An Investigation Into Baseband Techniques for Single-Channel Full-Duplex Wireless Communication Systems

Shenghong Li, *Student Member, IEEE*, and Ross D. Murch, *Fellow, IEEE*

***Abstract*—Full-duplex wireless communication is becoming an important research area because of its potential for increasing spectral efficiency. The challenge of such systems lies in cancelling the self-interference. In this paper, we focus on the design of digital cancellation schemes and use them to supplement RF/analog cancellation techniques. The performance of digital cancellation is limited by the non-ideal characteristics of different subsystems in the transceiver, such as analog/digital converter (ADC), power amplifier (PA), and phase noise. It is first shown that given the pre-cancellation achieved by existing RF/analog techniques, the effects of ADC, phase noise, and sampling jitter are not the bottleneck in the system. Instead, the performance of conventional digital cancellation approaches are mainly limited by nonlinearity of the PA and transmit I/Q imbalance. In addition, the output SINR of the desired signal is limited because the estimation precision of the self-interference channel is affected by the desired signal. To overcome these issues, we propose a two-stage iterative self-interference cancellation scheme based on the output signal of the power amplifier. Analytical and simulation results reveal that the proposed cancellation scheme substantially outperforms existing digital cancellation schemes for full-duplex wireless communication systems.**

***Index Terms*—Full-duplex, wireless communication, self-interference cancelation.**

I. Introduction

A KEY objective of future wireless communication systems is to use spectrum efficiently to achieve high capacity and data rates within a limited bandwidth. Various techniques have been proposed over the past decades. A different approach to increase spectrum efficiency that is attracting growing research interest is to consider full-duplex (FD) techniques in wireless communication systems [1]–[10]. Currently, most wireless systems require two separate channels for simultaneous uplink and downlink communication, and common techniques that can be employed are Time Division Duplexing (TDD) or Frequency Division Duplexing (FDD), which separate the uplink and downlink channels in the time/frequency domain. On the contrary, full-duplex allows users to transmit and receive signals simultaneously in the same frequency band and time slot, therefore the same channel can be used for the uplink and downlink communications. The motivation for developing full-duplex communication in wireless systems is to increase the overall capacity [11]. In addition, existing wireless communication systems can be redesigned dramatically to exploit the benefits of full-duplex communications. For example, full-duplex can be applied in Cognitive Radio and Relay Systems [12], [13], and significant changes can be made to the current Medium Access Control (MAC) layer protocols to improve the network throughput [2], [14].

The main challenge of full-duplex wireless communication is caused by the existence of self-interference, which needs to be cancelled from the received signal for the detection of the desired signal (signal of interest). Existing research on this topic can be found in [1]–[10], [15]–[18]. In [1], the authors propose a Radio Frequency (RF) analog echo cancellation method which provides 72 dB isolation for the self-interference signal. In [2], antenna cancellation is combined with RF interference cancellation to eliminate the self-interference, and a similar antenna cancellation technique for multi-antenna systems is presented in [7]. The authors of [3] present some experimental results on the performance of narrow-band self-interference cancellation, which is further characterized in [6]. A digitally controlled near-field cancellation scheme is demonstrated in [4], which provides up to 50 dB of isolation for the self-interference. An inverse signal generated using a transformer is utilized for cancellation in [5], which achieves 73 dB reduction in self-interference when combined with digital cancellation. The impact of phase noise on full-duplex wireless system is considered in [8]. In addition, a digital cancellation scheme is proposed in [10] which takes the effect of RF/analog impairments into consideration and utilizes iterative signal processing technique to improve the SINR performance; and similar techniques have also been used in [19], [20] to cancel the inter-subcarrier interference in subcarrier-based duplexing systems. The effect of ADC on digital cancellation is studied in [15], where a digital cancellation scheme based on the theory of sparse signal recovery is also proposed. The trade-off between ADC resolution, transmit power, physical isolation, and SNR is studied in [16], with the benefit of combining digital and analog cancellation highlighted. For systems with limited number of antennas, the tradeoff between using the antennas for full-duplex communication and using them for half-duplex MIMO communication is analyzed in [9]. In addition, beamforming

Manuscript received June 1, 2013; revised November 20, 2013 and April 9, 2014; accepted July 14, 2014. Date of publication July 22, 2014; date of current version September 8, 2014. This work was supported by Hong Kong RGC under Grant GRF 618209. The associate editor coordinating the review of this paper and approving it for publication was N. C. Sagias.

The authors are with Department of Electronic & Computer Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: eelsh@connect.ust.hk; eermurch@ee.ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2014.2341569

has been used in [13], [17] to suppress self-interference and it has been shown in [13] that receive beamforming can handle self-interference due to the transmit signal noise.

A detailed classification of techniques for cancelling the self-interference can be found in [21]. In this work, we categorize different techniques roughly into 'RF cancellation' and 'digital cancellation', depending on where the cancellation is carried out.[1] The advantage of RF cancellation is that the self-interference is suppressed prior to the ADC, so that the impact of various RF/analog impairments (quantization noise, nonlinearity, phase noise etc.) can be circumvented. In particular, RF cancellation prevents the saturation of ADC and thus reduces the requirement on the resolution of ADC. However, care needs to be taken and it can be expensive to implement RF cancellation techniques that are effective for wideband systems [16]. For example, the approach proposed in [2] is only effective for narrowband system, while the scheme of [5] can only handle the self-interference from a single signal path. The RF cancellation scheme of [22] is based on multiple analog delay lines which is expensive and difficult for adaptive adjusting. On the other hand, by using finite-impulse-response (FIR) filter to model the self-interference channel, digital cancellation can be applied in wideband systems effectively. Compared with RF techniques, digital cancellation is more flexible as the coefficients of FIR can be adjusted conveniently. In addition, advanced digital signal processing techniques can be employed to improve the system performance. However, due to the presence of various system impairments, it is impossible to use digital cancellation alone to achieve full-duplex wireless communication. In particular, since the self-interference is significantly stronger than the signal-of-interest, it is impossible to recover the signal-of-interest due to the limitation of ADC resolution. Given the fact that RF and digital cancellation can supplement each other, one can combine RF and digital cancellation techniques in wideband systems to improve the performance, as in [2], [3], [5], [15]. Specifically, RF techniques can be used to cancel the line-of-sight (LOS) component of the self-interference, so that the effects of some RF/analog impairments (e.g., quantization noise etc.) are reduced to a level that is not the bottleneck of the system. Digital baseband cancellation can be used subsequently to deal with the non-line-of-sight (NLOS) components.

While [1]–[7] has focused on RF techniques, no digital cancellation approaches are employed in [1], [4], [7], and the digital cancellation techniques used in [2], [3], [5], [6] are straightforward. In this paper, we will mainly focus on digital cancellation techniques and propose a novel scheme that can be combined with RF techniques to improve the performance of wideband full-duplex wireless communication systems. The contributions of this paper are as follows:

1) After observing that the effects of ADC, phase noise, and sampling jitter are not the bottleneck of WiFi-like systems given the pre-cancellation achieved by existing RF techniques, we show that PA nonlinearity and transmit I/Q imbalance are the most prominent factors that limit the precision of digital self-interference cancellation. As such, we propose to use the output signal of the power amplifier for more accurate self-interference cancellation.
2) We show that the output SINR of the desired signal using conventional digital cancellation schemes is limited because the estimation precision of the self-interference channel is affected by the desired signal. To address this issue, we propose a two-stage iterative cancellation scheme, which utilizes the detection result of the desired signal to improve the accuracy of self-interference cancellation.

The proposed digital cancellation scheme is evaluated by analysis and simulation, and is shown to substantially outperform conventional digital cancellation schemes.

This paper is organized as follows. In Section II we describe the implementation issues of full-duplex wireless communication systems and the impact of RF/analog impairments on the system; in Section III, we describe the proposed digital cancellation techniques, including self-interference cancellation based on the output signal of power amplifier and two-stage iterative cancellation scheme; the performance of the proposed scheme is analyzed in Section IV, and verified by computer simulations in Section V; finally, a conclusion is drawn in Section VI.

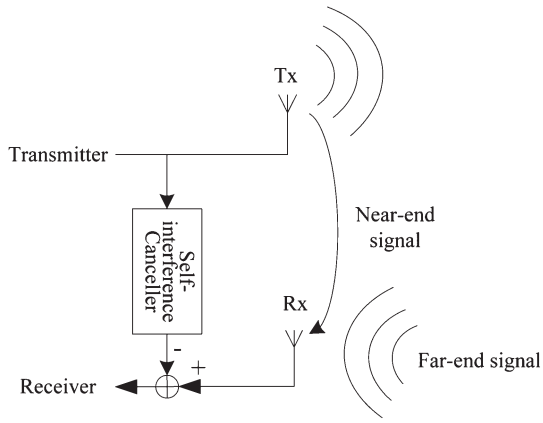

Fig. 1. Basic configuration of a full-duplex wireless device: transmission and reception happens in the same frequency band and time slot. For the purpose of exposition, here we assume that two different antennas are utilized for transmission and reception, respectively.

II. Full-Duplex Wireless Communication Systems

We consider a wireless communication device that transmits and receives signals simultaneously in the same frequency band, as shown in Fig. 1. The device is part of a communication system and its role depends on the application scenario. For example, it may be a full-duplex relay, or one of the two devices in a point to point full duplex system. The signal picked up by the device consists of three components: self-interference due to its own transmission (also denoted as near-end signal hereafter), desired signal transmitted from other device (also denoted as far-end signal), and noise. To facilitate the detection of the far-end signal, the self-interference is cancelled from the received signal, which is shown as 'Self-interference Cancellation' in Fig. 1, and may be carried out using RF techniques and/or digital techniques.

[1]We will not differentiate between RF-analog cancellation and baseband-analog cancellation since most of the existing analog cancellation schemes are based on RF techniques.

Since the near-end signal is significantly stronger than the far-end signal, the cancellation ratio needs to be sufficiently high to achieve satisfactory SINR for the desired signal. As a result, various aspects that limit the precision of self-interference cancellation need to be considered. In particular, the RF/analog impairments in wireless systems are important factors that require investigation, including quantization error, PA nonlinearity, I/Q imbalance, phase noise of local oscillator (LO), sampling jitter of ADC/DAC etc. In the following, we provide a brief introduction of the impact of RF/analog impairments on full-duplex wireless communication systems.

A. Impact of RF/Analog Impairments

Signals in wireless communication systems are distorted due to RF/analog impairments in different subsystems. When a complex baseband-equivalent signal $x(t)$ passes through a subsystem $\mathfrak{G}$, we write the resulting signal as $\mathfrak{G}[x(t)]$, where $\mathfrak{G}[\cdot]$ includes both the impairment effect and the linear characteristic of the subsystem. More specifically, we write

$$\overline{x}(t) = \mathfrak{G}\left[x(t)\right] = ax(t) + \mathfrak{F}\left[x(t)\right], \tag{1}$$

where $\overline{x}(t)$ denotes the output signal of the subsystem, $ax(t)$ denotes the ideally linear signal component in which a is a constant factor representing the signal gain, $\mathfrak{F}[x(t)]$ denotes the nonlinear spurious signal components caused by the RF/analog impairment. Detailed mathematical models for $\mathfrak{G}[\cdot]$ (and hence $\mathfrak{F}[\cdot]$) can be found in [23]–[25]. Note that the effects of $\mathfrak{G}[\cdot]$ depends on both $x(t)$ and the specific impairment. For example, the effect of PA nonlinearity is determined not only by the type of power amplifier used, but also by the relation between its operating point and the power of $x(t)$, i.e., the input/output backoff (IBO/OBO). It is also useful to define the normalized spurious signal as $\breve{x}(t) \triangleq \mathfrak{F}[x(t)]/a$ (i.e., $\overline{x}(t) = a[x(t) + \breve{x}(t)]$). Then the impact of the impairment on the signal quality can be characterized by 'SNR limitation' [24], which is defined as the ratio between the power of $x(t)$ and $\breve{x}(t)$, i.e., $\text{SNR}_{\text{lmt}} = P_x/\breve{P}_x$, where P_x and $\breve{P}_x$ denote the power of $x(t)$ and $\breve{x}(t)$, respectively. For ease of exposition, in this paper we will define γ as the inverse of SNR limitation, i.e.,

$$\gamma = \frac{1}{\text{SNR}_{\text{lmt}}} = \frac{\breve{P}_x}{P_x}, \tag{2}$$

which indicates the relative power of the spurious signal components to the ideally linear signal component. We note again that γ depends on both $x(t)$ and the specific impairment. In particular, for power amplifiers with fixed operating point, γ is dependant on P_x as it affects the IBO/OBO.

In this paper, we will consider the effects of quantization error, PA nonlinearity, I/Q imbalance, phase noise of local oscillator (LO), and sampling jitter of ADC/DAC, which are the most important RF/analog impairments in the system, as have been studied in [23], [24]. We first investigate the relative power (γ in (2)) of the spurious signal components caused by these impairments. An example of these values are listed in Table I. The result for PA nonlinearity is obtained from simulation based on a 802.11a/g system, where the PA nonlinearity is modeled by Rapp Model [24], [25] with the parameter $p = 3$ and an input back-off (IBO) of 8 dB. The other results are obtained by calculation according to [24]. The resolution of the ADC is 12 bits. The I/Q imbalance exhibits an amplitude mismatch of 0.5 dB and phase mismatch of 3 degrees. The power spectrum density (PSD) of the LO phase noise has a constant value of -110 dBc/Hz below 10 KHz, and decreases by 20 dB per decade beyond that [26]. Finally, the root mean square (rms) sampling jitter of the ADC/DAC is 2.5 ps (assuming 20 MHz sampling rate and identical oscillator as for the case of phase noise). Also note that the signal is assumed to be a single 10 MHz sinusoid wave for calculating the effects of quantization noise and sampling jitter.

TABLE I
RELATIVE POWER (γ) OF THE SPURIOUS SIGNAL COMPONENTS RESULTING FROM DIFFERENT RF/ANALOG IMPAIRMENTS

ADC	PA nonlinearity[(a)]	I/Q Imbalance	LO Phase Noise	Time Jitter of ADC/DAC
-74dB	-32dB	-28dB	-67dB	-73dB

[(a)] IBO=8dB

Since the received near-end signal is significantly stronger than the far-end signal in full-duplex wireless communication systems, it can be seen from Table I that the spurious components of the near-end signal also have significant impact on the far-end signal. For example, suppose the received near-end signal is 60 dB stronger than the far-end signal, the nonlinear signal components induced by PA nonlinearity would be 28 dB stronger than the far-end signal, making the detection of far-end signal impossible if not suppressed.

It is worth noting that there are also various other RF/analog impairments in wireless communication systems. However, the other impairments are much milder compared with those listed in Table I. For example, since power efficiency is not critical for the other amplifiers in the transceiver, they can be configured to have very low nonlinearity, which can be neglected compared with PA nonlinearity. In addition, by using one-stage down-converter and digital demodulation, the receiver can be treated as free of I/Q imbalance [27].[2]

B. Self-Interference Cancellation Using RF and Digital Techniques

As is described in Section I, RF cancellation techniques are immune to RF/analog impairments in the system, but care needs to be taken and it can be expensive to implement RF cancellation techniques that are effective for wideband systems. On the other hand, digital cancellation can be applied effectively in wideband systems, but is susceptible to RF/analog impairments (in fact, as will be shown in Section III-A, the cancellation ratio achieved by digital cancellation is limited by RF/analog impairments). Therefore, one can combine RF and digital cancellation techniques to improve the performance of wideband full-duplex wireless systems. Specifically, RF techniques can be used to cancel the line-of-sight (LOS) component of the

[2] Although similar techniques may be used at the transmitter side to mitigate transmit I/Q imbalance, we will not make such assumption in this paper because this approach is not adopted in [27].

self-interference, while digital baseband cancellation can be used to deal with the non-line-of-sight (NLOS) components.

For the effects of RF/analog impairments to not limit the system performance, it is desirable that the self-interference is suppressed by RF techniques to such a level that the spurious components of the self-interference is close to the noise floor. As an example, we consider a system with 10 dBm transmit power, 20 MHz bandwidth (corresponding to a noise floor of -101 dBm), and the same levels of spurious signal components as in Table I. It can be seen that about 40 dB attenuation is required for the effects of quantization noise, phase noise, and sampling jitter to reach the noise floor. However, to achieve the same objective for PA nonlinearity and transmit I/Q imbalance, the required attenuation is about 80 dB.

In this paper, unless otherwise stated, we will assume that the isolation between the transmitter and receiver (denoted as 'Tx-Rx isolation' hereafter) is 40 dB, i.e., the LOS component of near-end signal is attenuated by 40 dB using RF techniques. Note that this is a reasonable assumption for the following reasons:

1) The cancellation ratio achieved by RF techniques in [2], [4], [5], [7] is between 40 dB and 50 dB.
2) As is reported in [5], [7], 40 dB Tx-Rx isolation can be achieved when the antennas are separated by 15–20 cm.

In addition, since the NLOS components of the self-interference travel over distances that are much longer than 15–20 cm and also undergo reflections/diffractions, they are attenuated by more than 40 dB as well. Therefore the effects of quantization error, phase noise, and sampling jitter are low enough and are no longer the bottleneck of the system.

Nevertheless, as has been described in the previous example, the spurious signal components caused by PA nonlinearity and transmit I/Q imbalance are still much higher than the noise floor when the Tx-Rx isolation is 40 dB. Therefore they are the most prominent issues that limit the system performance. As such, we will introduce a digital self-interference canceller based on the output signal of the power amplifier in the following section. In addition, a two-stage iterative cancellation scheme is proposed to improve the output SINR of desired signal.

III. Baseband Self-Interference Cancellation for Full-Duplex Wireless Communication Systems

In this section, we present two digital cancellation techniques that can be applied in full-duplex wireless communication systems to improve the performance, i.e., self-interference cancellation based on the output signal of power amplifier and two-stage iterative cancellation.

We first consider a straightforward mathematical model of full-duplex systems in which the receeived signal can be written a

$$r(n) = x(n) \otimes h(l) + y(n) \otimes g(l) + z(n), \tag{3}$$

where $x(n)$ and $y(n)$ are the sample representations of the near-end and far-end transmit signal, respectively; $h(l)$ and $g(l)(l = 1, \ldots, L)$ denote the impulse response of near-end and far-end channel, respectively; L is the order of the channels; $\otimes$ denotes the operation of convolution; $z(n)$ is the received noise. For ease of exposition, we assume that the signal samples and channel responses are i.i.d.,[3] i.e., $\mathbb{E}[x(n)x^*(m)] = \mathbb{E}[y(n)y^*(m)] = \delta(n-m)$, $\mathbb{E}[h(l)h^*(k)] = \chi_{\text{near}}\delta(l-k)$, $\mathbb{E}[g(l)g^*(k)] = \chi_{\text{far}}\delta(l-k)$, $\mathbb{E}[z(n)z^*(m)] = \sigma_{\text{z}}^2\delta(n-m)$, where $\mathbb{E}[\cdot]$ denotes expectation, $\delta(\cdot)$ is the Kronecker delta function, χ_{near} and χ_{far} represent the gain of the near-end and far-end channel signal pathes, respectively, σ_{z}^2 is the power of received noise. Therefore the power of the received self-interference and desired signal are $P_{\text{near}} = \chi_{\text{near}}L$ and $P_{\text{far}} = \chi_{\text{far}}L$, respectively.

Considering a frame of N_f samples, (3) can be rewritten in vector form as

$$\mathbf{r} = \mathbf{X}\mathbf{h} + \mathbf{Y}\mathbf{g} + \mathbf{z}, \tag{4}$$

where $\mathbf{r} = [r(1); \ldots; r(N_f)]$ is the vertical concatenation of $r(n)$, $\mathbf{z} = [z(1); \ldots; z(N_f)]$, $\mathbf{h} = [h(0); \ldots; h(L-1)]$, $\mathbf{g} = [g(0); \ldots; g(L-1)]$. $\mathbf{X} = [\mathbf{x}_0, \ldots, \mathbf{x}_{L-1}]$, where $\mathbf{x}_l = [x(1-l); \ldots; x(N_f - l)]$. $\mathbf{Y}$ is defined similarly as $\mathbf{X}$ from $y(n)$.

During digital cancellation, the near-end channel $h(l)$ is estimated and used for generating estimates of the self-interference, which are then subtracted from the received signal. One approach of channel estimation is the least square method, which estimates the near-end channel according to

$$\widehat{\mathbf{h}} = \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{r}. \tag{5}$$

The self-interference cancellation is then carried out using

$$\mathbf{r}_c = \mathbf{r} - \mathbf{X}\widehat{\mathbf{h}}. \tag{6}$$

By substituting (4) into (5), we get

$$\begin{aligned}\widehat{\mathbf{h}} &= \mathbf{h} + \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H(\mathbf{Y}\mathbf{g} + \mathbf{z}) \\ &= \mathbf{h} + \Delta\mathbf{h},\end{aligned} \tag{7}$$

where $\Delta\mathbf{h} \triangleq \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H(\mathbf{Y}\mathbf{g} + \mathbf{z})$ is the channel estimation error. The residual self-interference after cancellation is then given by $\mathbf{r}_{\text{near},\Delta} = \mathbf{X}\Delta\mathbf{h}$, whose power can be shown to be (see Appendix A)

$$P_\Delta = \frac{L}{N_f}\left(P_{\text{far}} + \sigma_{\text{z}}^2\right). \tag{8}$$

Comparing P_Δ with the power of received self-interference, it is shown that the self-interference is suppressed by a factor of

$$\eta = \frac{P_{\text{near}}}{P_\Delta} = \frac{N_f}{L\left(P_{\text{far}} + \sigma_{\text{z}}^2\right)}P_{\text{near}}. \tag{9}$$

Furthermore, we can write the output SINR of the desired signal as

$$\begin{aligned}\text{SINR} &= \frac{P_{\text{far}}}{P_\Delta + \sigma_{\text{z}}^2} \\ &= \frac{P_{\text{far}}}{\frac{L}{N_f}\left(P_{\text{far}} + \sigma_{\text{z}}^2\right) + \sigma_{\text{z}}^2} < \frac{N_f}{L}.\end{aligned} \tag{10}$$

[3]Note that these assumptions do not compromise the insights of the analysis that follows. The same conclusions can be reached if more complex models are used.

It can be seen from (9) and (10) that for an ideal full-duplex system:

1) The suppression of self-interference increases with the power of received self-interference (9).
2) The output SINR of the desired signal is limited by a constant determined by the channel order L and frame length N_f (10).

The reason is that the estimation precision of $\mathbf{h}$ increases with P_{near}, but is affected by the existence of the desired signal (cf. (7)). Note that the second property relies on the validity of the first property. Suppose the suppression of self-interference is fixed, when P_{near} increases, the output SINR will decrease due to increasing P_{Δ}. Unfortunately, this situation will occur when RF/analog impairments are present, as will be shown in the following subsection.

A. Self-Interference Cancellation Based on the Output Signal of Power Amplifier

In this subsection, we show that the cancellation of self-interference is limited by RF/analog impairments, and propose a digital cancellation scheme based on the output signal of power amplifier to address this issue.

As is described in Section II-B, the effects of quantization error, phase noise, and sampling jitter can be made low enough given reasonable Tx-Rx isolation, while the spurious signals caused by PA nonlinearity and transmit I/Q imbalance are much stronger. Therefore we use the effect of PA nonlinearity and transmit I/Q imbalance as an example to demonstrate the impact of RF/analog impairments on digital cancellation. Note that the effect of other impairments (e.g., quantization error) are similar, but are omitted here for brevity (however, they are included in Sections IV and V).

We use $\overline{x}(n) = x(n) + \breve{x}(n)$ and $\overline{y}(n) = y(n) + \breve{y}(n)$ to denote the near-end and far-end transmit signal, respectively, where $\breve{x}(n)$ and $\breve{y}(n)$ represent the spurious signal components caused by PA nonlinearity and/or I/Q imbalance (cf. (1), note that a is assumed to be unity for simplicity). Then (3) becomes

$$\begin{aligned} r(n) &= \overline{x}(n) \otimes h(l) + \overline{y}(n) \otimes g(l) + z(n) \\ &= (x(n) + \breve{x}(n)) \otimes h(l) \\ &\quad + (y(n) + \breve{y}(n)) \otimes g(l) + z(n). \end{aligned} \tag{11}$$

Since the term $\breve{x}(n) \otimes h(l)$ in (11) cannot be removed by (6), the cancellation of self-interference in (9) becomes

$$\overline{\eta} \approx \frac{P_{\text{near}}}{\breve{P}_{\text{near}} + \frac{L}{N_f}\left(P_{\text{far}} + \sigma_z^2\right)}, \tag{12}$$

where $\breve{P}_{\text{near}}$ denotes the power of the term $\breve{x}(n) \otimes h(l)$. Following (2), we can write the power of $\breve{x}(t)$ as γ (note that the power of $x(n)$ is assumed to be unity in the above), then we have $\breve{P}_{\text{near}} \approx \gamma \chi_{\text{near}} L = \gamma P_{\text{near}}$. Therefore it can be seen from (12) that $\overline{\eta} < P_{\text{near}}/\breve{P}_{\text{near}} = 1/\gamma$, i.e., the cancellation ratio achieved by digital cancellation is limited, as opposed to the ideal case of (9) where the cancellation ratio increases with the received power of self-interference. In addition, since $N_f \gg L$, the denominator of (12) is usually dominated by $\breve{P}_{\text{near}}$.

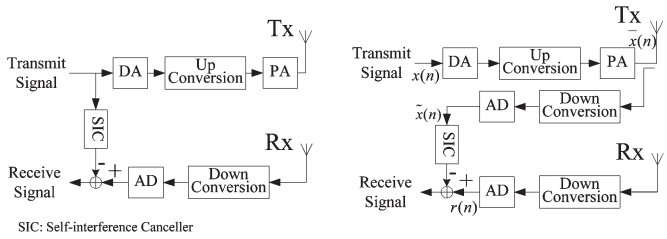

Fig. 2. Digital cancellation. Left: conventional method such as those used in [2], [3], [5], [6]. Right: proposed approach using the output signal of power amplifier.

Therefore we have $\overline{\eta} \approx 1/\gamma$, i.e., the cancellation ratio is a constant value determined by PA nonlinearity and transmit I/Q imbalance. This result is consistent with [5, Fig. 10]. Therefore we conjecture the constant cancellation ratio of digital cancellation in [5] is mainly caused by RF/analog impairments such as PA nonlinearity and transmit I/Q imbalance.

To improve the cancellation ratio, we propose to carry out digital cancellation using the output signal of the power amplifier,[4] which is obtained by attaching a coupler to the transmit antenna and use another RF front-end to convert the signal to digital samples,[5] as shown in Fig. 2. Denoting the resulting samples as $\widetilde{x}(n)$, the procedure of channel estimation and cancellation as in (5) and (6) can be denoted as $\widehat{\mathbf{h}} = \left(\widetilde{\mathbf{X}}^H \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^H \mathbf{r}$ and $\mathbf{r}_c = \mathbf{r} - \widetilde{\mathbf{X}}\widehat{\mathbf{h}}$, respectively, where $\widetilde{\mathbf{X}}$ is constructed from $\widetilde{x}(n)$ following the same structure as $\mathbf{X}$ in (4). This way, the spurious components of the self-interference caused by PA nonlinearity and transmit I/Q imbalance can also be suppressed, achieving higher cancellation ratio of the self-interference.

B. Two-Stage Decision Feedback Iterative Cancellation

As has been shown previously, due to the fact that the estimation precision of the near-end channel is affected by the desired signal, the output SINR of full-duplex systems is limited. According to (10), to achieve higher SINR, one need to increase the frame length N_f, i.e., to use more data samples for channel estimation. However, since the channel impulse response $h(l)$ and $g(l)$ are treated as static within each frame in (3)–(6), increasing N_f will decrease the cancellation ratio due to channel variation.

To address this issue, we propose a two-stage iterative cancellation scheme, where the far-end signal is removed from the received signal for more accurate channel estimation and cancellation [31]. In the first stage, the self-interference cancellation is carried out according to (5) and (6). In the second stage cancellation, the far-end signal is detected based on the result of (6) and used for generating an estimate of the far-end transmit signal $\widehat{y}(n)$. An estimate of the received far-end signal is subsequently generated and subtracted from the received signal as follows

$$\mathbf{r}' = \mathbf{r} - \widehat{\mathbf{Y}}\widehat{\mathbf{g}}, \tag{13}$$

[4]An alternative approach is nonlinear echo cancellation. For example, Volterra filter [28] can be employed to model the nonlinear signal path. We choose the proposed approach due to the high complexity in identifying the parameters of nonlinear models.

[5]Same approach has been used in [29], [30].

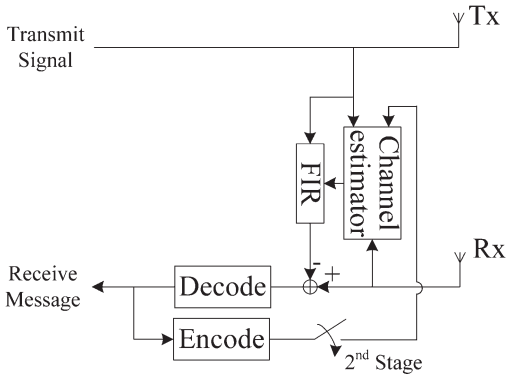

Fig. 3. Structure of the proposed two-stage iterative cancellation scheme.

where $\widehat{\mathbf{Y}}$ is constructed from $\widehat{y}(n)$ following the same structure as $\mathbf{Y}$ in (4), $\widehat{\mathbf{g}}$ is an estimate of the far-end channel. $\mathbf{r}'$ is then used together with $\mathbf{X}$ to estimate the near-end channel again according to (5). By taking the far-end signal into account, the near-end channel can be estimated more precisely, thus improving the accuracy of cancellation. After the cancellation, the far-end signal is detected again and used for a new iteration of near-end channel estimation and self-interference cancellation. This procedure is repeated until a satisfactory cancellation ratio and output SINR is achieved.

Note that in addition to the pre-known training symbols, we can also use $\widehat{y}(n)$ to improve the accuracy in estimating the far-end channel $\mathbf{g}$, i.e., decision-feedback channel estimation [32]. Therefore, we can combine the procedure of far-end channel estimation, far-end signal removing, and near-end channel estimation into one step as follows

$$\begin{bmatrix}\widehat{\mathbf{h}}\\ \widehat{\mathbf{g}}\end{bmatrix} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{r}, \tag{14}$$

where $\widehat{\mathbf{h}}$ and $\widehat{\mathbf{g}}$ are jointly estimated using $\mathbf{A} \triangleq [\mathbf{X}\ \widehat{\mathbf{Y}}]$.

The proposed two-stage iterative cancellation scheme is depicted graphically in Fig. 3. Note that the fundamental principle of this scheme is identical to that of multi-stage Parallel Interference Cancellation (PIC) in Multiuser Detection (MUD) [33]–[35], with each iteration in the second stage cancellation corresponding to a stage of PIC. In particular, one can show that the procedure of the proposed scheme is the same as that carried out for a two-user system in [34], [35], with the exception that in our case the near-end transmit signal is known. Therefore the performance of the two-stage iterative canceller must be as good or better than that of a two-user system in [34], [35]. In addition, by considering the near-end transmitter and far-end transmitter as two transmit antennas in a Multiple Input Multiple Output (MIMO) system, the system model of (4) is identical to that of a wide band 2×1 MIMO system. Therefore the proposed iterative cancellation scheme is the same as the iterative channel estimation and data detection algorithm of [36], though the soft decisions of data symbols are used for channel estimation in [36] while we use hard decisions here. We note that the decoding errors of the desired signal will affect the precision of near-end channel estimation in the next iteration, a phenomenon known as 'error propagation' in multiuser detection. To improve the resilience to error propagation, one can employ the technique of soft decision as in [32], [36]. However, as is shown in [32], [37], the system performance can be improved effectively using hard decision so long as the initial detection error is moderate. In Section V, it will be shown by simulation that the system performance in terms of SINR and BER is substantially improved by introducing the second stage cancellation.

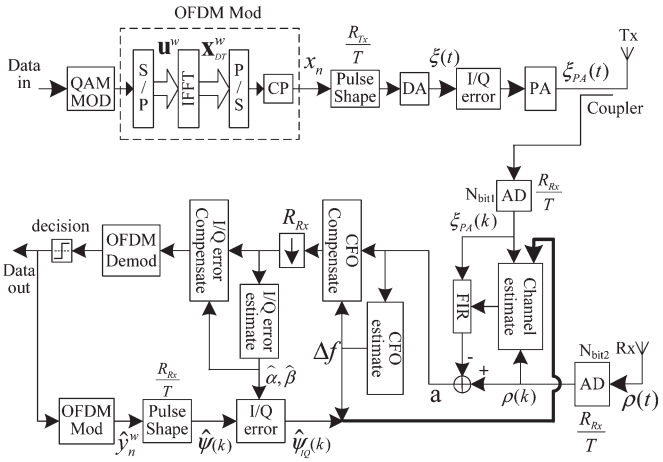

Fig. 4. Structure of the proposed transceiver for full-duplex wireless communication systems. RF cancellation is assumed to have been applied to cancel the LOS self-interference, but not shown in the figure.

C. Overall Structure of Proposed Full-Duplex Wireless Transceiver

Combining the two digital cancellation approaches described in Section III-A and B, we propose a transceiver structure for full-duplex wireless communication systems as shown in Fig. 4. We have considered a system based on 802.11a/g, an existing wireless standard where full-duplex communication could be applied because of its low transmitting power and low channel variation rate. The transceiver is depicted in baseband equivalent form, with the process of up/down-conversion omitted for simplicity of illustration. Note that RF cancellation has been applied in the system to suppress the LOS component of self-interference, but is not shown in Fig. 4 because one may also rely on antenna separation for the same purpose when the device is not constrained by size, as described in Section II-B.

As is shown in Fig. 4, the OFDM symbols are passed through a wave shaping filter and DAC, resulting in analog baseband signal $\xi(t)$, which is subsequently up-converted and amplified into $\xi_{\mathrm{PA}}(t)$ for transmission. The block labeled "I/Q error" represents the effect of I/Q imbalance introduced during up-conversion; while "PA" represents the nonlinear effect of the power amplifier. $\xi_{\mathrm{PA}}(t)$ is converted into digital samples for estimating the near-end channel and canceling the self-interference, and the detection of the far-end signal is carried out subsequently, after the carrier frequency offset (CFO) and transmitter I/Q imbalance of the far-end signal is estimated and compensated.

Since wide-band frequency selective channels are converted into multiple narrow-band sub-channels in OFDM systems.

One may also consider carrying out self-interference cancellation on each sub-channel, i.e., in the frequency domain. However, while the ideal linear components of the self-interference can be cancelled, the nonlinear components arising from PA nonlinearity and I/Q imbalance cannot be handled on a sub-channel basis. The reason is that the orthogonality between sub-channels provided by the feature of cyclic prefix etc. in ideal OFDM systems does not apply to the nonlinear signal components. Therefore the performance of frequency domain cancellation will be limited by PA nonlinearity and transmit I/Q imbalance as in Section III-A.

We denote the received signal as $\rho(t)$ (cf. Fig. 4) and the far-end transmit signal as $\psi_{\rm PA}(t)$. Using $\xi_{\rm PA}(k)$, $\psi_{\rm PA}(k)$, and $\rho(k)$ to denote the baseband equivalent samples of $\xi_{\rm PA}(t)$, $\psi_{\rm PA}(t)$, and $\rho(t)$, respectively, we have

$$\rho(k) \approx \sum_{l=0}^{\Lambda-1} \eta_l \xi_{\rm PA}(k-l) + e^{j(\omega_0+\omega k)} \sum_{l=0}^{\Lambda-1} \varsigma_l \psi_{\rm PA}(k-l) + \zeta(k), \tag{15}$$

which is similar to the model of (3) or (11), with η_l and ς_l representing the impulse response of the near-end and far-end channel respectively, Λ is the order of the channels and $\zeta(k)$ represents the received noise. Note that the term $e^{j(\omega_0+\omega k)}$ is included to model the carrier frequency offset (CFO) between the two communicating devices, where $\omega = 2\pi(\Delta f/f_s)$ with Δf and f_s represent the frequency offset and the sampling rate of $\rho(k)$ respectively, and ω_0 denotes the initial phase difference. In addition, since the signals in real systems are not strictly band-limited, the sampling rate f_s is chosen as $f_s = R_{\rm Rx}/T$ to ensure the accuracy of signal representation and cancellation, where T is the sampling period of OFDM symbols (i.e., system bandwidth), $R_{\rm Rx}$ is the oversampling ratio. We write (15) in vector form for a frame of N_f samples as follows

$$\boldsymbol{\rho} \approx \boldsymbol{\Xi}\boldsymbol{\eta} + e^{j\omega_0}\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\varsigma} + \boldsymbol{\zeta}, \tag{16}$$

where $\boldsymbol{\Phi} = diag([1, e^{j\omega}, \ldots, e^{j(N_f-1)\omega}])$, while $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$, $\boldsymbol{\eta}$, $\boldsymbol{\Psi}$, $\boldsymbol{\varsigma}$, and $\boldsymbol{\zeta}$ are constructed similar to the corresponding matrices/vectors in (4). The two-stage iterative cancellation scheme described in Section III-B can then be applied to (16), with $\boldsymbol{\Xi}$, $\boldsymbol{\Phi}\boldsymbol{\Psi}$, $\boldsymbol{\eta}$, and $e^{j\omega_0}\boldsymbol{\varsigma}$ corresponding to $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{h}$, and $\mathbf{g}$ of (4), respectively. In particular, the detection result of the far-end signal and the estimation of CFO are used for constructing $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Phi}}$ (i.e., estimates of $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$), respectively. In addition, the effect of transmit I/Q imbalance is also taken into consideration to improve the accuracy in generating estimates of the far-end signal. As is shown in Fig. 4, the detection result of the far-end signal is modulated again and passed through a pulse shaping filter to get $\widehat{\psi}(k)$. $\widehat{\boldsymbol{\Psi}}$ is then constructed from $\widehat{\psi}_{\rm IQ}(k) = \widehat{\alpha}\widehat{\psi}(k) + \widehat{\beta}\widehat{\psi}^*(k)$, where $\widehat{\alpha}$ and $\widehat{\beta}$ are the parameters characterizing the effect of I/Q imbalance [24].

IV. Performance Analysis of the Proposed Transceiver

This section investigates the analytical performance of the proposed full-duplex transceiver of Fig. 4 in terms of average signal to interference and noise ratio (SINR), which is defined as SINR $= \mathbb{E}[P_{\rm far}(t)]/(\sigma_{\rm z}^2 + \mathbb{E}[P_{\rm I}(t)])$, where $\mathbb{E}[\cdot]$ means taking the expectation over time (i.e., over different channel realizations). $P_{\rm far}(t)$ and $P_{\rm I}(t)$ denote the power of the received far-end signal and that of the residual near-end interference at time instant t, respectively. $\sigma_{\rm z}^2$ represents the noise power.

To simplify the analysis, we consider a model similar to (11), where $x(n)$ and $y(n)$ are OFDM symbols. We will denote $\gamma_{\rm PA}$, $\gamma_{\rm IQ}$, $\gamma_{\rm PN}$, and $\gamma_{\rm QN}$ as the relative power of the spurious signal components caused by PA nonlinearity, I/Q imbalance, phase noise, and quantization, respectively (cf. Table I). The analysis steps we follow are similar to those in (3)–(10). However, the analysis in this section is more involved due to the incorporation of signal distortions and the fact that the transmit signals $x(n)$ and $y(n)$ are not i.i.d. as assumed in Section III.

To facilitate the analysis, we use $\mathbf{u}^w = [u_1^w; \ldots; u_N^w]$ to denote the vector of near-end transmit symbols in the frequency-domain for the w-th OFDM symbol, where N is the length of the Inverse Fast Fourier Transform (IFFT), i.e., the total number of subcarriers. u_n^w is non-zero only if $n \in S$, which denotes the set of subcarriers in use. The time-domain samples $\mathbf{x}_{\rm DT}^w = [x_1^w; \ldots; x_N^w]$ are obtained by applying IFFT to $\mathbf{u}^w$, i.e., $\mathbf{x}_{\rm DT}^w = \mathbf{F}^H\mathbf{u}^w$, where the (m,n)-th entry of $\mathbf{F}$ is

$$[\mathbf{F}]_{(m,n)} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}(m-1)(n-1)} \quad m,n = 1,\ldots,N \tag{17}$$

Then the w-th OFDM symbol is given by $\mathbf{x}^w = [\mathbf{x}_{\rm CP}^w; \mathbf{x}_{\rm DT}^w]$, where $\mathbf{x}_{\rm CP}^w = [x_{N-M+1}^w; \ldots; x_N^w]$ is the last M samples of $\mathbf{x}_{\rm DT}^w$, i.e., the Cyclic Prefix (CP). Similarly, we can define the w-th OFDM symbol of the far-end signal as $\mathbf{y}^w = [\mathbf{y}_{\rm CP}^w; \mathbf{y}_{\rm DT}^w]$, with the corresponding frequency-domain symbols $\mathbf{v}^w = [v_1^w; \ldots; v_N^w]$. We then write the received signal as

$$\mathbf{r} = \overline{\mathbf{X}}\mathbf{h} + \overline{\mathbf{Y}}\mathbf{g} + \mathbf{z} + \mathbf{z}_{\rm q,Rx} + \mathbf{z}_{\rm PN}, \tag{18}$$

where the dimension of $\mathbf{r}$ is $W(M+N)\times 1$ since there are W OFDM symbols in each frame, $\mathbf{z}_{\rm q,Rx}$ and $\mathbf{z}_{\rm PN}$ denote the quantization noise and phase noise at the receiver side, respectively. $\overline{\mathbf{X}} = \mathbf{X} + \mathbf{X}_{\rm PA} + \mathbf{X}_{\rm IQ}$, $\overline{\mathbf{Y}} = \mathbf{Y} + \mathbf{Y}_{\rm PA} + \mathbf{Y}_{\rm IQ}$. $\mathbf{X}$ and $\mathbf{Y}$ are constructed from the $\mathbf{x}^w$ and $\mathbf{y}^w$ respectively. $\mathbf{X}_{\rm PA}$ and $\mathbf{Y}_{\rm PA}$ represent the nonlinear signal components resulting from PA nonlinearity, while $\mathbf{X}_{\rm IQ}$ and $\mathbf{Y}_{\rm IQ}$ represent the signal components resulting from Tx I/Q imbalance.

Note that different RF/analog impairments are coupled together in practical systems, and it is not rigorous to model the joint effects of multiple impairments by independent additive error terms (spurious signals). However, since the spurious signal caused by each impairment is much weaker than the ideally linear signal (cf. Table I), we can use (18) as an approximation. For example, when two RF/analog impairments are applied on a signal $x(t)$ sequentially, the resulting signal can be written as (cf. (1), note that a is assumed to be unity for simplicity)

$$\begin{aligned}\overline{x}(t) &= \mathfrak{G}_2\left\{\mathfrak{G}_1\left[x(t)\right]\right\} \\ &= \mathfrak{G}_2\left\{x(t) + \mathfrak{F}_1\left[x(t)\right]\right\} \\ &= x(t) + \mathfrak{F}_1\left[x(t)\right] + \mathfrak{F}_2\left\{x(t) + \mathfrak{F}_1\left[x(t)\right]\right\}\end{aligned} \tag{19}$$

Since $x(t)$ is much stronger than $\mathfrak{F}_1[x(t)]$, we have $\mathfrak{F}_2\{x(t)+\mathfrak{F}_1[x(t)]\}\approx\mathfrak{F}_2\{x(t)\}$, therefore $\overline{x}(t)\approx x(t)+\mathfrak{F}_1[x(t)]+\mathfrak{F}_2[x(t)]$, and the same approximation can be made for the effect of multiple impairments that are coupled together.

Because of the special structure of $\mathbf{x}^w$ and $\mathbf{y}^w$, we can write $\mathbf{X}$ as $\mathbf{X}=[\mathbf{X}^1;\ldots;\mathbf{X}^W]$, where $\mathbf{X}^w=[\mathbf{X}^w_{\rm CP};\mathbf{X}^w_{\rm DT}]$ with $\mathbf{X}^w_{\rm CP}$ and $\mathbf{X}^w_{\rm DT}$ defined as follows[6]

$$\mathbf{X}^w_{\rm CP}=\begin{bmatrix} x^w_{N-M+1} & x^{w-1}_N & \cdots & x^{w-1}_{N-L+2}\\ x^w_{N-M+2} & x^w_{N-M+1} & \cdots & x^{w-1}_{N-L+3}\\ \cdots & \cdots & & \cdots\\ x^w_N & x^w_{N-1} & \cdots & x^w_{N-L+1}\end{bmatrix}$$

$$\mathbf{X}^w_{\rm DT}=\begin{bmatrix} x^w_1 & x^w_N & \cdots & x^w_{N-L+2}\\ x^w_2 & x^w_1 & \cdots & x^w_{N-L+3}\\ \cdots & \cdots & & \cdots\\ x^w_N & x^w_{N-1} & \cdots & x^w_{N-L+1}\end{bmatrix}$$

Similarly, we can write $\mathbf{Y}$ as $\mathbf{Y}=[\mathbf{Y}^1;\ldots;\mathbf{Y}^W]$. $\mathbf{Y}^w$ may be written as $\mathbf{Y}^w=[[\mathbf{Y}^{w-1}_{\rm DT}]_{(N-q+1:N,1:L)};\mathbf{Y}^w_{\rm CP};[\mathbf{Y}^w_{\rm DT}]_{(1:N-q,1:L)}]$ or $\mathbf{Y}^w=[[\mathbf{Y}^{w-1}_{\rm CP}]_{(N-q+1:N,1:L)};\mathbf{Y}^w_{\rm DT};[\mathbf{Y}^w_{\rm CP}]_{(1:N-q,1:L)}]$, where $\mathbf{Y}^w_{\rm CP}$ and $\mathbf{Y}^w_{\rm DT}$ are defined similar to $\mathbf{X}^w_{\rm CP}$ and $\mathbf{X}^w_{\rm DT}$ respectively, and q is an integer that depends on the time offset between the near-end signal and the far-end signal. In addition, the signal obtained from the power amplifier is written in matrix form as

$$\begin{aligned}\widetilde{\mathbf{X}}&=\overline{\mathbf{X}}+\overline{\mathbf{X}}_{\rm q}+\overline{\mathbf{X}}_{\rm PN}\\&=\mathbf{X}+\mathbf{X}_{\rm PA}+\mathbf{X}_{\rm IQ}+\overline{\mathbf{X}}_{\rm q}+\overline{\mathbf{X}}_{\rm PN}.\end{aligned}\tag{20}$$

where $\overline{\mathbf{X}}_{\rm PN}$ and $\overline{\mathbf{X}}_{\rm q}$ denote the signal distortions caused by phase noise and quantization error.

A. SINR of the First Stage of Cancellation

In the first stage of cancellation, the near-end channel is estimated by

$$\widehat{\mathbf{h}}_{\rm s1}=(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H\mathbf{r}=\mathbf{h}+\Delta\mathbf{h}_{\rm s1},\tag{21}$$

where the channel estimation error is

$$\Delta\mathbf{h}_{\rm s1}=(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H(-\overline{\mathbf{X}}_{\rm q}\mathbf{h}-\overline{\mathbf{X}}_{\rm PN}\mathbf{h}+\overline{\mathbf{Y}}\mathbf{g}+\mathbf{z}+\mathbf{z}_{\rm q,Rx}+\mathbf{z}_{\rm PN}).\tag{22}$$

Therefore the received signal after self-interference cancellation is

$$\begin{aligned}\mathbf{s}_{\rm s1}&=\mathbf{r}-\widetilde{\mathbf{X}}\widehat{\mathbf{h}}_{\rm s1}\\&=\overline{\mathbf{Y}}\mathbf{g}-\widetilde{\mathbf{X}}\Delta\mathbf{h}_{\rm s1}-\overline{\mathbf{X}}_{\rm q}\mathbf{h}-\overline{\mathbf{X}}_{\rm PN}\mathbf{h}+\mathbf{z}+\mathbf{z}_{\rm q,Rx}+\mathbf{z}_{\rm PN},\end{aligned}\tag{23}$$

where $\widetilde{\mathbf{X}}\Delta\mathbf{h}_{\rm s1}$ represents the residual near-end signal. It can be noticed that the achievable SINR of far-end signal is limited by the last 6 terms of (23). In the following we will calculate the power of these terms to obtain the SINR value.

[6]We assume that when the received samples are divided into frames, the boundaries are chosen to be at the boundaries of transmit OFDM symbols.

We first look at the term $\widetilde{\mathbf{X}}\Delta\mathbf{h}_{\rm s1}$. Since in (22) the power of the received far-end signal ($\overline{\mathbf{Y}}\mathbf{g}$) is stronger than the other terms ($\overline{\mathbf{X}}_{\rm q}\mathbf{h}$, $\overline{\mathbf{X}}_{\rm PN}\mathbf{h}$, $\mathbf{z}$, $\mathbf{z}_{\rm q,Rx}$, and $\mathbf{z}_{\rm PN}$), as explained in Section II-B, $\Delta\mathbf{h}_{\rm s1}$ can be written approximately as $\Delta\mathbf{h}_{\rm s1}\approx(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H\overline{\mathbf{Y}}\mathbf{g}$. By following similar steps as (38)–(40), we can get the power of residual near-end signal as follows

$$P_{\Delta,\rm s1}=\frac{\mathrm{Tr}\left\{\mathbb{E}\left[(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H\overline{\mathbf{Y}}\mathbf{g}\mathbf{g}^H\overline{\mathbf{Y}}^H\widetilde{\mathbf{X}}\right]\right\}}{(N+M)W},\tag{24}$$

where the expectation can be firstly taken with respect to channel realizations $\mathbf{g}$ and then with respect to the transmit samples ($\mathbf{X}$ and $\mathbf{Y}$), which is valid if $\mathbf{g}$, $\mathbf{X}$, and $\mathbf{Y}$are mutually independent. As in Section III, we assume that the elements of $\mathbf{g}$ are i.i.d., i.e., $\mathbb{E}[\mathbf{g}\mathbf{g}^H]=\chi_{\rm far}\mathbf{I}$, then (24) changes into

$$P_{\Delta,\rm s1}=\frac{\chi_{\rm far}\mathrm{Tr}\left\{\mathbb{E}\left[(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H\overline{\mathbf{Y}\mathbf{Y}}^H\widetilde{\mathbf{X}}\right]\right\}}{(N+M)W}.\tag{25}$$

The difficulty of obtaining $P_{\Delta,\rm s1}$ then lies in the calculation of $\mathbb{E}[(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H\overline{\mathbf{Y}\mathbf{Y}}^H\widetilde{\mathbf{X}}]$ in (25), and we will provide an approximate result as follows.

Since $\mathbf{X}=[\mathbf{X}^1;\ldots;\mathbf{X}^W]$ as previously defined and $\widetilde{\mathbf{X}}$ features the same structure, we have $\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}}=\sum_{w=1}^W(\widetilde{\mathbf{X}}^w)^H\widetilde{\mathbf{X}}^w$. Since $\lim_{W\to\infty}(1/W)\sum_{w=1}^W(\widetilde{\mathbf{X}}^w)^H\widetilde{\mathbf{X}}^w=\mathbb{E}[(\widetilde{\mathbf{X}}^w)^H\widetilde{\mathbf{X}}^w]$, we can approximate $\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}}$ by $W\mathbb{E}[(\widetilde{\mathbf{X}}^w)^H\widetilde{\mathbf{X}}^w]$ for large W. Using a similar approach, $\widetilde{\mathbf{X}}^H\overline{\mathbf{Y}\mathbf{Y}}^H\widetilde{\mathbf{X}}$ can be written as

$$\widetilde{\mathbf{X}}^H\overline{\mathbf{Y}\mathbf{Y}}^H\widetilde{\mathbf{X}}=\left[\sum_{w=1}^W(\widetilde{\mathbf{X}}^w)^H\overline{\mathbf{Y}}^w\right]\left[\sum_{w=1}^W(\overline{\mathbf{Y}}^w)^H\widetilde{\mathbf{X}}^w\right]\tag{26}$$

$$=\sum_{w=1}^W(\widetilde{\mathbf{X}}^w)^H\overline{\mathbf{Y}}^w(\overline{\mathbf{Y}}^w)^H\widetilde{\mathbf{X}}^w+\sum_{\substack{w_1=1,w_2=1\\w_1\neq w_2}}^W(\widetilde{\mathbf{X}}^{w_1})^H\overline{\mathbf{Y}}^{w_1}(\overline{\mathbf{Y}}^{w_2})^H\widetilde{\mathbf{X}}^{w_2}\tag{27}$$

$$\overset{(a)}{\approx}W\times\mathbb{E}\left[(\widetilde{\mathbf{X}}^w)^H\overline{\mathbf{Y}}^w(\overline{\mathbf{Y}}^w)^H\widetilde{\mathbf{X}}^w\right]\tag{28}$$

$$\overset{(b)}{=}W\times\mathbb{E}\left\{(\widetilde{\mathbf{X}}^w)^H\mathbb{E}\left[\overline{\mathbf{Y}}^w(\overline{\mathbf{Y}}^w)^H\right]\widetilde{\mathbf{X}}^w\right\}\tag{29}$$

where (a) follows from the fact that $\widetilde{\mathbf{X}}^{w_1}$, $\widetilde{\mathbf{X}}^{w_2}$, $\overline{\mathbf{Y}}^{w_1}$, and $\overline{\mathbf{Y}}^{w_2}$ are independent with each other for $w_1\neq w_2$, thus the second term of (27) can be approximated by zero for large W; (b) holds if the expectation of (28) is taken firstly w.r.t. $\overline{\mathbf{Y}}^w$ and then w.r.t. $\widetilde{\mathbf{X}}^{w_1}$.

Therefore, (25) can be rewritten as

$$P_{\Delta,\rm s1}\approx\frac{\chi_{\rm far}}{(N+M)W}\times\mathrm{Tr}\left(\widetilde{\mathbf{R}}^{-1}_{\mathbf{X}^H\mathbf{X}}\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{Y}\mathbf{Y}^H\mathbf{X}}\right),\tag{30}$$

where $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{X}}\triangleq\mathbb{E}[(\widetilde{\mathbf{X}}^w)^H\widetilde{\mathbf{X}}^w]$, and $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{Y}\mathbf{Y}^H\mathbf{X}}\triangleq\mathbb{E}[(\widetilde{\mathbf{X}}^w)^H\overline{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}^H}\widetilde{\mathbf{X}}^w]$, $\overline{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}^H}\triangleq\mathbb{E}[(\overline{\mathbf{Y}}^w)^H\overline{\mathbf{Y}}^w]$.

According to (20), the signal distortions caused by RF/analog impairments need to be considered in calculating $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{X}}$. Since

$\gamma_{\rm QN}$ and $\gamma_{\rm PN}$ can be ignored compared with $\gamma_{\rm PA}$ and $\gamma_{\rm IQ}$ (cf. Table I), we can write $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{X}}$ as

$$\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{X}} \approx \mathbb{E}\left[(\mathbf{X}^w)^H\mathbf{X}^w\right] + \gamma_{\rm IQ}\mathbb{E}\left[(\mathbf{X}^w)^H\mathbf{X}^w\right]^* + \gamma_{\rm PA}\mathbf{I}, \tag{31}$$

where the term $\gamma_{\rm IQ}\mathbb{E}[(\mathbf{X}^w)^H\mathbf{X}^w]^*$ is due to the fact that the effect of I/Q imbalance on a signal $x(t)$ is characterized by $x_{\rm IQ}(t) = \alpha x(t) + \beta x^*(t)$, while the term $\gamma_{\rm PA}\mathbf{I}$ is due to the assumption that the samples of the nonlinear signal components resulting from PA nonlinearity are i.i.d. Similarly, $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{Y}\mathbf{Y}^H\mathbf{X}}$ can be written as

$$\begin{aligned}\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{Y}\mathbf{Y}^H\mathbf{X}} \approx{}& \mathbb{E}\left[(\mathbf{X}^w)^H\overline{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}^H}\mathbf{X}^w\right] \\ &+ \gamma_{\rm IQ}\mathbb{E}\left[(\mathbf{X}^w)^T\overline{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}^H}(\mathbf{X}^w)^*\right] + \gamma_{\rm PA}\overline{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}^H},\end{aligned} \tag{32}$$

where $\overline{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}^H} \approx \mathbb{E}[(\mathbf{Y}^w)^H\mathbf{Y}^w] + \gamma_{\rm IQ}\mathrm{E}[(\mathbf{Y}^w)^H\mathbf{Y}^w]^* + \gamma_{\rm PA}\mathbf{I}$. Therefore we can calculate $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{Y}\mathbf{Y}^H\mathbf{X}}$ and $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{X}}$ based on $\mathbf{u}^w$. We provide an example of calculating $\mathbb{E}[(\mathbf{X}^w)^H\mathbf{X}^w]$ in Appendix B. The other terms are similar and are omitted here.

We next look into the effects of the term $\overline{\mathbf{X}}_{\rm q}\mathbf{h}$ and $\mathbf{z}_{\rm q,RX}$ in (23). Note that $\overline{\mathbf{X}}_{\rm q}$ denotes the quantization error introduced when digitizing the output signal of power amplifier (cf. (20)), while $\mathbf{z}_{\rm q,RX}$ denotes the quantization error at the receiver side. According to the definition at the beginning of Section IV, the power of these two terms can be written as $P_{\rm QN,x} = \gamma_{\rm QN}P_{\rm Tx}$ and $P_{\rm QN,r} = \gamma_{\rm QN}P_{\rm Rx}$, where $P_{\rm Tx}$ and $P_{\rm Rx}$ denote the transmit and received signal power, respectively. Since $\overline{\mathbf{X}}_{\rm q}\mathbf{h}$ can be viewed as the quantization noise $\overline{\mathbf{X}}_{\rm q}$ going through the near-end channel, we can approximate its power by $P'_{\rm QN,x} = \mu P_{\rm QN,x}$, where μ represents the near-end channel gain. Denoting the power of received near-end signal as $P_{\rm near} = \mu P_{\rm Tx}$, we have $P'_{\rm QN,x} = \mu\gamma_{\rm QN}P_{\rm Tx} = \gamma_{\rm QN}P_{\rm near}$. Also, since the received is dominated by the near-end signal, $P_{\rm QN,r} \approx \gamma_{\rm QN}P_{\rm near}$. Therefore we can combine the power of $\overline{\mathbf{X}}_{\rm q}\mathbf{h}$ and $\mathbf{z}_{\rm q,RX}$ into one term as $P_{\rm QN} = P'_{\rm QN,x} + P_{\rm QN,r} \approx 2\gamma_{\rm QN}P_{\rm near}$. Similarly, we can write the combined power of the term $\overline{\mathbf{X}}_{\rm PN}\mathbf{h}$ and $\mathbf{z}_{\rm PN}$ in (23) as $P_{\rm PN} \approx 2\gamma_{\rm PN}P_{\rm near}$.

Therefore the SINR after the first stage cancellation can be obtained as follows

$$\mathrm{SINR}_{\rm s1} = \frac{P_{\rm far}}{\sigma_{\rm z}^2 + P_{\rm PN} + P_{\rm QN} + P_{\Delta,\rm s1}}, \tag{33}$$

where P_{far} and σ_z^2 represent the power of the received far-end signal and noise, respectively.

B. SINR of the Second Stage of Cancellation

Since the far-end signal is taken into consideration in the second stage of cancellation, the estimation precision of the near-end channel is improved over the first stage of cancellation. However, as has been described in Section III-B, there will be errors during the decoding of the desired signal, which will propagate into the next iteration and affect the precision of cancellation. Therefore it is very complicated to analyze the performance of the second stage cancellation. In this paper, we will give an analysis assuming that $\mathbf{Yg}$ and $\mathbf{Y}_{\rm IQ}\mathbf{g}$ are perfectly removed from the received signal of (18). Note that similar assumption has been made in [38], [39] to analyze the performance of SIC in MUD systems. In particular, the estimation error of the near-end channel will be

$$\begin{aligned}\Delta\mathbf{h}_{\rm s2} \approx (\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H(\mathbf{Y}_{\rm PA}\mathbf{g} + \mathbf{z} + \mathbf{z}_{\rm q,Rx} + \mathbf{z}_{\rm PN} \\ -\overline{\mathbf{X}}_{\rm q}\mathbf{h} - \overline{\mathbf{X}}_{\rm PN}\mathbf{h}).\end{aligned} \tag{34}$$

In the following, we will show that the terms $\overline{\mathbf{X}}_{\rm q}\mathbf{h}$, $\overline{\mathbf{X}}_{\rm PN}\mathbf{h}$, $\mathbf{z}$, $\mathbf{z}_{\rm q,RX}$, and $\mathbf{z}_{\rm PN}$ in (34) can be neglected in the analysis. Take the term $\mathbf{z}$ as an example, the estimation error of the near-end channel due to the existence of $\mathbf{z}$ is given by $\Delta\mathbf{h}_{\rm z} \approx (\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H\mathbf{z}$, and the resulting residual near-end signal is $\mathbf{s}_{\rm r,z} = \widetilde{\mathbf{X}}\Delta\mathbf{h}_{\rm z} = \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H\mathbf{z}$. Since pre-multiplying $\mathbf{z}$ by $\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^H\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^H$ means projecting $\mathbf{z}$ to the column space of $\widetilde{\mathbf{X}}$, i.e., from a space with a dimension of $(M+N)W$ to a subspace with a dimension of $L(L \ll (M+N)W)$, the norm of $\mathbf{s}_{\rm r,z}$ is much smaller than that of $\mathbf{z}$. Therefore, the power of the residual near-end signal caused by the existence of the term $\mathbf{z}$ is much smaller than that of $\mathbf{z}$, and can be neglected in the calculation of SINR since $\mathbf{z}$ also affects SINR (cf. (23). The same results apply to $\overline{\mathbf{X}}_{\rm q}\mathbf{h}$, $\overline{\mathbf{X}}_{\rm PN}\mathbf{h}$, $\mathbf{z}_{\rm q,RX}$, and $\mathbf{z}_{\rm PN}$ in (34), and we only need to account for the term $\mathbf{Y}_{\rm PA}\mathbf{g}$. Using a similar approach as (24)–(30), we can write the power of the residual near-end signal as follows

$$P_{\Delta,\rm s2} \approx \frac{\chi_{\rm far}}{(N+M)W} \times \mathrm{Tr}\left\{\widetilde{\mathbf{R}}^{-1}_{\mathbf{X}^H\mathbf{X}}\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{Y}_{\rm PA}\mathbf{Y}_{\rm PA}^H\mathbf{X}}\right\}, \tag{35}$$

where $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{Y}_{\rm PA}\mathbf{Y}_{\rm PA}^H\mathbf{X}} \triangleq \mathbb{E}[(\widetilde{\mathbf{X}}^w)^H\overline{\mathbf{R}}_{\mathbf{Y}_{\rm PA}\mathbf{Y}_{\rm PA}^H}\widetilde{\mathbf{X}}^w], \overline{\mathbf{R}}_{\mathbf{Y}_{\rm PA}\mathbf{Y}_{\rm PA}^H} = \mathbb{E}[\mathbf{Y}^w_{\rm PA}(\mathbf{Y}^w_{\rm PA})^H]$. Using the assumption again that the samples of the nonlinear signal components induced by PA nonlinearity are i.i.d., we have $\overline{\mathbf{R}}_{\mathbf{Y}_{\rm PA}\mathbf{Y}_{\rm PA}^H} = L \times \gamma_{\rm PA}P_{\rm Tx}\mathbf{I}$, therefore $\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{Y}_{\rm PA}\mathbf{Y}_{\rm PA}^H\mathbf{X}} = L \times \gamma_{\rm PA}P_{\rm Tx}\widetilde{\mathbf{R}}_{\mathbf{X}^H\mathbf{X}}$, and (35) will be

$$P_{\Delta,\rm s2} \approx \frac{\chi_{\rm far}\gamma_{\rm PA}P_{\rm Tx}L^2}{(N+M)W}. \tag{36}$$

Therefore the SINR in the second stage of cancellation is

$$\mathrm{SINR}_{\rm s2} = \frac{P_{\rm far}}{\sigma_{\rm z}^2 + P_{\rm PN} + P_{\rm QN} + P_{\Delta,\rm s2}}. \tag{37}$$

Note that $\mathrm{SINR}_{\rm s1}$ in (33) reduces to (10) when there are no RF/analog impairments and the transmit signals $x(n)$ and $y(n)$ are i.i.d. as assumed in Section III. Specifically, $P_{\Delta,\rm s1}$ reduces to $P_{\Delta,\rm s1} = (\chi_{\rm far}L^2/(N+M)W) = (P_{\rm far}L/N_f)$, which is consistent with P_Δ in (42) (note that the noise is neglected in calculating $P_{\Delta,\rm s1}$). With the introduction of second stage cancellation, the self-interference is cancelled more precisely. It can be shown from (36) that $P_{\Delta,\rm s2} \approx \gamma_{\rm PA}P_{\Delta,\rm s1}$ if the transmit signals $x(n)$ and $y(n)$ are i.i.d., therefore a higher SINR can be achieved. In addition, there is a tradeoff in choosing W as has been described in Section III-B: The precision of cancellation can be improved by increasing W since more samples can be used for estimating the near-end channel. On the other hand, since the channels are treated as static within each frame, increasing W will decrease the cancellation ratio due to channel variation.

V. Simulation Results

The performance of the proposed digital cancellation scheme is evaluated by computer simulation in this section. We consider a system with two users communicating with each other and the parameters used in the simulation are similar to those of IEEE 802.11a/g systems. The bandwidth of the system is 20 MHz. The FFT size and CP (Cyclic Prefix) length are 64 and 16, respectively. Fifty-two of the 64 subcarriers are occupied. The transmit symbols on each subcarrier are QPSK modulated. For simplicity, no coding or data scrambling are applied. The pulse shaping filter is a Hanning windowed sinc function, with oversampling factor $R_{\mathrm{Tx}} = 16$.

The far-end channel is modeled as multipath Rayleigh fading with standard power delay profile (PDP) as specified in the IEEE 802.11 standard [40]. Each signal path has a bell Doppler spectrum shape with the maximum Doppler shift of 6 Hz. The near-end channel is made up of two parts, i.e., line-of-sight and non-line-of-sight. The NLOS channel is modeled similar to the far-end channel with the same power delay profile and a path loss of 60 dB.[7] The LOS channel is modeled as Rician fading with a K-factor of 10, while the path loss depends on the Tx-Rx isolation, which is determined by the performance of the RF cancellation techniques applied. Unless otherwise stated, we will assume the Tx-Rx isolation to be 40 dB, as is described in Section II-B.

The transmit power is chosen to be 15 dBm. As in Table I, the PA nonlinearity is modeled by Rapp Model [24], [25] with the parameter $p = 3$. In addition, the operating point of the PA is adjusted so that the input back-off (IBO) is 8 dB. The I/Q imbalance exhibits an amplitude mismatch of 0.5 dB and phase mismatch of 3 degrees. The power spectrum density (PSD) of the LO phase noise has a constant value of -110 dBc/Hz below 10 KHz, and decreases by 20 dB per decade beyond that [26]. The carrier frequency offset between the near-end and far-end transmitter is set to 10 KHz. The resolution of ADCs is chosen to be 14 bits. The oversampling factor R_{Rx} at the receiver side is selected as 4, and a FIR filter with 40 taps is used for self-interference cancellation. Unless otherwise stated, 3 iterations are performed in the second stage of cancellation. The cancellation is carried out based on frames with length $N_f = 3200$, which corresponds to a duration of 10 OFDM symbols (note that the performance will be degraded if the frame is too long due to channel variation, as described in Section IV). The far-end channel state information for detection and the carrier frequency offset are acquired from training symbols. We notice that the time shift between the near-end signal and far-end signal has no obvious effect on the output SINR, therefore in the simulation we assume that the two users start transmitting signal frames at the same time, i.e., the time shift between the received near-end signal and far-end signal equals the propagation delay between the two users.

[7]This corresponds to a propagation distance of 5 m in free space for 5 GHz radio signal, which means that the nearest object that reflects the signal is 2.5 m away from the device. In fact, as described in Section II-B, due to the power loss in reflection/diffraction, the self-interference from NLOS pathes is much lower compared with the LOS component.

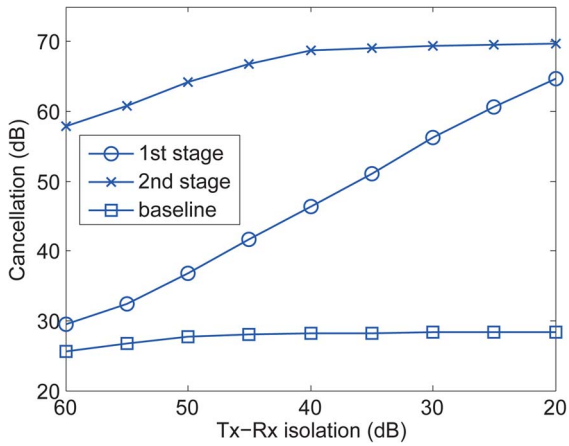

Fig. 5. Cancellation achieved by the proposed digital cancellation scheme under different levels of Tx-Rx isolation.

We use the conventional digital cancellation approach of [5] as a baseline for comparison, which features the same structure as the left hand side of Fig. 2 and cancels the self-interference according to (6) using near-end channel estimated in the frequency domain. The performance of our proposed scheme is compared with the baseline scheme [5] under the same Tx-Rx isolation and system parameters.

We first investigate the performance of the proposed scheme in terms of cancellation ratio, which is defined as the ratio between the power of the self-interference before and after the digital cancellation (cf. (9)). Fig. 5 shows the cancellation ratio for different levels of Tx-Rx isolation when the path loss experienced by the far-end signal is 70 dB. We can see that the proposed scheme achieves 46 dB cancellation in the first stage when the Tx-Rx isolation is 40 dB. In addition, the cancellation ratio increases as Tx-Rx isolation decreases, which is consistent with what has been observed in [6], [21]. This is because the power of the received self-interference increases as the Tx-Rx isolation decreases, which will increase the precision of near-end channel estimation, as described in Section III-A. The cancellation ratio achieved by the proposed scheme increases to 60–70 dB when the second stage cancellation is also included. On the other hand, the baseline digital cancellation scheme only achieves about 28 dB cancellation for almost all the cases of receive power, which is due to the effect of PA nonlinearity and transmit I/Q imbalance as described in Section III-A. Note that a similar phenomenon can be observed for the second stage of the proposed scheme. The reason is the quantization error and phase noise cannot be handled by digital cancellation. As the power of received self-interference increases, the effects of quantization error and phase noise become the bottleneck of the performance.

Fig. 6 depicts the output SINR of the desired signal corresponding to the result of Fig. 5. It can be seen that the SINR achieved in the first stage of cancellation is almost a constant value, while the introduction of second stage cancellation can improve the SINR performance. The simulation results matches well with the analysis. Moreover, the SINR values achieved by the proposed scheme is much higher than that of the baseline scheme. In fact, the performance of the proposed scheme is

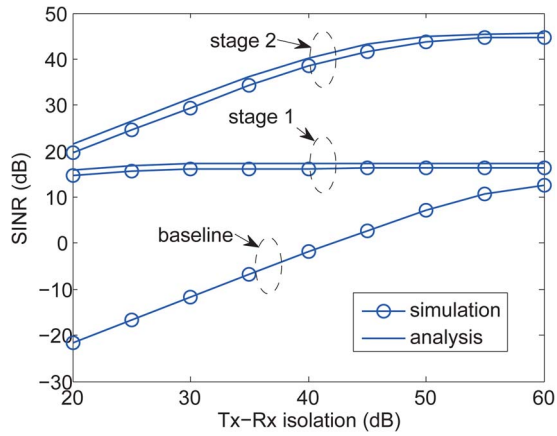

Fig. 6. SINR performance of the proposed scheme under different levels of Tx-Rx isolation. SNR $= 46$ dB.

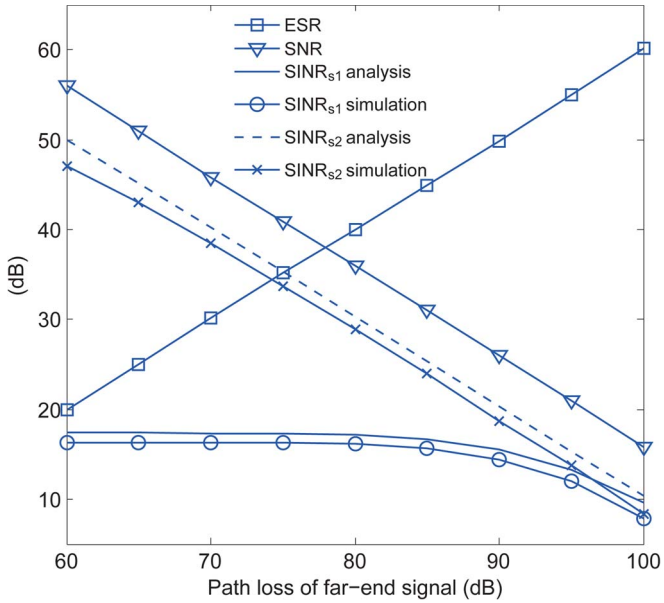

Fig. 7. SINR performance of the proposed full-duplex system for different distances between the near-end user and far-end user. Tx-Rx isolation is assumed to be 40 dB.

limited by quantization error and phase noise; while the baseline is limited by PA nonlinearity and I/Q imbalance, which generates much more signal distortions as is shown in Table I. As the Tx-Rx isolation increases, the receiver sees a decreased power in the spurious signal components caused by distortions (cf. $\check{P}_{\text{near}}$ in (12), P_{PN} and P_{QN} in (33) and (37)), therefore higher SINR values can be achieved. However, the SINR will finally saturate as Tx-Rx isolation keeps increasing, since it is bounded by the received far-end signal to noise ratio.

Fig. 7 shows the SINR performance of the proposed transceiver for different path losses experienced by the far-end signal. We define ESR and SNR as the received near-end signal to far-end signal ratio and the received far-end signal to noise ratio, respectively. The achieved SINR values achieved by the proposed schemes is much higher than that achieved by the baseline scheme, and is substantially increased by introducing the second stage of cancellation. In particular, the SINR achieved in the first stage of the proposed scheme is substantially lower than the SNR value when the path loss of far-end signal is small, while in the second stage cancellation this gap is reduced to about 8 dB. Note than this gap depends on the quantization error and phase noise, i.e., P_{PN} and P_{QN} in (37). Therefore the output SINR of the second stage cancellation can be further increased with higher Tx-Rx isolation, as can be seen in Fig. 6.

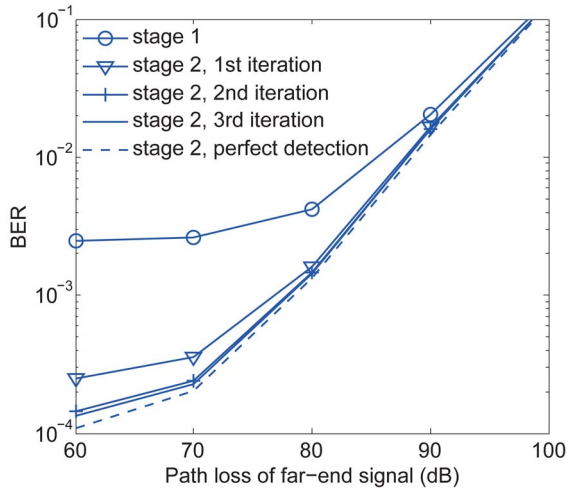

Fig. 8. BER performance of the proposed transceiver for different distances between the near-end user and far-end user.

To further demonstrate the effectiveness of the second stage cancellation, we show the bit error rate (BER) performance of the proposed system in Fig. 8, using the same configuration as for Fig. 7. It can be seen that the BER is decreased substantially by introducing the second stage cancellation, especially when the far-end path-loss is low. In addition, this improvement is mainly achieved in the first 2 iterations of the second stage cancellation, and is negligible after 3 iterations. The reason is that the near-end channel identification and far-end data detection have reached the highest achievable accuracy in 3 iterations, and the subsystem imperfections emerge as the bottleneck. Since the far-end transmit signal used in the second stage cancellation is generated based on the output of the decoder (cf. Fig. 3), the system performance is affected by the decoding errors. To investigate this effect, Fig. 8 also shows the BER performance when perfect detection is used for generating far-end transmit signals in the second stage cancellation (marked as 'perfect detection'). It can be seen that the BER performance for perfect detection is only slightly better than for the proposed scheme. The reason is that the effect of PA nonlinearity on the far-end signal cannot be reproduced when generating the estimate of the far-end transmit signal. Therefore the accuracy of self-interference cancellation in the second stage is mainly limited by the nonlinear components of the far-end signal, as can be seen in (34)–(36). It is also worth noting that the BER performance in Fig. 8 is different from that can be achieved in ideal Rayleigh-fading channels with the SINR values in Fig. 7. The reason for this is twofold: (1) the residual self-interference is not Gaussian distributed; (2) we have taken various RF/analog impairment effects into consideration in the simulation, including PA nonlinearity, LO phase noise, and CFO, each of which will degrade the BER performance.

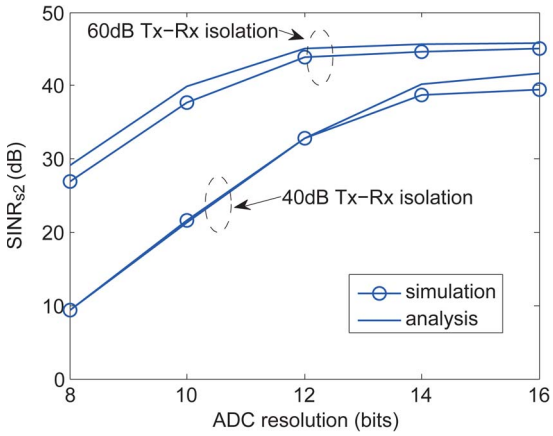

Fig. 9. Performance of the proposed cancellation scheme for different ADC resolutions. SNR $= 46$ dB.

Fig. 9 shows the performance of the proposed system for different ADC resolutions (N_{bit1} and N_{bit2} in Fig. 4). The far-end path-loss is chosen to be 70 dB, and we have considered two cases in which the Tx-Rx isolation is 40 dB and 60 dB, respectively. It can be seen that the output SINR in dB scale increases almost linearly with the resolution of ADC when the ADC resolution is low, and saturates as the resolution keeps increasing. This is because the quantization error is the major factor limiting the system performance when the ADC resolution is low. As the precision of ADCs increases, the power of quantization noise decreases and other effects will emerge as the bottleneck. It is also shown that when Tx-Rx isolation increases from 40 dB to 60 dB, the requirement on ADC resolution can be decreased by 3 bits to achieve the same SINR value. The reason is that the dynamic range of the received signal is decreased due to the reduction in the power of the self-interference.

VI. Conclusion

We have focused on digital cancellation techniques in full-duplex wireless communication systems. We show that the effects of ADC, phase noise, and sampling jitter are not the bottleneck of WiFi-like systems given the pre-cancellation achieved by existing RF techniques. However, PA nonlinearity and transmit I/Q imbalance are significant factors that limit the precision of digital self-interference cancellation, and as such we propose to use the signal at the output of the power amplifier to carry out the cancellation. In addition, to improve the output SINR, we propose a two-stage cancellation scheme which iteratively refines the cancellation using the detection result of the desired signal. The proposed approach is evaluated analytically and by simulation, and is shown to substantially outperform existing digital cancellation schemes.

Appendix A

The power of residual self-interference is given by

$$P_\Delta = \frac{1}{N_f} \times \text{Tr}\left\{\mathbb{E}[\mathbf{X}\Delta\mathbf{h}\Delta\mathbf{h}^H\mathbf{X}^H]\right\} \tag{38}$$

$$= \frac{1}{N_f} \times \mathbb{E}\left\{\text{Tr}\left[\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H \times(\mathbf{Y}\mathbf{g}\mathbf{g}^H\mathbf{Y}^H + \mathbf{z}\mathbf{z}^H)\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\right]\right\} \tag{39}$$

$$= \frac{\text{Tr}\left\{\mathbb{E}\left[\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H(\mathbf{Y}\mathbf{g}\mathbf{g}^H\mathbf{Y}^H + \mathbf{z}\mathbf{z}^H)\right]\right\}}{N_f}. \tag{40}$$

where the second equality is due to $\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$. Since $\mathbb{E}[g(l)g^*(k)] = \chi_{\text{far}}\delta(l-k)$ and $\mathbb{E}[z(n)z^*(m)] = \sigma_z^2\delta(n-m)$, we get $\mathbb{E}[\mathbf{g}\mathbf{g}^H] = \chi_{\text{far}}\mathbf{I}$ and $\mathbb{E}[\mathbf{z}\mathbf{z}^H] = \sigma_z^2\mathbf{I}$. By taking the expectation with respect to $\mathbf{g}$ and $\mathbf{z}$ first, (40) is written as

$$P_\Delta = \frac{\text{Tr}\left\{\mathbb{E}\left[\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\left(\mathbf{Y}\mathbf{Y}^H\chi_{\text{far}} + \sigma_z^2\mathbf{I}\right)\right]\right\}}{N_f}. \tag{41}$$

Moreover, since $\mathbb{E}[y(m)y^*(m)] = \delta(n-m)$, $\mathbb{E}[\mathbf{Y}\mathbf{Y}^H] = L\mathbf{I}$. By taking the expectation in (41) with respect to $y(n)$, we get

$$P_\Delta = \frac{\text{Tr}\left\{\mathbb{E}\left[\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\left(\chi_{\text{far}}L + \sigma_z^2\right)\mathbf{I}\right]\right\}}{N_f} \tag{42}$$

$$= \frac{\left(P_{\text{far}} + \sigma_z^2\right)\text{Tr}\left\{\mathbb{E}\left[(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{X}\right]\right\}}{N_f} \tag{43}$$

$$= \frac{L}{N_f}\left(P_{\text{far}} + \sigma_z^2\right). \tag{44}$$

Appendix B

Since $\mathbf{X}^w = [\mathbf{X}^w_{\text{CP}}; \mathbf{X}^w_{\text{DT}}]$, $\mathbb{E}[(\mathbf{X}^w)^H\mathbf{X}^w] = \mathbb{E}[(\mathbf{X}^w_{\text{DT}})^H\mathbf{X}^w_{\text{DT}}] + \mathbb{E}[(\mathbf{X}^w_{\text{CP}})^H\mathbf{X}^w_{\text{CP}}]$, we can then calculate $\mathbb{E}[(\mathbf{X}^w_{\text{DT}})^H\mathbf{X}^w_{\text{DT}}]$ and $\mathbb{E}[(\mathbf{X}^w_{\text{CP}})^H\mathbf{X}^w_{\text{CP}}]$ separately. Noticing that each column of $\mathbf{X}^w_{\text{DT}}$ contains the elements of $\mathbf{x}^w_{\text{DT}}$ with different orders and since $\mathbf{x}^w_{\text{DT}} = \mathbf{F}^H\mathbf{u}^w$, we can write the n-th column of $\mathbf{X}^w_{\text{DT}}$ as $[\mathbf{X}^w_{\text{DT}}]_{(:,n)} = \mathbf{F}^H_n\mathbf{u}^w$, where $\mathbf{F}^H_n (n = 1, \ldots, L)$ is obtained from IFFT matrix (cf. (17)) by changing the row orders according to the structure of $\mathbf{X}^w_{\text{DT}}$. Therefore the (i,j)-th element of $\mathbb{E}[(\mathbf{X}^w_{\text{DT}})^H\mathbf{X}^w_{\text{DT}}]$ can be written as

$$\mathbb{E}\left[(\mathbf{X}^w_{\text{DT}})^H\mathbf{X}^w_{\text{DT}}\right]_{(i,j)} = \mathbb{E}\left[(\mathbf{u}^w)^H\mathbf{F}_i\mathbf{F}_j^H\mathbf{u}^w\right] = \text{Tr}\left\{\mathbf{F}_i\mathbf{F}_j^H\mathbb{E}\left[\mathbf{u}^w(\mathbf{u}^w)^H\right]\right\},$$

where $\mathbb{E}[\mathbf{u}^w(\mathbf{u}^w)^H]$ can be calculated from the statistics of $\mathbf{u}^w$. For example, assuming the nonzero elements of $\mathbf{u}^w$ are i.i.d. with zero mean and variance of $\mathbb{E}[|u^w|^2]$, then $\mathbb{E}[\mathbf{u}^w(\mathbf{u}^w)^H] = \mathbb{E}[|u^w|^2]diag([a_1, , \ldots, a_N])$, where a_n equals 1 or 0 depending on whether the corresponding subcarrier is used or not. $\mathbb{E}[(\mathbf{X}^w_{\text{DT}})^H\mathbf{X}^w_{\text{DT}}]_{(i,j)}$ can then be obtained. $\mathbb{E}[(\mathbf{X}^w_{\text{CP}})^H\mathbf{X}^w_{\text{CP}}]$ can also be calculated similarly and therefore we can get $\mathbb{E}[(\mathbf{X}^w)^H\mathbf{X}^w]$.

Acknowledgment

The authors would like to thank the anonymous reviewers and the editors for helping improve this work. The authors also gratefully acknowledge the support of the Hong Kong RGC.

References

[1] S. Chen, M. A. Beach, and J. P. McGeehan, "Division-free duplex for wireless applications," *Electron. Lett.*, vol. 34, no. 2, pp. 147–148, Jan. 1998

[2] J. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. Annu. Int. Conf. MobiCom Netw.*, 2010, pp. 1–12.

[3] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1558–1562.
[4] T. Snow, C. Fulton, and W. Chappell, "Transmit-receive duplexing using digital beamforming system to cancel self-interference," *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 12, pp. 3494–3503, Dec. 2011.
[5] M. Jain *et al.*, "Practical, real-time, full duplex wireless," in *Proc. ACM Int. Conf. Mobile Comput. Netw.*, Las Vegas, NV, USA, Sep. 2011, pp. 301–312.
[6] 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, Dec. 2012.
[7] E. Aryafar, M. Khojastepour, K. Sundaresan, S. Rangarajan, and M. Chiang, "MIDU: Enabling MIMO full duplex," in *Proc. Annu. Int. Conf. MobiCom Netw.*, 2012, pp. 257–268.
[8] A. Sahai, G. Patel, C. Dick, and A. Sabharwal, "Understanding the impact of phase noise on active cancellation in wireless full-duplex," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2012, pp. 29–33.
[9] V. Aggarwal, M. Duarte, A. Sabharwal, and N. Shankaranarayanan, "Full- or half-duplex? A capacity analysis with bounded radio resources," in *Proc. IEEE ITW*, 2012, pp. 207–211.
[10] S. Li and R. D. Murch, "Full-duplex wireless communication using transmitter output based echo cancellation," in *Proc. IEEE GLOBECOM*, Dec. 2011, pp. 1–5.
[11] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Probability Stat.*, 1961, vol. 1, pp. 611–644, Berkeley, CA, USA: Univ. California Press.
[12] G. Zheng, I. Krikidis, and B. Ottersten, "Full-duplex cooperative cognitive radio with transmit imperfections," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2498–2511, May 2013.
[13] 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, Dec. 2011.
[14] N. Singh *et al.*, "Efficient and fair MAC for wireless networks with self-interference cancellation," in *Proc. Int. Symp. Modeling WiOpt Mobile, Ad Hoc Netw.*, May 2011, pp. 94–101.
[15] M. Khojastepour and S. Rangarajan, "Wideband digital cancellation for full-duplex communications," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 1300–1304.
[16] T. Riihonen and R. Wichman, "Analog and digital self-interference cancellation in full-duplex MIMO-OFDM transceivers with limited resolution in A/D conversion," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 45–49.
[17] Y. Hua, P. Liang, Y. Ma, A. Cirik, and Q. Gao, "A method for broadband full-duplex MIMO radio," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 793–796, Dec. 2012.
[18] M. Duarte *et al.*, "Design and characterization of a full-duplex multi-antenna system for WiFi networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1160–1177, Mar. 2014.
[19] S. Li and R. D. Murch, "Realizing up and downlink subcarrier allocation in Orthogonal Frequency Division Duplex (OFDD) systems," in *Proc. IEEE VTC*, May 2012, pp. 1–5.
[20] S. Li and R. D. Murch, "Realizing cooperative multiuser OFDMA systems with subcarrier resource allocation," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1923–1935, Apr. 2013.
[21] A. Sahai, G. Patel, C. Dick, and A. Sabharwal, "On the impact of phase noise on active cancelation in wireless full-duplex," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4494–4510, Nov. 2013.
[22] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *Proc. ACM SIGCOMM*, 2013, pp. 375–386.
[23] G. Fettweis *et al.*, "Dirty RF: A new paradigm," in *Proc. IEEE Int. Symp. PIMRC*, 2005, vol. 4, pp. 2347–2355.
[24] L. Smaini, *RF Analog Impairments Modeling for Communication Systems Simulation: Application to OFDM-Based Transceivers*. Hoboken, NJ, USA: Wiley, 2012.
[25] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology, Techniques*, 2nd ed. New York, NY, USA: Academic, 2002.
[26] Siversima, "SMT Frequency Synthesizer," [Online]. Available: http://www.siversima.com/wp-content/uploads/2010/05/smt-freguency-synth_1-2ghzaltb.pdf
[27] L. Ding, Z. Ma, D. R. Morgan, M. Zierdt, and G. T. Zhou, "Compensation of frequency-dependent gain/phase imbalance in predistortion linearization systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 1, pp. 390–397, Feb. 2008.
[28] A. Guerin, G. Faucon, and R. Le Bouquin-Jeannes, "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 672–683, Nov. 2003.
[29] L. Ding, Z. Ma, D. Morgan, M. Zierdt, and J. Pastalan, "A least-squares/Newton method for digital predistortion of wideband signals," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 833–840, May 2006.
[30] A. Wiewiorka and P. N. Noss, "Digital on-channel repeater for DAB," BBC, London, U.K., BBC R&D White Paper WHP120, Sep. 2005.
[31] D. Falconer, "Adaptive reference echo cancellation," *IEEE Trans. Commun.*, vol. 30, no. 9, pp. 2083–2094, Sep. 1982.
[32] C. Luschi *et al.*, "Advanced signal-processing algorithms for energy-efficient wireless communications," *Proc. IEEE*, vol. 88, no. 10, pp. 1633–1650, Oct. 2000.
[33] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[34] U. Fawer and B. Aazhang, "A multiuser receiver for code division multiple access communications over multipath channels," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 1556–1565, Feb.–Apr. 1995.
[35] D. Divsalar, M. K. Simon, and D. Raphaeli, "Improved parallel interference cancellation for CDMA," *IEEE Trans. Commun.*, vol. 46, no. 2, pp. 258–268, Feb. 1998.
[36] M. Lončar, R. R. Müller, J. Wehinger, C. F. Mecklenbräuker, and T. Abe, "Iterative channel estimation and data detection in frequency-selective fading MIMO channels," *Eur. Trans. Telecommun.*, vol. 15, no. 5, pp. 459–470, Sep./Oct. 2004.
[37] R. van der Hofstad and M. Klok, "Performance of DS-CDMA systems with optimal hard-decision parallel interference cancellation," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2918–2940, Nov. 2003.
[38] S. P. Weber, J. Andrews, X. Yang, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with successive interference cancellation," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2799–2814, Aug. 2007.
[39] R. Buehrer, "Equal BER performance in linear successive interference cancellation for CDMA systems," *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1250–1258, Jul. 2001.
[40] V. Erceg *et al.*, IEEE p802.11 Wireless LANs, May 2004, IEEE 802.11-03/940r4.

Shenghong Li (S'09) received the bachelor's degree in communication engineering from Nanjing University, Nanjing, Jiangsu, China, in 2008, and the Ph.D. degree in electronic and computer engineering from Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2013, respectively. He is now with the Commonwealth Scientific and Industrial Research Organisation (CSIRO) of Australia. His research interests include full-duplex techniques wireless communication systems, cooperative communication and MIMO.

Ross D. Murch (M'84–SM'98–F'09) received the bachelor's and Ph.D. degrees in electrical and electronic engineering from the University of Canterbury, New Zealand. He is a Chair Professor and Department Head of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST). His research contributions include more than 200 publications and 20 patents on wireless communication systems and antennas including MIMO, OFDM, propagation, MIMO antennas, and these have attracted over 9000 citations. His current research interests include the Internet-of-Things, energy harvesting, reconfigurable systems and autonomous systems. He is the founding and current Director of the Center for Wireless Information Technology, HKUST and also acts as a consultant for industry and government. He has been the publication editor, area editor and associate editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, is currently the Chair of the IEEE Communications Society technical committee on wireless communications and is also a distinguished lecturer for IEEE Vehicular Technology Society. He has been the David Bensted Fellow, Simon Fraser University, Canada, and was an HKTIIT Fellow at Southampton University, U.K. He was the Technical Program Chair for the IEEE Wireless Communications and Networking Conference in 2007, Keynote Chair for IEEE International Conference on Communications in 2010, and also the Advanced Wireless Communications Systems Symposium in IEEE International Communications Conference in 2002. He was also a keynote speaker at IEEE GCC 2007, IEEE WiCOM 2007, IEEE APWC 2008, and IEEE ICCT 2011. He has a strong interest in education and enjoys sharing his experiences with students. He is also a Fellow of IET and HKIE.