be applied in this setting, as well, to instruct nearby cognitive radios to not transmit.

Channel Sensing to Reduce the Effects of Interference. As a final example of full-duplex applied to medium access, we consider the case where an access point serves users in the presence of neighboring nodes that may inflict interference, as explored in [17]. For instance, one can consider two Wi-Fi access points operating independently but near one another. When one access point transmits to a user, successful reception at that user may be corrupted by neighboring interference, requiring the access point to retransmit the data. Naturally, this consumes radio resources and can lead to delays in communication. With full-duplex capability, the access point may sense the channel while transmitting, allowing it to potentially halt transmission to avoid collisions caused by neighboring interference and subsequently redirect resources to another user that may not be impacted by this interference [16, 17]. It is important to note that the requirements on mitigating SI are stricter when decoding data carried by a desired receive signal, compared to those for channel sensing, which is often merely detecting energy levels in a particular frequency band.

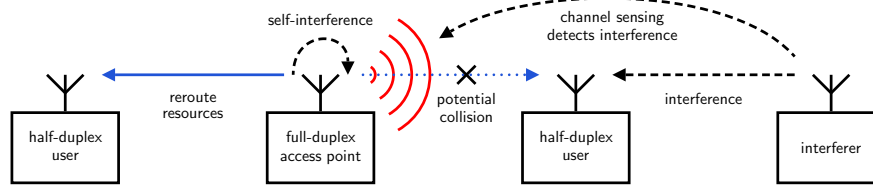


Fig. 6 An access point equipped with full-duplex capability can sense the channel while transmitting, allowing it to reroute resources in the presence of interference that would otherwise cause collisions [17].

2.3 Network-Level Enhancements

Upgrading transceivers with full-duplex capability can be felt at the network level in a few ways. Even when only some devices are equipped with full-duplex—while the remainder are half-duplex-capable—a wireless network can enjoy improvements in throughput and latency [18, 19]. In fact, network throughput can magnify beyond the doubling of spectral efficiency we are familiar with at the link level with full-duplex [18, 19]. This can be attributed to the fact that full-duplex can reduce multiplexing delays, reduce overhead associated with medium access control, and enable new ways to manage interference—all of which can improve network throughput. We elaborate more on network-level enhancements of full-duplex in Subsection 4.6.

Other applications. There are applications of full-duplex technology beyond what was highlighted herein, such as in joint communication and sensing [10, 26], secure communications [27], military communications [28, 29], radar [10], and more [5, 6].

3 Self-Interference Cancellation

Successfully equipping a device with full-duplex capability relies on mitigating—or *cancelling*—SI to levels that are sufficiently low [2]. The amount of self-interference cancellation (SIC) needed depends on the particular application. In most settings, full-duplex solutions aim to cancel SI to near or below the receiver noise floor (i.e., roughly $\text{INR}_{\text{rx}} \leq 0$ dB). This ensures that the full-duplex resource gains are not eroded by the presence of high SI, as highlighted in Fig. 4. The *residual* SI is that which remains after efforts of SIC. In this section, we outline methods of SIC in both the analog and digital domains. Regardless of domain, the motivation behind SIC can largely be summarized as leveraging the fact that a transceiver has knowledge of its own transmitted signal and can therefore potentially reconstruct SI and subtract it from the received signal, leaving the desired portion virtually free

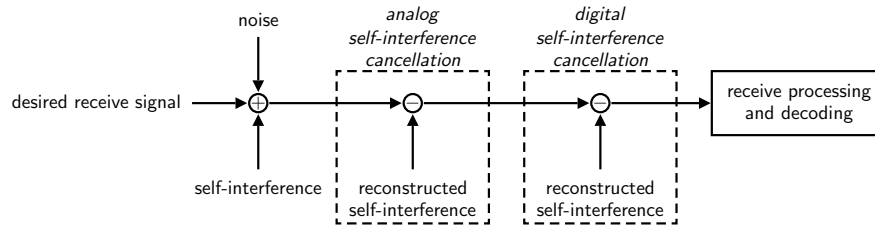


Fig. 7 A received signal undergoes analog and digital SIC before undergoing conventional receive processing to recover desired receive data.

of SI. In many cases, a staged approach to SIC is employed as illustrated in Fig. 7, where a portion of SI is cancelled using an analog filter and a significant portion of the remainder is cancelled using digital filtering. This staged approach is depicted in Fig. 8, where a full-duplex transceiver with separate antennas for transmission and reception employs an analog SIC filter between its antennas at RF and a digital SIC filter.

3.1 Analog Self-Interference Cancellation

As illustrated in Fig. 8, analog SIC typically exists as a digitally-controlled analog filter placed between the transmitter and receiver of a full-duplex transceiver. Analog SIC filters come in many forms, existing as RF, intermediate frequency, and baseband circuitry, and even as optical filters [2, 23, 30–32]. Analog SIC is often driven by the tapping off a small portion of the upconverted RF transmit signal. This transmit signal undergoes filtering within the analog SIC filter before being injected at the receiver. The injected signal is an inverted reconstruction of SI, which, when combined with the received signal, destructively combines with SI. After this combining, there is some degree of residual SI due to imperfect reconstruction, which may stem from estimation errors, hardware limitations, and hardware imperfections. By tapping off the transmit signal after the transmit chain, analog SIC will inherently incorporate transmit-side impairments unbeknownst to baseband, such as power amplifier (PA) nonlinearities, which have proven to be a dominant factor in SIC [24, 33]. Other approaches, sometimes called digitally-assisted approaches, use a dedicated transmit chain to drive analog SIC, as opposed to tapping off the transmit signal directly [2]. This approach cannot as well capture transmit-side impairments present in SI, however, since this second transmit chain naturally will not contain all artifacts of the true transmit chain.

Configuring an Analog Self-Interference Cancellation Filter. Tuning an analog SIC filter to effectively cancel SI largely consists of measuring SI and then configuring the filter to reconstruct its inverse. Analog SIC can be implemented as a time-domain filter or as a frequency-domain filter, meaning particular methods may

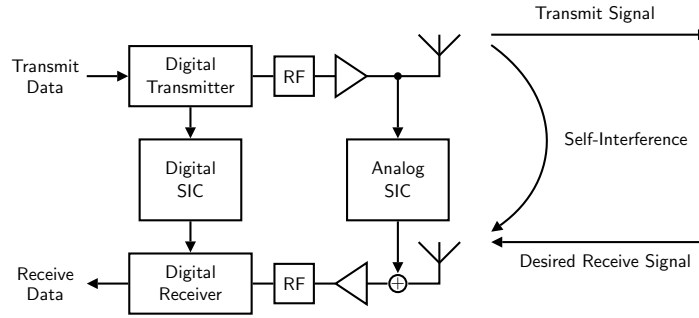


Fig. 8 SI manifests when a full-duplex transceiver attempts to simultaneously transmit and receive using the same frequency spectrum. Here, separate antennas are used for transmission and reception, with analog and digital SIC used to reconstruct and subsequently cancel SI incurred at the receiver.

vary but all tackle the same goal of reconstructing SI [2]. One method of time-domain analog SIC is to estimate the impulse response of the SI channel and then configure the analog filter to produce this (inverted) impulse response estimate, effectively equalizing SI. Estimation of the SI channel is typically executed by transmitting a pilot signal during a *quiet period*, when no desired receive signal is present. One difficulty with practically executing this method lay in the fact that estimation of the SI channel takes place digitally, meaning estimation of the channel of interest for analog SIC may be complicated by artifacts of the transmit and receive chains before and after analog SIC. This can be further complicated by the fact that an analog filter may not have an ideal impulse response itself, making it difficult to reliably produce the desired impulse response.

As an attempt to overcome these challenges, another approach is to measure SI and then measure the impulse response of the analog filter. Then, the filter can be configured to produce an inverted reconstruction of SI. For instance, consider a column vector of measured SI time-domain samples \mathbf{y} (during a quiet period) and a matrix \mathbf{A} whose i -th column is the measured impulse response of the i -th tap of the filter. Analog finite impulse response (FIR) filter weights \mathbf{x} can be computed to minimize the error in reconstructing an inverted copy of the measured SI as

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2, \quad (7)$$

which has the well-known closed form solution $\mathbf{x}^* = -(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{y}$. This approach has shown to be fairly robust, since it accounts for the imperfect impulse response of each of the filter's taps.

While it may seem fairly straightforward to implement analog SIC, it is practically quite challenging in most cases, especially outside of well-controlled lab settings. Most notably, there is small margin for error in SIC due to the overwhelming strength of SI, reinforcing the need for extremely accurate, adaptable, and low-overhead SI reconstruction. Another challenging aspect is the need to miniaturize analog SIC filters into form-factors that integrate into devices such as cell phones,

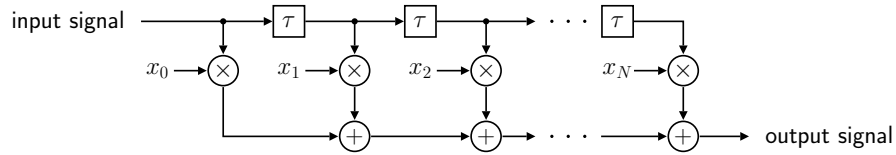


Fig. 9 An example analog SIC filter (an N -tap FIR filter) with tunable tap weights $\{x_i\}$ and fixed, uniform tap delay τ .

laptops, wireless routers, base stations, and the like [2]. Miniaturization is especially challenging in settings where the delay spread of SI has the potential to be high, since propagation delays need to be physically realized within the analog SIC filter.

3.2 Digital Self-Interference Cancellation

Cancelling SI through digital signal processing is naturally an attractive option in addition to analog SIC. The flexibility and sophistication of digital filtering can be applied to estimate and cancel SI and has had impressive success [24, 33, 34]. As depicted in Fig. 7 and Fig. 8, digital SIC is executed after analog SIC and therefore aims to cancel residual SI that remains after prior efforts of cancellation. Naturally, one may ask whether digital SIC can cancel all SI, rendering analog SIC unnecessary. In general, this is not possible for a few reasons, stemming from hardware limitations and nonidealities.

Limited Dynamic Range of Analog-to-Digital Converters. With reasonable resolution and appropriate gain control before analog-to-digital conversion, quantization noise is rarely an issue in traditional half-duplex systems. In full-duplex systems, on the other hand, the strength of SI has the potential to *saturate* analog-to-digital converters (ADCs), even with ideal automatic gain control (AGC) and a reasonable number of bits [1, 33]. Since AGC acts on the combination of a desired signal plus SI and noise at the ADC input, the strength of quantization noise is dictated largely by that of SI. In such cases, only a portion of the ADC's total dynamic range is effectively used to quantize a desired signal. Consequently, even if SI could be completely reconstructed and cancelled digitally, its effects may remain in the form of increased quantization noise, which can severely and irreversibly degrade the quality of a desired receive signal. This reinforces the need to sufficiently cancel SI before it reaches the ADC input, which is often most reliably done via analog SIC.

Transceiver Nonidealities. Practical transceivers introduce nonidealities in the transmit and receive chains, such as amplifier nonlinearities, I/Q imbalance, transmitter thermal noise, and phase noise, which complicate digital SIC since the digital domain does not have knowledge of these imperfections [24]. When these nonidealities are not negligible, this requires digital SIC to accurately estimate and subsequently cancel them, which can be computationally complex. To reduce this

burden, analog SIC can be well positioned to cancel transmit-side impairments using the RF transmit signal as input to its cancellation filter, which inherently will include nonidealities introduced by transmit PAs and transmit thermal noise, for example. In addition, analog SIC can importantly ensure that the power of residual SI is sufficiently low such that it does not overwhelm and saturate receive-chain components such as low noise amplifiers (LNAs), which practically have a limited dynamic range that can be exceeded by SI [33].

Recent Breakthroughs using Machine Learning. In addition to classical signal processing approaches for digital SIC, solutions based on machine learning have been gaining traction and have shown impressive results [11, 35–40]. The main motivation for the use of machine learning over classical approaches for digital SIC is to capture transceiver nonidealities with reduced complexity. Classical signal processing approaches have been able to effectively estimate and account for transceiver impairments when reconstructing and subsequently cancelling SI. This is done by modeling transceiver impairments with established, parameterized models, but the estimation of model parameters is computationally expensive with classical methods. Machine learning has shown to be able to offer comparable performance as classical methods in capturing transceiver impairments when reconstructing SI but with reduced complexity. Experimental validation of these digital SIC solutions based on machine learning has proven their effectiveness [11, 35, 37]. In addition, machine learning can be used to reduce the complexity of SIC in multi-antenna systems. Rather than merely replicating single-antenna SIC solutions as an extension to multi-antenna systems, researchers have shown that machine learning can reduce the size and complexity of digital SIC by learning correlations between antennas [39].

3.3 Circulators and Antenna Isolation

An RF component known as a *circulator* has been used in many monostatic radar and communication applications as a duplexer when a single antenna is used for simultaneous transmission and reception of RF signals [1, 2]. In its simplest form, a circulator is a three-port, passive device where a signal entering a given port is “circulated” to the next port in the rotation. An example of this device being used with a single antenna shared by transmission and reception can be seen in Fig. 10. Transmit signals enter Port 1 of the circulator and exit at Port 2, where they are radiated by the antenna. Signals received by the antenna enter Port 2 and are circulated to Port 3, where they exit the circulator and enter the receive chain. This establishes isolation between the transmitter and receiver of a full-duplex device using a single antenna for transmission and reception.

One may reason that with an ideal circulator and with two radios operating in free space, full-duplex operation is trivial since perfect isolation is achieved between a radio’s transmit signal and a desired receive signal. In reality, a circulator effectively offers limited RF isolation between its ports, which introduces SI at the receiver. This is due to a number of factors, most notably the leakage between ports, reflections

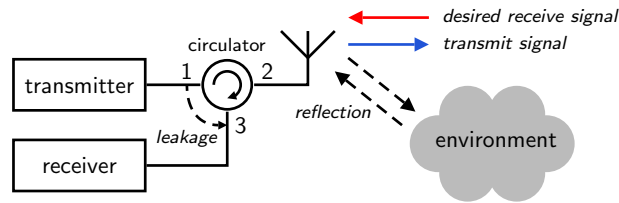


Fig. 10 A circulator can be used to establish isolation between a transmitter and receiver sharing an antenna. Leakage through the circulator and reflections off the environment give rise to SI, however.

caused by imperfect matching at the antenna, and reflections off the environment. Circulators with small form-factors that offer high isolation for full-duplex are an active area of research with immense potential [41, 42]. Nonetheless, analog and digital SIC can be used in conjunction with a circulator for single-antenna full-duplex transceivers. In such cases, SIC aims to cancel circulator leakage, as well as reflections off the environment and from the antenna.

Takeaways. Digital SIC is an effective and flexible route to enabling full-duplex, but it is bottlenecked in practice by imperfections and limitations of hardware. As such, it is often used in conjunction with analog SIC, which can inherently account for hardware impairments, relax the requirements of digital SIC, and prevent SI from saturating receive-chain components. Circulators and other duplexers can provide isolation between the transmitter and receiver of a full-duplex transceiver when using a single antenna, effectively weakening SI that analog and digital SIC must tackle.

4 A New Frontier: Full-Duplex Millimeter-Wave Systems

To meet the ever-growing demand for high-rate wireless access, cellular networks have turned to mmWave carrier frequencies, typically classified as ranging from around 30 GHz to 100 GHz [43]. Fifth-generation (5G) cellular networks and IEEE 802.11ad/ay, for instance, leverage frequency bands that span hundreds of megahertz. These wide swaths of spectrum facilitate higher data rates and enable new applications in entertainment, industry, and sensing. The widespread deployment of mmWave networks has faced hurdles thus far, but it is expected they will see greater success through the end of the 2020s. Both mmWave communication systems and full-duplex technology were explored concurrently during the 2010s and were proposed as core technologies for next-generation wireless networks. The combination of the two—full-duplex mmWave communication systems—has not been explored as extensively. Only recently has this topic garnered noteworthy attention from industry and academia [8, 9, 44, 45].

4.1 Prelude: Full-Duplex MIMO Transceivers

With multiple antennas at a transmitter and a receiver, there is the potential to multiplex more than one data stream over the resulting multiple-input multiple-output (MIMO) wireless channel via spatial signal processing [20]. MIMO communication transformed wireless networks forever by offering multiplicative rate gains over traditional single-input single-output communication systems. Given their prominence, the extension of full-duplex to MIMO transceivers was imperative. With multiple antennas at the transmitter and receiver of a full-duplex transceiver, SI is inflicted onto each receive antenna by each transmit antenna, leading to a MIMO SI channel. There is a quite natural extension of analog and digital SI to full-duplex MIMO transceivers [24, 46, 47]. Perhaps more exciting, though, is the potential to mitigate SI through precoding and combining (i.e., spatial processing) at the transmitter and receiver of the full-duplex transceiver [48–51]. By strategically transmitting energy into the SI channel and receiving energy from it, SI can be mitigated spatially while still communicating desired signals in a MIMO fashion. As the number of antennas grows—and especially in the massive MIMO regime—the prospects of spatial cancellation are even more promising. There is extensive literature on the subject of full-duplex MIMO systems; we encourage interested readers to explore [46–52] and references therein for more details. In the remainder of this section, we consider a particular class of full-duplex MIMO transceivers: those at mmWave frequencies. Solutions for full-duplex mmWave systems draw inspiration from those for traditional full-duplex MIMO systems at sub-6 GHz but face unique challenges and are subject to new transceiver- and network-level factors.

4.2 What’s New at Millimeter-Wave Frequencies?

Communication at mmWave is more than a mere shift in carrier frequency, as elegantly stated in [53]. In general, path loss increases with carrier frequency, which necessitates the use of dense antenna arrays to supply high beamforming gains that can deliver link margins that sustain high-rate communication. Antenna arrays on mmWave network infrastructure are typically equipped with 64–256 antennas, whereas user equipment may be equipped with 4–16 elements. Fortunately, antenna footprint shrinks as carrier frequency increases, allowing dense antenna arrays to fit in convenient form-factors. The severe path loss and susceptibility to blockage at mmWave frequencies, coupled with highly directional beamforming, reduces inter-user interference and facilitates base station deployments much denser than traditional sub-6 GHz macrocell deployments. All of this has led to new transceiver-level and network-level challenges and solutions at mmWave.

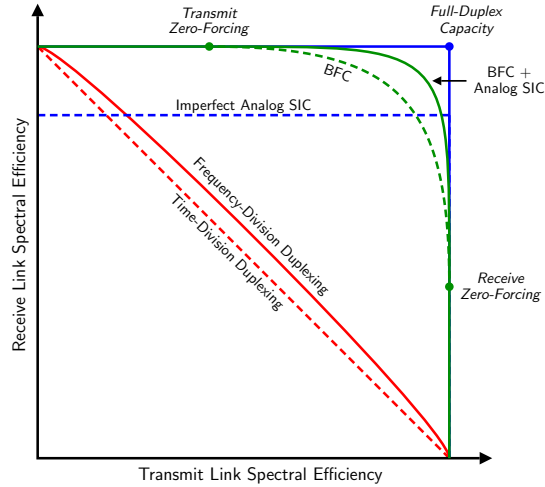


Fig. 11 The rate region boundaries for TDD, FDD, and full-duplex using analog SIC, beamforming cancellation (shown as BFC), and the combination of the two. Beamforming cancellation consumes spatial resources, which introduces a gap between it and the full-duplex capacity. This gap can be reduced via analog and/or digital SIC. Reproduced from [9] with permission.

4.3 Exciting Potential to Tackle Self-Interference via Beamforming

In Subsection 4.1, we mentioned that a multi-antenna full-duplex system can design its transmit precoder and receive combiner to reduce SI coupled over the MIMO SI channel. At mmWave, the prospect of such *spatial SIC* notably improves for a few reasons [8, 9, 44]. First of all, path loss and blockage increases at mmWave frequencies, compared to sub-6 GHz frequencies, which presumably weakens SI as it couples from a transmitter to receiver, both directly over-the-air and due to reflections off the environment; reflectivity increases at mmWave, however, which may lead to more SI from the environment. Second, with denser antenna arrays, the degrees-of-freedom available to cancel SI increases, compared to traditional sub-6 GHz MIMO systems, which typically only have 2-8 antennas. It is important to note that, as the number of antennas has increased immensely at mmWave, the number of data streams communicated has remained of similar order. Hybrid analog/digital beamforming and analog-only beamforming architectures, which are ubiquitous thus far in mmWave communication systems, fundamentally limit the number of data streams to the number of digital-to-analog converters (DACs) and ADCs, which is comparable to that in conventional sub-6 GHz MIMO systems. With one or few data streams communicated and tens or hundreds of antennas, many degrees-of-freedom are potentially available for cancelling SI via strategic design of transmit and receive beamformers at a full-duplex mmWave transceiver.

Excitedly, if beamforming can cancel SI sufficiently, the need for analog and/or digital SIC vanishes, meaning there is the potential for full-duplex mmWave systems

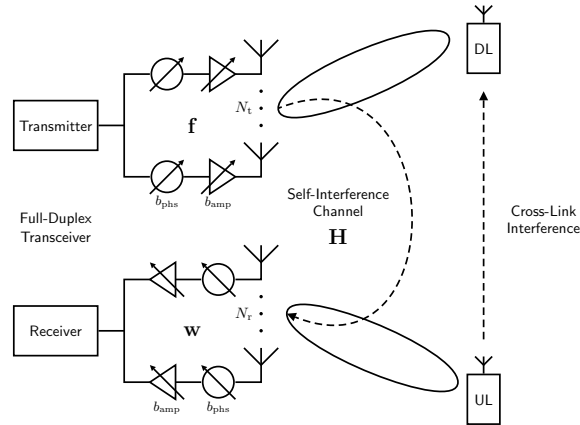


Fig. 12 A full-duplex mmWave transceiver transmits to a downlink user while receiving from an uplink user in-band. By strategically constructing its analog beamformers \mathbf{f} and \mathbf{w} , it can reduce the level of SI coupled across the MIMO channel \mathbf{H} while delivering service to the users.

to operate without any additional hardware or complex signal processing. In other words, analog and digital SIC may not be needed for full-duplex mmWave systems with sufficient spatial SIC. Plenty of existing work has highlighted this by designing transmit and receive beamformers in such a way that SI is mitigated while maintaining downlink and uplink to users [8, 9, 44, 48, 52, 54–65]. Interestingly, unlike traditional analog and digital SIC solutions, transmit beamforming introduces the unique opportunity to reduce the degree of SI that ever reaches the receive antennas. Like analog SIC, transmit and receive beamforming can be used to prevent SI from saturating receive chain components. For instance, in [64], a hybrid beamforming design is presented that guarantees SI is below some power level at each LNA and each ADC at the receiver of the full-duplex mmWave device, ensuring they do not saturate. In addition to using beamforming alone to mitigate SI, researchers have also considered analog and/or digital SIC in conjunction, which relaxes the cancellation requirements of beamforming, allowing it to better serve uplink and downlink at the cost of added hardware or signal processing [47, 65–71]. The rate region boundaries with spatial cancellation via beamforming (sometimes termed *beamforming cancellation*) and analog SIC are shown in Fig. 11 [9].

4.4 Example Beamforming Design Problems

We now overview two sample spatial SIC design problems for full-duplex mmWave systems, with the goal of introducing readers to the design objectives and considerations surrounding such research problems and those taking a similar form.

Analog Beamforming Design. Consider the system illustrated in Fig. 12. Suppose the full-duplex mmWave base station equipped with analog-only beamforming—as opposed to hybrid digital/analog beamforming [53]—is serving an uplink user and a downlink user simultaneously and in-band, both of which are single-antenna devices. It is currently practical to consider the use of separate, independently-controlled transmit and receive arrays at the base station equipped with N_t and N_r antennas, respectively [9]. Let $\mathbf{f} \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{w} \in \mathbb{C}^{N_r \times 1}$ be the transmit and receive beamforming vectors used at the full-duplex base station. Let $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ be the MIMO SI channel matrix manifesting between the transmit and receive arrays. Let $\mathbf{h}_{\text{tx}}^* \in \mathbb{C}^{1 \times N_t}$ be the channel vector from the base station to the downlink user. Let $\mathbf{h}_{\text{rx}} \in \mathbb{C}^{N_r \times 1}$ be the channel vector from the uplink user to the base station. The full-duplex base station will rely solely on beamforming to mitigate SI with no additional analog or digital SIC.

Practical analog beamforming networks are comprised of digitally-controlled phase shifters. In addition to phase control, some networks also offer quantized amplitude control through digitally-controlled attenuators or variable-gain amplifiers (VGAs) (see Subsection 4.5). Quantized phase and amplitude control confines all physically realizable analog beamformers to come from some discrete sets, which can be captured as simply $\mathbf{f} \in \mathcal{F}$ and $\mathbf{w} \in \mathcal{W}$. The transmit link and receive link SNRs are functions of their beamformers and can be expressed as

$$\text{SNR}_{\text{tx}}(\mathbf{f}) = \frac{P_{\text{tx}}^{\text{BS}} \cdot |\mathbf{h}_{\text{tx}}^* \mathbf{f}|^2}{P_{\text{noise}}^{\text{UE}}}, \quad \text{SNR}_{\text{rx}}(\mathbf{w}) = \frac{P_{\text{tx}}^{\text{UE}} \cdot |\mathbf{w}^* \mathbf{h}_{\text{rx}}|^2}{P_{\text{noise}}^{\text{BS}}}, \quad (8)$$

where $P_{\text{tx}}^{\text{BS}}$ and $P_{\text{tx}}^{\text{UE}}$ are the transmit powers of the base station and the uplink user, while $P_{\text{noise}}^{\text{BS}}$ and $P_{\text{noise}}^{\text{UE}}$ are the noise powers of the base station and the downlink user. The SI and cross-link interference terms of the system are

$$\text{INR}_{\text{rx}}(\mathbf{f}, \mathbf{w}) = \frac{P_{\text{tx}}^{\text{BS}} \cdot |\mathbf{w}^* \mathbf{H} \mathbf{f}|^2}{P_{\text{noise}}^{\text{BS}}}, \quad \text{INR}_{\text{tx}} = \frac{P_{\text{tx}}^{\text{UE}} \cdot |h_{\text{CL}}|^2}{P_{\text{noise}}^{\text{UE}}}, \quad (9)$$

where INR_{rx} is a function of the transmit and receive beams at the full-duplex base station and INR_{tx} is solely a function of the cross-link channel h_{CL} from the uplink user to the downlink user. Together, these desired and interference terms form the SINRs of the two links as

$$\text{SINR}_{\text{tx}}(\mathbf{f}) = \frac{\text{SNR}_{\text{tx}}(\mathbf{f})}{1 + \text{INR}_{\text{tx}}}, \quad \text{SINR}_{\text{rx}}(\mathbf{f}, \mathbf{w}) = \frac{\text{SNR}_{\text{rx}}(\mathbf{w})}{1 + \text{INR}_{\text{rx}}(\mathbf{f}, \mathbf{w})}, \quad (10)$$

which determine the achievable spectral efficiencies as

$$R_{\text{tx}}(\mathbf{f}) = \log_2(1 + \text{SINR}_{\text{tx}}(\mathbf{f})), \quad R_{\text{rx}}(\mathbf{f}, \mathbf{w}) = \log_2(1 + \text{SINR}_{\text{rx}}(\mathbf{f}, \mathbf{w})), \quad (11)$$

when treating interference as noise. Notice that the transmit link quality SINR_{tx} and therefore its spectral efficiency R_{tx} are solely functions of the transmit beam \mathbf{f} . The

fate of the receive link, however, is determined by both \mathbf{w} and \mathbf{f} due to SI. It would be desirable to design \mathbf{f} and \mathbf{w} such that they deliver high SNRs and couple low SI.

This motivates the following analog beamforming design problem, which aims to maximize the sum spectral efficiency of the system, while requiring the transmit and receive beams to be physically realizable.

$$\max_{\mathbf{f}, \mathbf{w}} R_{\text{tx}}(\mathbf{f}) + R_{\text{rx}}(\mathbf{f}, \mathbf{w}) \quad (12a)$$

$$\text{s.t. } \mathbf{f} \in \mathcal{F}, \mathbf{w} \in \mathcal{W} \quad (12b)$$

Several existing works on mmWave full-duplex aim to solve this problem or one of similar form [60, 61, 72]. In general, this problem is difficult to solve due to the non-convexity of the objective from the coupling of \mathbf{f} and \mathbf{w} and the fact that \mathcal{F} and \mathcal{W} are non-convex sets. Researchers typically instead tackle problems that are more readily solved and still yield high sum spectral efficiency.

One practical issue with solving analog beamforming problems of this type is that they are executed for each user pair, which can consume prohibitive amounts of radio resources (e.g., for channel estimation and over-the-air feedback) and computational resources. In fact, the time required to solve these sorts of problems may not translate to timescales of practical systems, even with modern computing power. Moreover, many existing solutions rely on unrealistic assumptions, such as real-time down-link/uplink MIMO channel knowledge (i.e., \mathbf{h}_{tx}^* and \mathbf{h}_{rx}), which is not obtainable in practical mmWave systems today.

Analog Beamforming Codebook Design. In the previous example, we considered the problem of designing transmit and receive beams that maximize sum spectral efficiency. Now, to circumvent and overcome some of the practical challenges mentioned in the previous example, let us consider the goal of designing transmit and receive *codebooks* that maximize sum spectral efficiency in full-duplex mmWave systems. Let us build on the previous example and the notation used therein. Suppose it is required that the transmit beam \mathbf{f} be selected from some set of M_{tx} transmit beams $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{M_{\text{tx}}}\}$, called a transmit codebook, which is common in practical systems. Likewise, suppose it is required that the receive beam \mathbf{w} be selected from some codebook of M_{rx} receive beams $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{M_{\text{rx}}}\}$. Here, M_{tx} and M_{rx} are on the order of tens or hundreds at most, meaning it is realistically the case that $M_{\text{tx}} \ll |\mathcal{F}|$ and $M_{\text{rx}} \ll |\mathcal{W}|$.

It is fairly straightforward to design transmit and receive codebooks for traditional half-duplex mmWave systems, typically done by tessellating beams to cover a desired service region to ensure that a user falling in this region can be served with high gain with at least one beam from the codebook (e.g., see Fig. 14). The design of codebooks for full-duplex mmWave systems, on the other hand, is far more involved and has only been investigated in [54, 73] thus far. Consider the following design problem, which aims to design codebook matrices $\mathbf{F} \in \mathbb{C}^{N_t \times M_{\text{tx}}}$ and $\mathbf{W} \in \mathbb{C}^{N_r \times M_{\text{rx}}}$ that maximize the expected sum spectral efficiency across a known user distribution for a given \mathbf{H} , for instance.

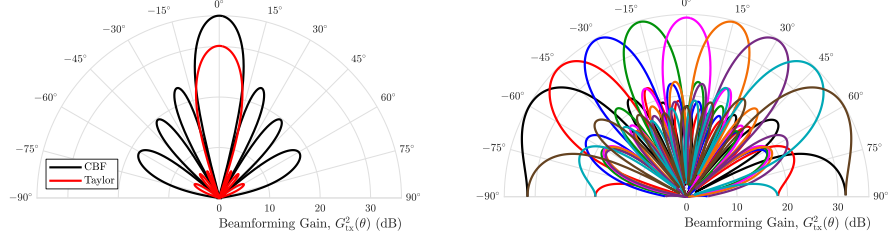


Fig. 13 (left) A broadside beam from two conventional beamforming codebooks. (right) Beams created by [54, 73] that span a coverage region with high gain while coupling low SI.

$$\max_{\mathbf{F}, \mathbf{W}} \mathbb{E} \left[\max_{\mathbf{f}, \mathbf{w}} R_{\text{tx}}(\mathbf{f}) + R_{\text{rx}}(\mathbf{f}, \mathbf{w}) \right] \quad (13a)$$

$$\text{s.t. } \mathbf{f} = [\mathbf{F}]_{:,i}, \mathbf{w} = [\mathbf{W}]_{:,j} \quad (13b)$$

$$[\mathbf{F}]_{:,i} \in \mathcal{F} \forall i = 1, \dots, M_{\text{tx}} \quad (13c)$$

$$[\mathbf{W}]_{:,j} \in \mathcal{W} \forall j = 1, \dots, M_{\text{rx}} \quad (13d)$$

Here, these codebook matrices are structured such that the i -th column of \mathbf{F} is \mathbf{f}_i and the j -th column of \mathbf{W} is \mathbf{w}_j , both of which are required to be physically realizable beamforming vectors, hence (13c) and (13d). To serve each user pair, some transmit and receive beams \mathbf{f} and \mathbf{w} are selected from their respective codebooks. With random user placement, any transmit and receive beam may be chosen from the codebooks, meaning desirable codebooks \mathbf{F} and \mathbf{W} would offer low $\text{INR}_{\text{rx}}(\mathbf{f}, \mathbf{w})$ for all possible \mathbf{f} and \mathbf{w} while still capable of delivering high beamforming gains. This is a difficult problem to solve, largely due to the daunting objective of aiming to maximize average sum spectral efficiency, but provides good direction for desirable codebooks \mathbf{F} and \mathbf{W} .

In [54, 73], a similar problem is tackled, as shown below in problem (14). Here, the objective is instead to minimize average SI coupled between possible transmit and receive beams across the channel \mathbf{H} , effectively minimizing average $\text{INR}_{\text{rx}}(\mathbf{f}, \mathbf{w})$. In doing so, high beamforming gain and broad coverage over some transmit and receive coverage regions are enforced by (14b) and (14c), where σ_{tx}^2 and σ_{rx}^2 are design parameters that throttle the so-called coverage variance of each codebook. In essence, these constraints ensure codebooks can reliably deliver high SNR_{tx} and high SNR_{rx} , while the objective aims to minimize INR_{rx} .

$$\min_{\mathbf{F}, \mathbf{W}} \|\mathbf{W}^* \mathbf{H} \mathbf{F}\|_{\text{F}}^2 \quad (14a)$$

$$\text{s.t. } \|N_{\text{t}} \cdot \mathbf{1} - \text{diag}(\mathbf{A}_{\text{tx}}^* \mathbf{F})\|_2^2 \leq \sigma_{\text{tx}}^2 \cdot N_{\text{t}}^2 \cdot M_{\text{tx}} \quad (14b)$$

$$\|N_{\text{r}} \cdot \mathbf{1} - \text{diag}(\mathbf{A}_{\text{rx}}^* \mathbf{W})\|_2^2 \leq \sigma_{\text{rx}}^2 \cdot N_{\text{r}}^2 \cdot M_{\text{rx}} \quad (14c)$$

$$[\mathbf{F}]_{:,i} \in \mathcal{F} \forall i = 1, \dots, M_{\text{tx}} \quad (14d)$$

$$[\mathbf{W}]_{:,j} \in \mathcal{W} \forall j = 1, \dots, M_{\text{rx}} \quad (14e)$$

A codebook of beams output by this design is shown in Fig. 13. Compared to traditional beams, the beams produced by [54, 73] make use of side lobes to cancel SI spatially while providing adequate coverage across the service region from -60° to 60° . The codebooks designed with this framework proved to offer far greater robustness to SI and similar beamforming gain when compared to conventional codebooks, which allowed them to deliver sum spectral efficiencies $R_{\text{tx}} + R_{\text{rx}}$ that approach the full-duplex capacity without analog or digital SIC. Designs like this are particularly exciting because they have the potential to accommodate codebook-based beam alignment while also mitigating SI through beamforming. For more details and more extensive evaluation of this design, please see [54, 73].

4.5 Key Practical Challenges and Considerations

We now outline important considerations when designing solutions for practical full-duplex mmWave systems, some of which have already been touched on in this chapter.

Digitally-Controlled Analog Beamforming Networks. Unlike digital beamforming, which takes place in software/logic, analog beamforming is executed using phase shifter components, potentially along with attenuators and/or amplifiers, all of which are digitally-controlled. In other words, some finite number of bits are dedicated to realizing the phase and amplitude of each beamforming weight. For instance, the discrete set of physically realizable phase settings by phase shifters with settings uniformly distributed between 0 and 2π with resolution b_{phs} bits can be expressed as

$$\mathcal{P} = \left\{ \theta_i = \frac{(i-1) \cdot 2\pi}{2^{b_{\text{phs}}}} : i = 1, \dots, 2^{b_{\text{phs}}} \right\}, \quad (15)$$

meaning it is practically required, for the i -th beamforming weight w_i , that $\text{angle}(w_i) \in \mathcal{P}$ for all i . In practice, it should be noted that both phase and amplitude control typically have some error associated with them, which is generally frequency-dependent.

Some phased arrays employ phase shifters based on a vector modulator architecture, where the in-phase and quadrature components of a signal can be scaled independently to produce a desired phase shift, in which case phase shifter settings are presumably no longer uniformly spaced. Note that such an architecture also offers a means to scale the amplitude of the output signal. Practical beamforming-based full-duplex solutions for mmWave systems should account for the limitations imposed by a particular analog beamforming architecture. Quantized control of each beamforming weight leads to non-convex sets that are difficult to optimize over. Blindly projecting a solution onto the set of physically realizable beamforming vectors can be detrimental, as small errors in this full-duplex setting are magnified by the sheer strength of SI relative to a desired signal. This motivates the use of high-

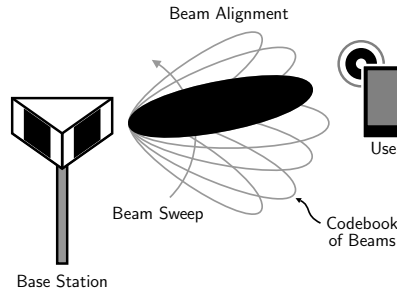


Fig. 14 A mmWave base station conducts beam alignment by sweeping a codebook of candidate beams, selecting that which maximizes SNR to serve a given user.

resolution phase shifters in full-duplex mmWave systems and/or ways to handle the non-convexity posed by quantized phase shifters.

Accommodating Codebook-Based Beam Alignment. Codebook-based analog beamforming is a critical component of mmWave communication systems [53, 74]. Rather than measure a high-dimensional MIMO channel and subsequently configure an analog beamforming network, modern mmWave systems instead rely on beam alignment procedures to identify promising beamforming directions, typically via exploration of a codebook of candidate beams [53, 74, 75]. This offers a simple and robust way to configure an analog beamforming network without downlink/uplink MIMO channel knowledge *a priori*, which is not obtainable in practice.

Let $\bar{\mathcal{F}}$ and $\bar{\mathcal{W}}$ be transmit and receive beamforming codebooks used at a full-duplex device. Executing beam alignment on each link independently (in a half-duplex fashion) would aim to solve (or approximately solve) the following problems or ones taking a similar form.

$$\mathbf{f}^* = \underset{\mathbf{f} \in \bar{\mathcal{F}}}{\operatorname{argmax}} \operatorname{SNR}_{\text{tx}}(\mathbf{f}), \quad \mathbf{w}^* = \underset{\mathbf{w} \in \bar{\mathcal{W}}}{\operatorname{argmax}} \operatorname{SNR}_{\text{rx}}(\mathbf{w}) \quad (16)$$

If selecting the transmit and receive beams independently to maximize each link's SNR, the selected beam pair may couple high SI when using traditional beamforming codebooks. In other words, $\operatorname{INR}_{\text{rx}}(\mathbf{f}^*, \mathbf{w}^*)$ may be much greater than 0 dB. In fact, we show this has been confirmed by recent measurements [76], which we cover shortly in Subsection 4.7. Therefore, one can imagine it would be preferable from a full-duplex perspective to jointly select transmit and receive beams that deliver high $\operatorname{SNR}_{\text{tx}}$ and $\operatorname{SNR}_{\text{rx}}$ and also couple low $\operatorname{INR}_{\text{rx}}$. This is precisely the motivation for [77], which is also introduced in Subsection 4.7. If codebooks could be designed such that all possible $(\mathbf{f}^*, \mathbf{w}^*)$ couple sufficiently low SI, then beam alignment may be conducted on the transmit and receive links independently, as shown in (16), with guarantees of low SI regardless of which beams are selected—the motivation for [54, 73] and the codebook design problem introduced in Subsection 4.3. Creating solutions like these that accommodate beam alignment will be critical to the deployment of full-duplex mmWave systems.

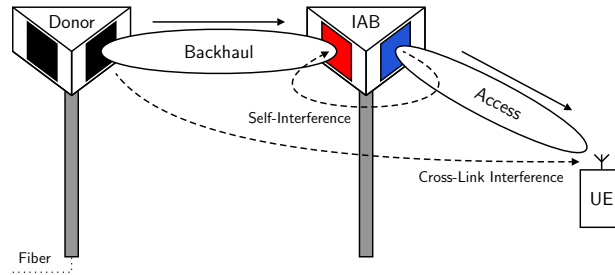


Fig. 15 A full-duplex mmWave IAB node receives wireless backhaul from a fiber-connected donor on one sector while transmitting access to a downlink user on another sector.

Self-Interference Channel Estimation and Limited Channel Knowledge. As mentioned just previously, current practical mmWave systems circumvent downlink/uplink channel estimation via beam alignment, meaning they do not have knowledge of the transmit and receive MIMO channels \mathbf{h}_{TX} and \mathbf{h}_{RX} . Practical beamforming-based solutions for full-duplex mmWave systems should account for this. Efficient and accurate estimation of the SI MIMO channel \mathbf{H} is a research problem still in its infancy. This is perhaps most largely due to the fact that modeling \mathbf{H} is still an open research problem itself, whose outcomes may inspire strategies for its estimation. MIMO channel estimation in mmWave transceivers is complicated by the sheer size of these channels and the fact that DACs and ADCs observe the channel through the lens of analog beamformers [53]. Routes to reduce estimation overhead would be valuable contributions, potentially by leveraging static portions of the SI channel (e.g., the direct coupling between the arrays) and/or by accurate channel modeling. Nonetheless, whatever SI channel estimation strategies are developed will naturally be imperfect to some degree, suggesting that practical designs should be robust to channel estimation error. Robustness is especially critical in full-duplex settings, since small errors in mitigating SI can be detrimental due to its overwhelming strength.

Leveraging User Selection. A full-duplex mmWave base station will likely serve multiple downlink users and uplink users over many time slots. As assumed thus far, let us consider the case where the base station can serve a single downlink-uplink user pair in a full-duplex fashion at any given time, multiplexing user pairs in time. The degree of SI coupled at the full-duplex base station depends on the transmit and receive beams when serving a particular downlink-uplink user pair. In addition, the degree of cross-link interference depends on the two users being served. When given a pool of user pairs needing service, one can therefore imagine that strategically selecting which user pair to serve has the potential to be a powerful tool to improve full-duplex performance [9]. This concept has not been fully fleshed out in existing literature and deserves future study. Interesting future work, for instance, would be the design of intelligent schedulers that incorporate full-duplex user selection to improve link-level spectral efficiency and gains in network-level throughput.

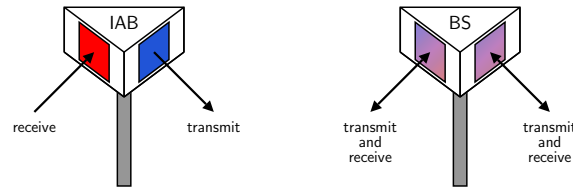


Fig. 16 (left) An inter-sector full-duplex IAB node full-duplexes transmission and reception across sectors. (right) An intra-sector full-duplex IAB node full-duplexes transmission and reception within each sector.

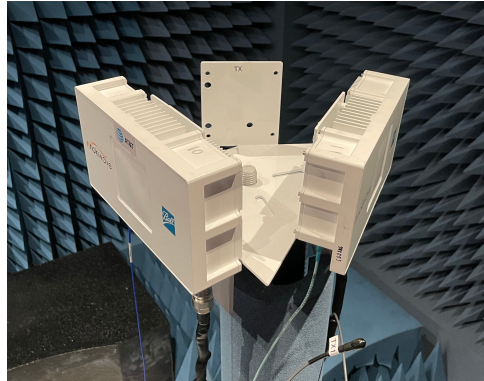
4.6 Full-Duplex Integrated Access and Backhaul

A particularly motivating application of full-duplex in mmWave cellular networks is in integrated access and backhaul (IAB) [45, 78, 79], where a fiber-connected base station serves nearby users and is responsible for maintaining a wireless backhaul to one or more nearby base stations, as illustrated in Fig. 15. By using the same pool of mmWave spectrum for access and wireless backhauling, dense mmWave networks can be deployed with fewer dedicated fiber connections, which reduces the cost, time, and permitting associated with deployment. Like other multi-hop wireless networks, however, IAB networks have faced scaling challenges due to degraded throughput and higher latency as the network grows. Recent work has investigated the use of full-duplex to overcome these obstacles [19, 71].

Inter-sector and intra-sector full-duplex. An IAB node can operate in a full-duplex fashion in two main ways, depending on scope. Consider a sectorized IAB node with three sectors, each equipped with a transceiver serving a 120° field-of-view. The first potential full-duplex operating mode, which we refer to as *inter-sector* full-duplex, allows each sector to either transmit or receive, meaning SI may be inflicted by one sector onto one or both of the other sectors. Fig. 15 depicts inter-sector full-duplex, for instance, where one sector receives while another transmits in-band. With inter-sector full-duplex, sectors are no longer required to collectively transmit or collectively receive but rather can be scheduled independently, and the transceiver on each sector can be merely half-duplex-capable. This full-duplex mode unlocks scheduling opportunities that are otherwise not available, allowing the network to achieve higher throughput, as we will highlight shortly.

The other potential full-duplex operating mode we refer to as *intra-sector* full-duplex, where each sector is equipped with a full-duplex transceiver, and we illustrate in Fig. 16. In this case, each sector may transmit and receive from a downlink-uplink user pair within its field-of-view. Notice that, when each sector operates simultaneously and in-band as its neighboring sectors, SI is inflicted from each transmitting sector onto each receiving sector. Naturally, intra-sector full-duplex has the potential to outperform inter-sector full-duplex, but the gains of such have not been thoroughly investigated. Early deployments of full-duplex mmWave base

Fig. 17 The 28 GHz phased array platform used for measurements of SI in [76, 77, 80]. Transmit array on right; receive array on left. Received SI power depends on the steering direction of the transmit and receive beams. The multi-panel triangular platform shown here is a relevant deployment option for mmWave small-cell base stations and IAB nodes. Reproduced from [76] with permission.



stations, especially for IAB, will likely be of the inter-sector form, since SI is likely less severe and half-duplex transceivers can be used per sector.

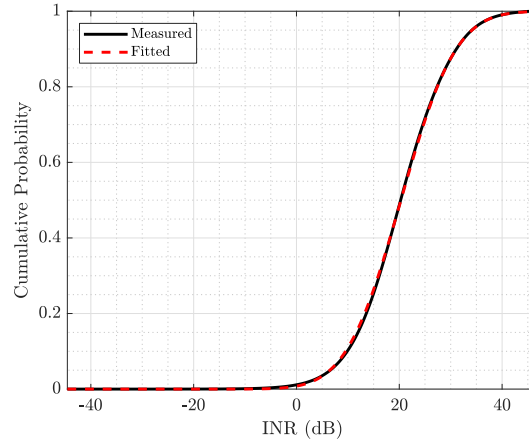
Recent Progress Validating Full-Duplex IAB. Full-duplex IAB networks are studied in [19] to characterize the throughput and latency gains when upgrading IAB nodes from half-duplex to full-duplex transceivers. Note that, in [19], users were kept as half-duplex devices, which is a realistic assumption for the foreseeable future. The authors show through analysis and simulation that, with full-duplex IAB nodes, user latency can reduce four-fold and user throughput can improve eight-fold for fourth-hop users—far transcending the familiar doubling of spectral efficiency offered by full-duplex at the link level. In general, [19] shows that users further from the donor enjoy greater performance improvements with full-duplex. This can be explained by the fact that full-duplex IAB networks can meet latency and throughput targets that half-duplex IAB networks cannot, yielding relative gains that can tend to infinity. Ultimately, the gains offered by full-duplex are thanks to the scheduling opportunities it unlocks: certain links that must be orthogonalized with half-duplex IAB nodes need not be with full-duplex, allowing packets to more quickly propagate through the multi-hop network. Compared to their half-duplex counterparts, full-duplex IAB networks can facilitate reduced latency, higher throughput, fairer service, and deeper networks—even with imperfect SIC [19].

4.7 Recent Experimental Research Outcomes

The majority of research on full-duplex mmWave systems has been theoretical in nature, using simulation to validate proposed solutions. Recently, there has been work experimentally investigating full-duplex mmWave systems, two efforts of which we introduce herein.

Measurements of mmWave Self-Interference. SI was studied quite extensively over the past decade or so, largely in the context of sub-6 GHz transceivers. To

Fig. 18 The empirical CDF of the nearly 6.5 million measurements of 28 GHz SI using 16×16 planar arrays in [76], along with a fitted log-normal distribution. Less than 1% of transmit-receive beam pairs yield $\text{INR}_{\text{rx}} \leq 0$ dB. Highly directional mmWave beams do not necessarily offer sufficient isolation for full-duplex but strategically selecting them can. Reproduced from [76] with permission.



progress development of full-duplex mmWave systems, a necessary first step is to better understand SI in mmWave systems. Measurements of mmWave SI in [71, 81–86] provide useful insights but do not offer a means to evaluate proposed mmWave full-duplex solutions since they provide neither a MIMO SI channel model nor adequate beam-based measurements; most of these were taken using horn or lens antennas, not phased arrays. To evaluate beamforming-based mmWave full-duplex solutions thus far, researchers have primarily used highly idealized channel models. To address these shortcomings, a measurement campaign of SI at 28 GHz was conducted in [76, 80] using a multi-panel 16×16 -element phased array platform. In this campaign [76, 80], a spatial inspection of SI was conducted in an anechoic chamber by electronically sweeping the beams of the transmit phased array and receive phased array across a number of combinations in azimuth and elevation. For each transmit direction and receive direction, SI power was measured, for a total of nearly 6.5 million measurements. This work showed that SI indeed tends to be well above the noise floor—even with highly directional mmWave beams—but select transmit-receive beam pairs coupled levels of SI below the noise floor without any additional cancellation. These measurements revealed large-scale trends based on steering direction, along with noteworthy small-scale phenomena when beams undergo small shifts (on the order of one degree). The authors provide a statistical characterization of their measurements, allowing researchers to draw realistic realizations of SI and conduct statistical analyses. A key takeaway from this work showed that a commonly-used idealized near-field channel model (i.e., the spherical-wave channel model [87]) is not a suitable one for practical mmWave full-duplex systems. This motivates the need for a new measurement-backed channel model for SI in full-duplex mmWave systems.

Beam Selection for Full-Duplex mmWave Communication Systems. A particularly exciting observation from the measurements in [76] was that slightly shifting the transmit and/or receive beams at the full-duplex transceiver (on the order of one degree) could significantly reduce SI, often by 20 dB or more. This motivated the

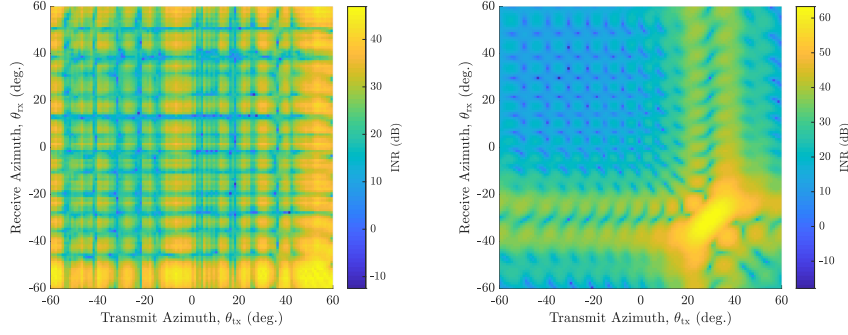


Fig. 19 (left) The azimuth cut of SI measurements from [76]. (right) The simulated counterpart of (left) using a popular idealized near-field channel model for SI [87]. The stark difference between the two motivates the need for a new, measurement-backed SI channel model. Reproduced from [76] with permission.

work of [77], in which the authors propose the first beam selection methodology for full-duplex mmWave communication systems. Traditional beam selection typically selects beams that maximize SNR via codebook-based beam alignment measurements, as highlighted in (16). In [77], the authors propose a measurement-driven beam selection methodology, called *STEER*, atop conventional beam alignment that incorporates SI into transmit and receive beam selection at a full-duplex mmWave transceiver.

Suppose a full-duplex base station serves a downlink user and uplink user as illustrated in Fig. 12. Taking the perspective of a full-duplex base station, let \mathcal{A}_{tx} be a set (a codebook) of N_{tx} candidate transmit beam steering directions (azimuth-elevation pairs) used during beam alignment, and let \mathcal{A}_{rx} be a codebook of N_{rx} candidate receive beam steering directions defined analogously (e.g., see Fig. 14).

$$\mathcal{A}_{\text{tx}} = \left\{ \left(\theta_{\text{tx}}^{(i)}, \phi_{\text{tx}}^{(i)} \right) : i = 1, \dots, N_{\text{tx}} \right\} \quad (17)$$

$$\mathcal{A}_{\text{rx}} = \left\{ \left(\theta_{\text{rx}}^{(j)}, \phi_{\text{rx}}^{(j)} \right) : j = 1, \dots, N_{\text{rx}} \right\} \quad (18)$$

Solving the following beam selection problems through conventional beam alignment yields initial beam selections at the full-duplex base station.

$$\left(\theta_{\text{tx}}^{(i^*)}, \phi_{\text{tx}}^{(i^*)} \right) = \underset{(\theta, \phi) \in \mathcal{A}_{\text{tx}}}{\operatorname{argmax}} \operatorname{SNR}_{\text{tx}}(\theta, \phi) \quad (19)$$

$$\left(\theta_{\text{rx}}^{(j^*)}, \phi_{\text{rx}}^{(j^*)} \right) = \underset{(\theta, \phi) \in \mathcal{A}_{\text{rx}}}{\operatorname{argmax}} \operatorname{SNR}_{\text{rx}}(\theta, \phi) \quad (20)$$

While these initial beam selections may yield high SNRs, they are likely to couple high levels of SI, shown by measurements in [76]. To identify transmit and receive beams that the base station can use to deliver high SNRs and low SI, the authors of [77] propose *STEER*.

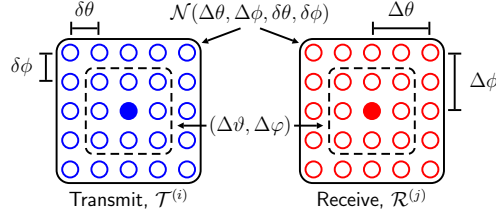


Fig. 20 The spatial neighborhoods surrounding a given transmit direction and receive direction (shown as filled circles). The size of the neighborhoods is dictated by $(\Delta\theta, \Delta\phi)$ and their resolution by $(\delta\theta, \delta\phi)$. The sub-neighborhood $(\Delta\vartheta, \Delta\varphi)$ is relevant in problem (26) [77]. Reproduced from [77] with permission.

STEER leverages the small-scale variability observed in the measurements of [76] to preserve high SNR_{TX} and high SNR_{RX} while reducing INR_{RX} . To identify attractive steering directions for full-duplex operation, STEER measures the SI incurred when transmitting and receiving around the *spatial neighborhoods* surrounding the initial transmit and receive steering directions, as described by Fig. 20. Quantifying the size of these spatial neighborhoods, let $\Delta\theta$ and $\Delta\phi$ be maximum absolute azimuthal and elevational deviations from the given transmit direction and receive direction. Discretizing these neighborhoods, let $\delta\theta$ and $\delta\phi$ be the measurement resolution in azimuth and elevation, respectively, which should not be larger than $(\Delta\theta, \Delta\phi)$. For instance, $(\delta\theta, \delta\phi) = (1^\circ, 1^\circ)$ while $(\Delta\theta, \Delta\phi) = (2^\circ, 2^\circ)$. The spatial neighborhood \mathcal{N} surrounding a transmit/receive direction can be expressed using the azimuthal neighborhood \mathcal{N}_θ and elevational neighborhood \mathcal{N}_ϕ defined as

$$\mathcal{N}_\theta(\Delta\theta, \delta\theta) = \left\{ m \cdot \delta\theta : m \in \left[-\left\lfloor \frac{\Delta\theta}{\delta\theta} \right\rfloor, \left\lfloor \frac{\Delta\theta}{\delta\theta} \right\rfloor \right] \right\} \quad (21)$$

$$\mathcal{N}_\phi(\Delta\phi, \delta\phi) = \left\{ n \cdot \delta\phi : n \in \left[-\left\lfloor \frac{\Delta\phi}{\delta\phi} \right\rfloor, \left\lfloor \frac{\Delta\phi}{\delta\phi} \right\rfloor \right] \right\} \quad (22)$$

where $\lfloor \cdot \rfloor$ is the floor operation and $[a, b] = \{a, a+1, \dots, b-1, b\}$. The complete neighborhood is the product of the azimuthal and elevational neighborhoods as

$$\mathcal{N}(\Delta\theta, \Delta\phi, \delta\theta, \delta\phi) = \{(\theta, \phi) : \theta \in \mathcal{N}_\theta(\Delta\theta, \delta\theta), \phi \in \mathcal{N}_\phi(\Delta\phi, \delta\phi)\}. \quad (23)$$

The spatial neighborhoods $\mathcal{T}^{(i^*)}$ and $\mathcal{R}^{(j^*)}$ surrounding the transmit and receive directions output by conventional beam selection are respectively written as

$$\mathcal{T}^{(i^*)} = \left(\theta_{\text{TX}}^{(i^*)}, \phi_{\text{TX}}^{(i^*)} \right) + \mathcal{N}(\Delta\theta, \Delta\phi, \delta\theta, \delta\phi) \quad (24)$$

$$\mathcal{R}^{(j^*)} = \underbrace{\left(\theta_{\text{RX}}^{(j^*)}, \phi_{\text{RX}}^{(j^*)} \right)}_{\text{initial selection}} + \underbrace{\mathcal{N}(\Delta\theta, \Delta\phi, \delta\theta, \delta\phi)}_{\text{neighborhood}}. \quad (25)$$

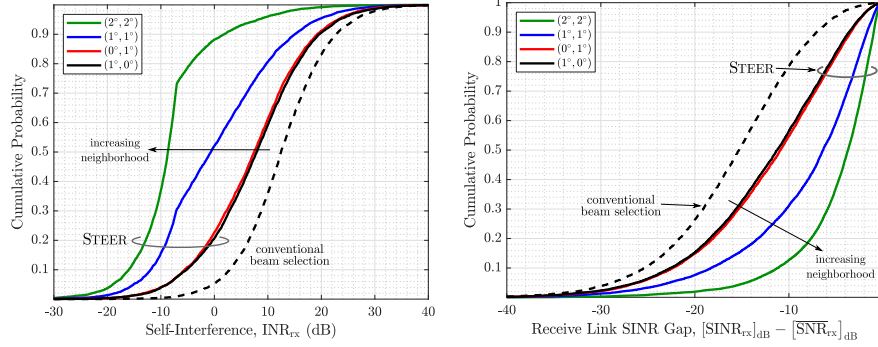


Fig. 21 (left) The CDF of INR_{rx} for various neighborhood sizes $(\Delta\theta, \Delta\phi)$. (right) The CDF of the gap between SINR_{rx} and its upper bound $\overline{\text{SNR}}_{\text{rx}}$ for various neighborhood sizes $(\Delta\theta, \Delta\phi)$. STEER reliably reduces INR_{rx} , as evident in (left), while maintaining high beamforming gain, shifting SINR_{rx} closer to $\overline{\text{SNR}}_{\text{rx}}$ as shown in (right) [77]. Reproduced from [77] with permission.

Let $\text{INR}_{\text{rx}}(\theta_{\text{tx}}, \phi_{\text{tx}}, \theta_{\text{rx}}, \phi_{\text{rx}})$ be the receive link INR due to SI when transmitting toward $(\theta_{\text{tx}}, \phi_{\text{tx}})$ and receiving toward $(\theta_{\text{rx}}, \phi_{\text{rx}})$ at the full-duplex base station. STEER solves the following beam selection problem to net a transmit direction $(\theta_{\text{tx}}^*, \phi_{\text{tx}}^*)$ and receive direction $(\theta_{\text{rx}}^*, \phi_{\text{rx}}^*)$ that the full-duplex transceiver will use.

$$(\theta_{\text{tx}}^*, \phi_{\text{tx}}^*), (\theta_{\text{rx}}^*, \phi_{\text{rx}}^*) = \underset{\substack{(\theta_{\text{tx}}, \phi_{\text{tx}}) \\ (\theta_{\text{rx}}, \phi_{\text{rx}})}}{\text{argmin}} \min_{(\Delta\vartheta, \Delta\varphi)} \Delta\vartheta^2 + \Delta\varphi^2 \quad (26a)$$

$$\text{s.t. } \text{INR}_{\text{rx}}(\theta_{\text{tx}}, \phi_{\text{tx}}, \theta_{\text{rx}}, \phi_{\text{rx}}) \leq \max(\text{INR}_{\text{rx}}^{\text{tgt}}, \text{INR}_{\text{rx}}^{\text{min}}) \quad (26b)$$

$$(\theta_{\text{tx}}, \phi_{\text{tx}}) \in \left(\theta_{\text{tx}}^{(i^*)}, \phi_{\text{tx}}^{(i^*)} \right) + \mathcal{N}(\Delta\vartheta, \Delta\varphi, \delta\theta, \delta\phi) \quad (26c)$$

$$(\theta_{\text{rx}}, \phi_{\text{rx}}) \in \left(\theta_{\text{rx}}^{(j^*)}, \phi_{\text{rx}}^{(j^*)} \right) + \mathcal{N}(\Delta\vartheta, \Delta\varphi, \delta\theta, \delta\phi) \quad (26d)$$

$$0 \leq \Delta\vartheta \leq \Delta\theta, 0 \leq \Delta\varphi \leq \Delta\phi \quad (26e)$$

Here, $\text{INR}_{\text{rx}}^{\text{min}}$ is the minimum INR over the entire $(\Delta\theta, \Delta\phi)$ spatial neighborhood and $\text{INR}_{\text{rx}}^{\text{tgt}}$ is an INR target the system aims for. By constraining the distance $\Delta\vartheta^2 + \Delta\varphi^2$, STEER minimizes the deviation of the selected beams make from the initial selections and thus can preserve high SNR_{tx} and SNR_{rx} . Reducing INR_{rx} , therefore, leads to SINR improvements over the initial beam selections. In [77], the authors present an algorithm to solve problem (26) with a minimal number of SI measurements. Evaluation of STEER with 28 GHz phased arrays highlights its ability to reduce SI while preserving high SNRs, courtesy of noteworthy variability of SI over small spatial neighborhoods. This can be seen in Fig. 21, which shows STEER's potential as a full-duplex solution without any supplemental analog or digital SIC. SI can be reliably reduced with STEER to below the noise floor with $(\Delta\theta, \Delta\phi) = (2^\circ, 2^\circ)$, and by preserving SNR while doing so, it can increase SINR toward its upper bound.

5 A Look Ahead: What is the Future of Full-Duplex?

We conclude this chapter by highlighting several key topics that need further research and development from engineers in industry and academia to advance and mature full-duplex technology.

Real-time, deployment-ready full-duplex solutions. Most full-duplex solutions are validated in simulation, lab settings, or controlled environments. Moreover, most evaluations ignore the overhead associated with configuring a full-duplex solution, which is a key hurdle in practical deployments. In practice, the radio resources consumed to configure a full-duplex solution must not outweigh the gains it offers. It is essential that full-duplex solutions are designed and evaluated with real-time deployments in mind—with strict overhead requirements, with the ability to adapt to dynamic environments, and with minimal form-factors and power consumption.

Further advancement of full-duplex mmWave and terahertz systems. While increased attention has been devoted to the research and development of full-duplex mmWave systems, there remain plenty of open problems that need addressing before such systems are brought to life. Continuing to develop means to mitigate SI is certainly welcome, along with characterizing the gains full-duplex can offer mmWave networks and prototyping proofs-of-concept. In addition, identifying and creating applications that are particularly beneficial from full-duplex would offer new directions and requirements for its solutions. For instance, the role of full-duplex in reconfigurable intelligent surface (RIS) applications has been investigated recently as it poses a means to better use RIS resources and aid in the cancellation of SI [88, 89]. Finally, exploring full-duplex terahertz systems, their applications, and how solutions for such may differ from those at mmWave would be valuable future work.

Further advancement of machine learning to enable full-duplex. To supplement existing work, the prospects of using machine learning for full-duplex are still quite open-ended, especially beyond digital SIC. Using machine learning to configure and adaptively update analog SIC filters, for instance, or to configure beamformers that mitigate SI in full-duplex mmWave systems are topics that have yet to be fully explored. In addition, machine learning may be able to reduce the effects of transceiver impairments in full-duplex through digital predistortion. There are also the prospects of using machine learning to intelligently schedule users and proactively manage cross-link interference within full-duplex networks.

Network-level studies comparing full-duplex to other duplexing technologies. To justify the deployment of wireless networks equipped with full-duplex, it is paramount that researchers conduct studies that prove its network-level gains over traditional multiplexing strategies, such as TDD, FDD, and space-division multiple access, as well as the recently-proposed cross-division duplexing (XDD) [90]. This has been examined fairly extensively for sub-6 GHz wireless networks but less so for mmWave networks and applications of IAB, in particular.

Full-duplex in joint communication and sensing systems. Joint communication and sensing is expected to play an important role in the next evolution of wireless networks. Full-duplex solutions have the opportunity to facilitate high-fidelity sensing while transmitting by eliminating SI [10, 26]. Sensing information may be used to directly improve communication performance [91] or for higher-level applications. In addition, there are opportunities for full-duplex to enable the sensing and jamming of eavesdroppers to establish more secure communication [27]. Finally, the relationship between SI channel estimation and environmental sensing poses an opportunity for the two to supplement and/or justify one another. For instance, accurate sensing of the environment may yield an estimate of the SI channel and, in turn, enable full-duplex operation.

Integrating full-duplex into wireless standards. Wireless networks of today have been built on decades of a half-duplex assumption at each transceiver. Full-duplex offers immediate upgrades to a transceiver but, to effectively make use of this powerful capability, wireless networks must support it. This motivates the need for research on seamlessly adopting full-duplex operation into existing wireless standards. Work items and studies in the 3rd Generation Partnership Project (3GPP) continue to investigate the merits and standardization of full-duplex in cellular systems, especially in the context of IAB in Releases 17, 18, and beyond [45].

Acknowledgements We would like to thank Jeffrey G. Andrews, Sriram Vishwanath, Aditya Chopra, Thomas Novlan, Manan Gupta, Rajesh K. Mishra, and Hardik B. Jain for the discussions, feedback, and collaborations that contributed to the preparation of this chapter. The work of Ian P. Roberts was supported by the U.S. National Science Foundation under Grant 1610403.

References

1. A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.
2. K. E. Kolodziej, B. T. Perry, and J. S. Herd, "In-band full-duplex technology: Techniques and systems survey," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3025–3041, Feb. 2019.
3. S. Chen, M. A. Beach, and J. P. McGeehan, "Division-free duplex for wireless applications," *Electronics Letters*, vol. 34, pp. 147–148, Jan. 1998.
4. J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. ACM MobiCom*, Sep. 2010, pp. 1–12.
5. S. Hong, J. Brand, J. I. Choi, M. Jain, J. Mehlman, S. Katti, and P. Levis, "Applications of self-interference cancellation in 5G and beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb. 2014.
6. H. Alves, T. Riihonen, and H. A. Suraweera, *Full-Duplex Communications for Future Wireless Networks*. Singapore: Springer, 2020.
7. M. Chung, M. S. Sim, J. Kim, D. K. Kim, and C.-B. Chae, "Prototyping real-time full duplex radios," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 56–63, Sep. 2015.
8. Z. Xiao, P. Xia, and X. Xia, "Full-duplex millimeter-wave communication," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 136–143, Dec. 2017.

9. I. P. Roberts, J. G. Andrews, H. B. Jain, and S. Vishwanath, "Millimeter-wave full duplex radios: New challenges and techniques," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 36–43, Feb. 2021.
10. C. B. Barneto, S. D. Liyanaarachchi, M. Heino, T. Riihonen, and M. Valkama, "Full duplex radio/radar technology: The enabler for advanced joint communication and sensing," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 82–88, Feb. 2021.
11. Y. Kurzo, A. Burg, and A. Balatsoukas-Stimming, "Design and implementation of a neural network aided self-interference cancellation scheme for full-duplex radios," in *Proc. Asilomar Conf. Signals, Sys., and Comput.*, Oct. 2018, pp. 589–593.
12. Y. Liao, K. Bian, L. Song, and Z. Han, "Full-duplex MAC protocol design and analysis," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1185–1188, Jul. 2015.
13. Y. Liao, T. Wang, L. Song, and Z. Han, "Listen-and-talk: Full-duplex cognitive radio networks," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 3068–3073.
14. Y. Liao, L. Song, Z. Han, and Y. Li, "Full duplex cognitive radio: a new design paradigm for enhancing spectrum usage," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 138–145, May 2015.
15. L. Song, Y. Liao, K. Bian, L. Song, and Z. Han, "Cross-layer protocol design for CSMA/CD in full-duplex WiFi networks," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 792–795, Apr. 2016.
16. K. M. Thilina, H. Tabassum, E. Hossain, and D. I. Kim, "Medium access control design for full duplex wireless systems: challenges and approaches," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 112–120, May 2015.
17. R. K. Mishra, Y. Chen, and I. P. Roberts, "Collision detection in dense Wi-Fi networks using self-interference cancellation," in *Proc. IEEE ICC Wkshp.*, Jun. 2020, pp. 1–6.
18. S. Haddad, A. Özgür, and E. Telatar, "Can full-duplex more than double the capacity of wireless networks?" in *Proc. IEEE ISIT*, Jun. 2017, pp. 963–967.
19. M. Gupta, I. P. Roberts, and J. G. Andrews, "System-level analysis of full-duplex self-backhauled millimeter wave networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1130–1144, Feb. 2023.
20. R. W. Heath Jr. and A. Lozano, *Foundations of MIMO Communication*. Cambridge, UK: Cambridge University Press, 2018.
21. M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Nov. 2012.
22. M. Jain, J. I. Choi, T. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. ACM MobiCom*, Sep. 2011, pp. 301–312.
23. D. Bharadia, E. McMillin, and S. Katti, "Full duplex radios," in *Proc. ACM SIGCOMM*, Aug. 2013, pp. 375–386.
24. D. Korpi, "Full-duplex wireless: Self-interference modeling, digital cancellation, and system studies," Ph.D. dissertation, Tampere University of Technology, Dec. 2017.
25. S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
26. Z. Xiao and Y. Zeng, "Waveform design and performance analysis for full-duplex integrated sensing and communication," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1823–1837, Mar. 2022.
27. X. Wang, Z. Fei, J. A. Zhang, and J. Huang, "Sensing-assisted secure uplink communications with full-duplex base station," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 249–253, Dec. 2022.
28. G. Zheng, I. Krikidis, J. Li, A. P. Petropulu, and B. Ottersten, "Improving physical layer secrecy using full-duplex jamming receivers," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4962–4974, Oct. 2013.
29. T. Riihonen, D. Korpi, O. Rantula, H. Rantanen, T. Saarelainen, and M. Valkama, "Inband full-duplex radio transceivers: A paradigm shift in tactical communications and electronic warfare?" *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 30–36, Oct. 2017.
30. J. Suarez, K. Kravtsov, and P. R. Prucnal, "Incoherent method of optical interference cancellation for radio-frequency communications," *IEEE J. Quantum Electron.*, vol. 45, no. 4, pp. 402–408, Mar. 2009.
31. X. Han, B. Huo, Y. Shao, and M. Zhao, "Optical RF self-interference cancellation by using an integrated dual-parallel MZM," *IEEE Photon. J.*, vol. 9, no. 2, pp. 1–8, Apr. 2017.

32. X. Su, X. Han, S. Fu, S. Wang, C. Li, Q. Tan, G. Zhu, C. Wang, Z. Wu, Y. Gu, and M. Zhao, "Optical multipath RF self-interference cancellation based on phase modulation for full-duplex communication," *IEEE Photon. J.*, vol. 12, no. 4, pp. 1–14, Jun. 2020.
33. D. Korpi, T. Riihonen, V. Syrjälä, L. Anttila, M. Valkama, and R. Wichman, "Full-duplex transceiver system calculations: Analysis of ADC and linearity challenges," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3821–3836, Jul. 2014.
34. D. Korpi, L. Anttila, V. Syrjälä, and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion full-duplex transceiver," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1674–1687, Sep. 2014.
35. H. Guo, J. Xu, S. Zhu, and S. Wu, "Realtime software defined self-interference cancellation based on machine learning for in-band full duplex wireless communications," in *Proc. IEEE ICNC*, Mar. 2018, pp. 779–783.
36. A. Balatsoukas-Stimming, "Non-linear digital self-interference cancellation for in-band full-duplex radios using neural networks," in *Proc. IEEE SPAWC*, Jun. 2018, pp. 1–5.
37. Y. Kurzo, A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Hardware implementation of neural self-interference cancellation," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 2, pp. 204–216, Jun. 2020.
38. H. Guo, S. Wu, H. Wang, and M. Daneshmand, "DSIC: Deep learning based self-interference cancellation for in-band full duplex wireless," in *Proc. IEEE GLOBECOM*, Dec. 2019, pp. 1–6.
39. Y. Chen, R. K. Mishra, D. Schwartz, and S. Vishwanath, "MIMO full duplex radios with deep learning," in *Proc. IEEE ICC Wkshp.*, Jun. 2020, pp. 1–6.
40. A. Balatsoukas-Stimming, "Joint detection and self-interference cancellation in full-duplex systems using machine learning," in *Proc. Asilomar Conf. Signals, Sys., and Comput.*, Oct. 2021, pp. 989–992.
41. L. Laughlin, M. A. Beach, K. A. Morris, and J. L. Haine, "Optimum single antenna full duplex using hybrid junctions," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1653–1661, Jun. 2014.
42. T. Dinc, M. Tymchenko, A. Nagulu, D. Sounas, A. Alu, and H. Krishnaswamy, "Synchronized conductivity modulation to realize broadband lossless magnetic-free non-reciprocity," *Nature Commun.*, vol. 8, no. 11, pp. 1–9, Oct. 2017.
43. J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
44. X. Liu, Z. Xiao, L. Bai, J. Choi, P. Xia, and X.-G. Xia, "Beamforming based full-duplex for millimeter-wave communication," *Sensors*, vol. 16, no. 7, p. 1130, Jul. 2016.
45. 3GPP, "3GPP TS 38.174: New WID on IAB enhancements," 2021. [Online]. Available: <https://www.3gpp.org/dynareport/38174.htm>
46. D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. USENIX NSDI*, Apr. 2014, pp. 359–372.
47. S. Huberman and T. Le-Ngoc, "MIMO full-duplex precoding: A joint beamforming and self-interference cancellation structure," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2205–2217, Apr. 2015.
48. T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Aug. 2011.
49. H. A. Suraweera, I. Krikidis, G. Zheng, C. Yuen, and P. J. Smith, "Low-complexity end-to-end performance optimization in MIMO full-duplex relay systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 913–927, Jan. 2014.
50. H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Jun. 2014.
51. B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1541–1553, Sep. 2012.

52. E. Everett, C. Shepard, L. Zhong, and A. Sabharwal, "SoftNull: Many-antenna full-duplex wireless via digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8077–8092, Dec. 2016.
53. R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
54. I. P. Roberts, H. B. Jain, S. Vishwanath, and J. G. Andrews, "Millimeter wave analog beamforming codebooks robust to self-interference," in *Proc. IEEE GLOBECOM*, Dec. 2021, pp. 1–6.
55. R. López-Valcarce and N. González-Prelcic, "Beamformer design for full-duplex amplify-and-forward millimeter wave relays," in *Proc. ISWCS*, Aug. 2019, pp. 86–90.
56. R. López-Valcarce and N. González-Prelcic, "Analog beamforming for full-duplex millimeter wave communication," in *Proc. ISWCS*, Aug. 2019, pp. 687–691.
57. J. Palacios, J. Rodríguez-Fernández, and N. González-Prelcic, "Hybrid precoding and combining for full-duplex millimeter wave communication," in *Proc. IEEE GLOBECOM*, Dec. 2019, pp. 1–6.
58. K. Satyanarayana, M. El-Hajjar, P. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1394–1404, Feb. 2019.
59. L. Zhu, J. Zhang, Z. Xiao, X. Cao, X. Xia, and R. Schober, "Millimeter-wave full-duplex UAV relay: Joint positioning, beamforming, and power control," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2057–2073, Sep. 2020.
60. J. M. B. da Silva, A. Sabharwal, G. Fodor, and C. Fischione, "1-bit phase shifters for large-antenna full-duplex mmWave communications," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6916–6931, Oct. 2020.
61. R. López-Valcarce and M. Martínez-Cotelo, "Full-duplex mmWave MIMO with finite-resolution phase shifters," *IEEE Trans. Wireless Commun.*, May 2022, (early access).
62. A. Koc and T. Le-Ngoc, "Full-duplex mmWave massive MIMO systems: A joint hybrid precoding/combining and self-interference cancellation design," *IEEE Open J. Commun. Society*, vol. 2, pp. 754–774, Mar. 2021.
63. Y. Cai, Y. Xu, Q. Shi, B. Champagne, and L. Hanzo, "Robust joint hybrid transceiver design for millimeter wave full-duplex MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1199–1215, Feb. 2019.
64. I. P. Roberts, J. G. Andrews, and S. Vishwanath, "Hybrid beamforming for millimeter wave full-duplex under limited receive dynamic range," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7758–7772, Dec. 2021.
65. I. P. Roberts, H. B. Jain, and S. Vishwanath, "Equipping millimeter-wave full-duplex with analog self-interference cancellation," in *Proc. IEEE ICC Wkshp.*, Jun. 2020, pp. 1–6.
66. G. C. Alexandropoulos, M. A. Islam, and B. Smida, "Full duplex hybrid A/D beamforming with reduced complexity multi-tap analog cancellation," in *Proc. IEEE SPAWC*, May 2020, pp. 1–5.
67. G. C. Alexandropoulos and M. Duarte, "Joint design of multi-tap analog cancellation and digital beamforming for reduced complexity full duplex MIMO systems," in *Proc. IEEE ICC*, May 2017, pp. 1–7.
68. V. Singh, S. Mondal, A. Gadre, M. Srivastava, J. Paramesh, and S. Kumar, "Millimeter-wave full duplex radios," in *Proc. ACM MobiCom*, Apr. 2020.
69. C. Zhang and X. Luo, "Adaptive digital self-interference cancellation for millimeter-wave full-duplex backhaul systems," *IEEE Access*, vol. 7, pp. 175 542–175 553, Dec. 2019.
70. T. Dinc, A. Chakrabarti, and H. Krishnaswamy, "A 60 GHz CMOS full-duplex transceiver and link with polarization-based antenna and RF cancellation," *IEEE J. Solid-State Circuits*, vol. 51, no. 5, pp. 1125–1140, May 2016.
71. G. Y. Suk, S.-M. Kim, J. Kwak, S. Hur, E. Kim, and C.-B. Chae, "Full duplex integrated access and backhaul for 5G NR: Analyses and prototype measurements," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 40–46, Aug. 2022.

72. R. López-Valcarce and M. Martínez-Cotelo, "Analog beamforming for full-duplex mmWave communication with low-resolution phase shifters," in *Proc. IEEE ICC*, Jun. 2021, pp. 1–6.
73. I. P. Roberts, S. Vishwanath, and J. G. Andrews, "LONESTAR: Analog beamforming codebooks for full-duplex millimeter wave systems," *IEEE Trans. Wireless Commun.*, Jan. 2023, (early access).
74. Y. Heng, J. G. Andrews, J. Mo, V. Va, A. Ali, B. L. Ng, and J. C. Zhang, "Six key challenges for beam management in 5.5G and 6G systems," *IEEE Commun. Mag.*, vol. 59, no. 7, pp. 74–79, Jul. 2021.
75. J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.
76. I. P. Roberts, A. Chopra, T. Novlan, S. Vishwanath, and J. G. Andrews, "Beamformed self-interference measurements at 28 GHz: Spatial insights and angular spread," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9744–9760, Jun. 2022.
77. I. P. Roberts, A. Chopra, T. Novlan, S. Vishwanath, and J. G. Andrews, "STEER: Beam selection for full-duplex millimeter wave communication systems," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6902–6917, Oct. 2022.
78. M. Cudak, A. Ghosh, A. Ghosh, and J. G. Andrews, "Integrated access and backhaul: A key enabler for 5G millimeter-wave deployments," *IEEE Commun. Mag.*, vol. 59, no. 4, pp. 88–94, Apr. 2021.
79. C. Dehos, J. L. González, A. D. Domenico, D. Kténas, and L. Dussot, "Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5G mobile communications systems?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 88–95, Sep. 2014.
80. A. Chopra, I. P. Roberts, T. Novlan, and J. G. Andrews, "28 GHz phased array-based self-interference measurements for millimeter wave full-duplex," in *Proc. IEEE WCNC*, Apr. 2022, pp. 2583–2588.
81. S. Rajagopal, R. Taori, and S. Abu-Surra, "Self-interference mitigation for in-band mmWave wireless backhaul," in *Proc. IEEE CCNC*, Jan. 2014, pp. 551–556.
82. Y. Kohda, K. Takano, D. Nakano, N. Ohba, T. Yamane, and Y. Katayama, "Single-channel full-duplex mmWave link using phased-array for Ethernet," in *Proc. IEEE CCNC*, Jan. 2015, pp. 400–405.
83. B. Lee, J. Lim, C. Lim, B. Kim, and J. Seol, "Reflected self-interference channel measurement for mmWave beamformed full-duplex system," in *Proc. IEEE GLOBECOM Wkshp.*, Dec. 2015, pp. 1–6.
84. H. Yang, Y. He, C. Jen, C. Liu, S. Jou, X. Yin, M. Ma, and B. Jiao, "Interference measurement and analysis of full-duplex wireless system in 60 GHz band," in *Proc. IEEE APCCAS*, Oct. 2016, pp. 273–276.
85. Y. He, X. Yin, and H. Chen, "Spatiotemporal characterization of self-interference channels for 60-GHz full-duplex communication," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 2220–2223, May 2017.
86. K. Haneda, J. Järveläinen, A. Karttunen, and J. Putkonen, "Self-interference channel measurements for in-band full-duplex street-level backhaul relays at 70 GHz," in *Proc. IEEE PIMRC*, Sep. 2018, pp. 199–204.
87. J.-S. Jiang and M. A. Ingram, "Spherical-wave model for short-range MIMO," *IEEE Trans. Commun.*, vol. 53, no. 9, pp. 1534–1541, Sep. 2005.
88. Y. Liu, Q. Hu, Y. Cai, G. Yu, and G. Y. Li, "Deep-unfolding beamforming for intelligent reflecting surface assisted full-duplex systems," *IEEE Trans. Wireless Commun.*, Dec. 2021, (early access).
89. P. P. Perera, V. G. Warnasooriya, D. Kudathanthirige, and H. A. Suraweera, "Sum rate maximization in STAR-RIS assisted full-duplex communication systems," in *Proc. IEEE ICC*, May 2022, pp. 1–6.
90. H. Ji, Y. Kim, K. Muhammad, C. Tarver, M. Tonnemacher, T. Kim, J. Oh, B. Yu, G. Xu, and J. Lee, "Extending 5G TDD coverage with XDD: Cross division duplex," *IEEE Access*, vol. 9, pp. 51 380–51 392, Mar. 2021.

91. A. Ali, N. González-Prelcic, R. W. Heath, and A. Ghosh, "Leveraging sensing at the infrastructure for mmWave communication," *IEEE Commun. Mag.*, vol. 58, no. 7, pp. 84–89, Jul. 2020.