Signal Processing 108 (2015) 322–340

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Radio self-interference cancellation by transmit beamforming, all-analog cancellation and blind digital tuning ☆

Yingbo Hua *, Yiming Ma, Armen Gholian, Yifan Li, Ali Cagatay Cirik, Ping Liang

Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, United States

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form
16 September 2014
Accepted 17 September 2014
Available online 28 September 2014

Keywords:
Radio self-interference cancellation
Full-duplex radio
Transmit beamforming
Receive beamforming
All-analog cancellation
Blind digital tuning
IQ imbalance

ABSTRACT

Radio self-interference cancellation has been a technological challenge for more than a century while it is the most critical enabler for full-duplex radios. The eventual success of radio self-interference cancellation may well depend on not only improved hardware technology but also innovative signal processing schemes. In this paper, we present a few latest discoveries on such schemes. The first is an improvement of time-domain transmit beamforming with robustness against the IQ imbalances in radio circuits, which is supported by both simulation and hardware experimental results. A key innovation here is due to the use of real-valued linear model instead of complex-valued linear (or widely linear) model. The second is a numerical investigation of the performance limits of an all-analog cancellation channel based on clustered-taps of attenuators when the interference channel has a large number of random multipaths. The third is a blind digital tuning method which uses only the baseband waveforms to determine the values of the variable attenuators embedded in the all-analog cancellation channel. This method is robust against imperfections in the knowledge of the transfer function of any component in the system provided that a real-valued linearity property holds (except for the transmit chain).

1. Introduction

Everyone in the modern society has been using radio in one way or another. Radio devices such as smart phones have become ubiquitous. Current and future wireless mobile applications on these radio devices, including Facetime and Cloud Computing, are rapidly increasing the burden on the radio spectrum which is a limited natural resource at any given location. There is clearly a need to develop full-duplex radios to double or nearly double the spectral efficiency in certain applications. As cellular mobile networks become denser to increase the network throughput, inter-cell interferences have become an important concern and inter-cell interference cancellation is desired to realize a full-duplex radio network. A full-duplex radio that can transmit and receive at the same time and same frequency is also important for military applications where for example jamming signals against enemies need to be removed for selves and friends.

Since the first radio was demonstrated by Guglielmo Marconi in 1895, the world has not yet seen a commercially viable full-duplex radio. The fundamental challenge and enabler for a full-duplex radio is radio self-interference cancellation. Self-interference cancellation refers to cancellation of interference caused by a known source. In the fields of speech processing and wireline data communication, it is also known as echo cancellation and has a rich history of many decades. For example, see [1].

However, for radio self-interference cancellation, the earliest work we can identify is [2] published in late 1990s where a testbed was demonstrated for radio self-interference

☆ This work was supported in part by ARO under Contract no. W911NF1210432 and UCOP under Grant no. PC-12-247260.

* Corresponding author.
E-mail addresses: yhua@ee.ucr.edu (Y. Hua), yma@ee.ucr.edu (Y. Ma), armen.gholian@email.ucr.edu (A. Gholian), yli068@ucr.edu (Y. Li), acirik@ee.ucr.edu (A.C. Cirik), liang@ee.ucr.edu (P. Liang).

http://dx.doi.org/10.1016/j.sigpro.2014.09.025
0165-1684/

cancellation of a relatively narrow bandwidth (200 kHz). For about ten years after that, there were very few reports on radio self-interference cancellation besides [3–6]. A reason behind such a lack of progress is apparently due to the difficulty of the problem.

In the past five years or so, however, there have been a flurry of papers on the topic of full-duplex radio wireless communications. They are addressing a wide range of aspects of full-duplex radios. In the following, we provide a brief review of some selected papers to motivate the theme of this paper.

Ten years after [2], there came a trend of research activities on various forms of transmit beamforming for radio self-interference cancellation, e.g., [6–19]. In theory, the principle of beamforming can be applied in any of the four possible hardware settings: baseband to baseband, baseband to RF band, RF band to baseband, and RF band to RF band. Transmit beamforming refers to beamforming from baseband to RF band. The motivation of using transmit beamforming is that one could cancel the radio self-interference at the RF frontend of a receiver by generating an accurate RF cancellation signal based on the source of interference known in the baseband. Cancellation of interference at the RF frontend leaves the receive chain less burdened with potential saturations. Saturation happens when the desired signal is too low compared to the (undesired) interference. The idea of transmit beamforming was tested on a programmable radio board (WARP radio http://warpproject.org) as shown in [12,13] where a frequency-domain transmit beamforming (FDTB) method (although it was not so-called there) was used. Having several advantages over FDTB, time-domain transmit beamforming (TDTB) was introduced in [19]. All the prior transmit beamforming methods are based on a complex-valued linear system model which neglects the effect of IQ imbalances in practical radios. A simple way to address IQ imbalances is to treat each pair of IQ components as a vector of two real numbers instead of a complex number. In this way, the linearity of the system model is no longer affected by IQ imbalances. In Section 2, we will illustrate the performance gains of using the real-valued system model over using the complex-valued system model via both simulation and hardware-based experiments. This experimental contribution on TDTB is new and important.

It should be noted that the IQ imbalances have been frequently handled in the past two decades via a complex-valued widely linear model where an observed complex sequence is expressed as a complex linear combination of the source complex sequence and its complex conjugate. See [27] for its early introduction and [28] for its recent application in self-interference cancellation. But we believe that the real-valued linear model as discussed in this paper is much more straightforward, and hence potentially more powerful, than the complex-valued widely linear model. A brief explanation is provided in Appendix C.

The idea of transmit beamforming overlooks a problem known as transmission noise. A typical radio transmitter has a transmission SNR at around 30 dB. This sets an upper bound on the amount of radio self-interference cancellation achievable using transmit beamforming. In many practical situations, the required amount of radio self-interference cancellation is much larger than 30 dB. Until the radio hardware technology is much improved so that the transmission SNR is much increased, transmit beamforming will not be practical for situations where the transmit and receive power ratio is much larger than 30 dB. The limitation caused by radio transmission noise was recognized in [20,21]. A key remedy to resolve this limitation is to directly tap the output of the transmit power amplifier to produce the cancellation signal. In [20], the cancellation is done in the baseband which leaves the frontend of the receiver vulnerable to saturation by strong self-interference and quantization noise. In a way, the principle used in [20] can be called receive beamforming [9], i.e., beamforming from RF band to baseband.

In parallel to transmit beamforming, there has been another trend of research activities focused entirely on the RF frontend self-interference cancellation [22,23,25,24,26], which we call all-analog radio self-interference cancellation. The all-analog approach uses only passive devices to turn an analog signal tapped from the output of transmit power amplifier into an analog cancellation signal that is combined with the received RF signal containing a strong interference. The passive devices generate virtually no noise themselves. Unlike that of transmit beamforming, the performance of the all-analog approach is not limited asymptotically by the radio transmission SNR. Unlike the method in [20], the all-analog approach cancels the interference before it reaches the low noise amplifier (LNA) in the receive chain. The early work in [2] also belongs to the all-analog category. From a signal processing point of view, however, the all-analog cancellation method can be viewed as an implementation of beamforming completely at the RF frontend to avoid the limitation by radio transmission noise.

While the prior works on all-analog radio self-interference cancellation contain useful and/or important contributions, they all have limitations. The work in [2] is limited to a narrow band (200 kHz) system. The work in [22] attempted two transmit antennas with a half-wavelength spacing difference relative to a receiving antenna to create a null at the receiver. This approach only works for the narrowband case and is also sensitive to the actual spacing between antennas. The work in [23] used two delay paths each with an attenuator. This approach does not handle well a frequent situation where the interference channel comprises multiple random paths. The recent work in [24] extended the two-paths cancellation channel to a multiple-path cancellation channel. But they did not provide an effective choice of the distribution of the tap delays for arbitrary environment. No statistical effect of a random (due to change of environment) interference channel on the cancellation performance was reported there. Note that even the self-interference (imperfect isolation) channel of a RF circulator is in practice affected by objects near the antenna(s). The work in [25] assumes the use of both an attenuator and a phase shifter for each tap of an all-analog cancellation channel. But the phase of a typical phase shifter is not easy to adjust, and its insertion loss is significantly coupled

with its phase. Even when the optimal value of the complex coefficient of a tap is known, there remains a major hardware challenge to implement it on a pair of attenuator and phase shifter.

A latest work shown in [26] is also limited to a small number of multipaths. They proposed a least-mean-square (LMS) based algorithm for implementation at the RF frontend. They assume the use of a down-converter for the RF input to each and every attenuator in an all-analog cancellation channel, which is very costly and would cause a very large form-factor in hardware implementation. Their computer simulation demonstrates that their method has some level of tolerance against system model errors. Our scheme shown in Sections 3 and 4 is very different from theirs. We use real-valued system model instead of complex-valued system model. We consider a random interference channel with a large number of random multipaths instead of a few fixed multipaths. We assume no direct measurement of the RF input to any attenuator in the all-analog cancellation channel.

In Section 3, we will propose new architectures of the all-analog cancellation channel. Unlike the one in [25], the proposed architectures do not use variable phase shifters. The only required adjustable components are attenuators. Unlike [24], we propose delay distributions in the cancellation channel that are applicable for a wide range of interference channels. We will present a statistical evaluation of the performance limits of the cancellation channel subject to a large number of random interference channel realizations.

In Section 4, we will present a blind digital tuning method for all-analog cancellation channels. This method does not require a direct access to either the analog input or the analog output of the interference channel. Instead, it only uses the baseband output from a radio receiver (i.e., a single down-converter) to determine the desired values of the adjustable attenuators used in the all-analog cancellation channel. This method does not need to know the exact knowledge of the transfer function of any part of the system. It is also inherently robust to I/Q imbalances. The main assumption required by this method is that the observed data, when treated as real-valued instead of complex-valued, are linearly affected by some controllable values of the variable attenuators and linearly affected by all independent sources of noise. Simulations shown in Section 5 illustrate that the blind digital tuning method combined with an all-analog cancellation channel can break the barrier of radio transmission noise even when it is subject to other hardware imperfections.

The rest of this paper is organized as follows. In Section 2, we review the TDTB method and present an IQ-imbalance resistent extension of the TDTB method. The performance improvement of the extension (real-valued TDTB) over the original (complex-valued TDTB) is supported by both simulation and hardware results. In Section 3, we present new architectures of the all-analog cancellation channel and their performance limits subject to a random interference channel model. We also show how step attenuators with a finite step size and a finite number of control bits affect the performance limits of the all-analog cancellation channel. In Section 4, we present the blind digital tuning method which only uses the baseband output from a radio receiver to determine the optimal values of the variable attenuators used in the all-analog cancellation channel. In Section 5, we present how the blind digital tuning method performs in simulation. The conclusion of this paper is given in Section 6.

Notations: Vectors are represented by bold lowercase letters and matrices by bold uppercase letters unless otherwise mentioned. $\mathbf{A}\in C^{n\times m}$ (or $\mathbf{A}\in R^{n\times m}$) means that $\mathbf{A}$ belongs to the set of $n\times m$ complex matrices (or real matrices respectively). The convolution is denoted by $*$. The Kronecker product is denoted by $\otimes$. All random variables are uniformly distributed unless otherwise mentioned.

2. Transmit beamforming

2.1. Using complex-valued linear model

Shown below is a compact form of the TDTB method shown in [19]. Consider a MIMO radio system with n_r receivers (receive chains) and $n_t > n_r$ transmitters (transmit chains). This system can be as small as a single radio node with multiple antennas or as large as a network of interconnected radio base stations each with multiple antennas. In practice, n_r of the n_t transmitters' RF outputs can be directly connected via cables or circuits to the n_r receivers at the RF frontend for more effective cancellation (and less overall channel dispersion). Furthermore, in the case of a single MIMO full-duplex radio, $\max(n_r, n_t-n_r)$ antennas can be shared via circulators for both reception and transmission.

We assume that the baseband sources of the radio interferences are accessible and controllable. These source waveforms can be represented by $\mathbf{x}[n]=[x_1[n],\ldots,x_{n_t}[n]]^T\in C^{n_t\times 1}$ where n is the sampling time index. The baseband-equivalent waveforms of the received radio interferences can be represented by $\mathbf{y}[n]=[y_1[n],\ldots,y_{n_r}[n]]^T\in C^{n_r\times 1}$. Here, ^{T} denotes transpose. If IQ imbalances are ignored (along with other imperfections in practical radio), the following complex-valued linear model holds:

$$\mathbf{y}[n]=\mathbf{H}[n]*\mathbf{x}[n]+\mathbf{w}[n]=\sum_{l=0}^{L_H}\mathbf{H}[l]\mathbf{x}[n-l]+\mathbf{w}[n] \tag{1}$$

where $\mathbf{H}[n]\in C^{n_r\times n_t}$ is the impulse response (matrix sequence) of the interference channel, and $\mathbf{w}[n]\in C^{n_r\times 1}$ is noise. In the context of radio self-interference cancellation, for most practical radios, this noise is dominated by the radio transmission noise. This is because of the high power of self-interference.

Throughout this paper, we focus on radio self-interference cancellation in the absence of desired signals. The methods so derived require a time period of "silence" from remote nodes. But such a required period could be as small as a few microseconds to one millisecond in broadband applications, which depends on the choice of cancellation strategy and its implementation.

To eliminate the interference, we first choose $\mathbf{x}[n]=\sum_{l=0}^{L_P}\mathbf{P}[l]\mathbf{s}[n-l]$ where $\mathbf{P}[n]\in C^{n_t\times(n_t-n_r)}$ is the complex-valued response of the transmit beamforming prefilters, and $\mathbf{s}[n]\in C^{(n_t-n_r)\times 1}$ is the information-carrying signal

that is meant for a remote radio but causes the self-interference. Then, we need to determine $\mathbf{P}[l], l=0,\ldots,L_P$ such that $\sum_{l=0}^{L_H}\mathbf{H}[l]\mathbf{x}[n-l]=0$ for any $\mathbf{s}[n]$, or equivalently,

$$\mathbf{H}[n]*\mathbf{P}[n]=0 \tag{2}$$

To show a compact form of a solution to (2), we first let $\mathbf{H}[n]=[\mathbf{H}_a[n],\mathbf{H}_b[n]]$ and $\mathbf{P}[n]=[\mathbf{P}_a^T[n],\mathbf{P}_b^T[n]]^T$ where $\mathbf{H}_a[n]\in C^{n_r\times n_r}$, $\mathbf{H}_b[n]\in C^{n_r\times(n_t-n_r)}$, $\mathbf{P}_a[n]\in C^{n_r\times(n_t-n_r)}$ and $\mathbf{P}_b[n]\in C^{(n_t-n_r)\times(n_t-n_r)}$. It follows from (2) that

$$\mathbf{H}_a[n]*\mathbf{P}_a[n]+\mathbf{H}_b[n]*\mathbf{P}_b[n]=0 \tag{3}$$

Similar to a discussion in [19], a compact form of a solution to (3) is

$$\mathbf{P}_a[n]=-adj\{\mathbf{H}_a[n]\}*\mathbf{H}_b[n] \tag{4}$$

$$\mathbf{P}_b[n]=det\{\mathbf{H}_a[n]\}\mathbf{I}_{n_t-n_r} \tag{5}$$

where $\mathbf{I}_n$ denotes the $n\times n$ identity matrix, $adj\{\mathbf{H}_a[n]\}$ is the functional adjoint of the functional matrix $\mathbf{H}_a[n]$, and $det\{\mathbf{H}_a[n]\}$ is the functional determinant of the (same) matrix $\mathbf{H}_a[n]$. For example,

$$det\begin{Bmatrix} a_1[n] & a_2[n] \\ a_3[n] & a_4[n] \end{Bmatrix}=a_1[n]*a_4[n]-a_2[n]*a_3[n].$$

The definition of the functional adjoint and determinant is the same as that of the numerical adjoint and determinant of a numerical matrix as given in [35] except that the multiplications used in the numerical adjoint and determinant are replaced here by convolutions.

Note that we have chosen to use the time-domain representations. If one applies the Fourier transform to the above expressions, all convolutions in the time-domain become multiplications in the frequency-domain as is well-known.

A couple of examples of the solution to (2) are shown next. If $n_r=1$ and $n_t=2$, we can write the matrices $\mathbf{H}[n]$ and $\mathbf{P}[n]$ in terms of their elements as $\mathbf{H}[n]=[\mathbf{H}_a[n],\mathbf{H}_b[n]]=[h_a[n],h_b[n]]$ and $\mathbf{P}[n]=[\mathbf{P}_a^T[n],\mathbf{P}_b^T[n]]^T=[p_a[n],p_b[n]]^T$, respectively. And (4) and (5) imply that $p_a[n]=-h_b[n]$ and $p_b[n]=h_a[n]$.

If $n_r=2$ and $n_t=3$, we can write similarly

$$\mathbf{H}[n]=[\mathbf{H}_a[n]|\mathbf{H}_b[n]]=\begin{bmatrix} h_{1,1}[n] & h_{1,2}[n] & h_{1,3}[n] \\ h_{2,1}[n] & h_{2,2}[n] & h_{2,3}[n] \end{bmatrix}$$

and $\mathbf{P}[n]=[\mathbf{P}_a^T[n];\mathbf{P}_b^T[n]]^T=[p_{1,1}[n],p_{2,1}[n];p_{3,1}[n]]^T$. It is easy to verify from (4) and (5) that $p_{1,1}[n]=-h_{2,2}[n]*h_{1,3}[n]+h_{1,2}[n]*h_{2,3}[n]$, $p_{1,2}[n]=h_{2,1}[n]*h_{1,3}[n]-h_{1,1}[n]*h_{2,3}[n]$, $p_{1,3}[n]=h_{1,1}[n]*h_{2,2}[n]-h_{1,2}[n]*h_{2,1}[n]$.

A useful property of the above forms of solutions is that the operations involved are no more than sums of convolutions. And there is no inverse required. All of these are useful for hardware implementation.

The (rank) condition of $\mathbf{H}_a[n]$ can be made strong by connecting the first n_r transmit chains to the n_r receive chains via wires/circuits. In this case, $\mathbf{H}_a[n]$ is approximately an all-pass MIMO channel response and hence it causes virtually no additional dispersion on the channel response between this node and a remote node.

2.1.1. Experiment

The above method with $n_r=1$ and $n_t=2$ has been tested on a hardware platform shown in Fig. 1 which uses the carrier frequency at 2.4 GHz. The bandwidth of the source waveforms we used (i.e., sequences of hamming windowed *sinc* functions each of 1 μs sampled at 40 MHz) is about 15 MHz.

The quality of cancellation depends on many factors including the accuracy of the model (1) as well as the power of the noise $\mathbf{w}[n]$. These factors affect the accuracy of the estimates of $h_a[n]$ and $h_b[n]$ and hence the accuracy of the chosen $p_a[n]$ and $p_b[n]$.

The amount of the interference cancellation is also a random quantity which depends on many time-varying factors in both the wireless channel environment and the RF circuit. This was also recognized in [14]. To show a statistical performance, we use the cumulative distribution functions (CDF) of the interference-(plus noise)-to-noise ratio (INR). We compute INR as the ratio of the power of the peak of a received interference waveform (plus noise) of length 1 μs (including the inherent receiver noise) over the average power of the receiver noise (in the absence of interference). INR=0 dB would indicate the absence of interference. The peak of each (actually received) waveform is random, and so is INR. Notice that the INR is meaningful only when the receiver is not saturated. So, in obtaining the meaningful INR, we purposely kept the transmitted power low enough so that no saturation happened at the receiver.

In Fig. 2, we show the CDFs of the INR for the four different antenna configurations. In this figure, "RF" denotes the RF frontend cancellation using the TDTB method, and "RF+BB" denotes the RF frontend cancellation followed by a baseband-only cancellation. The baseband-only method follows the conventional strategy where the received baseband signal (self-interference) is subtracted by a least-square linear transformation of the transmitted baseband signal.

From Fig. 2, we observe that (1) the RF frontend interference cancellation based on the TDTB method achieved an average 25 dB reduction of interference, (2) the residual

Fig. 1. A programmable radio board (WARP) where the TDTB method with $n_r=1$ and $n_t=2$ was tested. The FPGA on radio board was reprogrammed to accommodate the time domain approach.

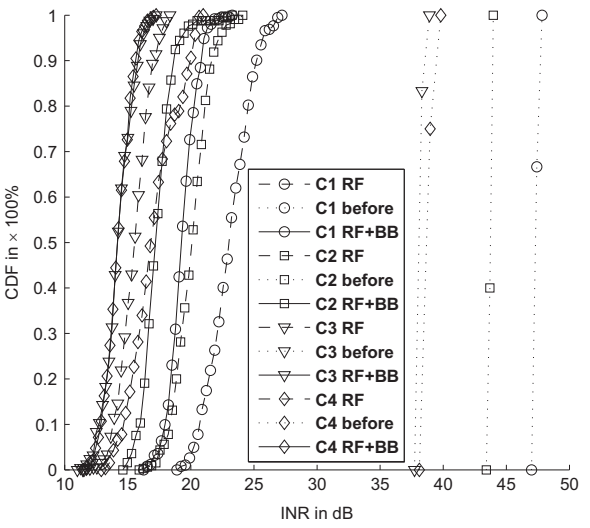

Fig. 2. The CDFs of the INR based on the hardware platform shown in Fig. 1 with four different antenna configurations "C1, C2, C3, C4". Here, "before" means that there is no interference cancellation, "RF" means that there is interference cancellation at the RF frontend, and "RF+BB" means that the RF frontend cancellation is combined with a baseband-only cancellation. Each curve is generated from hundreds of samples regardless of the number of markers shown on the curve. The samples for each curve were collected from multiple runs of a sequence of waveforms, and these runs span multiple seconds.

interference after the RF frontend cancellation was further reduced by only about 5 dB by the baseband-only method, and (3) the residual interference after both RF frontend cancellation and baseband-only cancellation is about 14–17 dB above the receiver noise. A similar level of interference cancellation using the FDTB was reported in [12–14]. These results of cancellation are not good enough for most applications.

2.2. Using real-valued linear model

One of the causes of the residual interference is IQ imbalances in practical radio [38]. IQ imbalances destroy the linearity of the conventionally adopted complex-valued system model. For example, see Appendix B. To reduce the impact of IQ imbalances, shown below is the TDTB method based on a real-valued linear system model.

To handle IQ imbalances, it is effective to treat each IQ waveform pair as a sequence of 2×1 real-valued vectors instead of a sequence of complex-valued numbers. We will use a "bar" on a symbol to highlight its real-valued representation. For example,

$$\overline{x}[n]=\begin{bmatrix}\mathrm{Re}\{x[n]\}\\ \mathrm{Im}\{x[n]\}\end{bmatrix}.$$

In general, one can replace the complex-valued linear model (1) by the following real-valued linear model:

$$\overline{\mathbf{y}}[n]=\overline{\mathbf{H}}[n]*\overline{\mathbf{x}}[n]+\overline{\mathbf{w}}[n]=\sum_{l=0}^{L_H}\overline{\mathbf{H}}[l]\overline{\mathbf{x}}[n-l]+\overline{\mathbf{w}}[n] \tag{6}$$

where $\overline{\mathbf{y}}[n]=[\overline{y}_1[n],\ldots,\overline{y}_{n_r}[n]]^T\in R^{2n_r\times 1}$, $\overline{\mathbf{x}}[n]=[\overline{x}_1[n],\ldots,\overline{x}_{n_t}[n]]^T\in R^{2n_t\times 1}$, and $\overline{\mathbf{H}}[n]\in R^{2n_r\times 2n_t}$ is the real-valued impulse response of the interference channel. More specifically, we can write the (i,j)th 2×2 block of $\overline{\mathbf{H}}[n]$ as $\overline{H}_{i,j}[n]=\overline{C}(\delta_{r,i},\phi_{r,i})\overline{H}(h_{i,j}[n])\overline{C}(\delta_{t,j},\phi_{t,j})$ where $h_{i,j}[n]$ is the (i,j)th element of $\mathbf{H}[n]$. Here,

$$\overline{C}(\delta,\phi)=\begin{bmatrix}(1+\delta)\cos(\phi) & (1-\delta)\sin(\phi)\\ (1+\delta)\sin(\phi) & (1-\delta)\cos(\phi)\end{bmatrix} \tag{7}$$

$$\overline{H}(h[n])=\begin{bmatrix}\mathrm{Re}\{h[n]\} & -\mathrm{Im}\{h[n]\}\\ \mathrm{Im}\{h[n]\} & \mathrm{Re}\{h[n]\}\end{bmatrix} \tag{8}$$

Although we will need to know $\overline{\mathbf{H}}[n]$ which can be estimated based on the model (6), we do not need to know any of these individual components: $\delta_{r,i}$, $\phi_{r,i}$, $\delta_{t,j}$, $\phi_{t,j}$ and $h_{i,j}[n]$. The matrix $\overline{C}(\delta_{r,i},\phi_{r,i})$ accounts for the ith receiver IQ imbalance, and the matrix $\overline{C}(\delta_{t,j},\phi_{t,j})$ accounts for the tth transmitter IQ imbalance.

It is important to stress that the real-valued linear model (6) is more general than the complex-valued linear model (1). While the IQ imbalances destroy the validity of (1), they do not affect the validity of (6).

To eliminate the interference given by (6), we now let $\overline{\mathbf{x}}[n]=\sum_{l=0}^{L_P}\overline{\mathbf{P}}[l]\overline{\mathbf{s}}[n-l]$ where $\overline{\mathbf{P}}[n]\in R^{2n_t\times 2(n_t-n_r)}$ is the real-valued response of the transmit beamforming prefilters.

Assume that $\overline{\mathbf{H}}[l], l=0,\ldots,L_H$, are known. Then, $\overline{\mathbf{P}}[l], l=0,\ldots,L_P$, should be such that $\sum_{l=0}^{L_H}\overline{\mathbf{H}}[l]\overline{\mathbf{x}}[n-l]=0$ for any $\overline{\mathbf{s}}[n]$, which is equivalent to

$$\overline{\mathbf{H}}[n]*\overline{\mathbf{P}}[n]=0 \tag{9}$$

Then, let $\overline{\mathbf{H}}[n]=[\overline{\mathbf{H}}_a[n],\overline{\mathbf{H}}_b[n]]$ and $\overline{\mathbf{P}}[n]=[\overline{\mathbf{P}}_a^T[n],\overline{\mathbf{P}}_b^T[n]]^T$ where $\overline{\mathbf{H}}_a[n]\in R^{2n_r\times 2n_r}$, $\overline{\mathbf{H}}_b[n]\in R^{2n_r\times 2(n_t-n_r)}$, $\overline{\mathbf{P}}_a[n]\in R^{2n_r\times 2(n_t-n_r)}$ and $\overline{\mathbf{P}}_b[n]\in R^{2(n_t-n_r)\times 2(n_t-n_r)}$. It follows from (9) that

$$\overline{\mathbf{H}}_a[n]*\overline{\mathbf{P}}_a[n]+\overline{\mathbf{H}}_b[n]*\overline{\mathbf{P}}_b[n]=0 \tag{10}$$

It then follows from (10) that

$$\overline{\mathbf{P}}_a[n]=-adj\{\overline{\mathbf{H}}_a[n]\}*\overline{\mathbf{H}}_b[n] \tag{11}$$

$$\overline{\mathbf{P}}_b[n]=det\{\overline{\mathbf{H}}_a[n]\}\mathbf{I}_{2(n_t-n_r)} \tag{12}$$

Note that unlike (4) and (5), the solutions (11) and (12) are robust against the IQ imbalances.

2.2.1. Analytical comparison

To compare the complex-valued TDTB with the real-valued TDTB, we now consider the simple case of $n_r=1$ and $n_t=2$.

With the complex-valued linear model, as shown before, we have $\mathbf{H}[n]=[h_a[n],h_b[n]]\in C^{1\times 2}$ and

$$\mathbf{P}[n]=[-h_b[n],h_a[n]]^T \tag{13}$$

Without loss of generality, we can refer to $h_b[n]$ as the complex-valued impulse response of the interference channel and $h_a[n]$ as the complex-valued impulse response of the cancellation channel. Then, the prefiltered (complex) signal is

$$\mathbf{x}[n]_c=\begin{bmatrix}-h_b[n]\\ h_a[n]\end{bmatrix}*s[n] \tag{14}$$

and its equivalent real-valued form is

$$\overline{\mathbf{x}}[n]_c=\begin{bmatrix}-\overline{H}(h_b[n])\\ \overline{H}(h_a[n])\end{bmatrix}*\overline{s}[n] \tag{15}$$

With the real-valued model, we can first define

$$\overline{\mathbf{H}}_a[n]=\begin{bmatrix} h_{r,1,1}[n] & h_{r,1,2}[n] \\ h_{r,2,1}[n] & h_{r,2,2}[n] \end{bmatrix} \in R^{2\times 2} \tag{16}$$

$$\overline{\mathbf{H}}_b[n]=\begin{bmatrix} h_{r,1,3}[n] & h_{r,1,4}[n] \\ h_{r,2,3}[n] & h_{r,2,4}[n] \end{bmatrix} \in R^{2\times 2} \tag{17}$$

and then write from (11) and (12) that

$$\overline{\mathbf{P}}_a[n]=\begin{bmatrix} -h_{r,2,2}[n] & h_{r,1,2}[n] \\ h_{r,2,1}[n] & -h_{r,1,1}[n] \end{bmatrix} * \begin{bmatrix} h_{r,1,3}[n] & h_{r,1,4}[n] \\ h_{r,2,3}[n] & h_{r,2,4}[n] \end{bmatrix}$$

$$\overline{\mathbf{P}}_b[n]=(h_{r,2,2}[n]*h_{r,1,1}[n]-h_{r,1,2}[n]*h_{r,2,1}[n])diag[1,1]$$

And the corresponding prefiltered (real) signal is

$$\overline{\mathbf{x}}[n]_r=\begin{bmatrix} \overline{\mathbf{P}}_a[n] \\ \overline{\mathbf{P}}_b[n] \end{bmatrix} * \overline{s}[n] \tag{18}$$

Note that $\overline{\mathbf{x}}[n]_c$ in (15) and $\overline{\mathbf{x}}[n]_r$ in (18) are both real-valued prefiltered waveforms but they are based on two different models. As long as there is an IQ imbalance, the two waveforms should yield different results for interference cancellation. But if there is no IQ imbalance, one can verify that the two waveforms are equivalent to each other (up to a complex-valued scalar filter), i.e., $\mathbf{x}[n]_r=h_a^*[n]*\mathbf{x}[n]_c$ where $\mathbf{x}[n]_r$ is the complex-valued equivalent of $\overline{\mathbf{x}}[n]_r$ and $h_a^*[n]$ is the complex conjugate of $h_a[n]$.

2.2.2. Numerical comparison

We have not yet been able to test the real-valued TDTB on a hardware platform. But we have obtained the numerical results shown below. With $n_r=1$ and $n_t=2$, (6) implies

$$\overline{y}[n]=H_a[n]*\overline{x}_a[n]+H_b[n]*\overline{x}_b[n]+\overline{w}[n] \tag{19}$$

where $[\overline{x}_a^T[n],\overline{x}_b^T[n]]=\overline{\mathbf{x}}^T[n]$. And we model the cancellation channel by

$$\overline{H}_a[n]=C(\delta_r,\phi_r)\overline{H}(h_a[n])C(\delta_{t,a},\phi_{t,a}) \tag{20}$$

and the interference channel by

$$\overline{H}_b[n]=C(\delta_r,\phi_r)\overline{H}(h_b[n])C(\delta_{t,b},\phi_{t,b}) \tag{21}$$

where each of δ_r, $\delta_{t,a}$, $\delta_{t,b}$, ϕ_r, $\phi_{t,a}$ and $\phi_{t,b}$ is randomly chosen within $[-\delta_{max},\delta_{max}]$.

For channel estimation, we need four training slots. In slot 1, we let $\overline{\mathbf{x}}[n]=[\delta[n],0,\delta[n],0]^T$. In slot 2, we let $\overline{\mathbf{x}}[n]=[0,\delta[n],0,\delta[n]]^T$. In slot 3, we let $\overline{\mathbf{x}}[n]=[\delta[n],0,-\delta[n],0]^T$. In slot 4, we let $\overline{\mathbf{x}}[n]=[0,\delta[n],0,-\delta[n]]^T$.

Then, based on (19), the received signals $\overline{y}_1[n]$ and $\overline{y}_2[n]$ in the first two slots can be written as

$$\overline{Y}_a[n]=[\overline{y}_1[n],\overline{y}_2[n]]=(\overline{H}_a[n]+\overline{H}_b[n])*\overline{P}[n]+\overline{W}_1[n] \tag{22}$$

and $\overline{y}_3[n]$ and $\overline{y}_4[n]$ in the second two slots can be written as

$$\overline{Y}_b[n]=[\overline{y}_3[n],\overline{y}_4[n]]=(\overline{H}_a[n]-\overline{H}_b[n])*\overline{P}[n]+\overline{W}_2[n] \tag{23}$$

where $\overline{P}[n]=diag(\delta[n],\delta[n])$. Then, we compute the estimates $\hat{\overline{H}}_a[n]=\frac{1}{2}(\overline{Y}_a[n]+\overline{Y}_b[n])$ and $\hat{\overline{H}}_b[n]=\frac{1}{2}(\overline{Y}_a[n]-\overline{Y}_b[n])$. It is easy to verify that these two estimates are the least-square estimates. With the estimates of $\overline{H}_a[n]$ and $\overline{H}_b[n]$, we can compute $\overline{\mathbf{x}}[n]_r$ from (18)and then the residue interference from (19).

Using the complex-valued linear model (1) (instead of the actual model (19)), the same four training signals computed from (19) are modelled as

$$\overline{y}_1[n]\rightarrow y_1[n]=h_a[n]+h_b[n]+w_1[n] \tag{24}$$

$$\overline{y}_2[n]\rightarrow y_2[n]=jh_a[n]+jh_b[n]+w_2[n] \tag{25}$$

$$\overline{y}_3[n]\rightarrow y_3[n]=h_a[n]-h_b[n]+w_3[n] \tag{26}$$

$$\overline{y}_4[n]\rightarrow y_4[n]=jh_a[n]-jh_b[n]+w_4[n] \tag{27}$$

Hence, the estimates of $h_a[n]$ and $h_b[n]$ based on the complex-valued linear model should be obtained as $\hat{h}_a[n]=\frac{1}{4}(y_1[n]+y_3[n]-jy_2[n]-jy_4[n])$ and $\hat{h}_b[n]=\frac{1}{4}(y_1[n]-y_3[n]-jy_2[n]+jy_4[n])$. Then, the residual interference is

$$\overline{y}[n]=A[\overline{H}_a[n],\overline{H}_b[n]]*\begin{bmatrix} -\overline{H}(\hat{h}_b[n]) \\ \overline{H}(\hat{h}_a[n]) \end{bmatrix} * \overline{s}[n]+\overline{w}[n] \tag{28}$$

where A is such that the power of $\overline{\mathbf{x}}[n]$ is the same for both models.

The performance advantage of using the real-valued linear model can be measured by the following ratio of the power of residual interference using the real-valued linear model over the power of residual interference using the complex-valued linear model:

$$\beta_{dB}=10\log_{10}\frac{\sum_n\|\overline{y}[n]\|^2 \text{ based on (19) with (18)}}{\sum_n\|\overline{y}[n]\|^2 \text{ based on (28)}} \tag{29}$$

Shown in Fig. 3 are the cumulative distribution functions (CDF) of β_{dB} based on the following channel models:

$$h_a[n]=w[n]\sum_{i=0}^{100}a_i sinc\left(\frac{nT_s-\tau_{a,i}}{2T_s}\right)$$

$$h_b[n]=w[n]\sum_{i=0}^{100}b_i sinc\left(\frac{nT_s-\tau_{b,i}}{2T_s}\right)$$

where $-M\le n\le M$, $M=20$, $w[n]=0.54+0.46\cos(2\pi n/(2M+1))$, $T_s=25\times 10^{-9}$, $\tau_{a,0}=\tau_{b,0}=0$, $a_0=e^{j\phi_a}/D^2$, $b_0=e^{j\phi_b}/D^2$, $-\pi<\phi_a\le\pi$, $-\pi<\phi_b\le\pi$, $D=0.3$. And for $i>0$, $0<\tau_{a,i}<10^{-8}$, $0<\tau_{b,i}<10^{-8}$, $a_i=\alpha_{a,i}/(D+\tau_{a,i}c)^2$, $b_i=\alpha_{b,i}/(D+\tau_{b,i}c)^2$, $c=3\times 10^8$, $\alpha_{a,i}$ and $\alpha_{b,i}$ are complex Gaussian random variables with zero mean and unit variance. The nominal baseband-equivalent bandwidth of both the interference channel ($h_b[n]$) and the cancellation channel ($h_a[n]$) is 10 MHz. The sums in $h_a[n]$ and $h_b[n]$ represent random scattering within a reasonable range relative to the distance betweentransmitter and receiver. Furthermore, the following source signal is used to determine the residual interference:

$$s[n]=\sum_{i=0}^{100}s_i c[n-4i]$$

where s_i is 4-QAM symbol, $c[n]=w[n]sinc[n/4]$ with $-20\le n\le 20$. The symbol rate of the source signal is 10 Mega symbols per second.

Each curve in Fig. 3 is based on 500 runs and a given value of transmission SNR. For each run, all parameters of the IQ imbalances in (20) and (21) and all parameters in

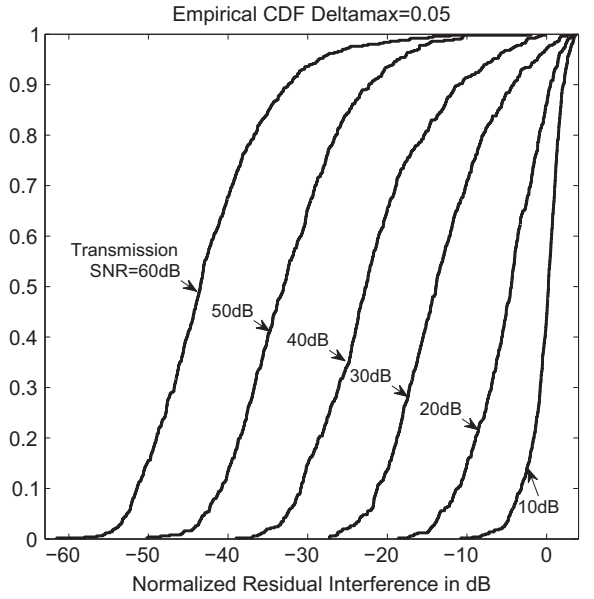

Fig. 3. CDF of β_{dB} for a simulated multipath random interference channel. The upper bound on random IQ imbalances is $\delta_{max} = 0.05$. Each curve corresponds to a given transmission SNR.

the above channel models are realized independently with uniform distribution unless otherwise mentioned.

From Fig. 3, we see that the real-valued TDTB can yield much better cancellations than the complex-valued TDTB even if the radio has a moderate transmission SNR (e.g., 30 dB).

2.2.3. Hardware experiment

To do a test on the same WARP radio board shown in Fig. 1, there was a difficulty to reprogram the FPGA. To simplify the test, we used two Agilent signal generators (MXG) to simulate two transmit chains and one Agilent signal analyzer (MXA) to simulate one receive chain. Since the data transfer from the signal analyzer to the signal generators cannot be done in real time (unlike that on the WARP), we put the transmit and receive antennas inside a radio anechoic chamber. In this way, we can measure the channel responses and then do the beamforming computations offline without affecting the quality of cancellation. We purposely maximized the IQ imbalances in the two generators. One has 10° IQ phase imbalance, and the other has -10° IQ phase imbalance. For the IQ gain imbalances, one is set to have 1 dB, and the other is set to have -1 dB. The waveform transmitted is a sequence of Gaussian pulses of bandwidth equal to 35 MHz with the carrier frequency 2.4 GHz.

Shown in Fig. 4 are the CDF of the amount of interference cancellations in dB for four cases: (1) using the complex-valued TDTB (denoted as RF in Fig. 4), (2) using the complex-valued TDTB and additional complex-valued baseband/digital cancellation (denoted as RF-BB in Fig. 4), (3) using the real-valued TDTB (denoted as IQRF in Fig. 4), and (4) using the real-valued TDTB and additional real-valued baseband/digital cancellation (denoted as IQRF-IQBB in Fig. 4). The right most curve in the figure

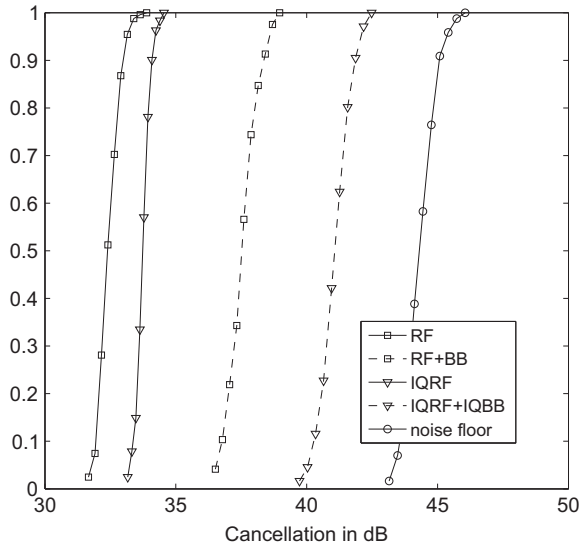

Fig. 4. The CDF of the amount of cancellation in dB in four cases of TDTB using Agilent's signal generators and signal analyzer.

corresponds to the amount of cancellation that would be needed to reach the noise floor. The advantage of using the real-valued linear model over the complex-valued linear model is obvious from this figure. For the above-mentioned IQ imbalances set on the signal generators, the real-valued TDTB has resulted in 3 dB improvement over the complex-valued TDTB. See Fig. 4. Even when the IQ imbalances were set to zero on the signal generators, a performance difference about 1–2 dB was also observed. This improvement based on hardware experiment is not as high as one would expect from computer simulation. This is because of hardware impairments other than IQ imbalances.

According to our hardware experiment as reported in this paper as well as those reported elsewhere such as [12], using the conventional radio chains to implement the transmit beamforming based methods can only yield about 30–40 dB cancellation at the best. This limits the range of applications of transmit beamforming for radio self-interference cancellation. But nevertheless transmit beamforming is still potentially useful. One such application is inter-cell radio interference cancellation between (small cell) base stations. Deploying small cells is useful for improving network spectral efficiency but also causes a problem of inter-cell interference. See Figs. 5 and 6. Unlike self-interference at the same site or even on the same unit, the power ratio of inter-cell interference over a desired signal may well be within 30 dB and hence the performance of 30 dB cancellation may be sufficient. For wireless infrastructure companies, it is feasible to use cables between stations (that are not too far from each other) to perform inter-cell interference cancellation.

For military applications such as "jamming while receiving" where a jamming signal is transmitted at the same time and same frequency as a (weak) desired signal is being received, the dynamic range or power ratio of interference over desired signal can be more than 100 dB. In this case, transmit beamforming from baseband to RF

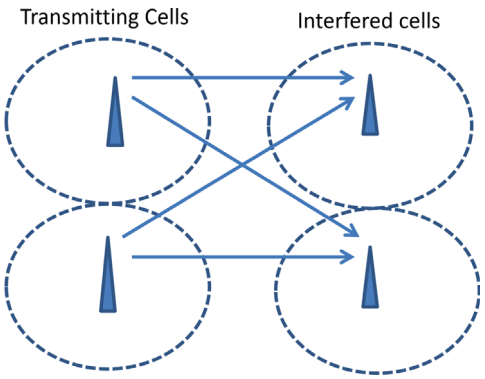

Fig. 5. Illustration of inter-cell interference between small base stations (or self-interference between radio units on the same radio board). For a given time slot, the transmitting stations are interfering the receiving stations although the receiving stations may be one or more tiers away from the transmitting stations.

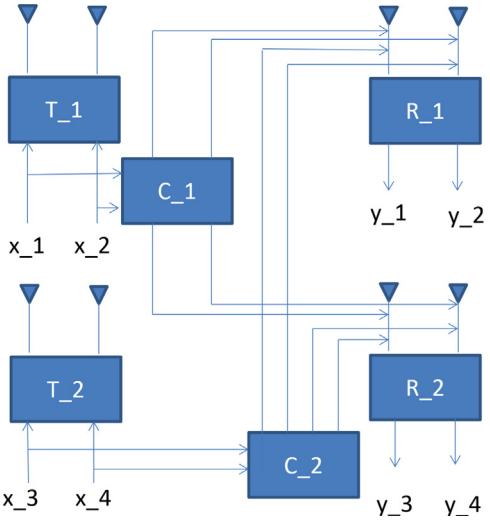

Fig. 6. Illustration of cabling between small base stations (or radio units) for transmit beamforming to cancel interferences at the interfered stations (or units). Here T_1 and T_2 denote the transmit radio chains, R_1 and R_2 the receive radio chains, and C_1 and C_2 the cancellation channels. Its dual form is receive beamforming as illustrated in Fig. 7.

band is inadequate unless there will be a radio transmitter that has a transmission SNR at that level.

3. All-analog cancellation

As discussed previously, the performance of the TDTB method (like all other transmit beamforming based methods) is fundamentally limited by the transmission SNR. In this section, we consider all-analog cancellation channels and present some of their performance limits. One such channel is illustrated in Fig. 8, which consists of N clustered taps to be called c-taps. Each c-tap has a cluster of four (variable) attenuators connected together via power splitter/dividers. A power divider is also often called power splitter. A power divider is also a power combiner when connected in reverse order [38]. For each cluster of attenuators, three cascaded two-way 90° power dividers are used. A two-way 90° power

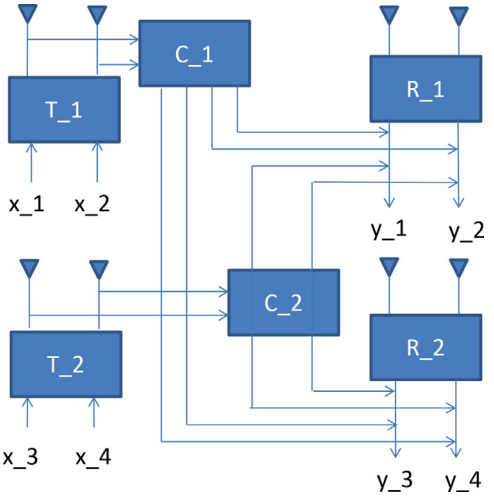

Fig. 7. Illustration of cabling between small base stations for receive beamforming to cancel interferences at the interfered stations (or units). This is the dual form of Fig. 6.

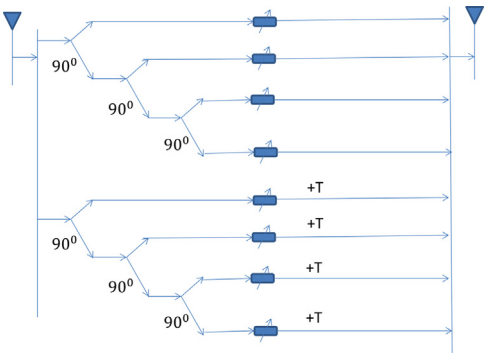

Fig. 8. An architecture of all-analog cancellation channel using cables, power splitters/dividers and variable attenuators. When two or more arrows diverge from each other, it is a power divider/splitter. When two or more arrows merge together, it is a power combiner. A (fixed) delay can be easily implemented by choosing the length of a RF cable/circuit. The relative delay between two adjacent clusters (c-taps) of attenuators is $T=\sigma_T/W$ with W being the bandwidth and $\sigma_T<1$. Illustrations with more than two c-taps follow obviously from this illustration. In practice, the insertion loss and phase change of each of the power splitters can be compensated with simple modifications.

divider splits a RF signal into two which differ from each other by 90° in phase. The delay between any two adjacent c-taps is T which is fixed be $T=\sigma_T/W$ where W is the bandwidth of interest and $\sigma_T<1$. Our experiment has shown that $\sigma_T=0.1$ is a good choice and the performance is not sensitive to σ_T when it is around 0.1.

Ideally, the frequency response of this cancellation channel is

$$H(f)=e^{-j2\pi fT_0}\sum_{n=0}^{N-1}[(g_{n,1}-g_{n,3})-j(g_{n,2}-g_{n,4})]e^{-j2\pi fTn} \tag{30}$$

where T_0 accounts for the delay of the 0th c-tap (which should be determined based on a given system), and $g_{n,i}$, $i=1,2,3,4$, are the attenuation factors of the four attenuators in the nth c-tap.

It is clear from (30) that the attenuators in each c-tap can produce any complex value within a 2×2 square centered at zero in the form of $(g_{n,1}-g_{n,3})-j(g_{n,2}-g_{n,4})$. The intuition is that each c-tap with only four attenuators can match a large number of clustered multipaths in the interference channel. And if we also choose T small enough, the frequency response $H(f)$ can be tuned to match a very wide range of those of the interference channel provided that the delay spread of the interference channel is no larger than NT.

The structure of this cancellation channel differ from that shown in [25] where an adjustable phase shifter is required along with an adjustable attenuator in each tap. Accurate adjustable phase shifters seem harder to find than accurate adjustable attenuators. Typically, the phase of a phase shifter is also coupled with its insertion loss. Hence, simultaneously setting both an attenuator and an cascaded phase shifter accurately seems a hard hardware problem. For the structure shown in Fig. 8, there is no such a problem. Each $g_{n,i}$ can be represented by a high precision step attenuator. More discussion on this will be shown later.

While an all-analog cancellation channel of multiple taps is also mentioned in [24], they did not provide an effective choice of the delay distribution applicable for arbitrary environment. The prior works on all-analog cancellation channels such as [22–25] did not address the statistical behaviors of their cancellation performance in the presence of a random interference channel (corresponding to varying environment).

We will next consider the performance limits of (30) by assuming the following multipath interference channel impulse response:

$$h_{int}(t)=\sum_{i=0}^{I-1} a_i\delta(t-\tau_i-T_0) \tag{31}$$

where I is the number of radio paths, a_i is the attenuation of the ith path, and τ_i+T_0 is the total delay of the ith path.

Note that in the RF domain (instead of in the baseband) of a single polarization measurement, the impulse response is real-valued (instead of complex-valued or 2×1 real-valued vectors).

To simulate a wide range of interference channels, we assume that a_i for each i represents the attenuation of an *absolutely single* path (instead of a cluster of paths) and has the following structure:

$$a_i=\frac{\epsilon\alpha_i d^{\alpha_p}}{(d+c\tau_i)^{\alpha_p}} \tag{32}$$

where α_i is a uniform random number within $(0,1)$, $d=0.03$ m, $c=3\times 10^8$ m/s, τ_i is a uniform random number within $(0,1)$ μs, and $\alpha_p\geq 1$ is the amplitude path loss exponent. We also choose $\tau_0=0$, $a_0=\epsilon$, $I=1000$ and $\epsilon=\frac{1}{1000}$. This choice of ϵ is to make sure that even if all the delays of the multipaths in the interference channel are the same (or nearly the same), there is still no effective gain through the interference channel. Obviously, if there is an effective gain through the interference channel, then the interference cancellation channel consisting of attenuators (with no amplification) can never produce a good cancellation signal.

The frequency response of $h_{int}(t)$ is

$$H_{int}(f)=e^{-j2\pi f_c T_0}\sum_{i=0}^{I-1} a_i e^{-j2\pi f\tau_i} \tag{33}$$

For each realization of $H_{int}(f)$, the best cancellation is achieved if $H(f)$ in (30) is obtained from the following:

$$\min_{0\leq g_{n,i}\leq 1\,\forall n,i}\int_{f_c-W/2}^{f_c+W/2}|H_{int}(f)-H(f)|^2 \tag{34}$$

where f_c is the center frequency and W the bandwidth of interest. We will choose $f_c=2.4$ GHz and $W=100$ MHz. (We have observed that when the bandwidth W is reduced the performance improves.) To solve the above problem, we approximate each of $H_{int}(f)$ and $H(f)$ by $M=1000$ uniformly distributed samples within $(f_c-W/2, f_c+W/2)$ and then apply the CVX software (http://cvxr.com/cvx/) to perform the convex optimization.

To evaluate the performance of the all-analog canceller, we define as follows a relative power (in dB) of the residual interference for the rth realization of the interference channel and the mth frequency:

$$e^{(r)}(f_m)=10\log_{10}\frac{|H_{int}^{(r)}(f_m)-H^{(r)}(f_m)|^2}{|H_{int}^{(r)}(f_m)|^2} \tag{35}$$

where $H^{(r)}(f_m)$ follows from (30) with all $g_{n,i}$ determined by (34) for the rth realization. One can treat the negative of $e^{(r)}(f_m)$ as the dB amount of cancellation for the rth run and the mth frequency. Also define three averaged residuals:

$$E_1^{(r)}=\frac{1}{M}\sum_{m=1}^{M} e^{(r)}(f_m) \tag{36}$$

$$E_2(f_m)=\frac{1}{N_r}\sum_{r=1}^{N_r} e^{(r)}(f_m) \tag{37}$$

$$E_3=\frac{1}{MN_r}\sum_{r=1}^{N_r}\sum_{m=1}^{M} e^{(r)}(f_m) \tag{38}$$

where we choose $N_r=1000$.

Shown in Fig. 10 is the distribution of the relative residual interference over frequency after all-analog cancellation. We see the impact of the number of c-taps N. With only $N=2$, one can see that the ideal performance of cancellation can be above 70 dB. Note that even for $N=5$, the delay spread of the cancellation channel is only half of the inverse of the bandwidth. It means that the interference channel is essentially all-pass. This is because of the short distance between the transmitter and the receiver. If the path loss exponent α_p increases, the amount of cancellation also increases (almost linearly as shown in simulation) subject to the same N.

Shown in Fig. 11 is the CDF of the relative residual interference after all-analog cancellation. It is clear that there is generally a large gap between the best case (the top point of a curve) and the worse case (the bottom point of the curve) for each given N. One can see that the worst case decreases substantially as N increases. The randomness of the amount of cancellation is due to the random nature of the interference channel.

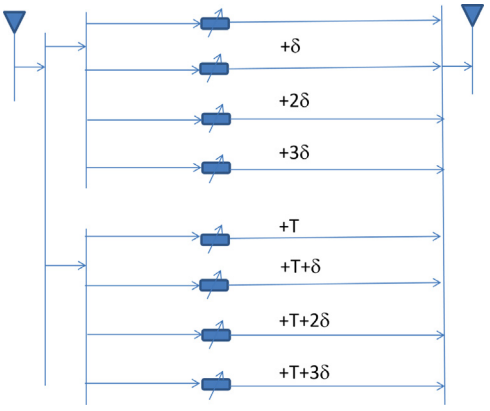

Fig. 9. An alternative of all-analog cancellation channel where the attenuators are also clustered as in Fig. 8 but the fixed 90° phase of the power divider is replaced by a small delay δ satisfying $f_c\delta=\frac{1}{4}$.

An alternative architecture of all-analog cancellation channel is shown in Fig. 9 where each 90° phase change is replaced by a small delay $\delta=1/4f_c$. For this architecture, the ideal frequency response has the form:

$$H_\delta(f)=e^{-j2\pi fT_0}\sum_{n=0}^{N-1}[g_{n,1}+g_{n,2}e^{-j2\pi f\delta}+g_{n,3}e^{-j2\pi f2\delta}+g_{n,4}e^{-j2\pi f3\delta}]e^{-j2\pi fTn} \tag{39}$$

When $W>f_c$, then within $(f_c-W/2,f_c+W/2)$, $e^{-j2\pi f\delta}\approx -j$ and hence $H_\delta(f)\approx H(f)$ even with the same set (except for a trivial permutation) of $g_{n,i}\,\forall n,i$. From our simulation, this δ-delayed architecture has a similar (although slightly worse) performance limit than the previous one using 90°-phase shifters.

We also considered an architecture with uniformly distributed attenuators (instead of clustered attenuators) with the time delay between two adjacent taps to be $T_u=T/4$. With this choice of T_u, the uniform architecture and those in Figs. 8 and 9 span the same delay spread with the same number of attenuators. For the uniform architecture, the ideal frequency response is

$$H_{uniform}(f)=e^{-j2\pi fT_0}\sum_{l=0}^{4N-1}g_l e^{-j2\pi fT_u l} \tag{40}$$

It is easy to see that if T_u is such that f_cT_u is an integer, then all terms inside the sum are real and nonnegative. In this case, the uniform architecture loses all its capacity to match a random interference channel. A necessary condition on T_u is to make sure that $e^{-j2\pi f_cT_ul}$ for all l span evenly in the complex space. If T_u is chosen arbitrarily, there is no guarantee of good performance, which has been supported by our simulation.

Unlike the first architecture, the second and third need to choose the delays (or some of the delays) based on a known carrier frequency. Other architectures we have considered are omitted here.

Next, let us consider the effect of step attenuators. A step attenuator typically has two important parameters: the step size Δ (in dB typically) and the number of control bits n_b, e.g., see http://www.minicircuits.com. Although the currently available commercial step attenuators have attenuation-dependent phases, we assume in this paper that such phases are already pre-compensated. The gain factor g (still called "gain" although it is less than one) of a step attenuator meet the following condition:

$$20\log_{10}g=-\Delta m \tag{41}$$

where $0\le m\le 2^{n_b}-1\doteq m_{max}$.

For the architecture in Fig. 8, the optimal solutions for the gains $g_{n,i}$ are not unique, which is easy to see from (30). Namely, as long as the differences $g_{n,1}-g_{n,3}$ and $g_{n,2}-g_{n,4}$ do not change, nor does the frequency response $H(f)$. This property turns out to be important in utilizing the step attenuators, which is shown next.

Assuming a desired (positive) difference D_g between a pair of gains g_a and g_b to be quantized into $\hat{g}_a$ and $\hat{g}_b$, we need to choose two integers m_a and m_b such that

$$\min_{0\le m_a\le m_b\le m_{max}}|D_g-f(m_a,m_b)| \tag{42}$$

where $f(m_a,m_b)=\hat{g}_a-\hat{g}_b=10^{-\Delta m_a/20}-10^{-\Delta m_b/20}$.

The problem (42) can be solved efficiently as follows.[1] Initialize $a=0$ and $b=m_{max}$. For each integer m_a with $0\le m_a\le m_{max}$, determine $\hat{m}_b=-(20/\Delta)\log_{10}(D_g-10^{-\Delta m_a/20})$, pick the best among the three pairs of integers: (a,b), $(m_a,\lfloor\hat{m}_b\rfloor)$ and $(m_a,\lceil\hat{m}_b\rceil)$, and reassign it as (a,b). The final pair of (a,b) is the solution to (42). The complexity of this algorithm is in the order of m_{max}.

The problem (42) inspires a more general problem which is the same as (42) except that $f(m_a,m_b)$ does not have any structure other than the following properties:

$$f(m_a,m_b)>f(m_a+1,m_b) \tag{43}$$

$$f(m_a,m_b)>f(m_a,m_b-1) \tag{44}$$

$$f(m_a,m_b)>f(m_a+1,m_b+1) \tag{45}$$

Shown in Fig. 12 is the grid of m_a and m_b, which is useful for the derivation and understanding of an efficient search algorithm. The basic idea behind the algorithm is to find a bounding pair of points (a_{low},b_{low}) and (a_{up},b_{up}) such that the inequality $f(a_{low},b_{low})<D_g<f(a_{up},b_{up})$ is as tight as possible before the final solution of the grid point (a_{opt},b_{opt}) is determined. Such an algorithm is given in Appendix A. This algorithm has been tested to consistently yield the identical results as the 2-D exhaustive search for thousands of random choices of D_g. While always smaller than that using the exhaustive 2-D search, the actual amount of the time needed to find the optimal solution using the efficient algorithm is random, which depends on the given D_g. Our simulation shows that for $n_b=12$ and $\Delta=0.5$ dB, the efficient algorithm is on average (over 1000 random choices of D_g between $(0,1)$) eighteen times faster than the 2-D exhaustive search.

Compared to the first algorithm which uses the detailed structure $f(m_a,m_b)=\hat{g}_a-\hat{g}_b=10^{-\Delta m_a/20}-10^{-\Delta m_b/20}$, the second algorithm which does not use the detailed structure has been found less efficient when $n_b>8$.

It is useful to note that the smallest non-zero D_g achievable is $D_{g,min}\doteq f(m_{max}-1,m_{max})=10^{-\Delta(m_{max}-1)/20}(1-10^{-\Delta/20})$. For example, if $\Delta=0.5$ dB and $n_b=6$, then $D_{g,min}=0.0016$.

[1] This was suggested by an anonymous reader.

And, in contrast, the smallest achievable value for individual g is $10^{-\Delta(m_{max}-1)/20} = 0.0266$. The resolution for D_g increases rapidly as n_b increases.

Shown in Fig. 13 is the residual interference after all-analog cancellation with step attenuators and $N=3$. We see that for each given step size Δ, there is a threshold point for the number of bits used. This is because when n_b is large enough, the noise becomes dominant. The noise is added to $H_{int}(f)$ here but not for Figs. 10 and 11 which show the ideal potential of all-analog cancellation in the absence of noise. It is useful to note that for $\Delta = 0.5$ dB (a typical value for commercially available step attenuators), the amount of cancellation with 8 control bits is nearly 70 dB, which is an encouraging number. Additional results can be found in [31].

It should be noted that for inter-cell radio interference cancellation between base stations, all-analog cancellation channels can also be installed as shown in Fig. 14. This is similar to Fig. 6 except that all the cancellation signals here are tapped from and combined at the RF frontend.

The actual performance of the all-analog cancellation channel is also affected by many other parameters (imperfections of hardware) other than those of the step attenuators. In the next section, we present a method that is robust to many of these imperfections.

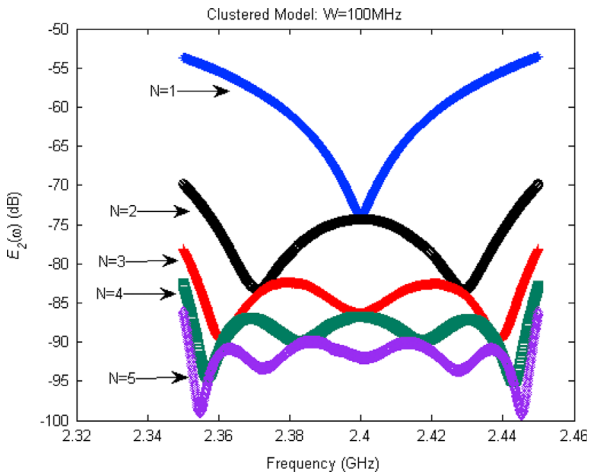

Fig. 10. The distribution of the residual interference over frequency after all-analog cancellation, i.e., $E_2(f_m)$. $T = 1/10W$. $\alpha_p = 2$.

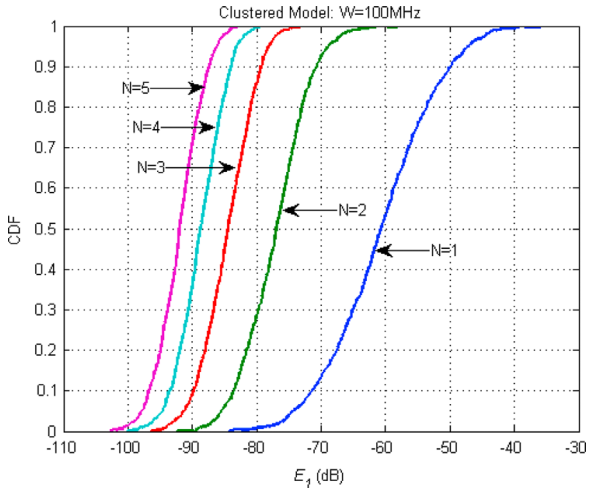

Fig. 11. The CDF of the residual interference after all-analog cancellation, i.e., the CDF of $E_1^{(r)}$. $T = 1/10W$. $\alpha_p = 2$.

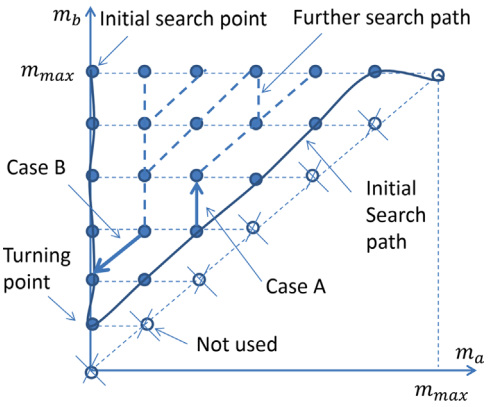

Fig. 12. The quantization grid for $f(m_a, m_b)$ satisfying (43)–(45). For an efficient search algorithm, the solid path of the connected dots is the initial search path to find an initial bounding pair of adjacent points. Throughout the entire search process, there are two special cases of a bounding pair of (not necessarily adjacent) points called Case A and Case B, the examples of which are illustrated (the point being pointed to corresponds to an upper bound). An efficient search algorithm to solve (42) without using the structure of $f(m_a, m_b)$ other than (43)–(45) is given in Appendix A.

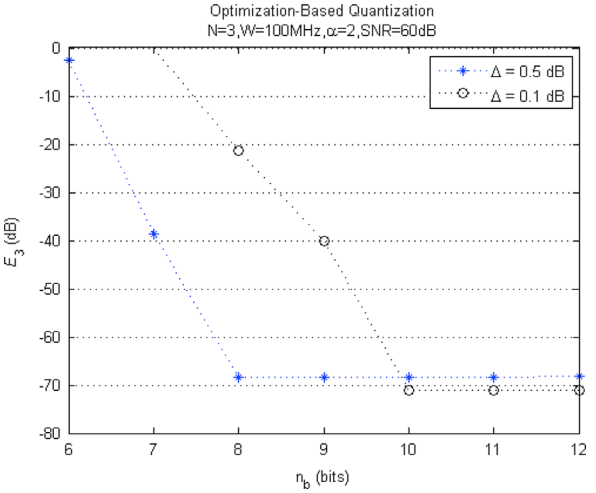

Fig. 13. The residual interference after all-analog cancellation, i.e., E_3, versus n_b the number of quantization bits of step attenuators. SNR = 60 dB due to a noise added to $H_{int}(f)$.

4. Blind digital tuning

The previous section has presented an ideal situation of all-analog cancellation. In reality, it is impossible to obtain the exact transfer function of the interference channel $H_{int}(f)$. Even if an exact $H_{int}(f)$ is available and if the optimal values of the attenuators are computed based on (34), the

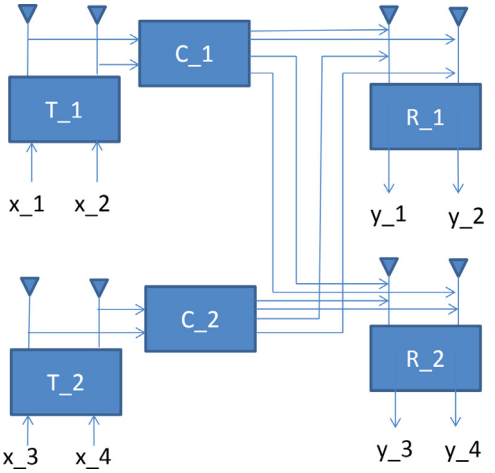

Fig. 14. Illustration of cabling for all-analog cancellation. The conventional transmit-beams and/or receive-beams are difficult to implement completely at the RF frontend. But the all-analog cancellation can be viewed as a special case of the implementation of a high-level notion of beamforming for self-interference cancellation.

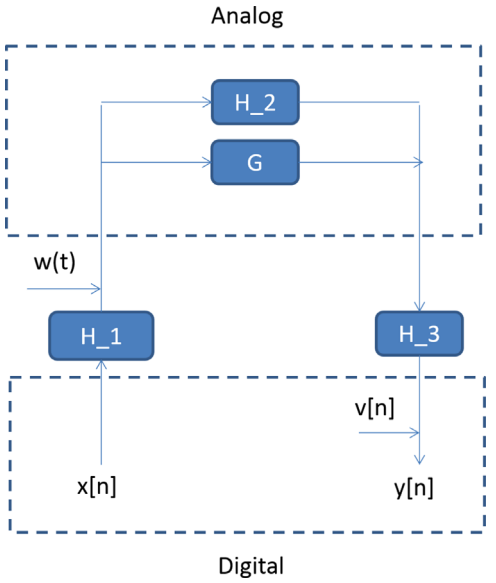

Fig. 15. System configuration of blind digital tuning for self-interference cancellation. One type of the G channel is shown in Fig. 8. The G channel is tuned only based on the knowledge of $x[n]$ and $y[n]$. The block of H_1 denotes the transmit chain (from digital baseband to analog RF) and the block of H_3 denotes the receive chain (from analog RF to digital baseband). An alternative form of this configuration is shown by Fig. 4 in [21].

hardware implementation of the all-analog channel $H_\delta(f)$ will introduce relatively large errors (likely 30 dB SNR). In other words, one never knows an exact $H_{int}(f)$ and never be able to implement an exact $H_\delta(f)$.

In this section, we introduce a blind digital tuning method which exploits a real-valued linearity (instead of a complex-valued linearity) but otherwise assumes virtually nothing about the transfer function of any component in the system. It should be noted that if the nonlinearity of a device is known then it can typically be pre-compensated such that the resulting effective transfer function becomes linear. Unknown nonlinearity would require online estimation. The recent works shown in [29,30] attempted a complex-valued nonlinearity estimation (instead of a real-valued nonlinearity estimation) where the real-valued linearity of IQ imbalances is not exploited.

Consider a system configuration for self-interference cancellation shown in Fig. 15. In this system, H_2 denotes the interference channel, G could be an all-analog interference cancellation channel as shown in Fig. 8, H_1 denotes the transmit chain, and H_3 the receive chain.

All noises from the transmit chain are lumped into $w(t)$, and all noises from the receive chain are lumped into $v[n]$. The source signal is $x[n]$ and the observation is $y[n]$. Note that we use t to denote the continuous-time variable and n the discrete-time variable. For example, $w(t)$ is continuous in time while $x[n]$, $y[n]$ and $v[n]$ are discrete in time. Also note that $x[n]$ and $y[n]$ are baseband digital waveforms.

For a given sequence $x[n]$, $n=0,\ldots,N_x-1$, there is a corresponding (observed) sequence $y[n]$, $n=0,\ldots,N_y-1$ where $N_y \geq N_x$. For each n, $x[n]$ has two components (known as I and Q components) and so does $y[n]$. For robust linearity against IQ imbalances, we will handle $x[n]$ by its real-valued form $\overline{x}[n]$ (i.e., a 2×1 real-valued vector). The same applies to $y[n]$, $v[n]$ and $w(t)$.

Subject to the real-valued representations, we will treat all blocks except for H_1 in the system to be linear.

The (inherently real) gains of the variable attenuators in Fig. 8 are represented by the elements of the real-valued vector $\mathbf{g}=[g_0,\ldots,g_{N_g-1}]^T$ where $0\leq g_i\leq 1$ for $i=0,\ldots,N_g-1$. These gains are the only variables in the cancellation channel and assumed to be controllable accurately.

Define the real-valued vectors: $\overline{\mathbf{x}}=[\overline{x}[0]^T,\ldots,\overline{x}[N_x-1]^T]^T$, $\overline{\mathbf{y}}=[\overline{y}[0]^T,\ldots,\overline{y}[N_y-1]^T]^T$, and $\overline{\mathbf{v}}=[\overline{v}[0]^T,\ldots,\overline{v}[N_y-1]^T]^T$. Also denote $\overline{\mathbf{w}}=\{\overline{w}(t),0\leq t\leq N_yT_s\}$ where T_s is the sampling interval associated with $x[n]$ and $y[n]$. Without loss of practical significance, one can view $\overline{\mathbf{w}}$ as a vector of the samples of $\overline{w}(t)$ with a sampling interval T_L much smaller than T_s.

Then, based on Fig. 15, an important relationship among these vectors can be expressed as

$$\overline{\mathbf{y}}=\overline{\mathbf{y}}_1+\overline{\mathbf{y}}_2 \tag{46}$$

with

$$\overline{\mathbf{y}}_1=\mathbf{T}_{H_3}\mathbf{T}_{\mathbf{g}}\overline{\mathbf{w}}+\mathbf{T}_{H_3}\mathbf{T}_{\mathbf{g}}\mathbf{T}_{H_1}\overline{\mathbf{x}} \tag{47}$$

$$\overline{\mathbf{y}}_2=\overline{\mathbf{v}}+\mathbf{T}_{H_3}\mathbf{T}_{H_2}\overline{\mathbf{w}}+\mathbf{T}_{H_3}\mathbf{T}_{H_2}\mathbf{T}_{H_1}\overline{\mathbf{x}} \tag{48}$$

Here, $\mathbf{T}_{H_i}$, $i=2,3$, are real-valued linear operators corresponding to the channels H_i, $i=2,3$, respectively. (All the following operators are real-valued.) We do not need to know these linear operators. Also, $\mathbf{T}_{\mathbf{g}}$ is a linear operator corresponding to the cancellation channel. Furthermore, we assume that the operator $\mathbf{T}_{\mathbf{g}}$ is a linear function of $\mathbf{g}$. However, we do not need to know anything beyond that.

Furthermore, we can write $\mathbf{T}_{\mathbf{g}}\overline{\mathbf{w}}=\mathbf{T}_{\mathbf{w}}\mathbf{g}$ where the linear operator $\mathbf{T}_{\mathbf{w}}$ is a linear function of $\overline{\mathbf{w}}$. Also, we can write $\mathbf{T}_{\mathbf{g}}\mathbf{T}_{H_1}\overline{\mathbf{x}}=\mathbf{T}_{H_1,\mathbf{x}}\mathbf{g}$ where the linear operator $\mathbf{T}_{H_1,\mathbf{x}}$ is governed by both the H_1-channel and the source vector $\overline{\mathbf{x}}$. And $\mathbf{T}_{H_1,\mathbf{x}}$ is not necessarily a linear function of $\overline{\mathbf{x}}$. Then, we can

rewrite (47) as

$$\overline{\mathbf{y}}_1 = \mathbf{T}_{H_3}\mathbf{T}_{\mathbf{w}}\mathbf{g} + \mathbf{T}_{H_3}\mathbf{T}_{H_1,\mathbf{x}}\mathbf{g} \tag{49}$$

Ideally, we want to find such $\mathbf{g}$ that $\overline{\mathbf{y}} = \overline{\mathbf{v}}$, i.e., the outputs from the H_2-channel and the G-channel cancel each other. We will show next an algorithm that can yield such $\mathbf{g}$ based on the measurements of $\overline{\mathbf{y}}$.

Although $\overline{\mathbf{x}}$ can be chosen to be deterministic, $\overline{\mathbf{y}}$ is random due to the noise vectors $\overline{\mathbf{w}}$ and $\overline{\mathbf{v}}$. To average out the noise, we need to collect a large number of observations of $\overline{\mathbf{y}}$ for a common $\overline{\mathbf{x}}$ and then determine a good estimate of

$$e = E\{\|\overline{\mathbf{y}}\|^2\} \tag{50}$$

where E denotes expectation.

The connection between e and $\mathbf{g}$ can be shown to be

$$e = \frac{1}{2}\mathbf{g}^T\mathbf{A}\mathbf{g} + \mathbf{b}^T\mathbf{g} + c \tag{51}$$

where $\mathbf{A}$, $\mathbf{b}$ and c are constant and independent of $\mathbf{g}$. To show this, we can first write

$$e = E\{\|\overline{\mathbf{y}}_1\|^2\} + 2E\{\overline{\mathbf{y}}_1^T\overline{\mathbf{y}}_2\} + E\{\|\overline{\mathbf{y}}_2\|^2\} \tag{52}$$

where

$$E\{\|\overline{\mathbf{y}}_1\|^2\} = \mathbf{g}^T\left(E\{\mathbf{T}_{\mathbf{w}}^T\mathbf{T}_{H_3}^T\mathbf{T}_{H_3}\mathbf{T}_{\mathbf{w}}\} + \mathbf{T}_{H_1,\mathbf{x}}^T\mathbf{T}_{H_3}^T\mathbf{T}_{H_3}\mathbf{T}_{H_1,\mathbf{x}}\right)\mathbf{g} \tag{53}$$

$$E\{\overline{\mathbf{y}}_1^T\overline{\mathbf{y}}_2\} = \left(E\{\overline{\mathbf{w}}^T\mathbf{T}_{H_2}^T\mathbf{T}_{H_3}^T\mathbf{T}_{H_3}\mathbf{T}_{\mathbf{w}}\} + \overline{\mathbf{x}}^T\mathbf{T}_{H_1}^T\mathbf{T}_{H_2}^T\mathbf{T}_{H_3}^T\mathbf{T}_{H_3}\mathbf{T}_{H_1,\mathbf{x}}\right)\mathbf{g} \tag{54}$$

$$E\{\|\overline{\mathbf{y}}_2\|^2\} = E\{\|\overline{\mathbf{v}}\|^2\} + E\{\|\mathbf{T}_{H_3}\mathbf{T}_{H_2}\overline{\mathbf{w}}\|^2\} + \|\mathbf{T}_{H_3}\mathbf{T}_{H_2}\mathbf{T}_{H_1}\overline{\mathbf{x}}\|^2 \tag{55}$$

Then, the exact forms of $\mathbf{A}$, $\mathbf{b}$ and c follow readily from the above expressions. It is important to note that we do not need to use these structures of $\mathbf{A}$, $\mathbf{b}$ and c. We will only need to use (51).

Based on the model (51), our algorithm to find the optimal $\mathbf{g}$ for self-interference cancellation has two phases. In phase 1 (learning), we estimate $\mathbf{A}$, $\mathbf{b}$ and c by using some training vectors of $\mathbf{g}$. In phase 2 (optimization), we use the estimates of $\mathbf{A}$, $\mathbf{b}$ and c to find the optimal $\mathbf{g}$ which minimizes e in (51) subject to the attenuation constraint (i.e., the gain of each attenuator is between zero and one).

4.1. Learning

To learn $\mathbf{A}$, $\mathbf{b}$ and c of the system, we need to use a training set of $\mathbf{g}$ denoted as $\mathbf{g}_1, \ldots, \mathbf{g}_{M_g}$. Corresponding to this training set, there is the measurement set of e denoted as $e_1, \ldots, e_{M_g}$. For each e, we can rewrite (51) as

$$e = \frac{1}{2}\left[\mathbf{g}^T \otimes \mathbf{g}^T\right]\mathbf{a} + \mathbf{b}^T\mathbf{g} + c \tag{56}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{a} = vec(\mathbf{A})$. Furthermore,

$$e = \mathbf{g}_{big}^T\mathbf{p} \tag{57}$$

with $\mathbf{g}_{big}^T = [1, \mathbf{g}^T, \mathbf{g}^T \otimes \mathbf{g}^T]$ and $\mathbf{p}^T = \left[c, \mathbf{b}^T, \frac{1}{2}\mathbf{a}^T\right]$.

Without loss of generality, we can let $\mathbf{A}$ be symmetric. Hence, alternatively, we can write

$$e = \mathbf{g}_{big,T}^T\mathbf{p}_T \tag{58}$$

with $\mathbf{g}_{big,T}^T = [1, \mathbf{g}^T, (\mathbf{g}^T \otimes \mathbf{g}^T)\mathbf{S}^T]$ and $\mathbf{p}_T^T = \left[\frac{1}{2}\mathbf{a}^T\mathbf{S}^T\mathbf{D}, \mathbf{b}^T, c\right]$. Here, $\mathbf{S}$ is a selection matrix such that $\mathbf{Sa}$ selects the lower triangular elements of $\mathbf{A}$, and $\mathbf{D}$ is a diagonal matrix which scales the selected off-diagonal elements of $\mathbf{A}$ by the factor of two. The purpose of $\mathbf{S}$ and $\mathbf{D}$ is to have the simple expressions for $\mathbf{g}_{big,T}$ and $\mathbf{p}_T$.

For example, if $\mathbf{g}$ has the dimension 4×1, then

$$\mathbf{S} = \begin{bmatrix} \mathbf{J}_{4,4} & & & \\ & \mathbf{J}_{4,3} & & \\ & & \mathbf{J}_{4,2} & \\ & & & \mathbf{J}_{4,1} \end{bmatrix} \tag{59}$$

where $\mathbf{J}_{N,M}$ with $M \le N$ is the last M rows of the $N \times N$ identity matrix, and

$$\mathbf{D} = diag[1\ \ 2\ \ 2\ \ 2\ \mid\ 1\ \ 2\ \ 2\ \mid\ 1\ \ 2\ \mid\ 1]. \tag{60}$$

The blank entries in the above expression of $\mathbf{S}$ are zeros.

Obviously, $\mathbf{A}$, $\mathbf{b}$ and c can be easily found once $\mathbf{p}_T$ is available. Note that $\mathbf{g}$ has the dimension $N_g \times 1$, and $\mathbf{p}_T$ has the dimension $N_p \times 1$ with $N_p = N_g(N_g+1)/2 + N_g + 1$. To find the $\mathbf{p}_T$, we need to choose a set of $M_g \ge N_p$ attenuation training vectors $\mathbf{g}_1, \ldots, \mathbf{g}_{M_g}$ of $\mathbf{g}$ so that we can write the M_g realizations of (51) into

$$\mathbf{e} = \mathbf{G}\mathbf{p}_T \tag{61}$$

where $\mathbf{e} = [e_1, \ldots, e_{M_g}]^T$ and

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_{big,T,1}^T \\ \vdots \\ \mathbf{g}_{big,T,M_g}^T \end{bmatrix} \tag{62}$$

and $\mathbf{g}_{big,T,m}$ is $\mathbf{g}_{big,T}$ obtained from the mth training vector of $\mathbf{g}$.

In order to have a unique solution from (61), a necessary condition for $\mathbf{G}$ is that it has the full column rank N_p. In addition, it is desirable to choose such a $\mathbf{G}$ that is sparse and well conditioned. One such $\mathbf{G}$ follows from the following $M_g = N_p$ training vectors of $\mathbf{g}$:

- Let $\mathbf{g}_1 = 0$.
- For $i = 2, \ldots, N_g + 1$, let $\mathbf{g}_i = \alpha\mathbf{e}_{N_g, i-1}$ where $0 < \alpha \le 1$ and $\mathbf{e}_{N_g, j}$ is the jth column of the $N_g \times N_g$ identity matrix.
- For $i = N_g + 2, \ldots, 2N_g + 1$, let $\mathbf{g}_i = \beta\mathbf{e}_{N_g, i-N_g-1}$ with $0 < \beta < 1$ and $\beta \ne \alpha$.
- For $i = 2N_g + 2, \ldots, M_g = N_p$, let $\mathbf{g}_i = \alpha\mathbf{e}_{N_g,k} + \alpha\mathbf{e}_{N_g,l}$ where $1 \le k < l \le N_g$.

One can verify that the corresponding $N_p \times N_p$ matrix $\mathbf{G}$ has the property $|\det(\mathbf{G})| = \alpha^{N_g}\beta^{N_g}(\alpha-\beta)^{N_g}\alpha^{N_g(N_g-1)}$.

It follows from (61) that $\mathbf{p}_T = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{e}$ if $M_g > N_p$, or $\mathbf{p}_T = \mathbf{G}^{-1}\mathbf{e}$ if $M_g = N_p$. Note that $\mathbf{G}^+$ or $\mathbf{G}^{-1}$ can be precomputed offline for any given set of $\mathbf{g}_1, \ldots, \mathbf{g}_{M_g}$.

To show an example of how to find the solution to (61) efficiently, let us assume $N_g = 3$ and denote a training vector

of $\mathbf{g}$ as $\mathbf{g}=[g_1,g_2,g_3]^T$. Then the structure of a row of $\mathbf{G}$ is

$$[1|g_1,g_2,g_3|g_1^2,g_1g_2,g_1g_3|g_2^2,g_2g_3|g_3^2]$$

Hence, the training vectors described above yield the following 10×10 matrix $\mathbf{G}$:

$$\mathbf{G}=\begin{bmatrix}1&0&0&0&0&0&0&0&0&0\\1&\alpha&0&0&\alpha^2&0&0&0&0&0\\1&\beta&0&0&\beta^2&0&0&0&0&0\\1&0&\alpha&0&0&0&0&\alpha^2&0&0\\1&0&\beta&0&0&0&0&\beta^2&0&0\\1&0&0&\alpha&0&0&0&0&0&\alpha^2\\1&0&0&\beta&0&0&0&0&0&\beta^2\\1&\alpha&\alpha&0&\alpha^2&\alpha^2&0&\alpha^2&0&0\\1&\alpha&0&\alpha&\alpha^2&0&\alpha^2&0&0&\alpha^2\\1&0&\alpha&\alpha&0&0&0&\alpha^2&\alpha^2&\alpha^2\end{bmatrix} \tag{63}$$

Eq. (61) can be written as $\mathbf{Ie}=\mathbf{Gp}_T$ where $\mathbf{I}$ is the identity matrix. To find the inverse of $\mathbf{G}$, we can apply a sequence of elementary row operations identically on both $\mathbf{G}$ and $\mathbf{I}$ until the original $\mathbf{G}$ becomes a permutation matrix $\mathbf{P}$ and consequently the original $\mathbf{I}$ becomes the inverse of the original $\mathbf{G}$ up to the permutation matrix $\mathbf{P}$. By using this approach, one can verify that after a sequence of row operations (i.e., via elimination of elements by row operations [36]), (61) becomes

$$\mathbf{He}=\mathbf{Pp}_T \tag{64}$$

where

$$\mathbf{P}=\begin{bmatrix}1&0&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0&0\\0&0&0&0&1&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&1&0&0\\0&0&0&1&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&1\\0&0&0&0&0&1&0&0&0&0\\0&0&0&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&0&1&0\end{bmatrix} \tag{65}$$

$$\mathbf{H}=\begin{bmatrix}1&0&0&0&0&0&0&0&0&0\\\frac{\Delta_1}{\Delta}&\frac{\beta^2}{\Delta}&\frac{-\alpha^2}{\Delta}&0&0&0&0&0&0&0\\\frac{\Delta_2}{\Delta}&\frac{-\beta}{\Delta}&\frac{\alpha}{\Delta}&0&0&0&0&0&0&0\\\frac{\Delta_1}{\Delta}&0&0&\frac{\beta^2}{\Delta}&\frac{-\alpha^2}{\Delta}&0&0&0&0&0\\\frac{\Delta_2}{\Delta}&0&0&\frac{-\beta}{\Delta}&\frac{\alpha}{\Delta}&0&0&0&0&0\\\frac{\Delta_1}{\Delta}&0&0&0&0&\frac{\beta^2}{\Delta}&\frac{-\alpha^2}{\Delta}&0&0&0\\\frac{\Delta_2}{\Delta}&0&0&0&0&\frac{-\beta}{\Delta}&\frac{\alpha}{\Delta}&0&0&0\\1&-1&0&-1&0&0&0&1&0&0\\1&-1&0&0&0&-1&0&0&1&0\\1&0&0&-1&0&-1&0&0&0&1\end{bmatrix} \tag{66}$$

where $\Delta=\alpha\beta^2-\alpha^2\beta$, $\Delta_1=\alpha^2-\beta^2$, $\Delta_2=\beta-\alpha$. We see that the solution $\mathbf{p}_T$ to (61) can be found very efficiently by using (64), i.e., computing the sparse multiplication $\mathbf{He}$ followed by a permutation.

The above approach applies straightforwardly to any N_g although care is needed to keep track of the positions (or indices) of the nonzero entries in the matrices. There is an obvious pattern of the nonzero entries in $\mathbf{H}$ and $\mathbf{P}$ as shown above, which holds for any N_g. A simple way to verify this pattern is to try another value of N_g and compare the result with the above. We will omit this verification as it is straightforward.

4.2. Optimization

Given $\mathbf{A}$, $\mathbf{b}$ and c, the optimal $\mathbf{g}$ should follow from

$$\min_{0\le g_i\le 1,\ i=0,\ldots,N_g-1}\left\{\tfrac{1}{2}\mathbf{g}^T\mathbf{Ag}+\mathbf{b}^T\mathbf{g}+c\right\} \tag{67}$$

This is a constrained convex optimization problem which can be easily solved by a general optimization software such as CVX. For hardware implementation, a specialized (efficient) algorithm can also be found by following the interior-point optimization approach [37].

5. Simulation of blind digital tuning

We have simulated the digital tuning algorithm shown in Section 4. In this simulation, we model the RF input/output relationship of the H_2 channel as

$$\tilde{y}_2(t)=\sum_{i=0}^{I-1}a_i\tilde{x}_2(t-\tau_i) \tag{68}$$

where $\tilde{x}_2(t)$ is the (real-valued) RF input signal and $\tilde{y}_2(t)$ is the (real-valued) RF output signal. All the attenuation factors a_i, $i=0,1,\ldots,I-1$, are also real. The complex-valued baseband equivalents of $\tilde{x}_2(t)$ and $\tilde{y}_2(t)$ are denoted as $x_2(t)$ and $y_2(t)$, and by definition they satisfy $\tilde{x}_2(t)=\mathrm{Re}\{x_2(t)e^{j2\pi f_c t}\}$ and $\tilde{y}_2(t)=\mathrm{Re}\{y_2(t)e^{j2\pi f_c t}\}$. Then, (68) is equivalent to

$$y_2(t)=\sum_{i=0}^{I-1}a_i x_2(t-\tau_i)e^{-j2\pi f_c\tau_i} \tag{69}$$

The 2×1 real-valued equivalents of $x_2(t)$ and $y_2(t)$ are denoted by $\overline{x}_2(t)$ and $\overline{y}_2(t)$. It follows that (69) is equivalent to

$$\overline{y}_2(t)=\sum_{i=0}^{I-1}a_i\overline{H}_i\overline{x}_2(t-\tau_i) \tag{70}$$

where

$$\overline{H}_i=\begin{bmatrix}\cos(2\pi f_c\tau_i)&\sin(2\pi f_c\tau_i)\\-\sin(2\pi f_c\tau_i)&\cos(2\pi f_c\tau_i)\end{bmatrix} \tag{71}$$

The attenuation factors of the multipaths have the form

$$a_i=\frac{\epsilon\alpha_i}{(d+c\tau_i)^2} \tag{72}$$

where $a_0=\epsilon/d^2$, $c=3\times 10^8$ m/s, $d=0.3$ m, $0\le\tau_i\le 10$ ns (random), $0\le\alpha_i\le 1$ (random) and $\epsilon=8\times 10^{-4}$. Note that some of these parameters are slightly different from those used in Section 3, which however does not have any important consequence.

For the cancellation channel, we choose the one shown in Fig. 9. The RF input/output relationship of the cancellation

channel (or the G-channel) is

$$\tilde{y}_G(t)=\sum_{n=0}^{N-1}\sum_{l=0}^{3} g_{n,l}\tilde{x}_G(t-nT-l\delta) \tag{73}$$

where $\tilde{x}_G(t)$ is the RF input to the G channel and $\tilde{y}_G(t)$ is the RF output of the G channel. Then, its complex-valued baseband equivalent is

$$y_G(t)=\sum_{n=0}^{N-1}\sum_{l=0}^{3} g_{n,l}x_G(t-nT-l\delta)e^{-j2\pi f_c Tn-j2\pi f_c\delta l} \tag{74}$$

and its 2×1 real-valued equivalent is

$$\overline{y}_G(t)=\sum_{n=0}^{N-1}\sum_{l=0}^{3} g_{n,l}\overline{G}_{n,l}\overline{x}_G(t-nT-l\delta) \tag{75}$$

where

$$\overline{G}_{n,l}=\begin{bmatrix}\cos(2\pi f_c Tn+2\pi f_c\delta l) & \sin(2\pi f_c Tn+2\pi f_c\delta l)\\ -\sin(2\pi f_c Tn+2\pi f_c\delta l) & \cos(2\pi f_c Tn+2\pi f_c\delta l)\end{bmatrix} \tag{76}$$

We choose $f_c=2.5$ GHz, $T=1/20W$ (delay between two c-taps), $W=40$ MHz (bandwidth of interest), and $\delta=1/4f_c$.

We also let $x_2(t)=x_G(t)$ and choose the input to H_3 to be $y_2(t)+y_G(t)$. However, we do not assume any direct access to these waveforms.

To simulate the analog channels, we then apply the sampling interval $T_L=T_s/L$ with a large L to approximate (70) and (75) by discrete operations. We let $1/T_s=W$ (the sampling rate of the digital parts). Accordingly, the H_1 channel is modelled by a discrete-time channel $H_{1,D}$ with the rate $1/T_s$ followed by an interpolator of the factor L. Similarly, the H_3 channel is modelled by a discrete-time channel $H_{3,D}$ with the rate $1/T_s$ preceded by a decimator of the factor L. We choose $L=500$. Note that we use the baseband equivalents to model the RF system.

We model $H_{1,D}$ by an FIR lowpass filter (of the double-sided bandwidth W) subject to a transmit IQ imbalances. Specifically, the 2×1 real-valued input/output relationship of $H_{1,D}$ is

$$\overline{y}_1[n]=\sum_{l=-M_h}^{M_h}\overline{H}_{1,D}[l]\overline{x}_1[n-l] \tag{77}$$

with

$$\overline{H}_{1,D}[n]=\begin{bmatrix}h_{t,r}[n] & -h_{t,i}[n]\\ h_{t,i}[n] & h_{t,r}[n]\end{bmatrix}\times\begin{bmatrix}(1+\delta_t)\cos(\phi_t) & (1-\delta_t)\sin(\phi_t)\\ (1+\delta_t)\sin(\phi_t) & (1-\delta_t)\cos(\phi_t)\end{bmatrix} \tag{78}$$

where $h_{t,r}[n]=h_{t,i}[n]=w_h[n]c_h(nT_s)$, $c_h(t)=sinc(Wt)=\sin(\pi Wt)/\pi Wt$ and $w_h[n]=0.54+0.46\cos(2\pi n/(2M_h+1))$. We choose $M_h=20$. Furthermore, $-0.05\le\delta_t\le 0.05$ (random) and $-0.05\le\phi_t\le 0.05$ (random).

We model $H_{3,D}$ by an FIR lowpass filter (same as above) but subject a receive IQ imbalances. Namely, the 2×1 real-valued input/output relationship of $H_{3,D}$ is

$$\overline{y}_3[n]=\sum_{l=-M_h}^{M_h}\overline{H}_{3,D}[l]\overline{x}_3[n-l] \tag{79}$$

with

$$\overline{H}_{3,D}[n]=\begin{bmatrix}(1+\delta_r)\cos(\phi_r) & (1-\delta_r)\sin(\phi_r)\\ (1+\delta_r)\sin(\phi_r) & (1-\delta_r)\cos(\phi_r)\end{bmatrix}\times\begin{bmatrix}h_{r,r}[n] & -h_{r,i}[n]\\ h_{r,i}[n] & h_{r,r}[n]\end{bmatrix} \tag{80}$$

where $h_{r,r}(n)=h_{r,i}(n)=w_h[n]c_h(nT_s)$, and $-0.05\le\delta_r\le 0.05$ (random) and $-0.05\le\phi_r\le 0.05$ (random).

During the learning phase of a fixed H_2, we choose the training vectors $\mathbf{g}_1,\ldots,\mathbf{g}_{N_p}$ as described below (62). For each $\mathbf{g}_i$, we choose $x[n]=\delta[n]$, apply N_r realizations of $w(mT_L)$ and $v[n]$, and compute the corresponding N_r realizations of $y[n]$. Then, we compute

$$e_i=\frac{1}{N_r}\sum_{r=1}^{N_r}\sum_n\|y^{(r,i)}[n]\|^2 \tag{81}$$

where $y^{(r,i)}[n]$ is the rth realization of $y[n]$ corresponding to $\mathbf{g}=\mathbf{g}_i$. With e_i for $i=1,\ldots,N_p$, we compute $\mathbf{p}_T=\mathbf{G}^{-1}\mathbf{e}$. From $\mathbf{p}_T$, we determine $\mathbf{A}$, $\mathbf{b}$ and c.

With the estimates of $\mathbf{A}$, $\mathbf{b}$ and c for a fixed H_2, we follow Section 4.2 to determine the optimal $\mathbf{g}$. With the optimal $\mathbf{g}$, we compute $y[n]$ for a new (single) realization of $w(mT_L)$ and $v[n]$.

We then compute the interference to noise ratio (INR):

$$\mathrm{INR}_{dB}=10\log_{10}\frac{\sum_n\|y[n]\|^2}{\sum_n\|v[n]\|^2} \tag{82}$$

The variance of the receiver noise is a constant chosen to make INR_{dB} to be around 100 dB in the absence of cancellation.

Obviously, INR_{dB} depends on the realization of H_2 as well as the transmission SNR defined as

$$\mathrm{SNR}_{T,dB}=10\log_{10}\frac{\sum_m\|y_1(mT_L)\|^2}{\sum_m\|w(mT_L)\|^2} \tag{83}$$

where $y_1(mT_L)$ is the output of H_1 (i.e., the output of the factor-of-L interpolator following the filter $H_{1,D}$). Random realizations of H_2 are based on (70) with random α_i and τ_i.

To measure the performance, we define γ_{dB} as "INR_{dB} after cancellation" minus "INR_{dB} before cancellation".

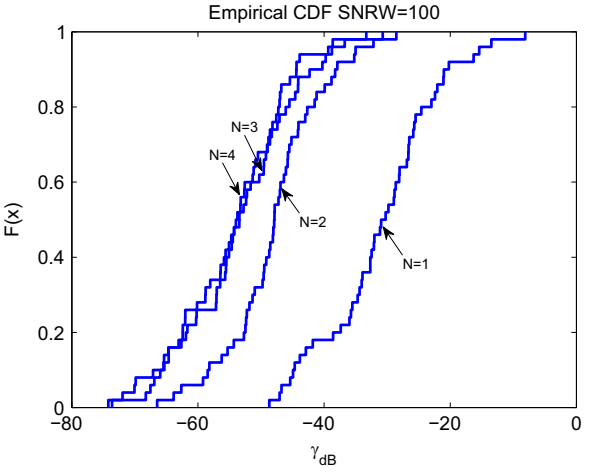

Fig. 16. CDF of γ_{dB} under $\mathrm{SNR}_{T,dB}=100$ dB. It shows the effect of the number of c-taps N.

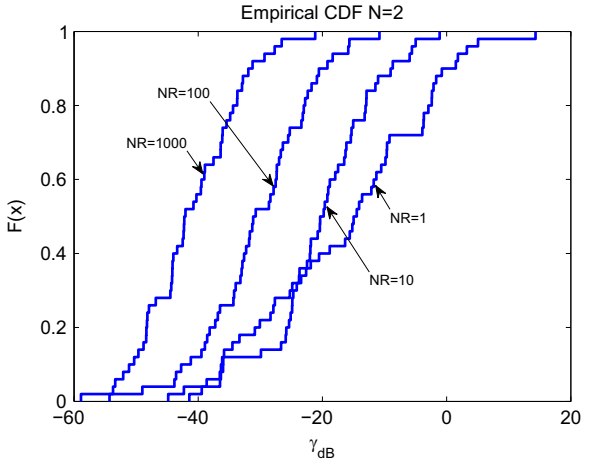

Fig. 17. CDF of γ_{dB} under $\mathrm{SNR}_{T,dB} = 30$ dB. It shows the effect of the number of realizations N_R used for training. The two curves for $N_R = 1$ and $N_R = 10$ crosses at lower part. (N_R is shown as NR in the plot.)

Shown in Fig. 16 is the CDF of γ_{dB} subject to a large $\mathrm{SNR}_{T,dB}$, which illustrates the effect of the number of c-taps N.

Shown in Fig. 17 is the CDF of γ_{dB} subject to a typical value of $\mathrm{SNR}_{T,dB}$, which shows that when N_R (the number of realizations used for training) is large enough, we can break the barrier of the transmission noise. Specifically, we see that while $\mathrm{SNR}_{T,dB} = 30$ dB, the average amount of cancellation is about 40 dB when $N = 2$ and $N_R = 1000$.

From simulation, we also notice that with a larger N, the required value of N_R becomes larger unless $\mathrm{SNR}_{T,dB}$ is high. This is because as N increases, the number of variables (the attenuators) increases which increases their sensitivity to noise. A good tradeoff value for N appears to be two.

6. Conclusions

We have presented several important results on radio self-interference cancellation. We have reviewed a time-domain transmit beamforming (TDTB) method and presented experimental results based on a programmable radio board. We have revised the TDTB method by adopting the real-valued system model so that it becomes more robust against IQ imbalances. We have provided both simulation and hardware experimental results to illustrate the improved robustness.

Transmit beamforming requires additional transmit radio chains for interference cancellation. Its performance is limited by the quality of the transmit radio chains. Unlike transmit beamforming, an all-analog cancellation scheme uses an analog cancellation channel where both its input and output are in the RF frontend. Theoretically, the performance of an all-analog scheme is not as much constrained by the quality of the transmit/receive radio chains as that of transmit beamforming. But the configuration and control of the analog devices in an all-analog cancellation channel are an important issue.

In this paper, we have presented new architectures of all-analog cancellation channels where the only adaptively variable devices are attenuators. The novelty of these architectures is that we use clustered-taps (or called c-taps) of attenuators where each c-tap effectively models a complex-valued tap in an equivalent baseband representation. Using multiple c-taps ensures a wide coverage of possible interference channels without the need to change the delays. Assuming ideal attenuators, we have presented the statistical performance (cancellation) limits of such an architecture with varying numbers of c-taps. We have also considered the effect of ideal step attenuators.

In addition, we have presented a blind digital tuning method for determining the attenuations of the attenuators in an all-analog cancellation channel without any direct access to the RF frontend signals. This method does not require the knowledge of the transfer function of any component in the system. The desired attenuations are determined through training and optimization all based on the baseband output of the receive radio chain. Simulation shows that such a scheme can achieve an amount of cancellation larger than the transmission SNR of the radio chain. All of the contributions shown in this paper can be viewed as specific ways for implementing a high-level notion called transmit and/or receive beamforming for radio self-interference cancellation.

Mostly inspired and challenged by hardware impairments, the theoretical foundation for radio self-interference cancellation differs from those for the more conventional problems of interference cancellation such as in [32,39]. Research with only hardware experiments can be too constraining and less productive in scientific understanding. We have taken a hybrid approach where we have attempted to run hardware based experiment as much as we can afford and also developed theoretical insights and ideas which are useful in guiding future implementations. Extraction of theoretical problems from radio self-interference cancellation may not always be useful to achieve its original goal (due to imprecision of modelling) but such an intellectual exercise may well lead to useful ideas for future efforts. As a by-product, the blind digital tuning method presented in this paper differs from and complements well the existing theories of blind system identification such as in [33,34].

Appendix A. An algorithm to solve (42)

In this section, an efficient algorithm to solve the integer optimization problem (42) assuming no knowledge of the structure of $f(m_a, m_b)$ other than the properties (43)–(45) is presented. The optimal solutions for m_a and m_b are denoted by a_{opt} and b_{opt}.

(1) Let $\overline{a}_{low} = 0$ and $\overline{b}_{low} = m_{max}$. Choose a small number ξ as a tolerance in treating two real numbers as equal. Set $marker = 0$.
(2) If $f(\overline{a}_{low}, \overline{b}_{low}) \le D_g$, set $a_{opt} = \overline{a}_{low}$ and $b_{opt} = \overline{b}_{low}$ and stop.
(3) Entry A:
(4) Set $\overline{b}_{low} = \overline{b}_{low} - 1$.
(5) If $\overline{b}_{low} = 0$, set $\overline{b}_{low} = 1$ and go to Turning Point.
(6) Set $\overline{f}_{low} = f(\overline{a}_{low}, \overline{b}_{low})$.

(7) If $|\overline{f}_{low}-D_g|<\xi$, set $a_{opt}=\overline{a}_{low}$ and $b_{opt}=\overline{b}_{low}$ and stop.
(8) If $\overline{f}_{low}>D_g$, go to Entry A.
(9) Set $a_{low}=\overline{a}_{low}$, $b_{low}=\overline{b}_{low}$, $a_{up}=a_{low}$, $b_{up}=b_{low}+1$, $f_{up}=f(a_{up},b_{up})$, $f_{low}=f(a_{low},b_{low})$ and go to Case A.
(10) **Turning Point** (from vertical downward initial search to diagonal upper right initial search):
(11) Entry B:
(12) Set $\overline{a}_{low}=\overline{a}_{low}+1$ and $\overline{b}_{low}=\overline{b}_{low}+1$.
(13) If $\overline{b}_{low}=m_{max}+1$, set $a_{up}=\overline{a}_{low}-1$, $b_{up}=\overline{a}_{low}$, $a_{low}=m_{max}$, $b_{low}=m_{max}$, $f_{up}=f(a_{up},b_{up})$, $f_{low}=0$ and go to Entry E.
(14) Set $\overline{f}_{low}=f(\overline{a}_{low},\overline{b}_{low})$.
(15) If $|\overline{f}_{low}-D_g|<\xi$, set $a_{opt}=\overline{a}_{low}$ and $b_{opt}=\overline{b}_{low}$ and stop.
(16) If $\overline{f}_{low}>D_g$, go to Entry B.
(17) Set $a_{low}=\overline{a}_{low}$, $b_{low}=\overline{b}_{low}$, $a_{up}=a_{low}-1$, $b_{up}=b_{low}-1$, $f_{up}=f(a_{up},b_{up})$, $f_{low}=f(a_{low},b_{low})$ and go to Case B.
(18) **Case A** (The upper bound point is above the lower bound point)
(19) Entry C:
(20) Set $\overline{a}_{up}=a_{up}+1$ and $\overline{b}_{up}=b_{up}+1$.
(21) Entry C2:
(22) If $\overline{b}_{up}=m_{max}+1$ and $\overline{b}_{up}-\overline{a}_{up}=1+b_{low}-a_{low}$, go to Entry E.
(23) If $\overline{b}_{up}=m_{max}+1$ and $\overline{b}_{up}-\overline{a}_{up}>1+b_{low}-a_{low}$, set $\overline{b}_{up}=\overline{b}_{up}-1$ and *marker*=1.
(24) Set $\overline{f}_{up}=f(\overline{a}_{up},\overline{b}_{up})$.
(25) If $\overline{f}_{up}\geq f_{up}$, set $\overline{a}_{up}=\overline{a}_{up}+1$ and $\overline{b}_{up}=\overline{b}_{up}+1$ and go to Entry C2.
(26) If $\overline{f}_{up}\leq f_{low}$, go to Entry C3
(27) If $|\overline{f}_{up}-D_g|<\xi$, set $a_{opt}=\overline{a}_{up}$ and $b_{opt}=\overline{b}_{up}$ and stop.
(28) Set *marker*=0.
(29) If $\overline{f}_{up}<D_g$, set $a_{low}=\overline{a}_{up}$, $b_{low}=\overline{b}_{up}$, $f_{low}=f(a_{low},b_{low})$ and go to Case B.
(30) Set $a_{up}=\overline{a}_{up}$, $b_{up}=\overline{b}_{up}$, $f_{up}=f(a_{up},b_{up})$ and go to Entry C.
(31) Entry C3:
(32) If *marker*=1, go to Entry E.
(33) Set $\overline{b}_{up}=\overline{b}_{up}+1$, go to Entry C2.
(34) **Case B** (The upper bound point is below the lower bound point):
(35) Set $\overline{a}_{low}=a_{low}$ and $\overline{b}_{low}=b_{low}$.
(36) Entry D:
(37) Set $\overline{b}_{low}=\overline{b}_{low}+1$.
(38) If $\overline{b}_{low}=m_{max}+1$, go to Entry E.
(39) Entry D2:
(40) $\overline{f}_{low}=f(\overline{a}_{low},\overline{b}_{low})$.
(41) If $\overline{f}_{low}<f_{low}$, go to Entry D.
(42) If $\overline{f}_{low}\geq f_{up}$, go to Entry D3.
(43) If $|\overline{f}_{low}-D_g|<\xi$, set $a_{opt}=\overline{a}_{low}$ and $b_{opt}=\overline{b}_{low}$ and stop.
(44) If $\overline{f}_{low}<D_g$, set $a_{low}=\overline{a}_{low}$, $b_{low}=\overline{b}_{low}$, $f_{low}=f(a_{low},b_{low})$ and go to Case B.
(45) Set $a_{up}=\overline{a}_{low}$ and $b_{up}=\overline{b}_{low}$, $f_{up}=f(a_{up},b_{up})$ and go to Case A.
(46) Entry D3:
(47) Set $\overline{a}_{low}=\overline{a}_{low}+1$ and $\overline{b}_{low}=\overline{b}_{low}+1$.
(48) If $\overline{b}_{low}=m_{max}+1$ and $\overline{b}_{low}-\overline{a}_{low}=1+b_{low}-a_{low}$, go to Entry E.
(49) If $\overline{b}_{low}=m_{max}+1$ and $\overline{b}_{low}-\overline{a}_{low}>1+b_{low}-a_{low}$, set $\overline{b}_{low}=\overline{b}_{low}-1$.
(50) Go to Entry D2.
(51) Entry E (finalizing the results):
(52) If $|f_{up}-D_g|\leq|f_{low}-D_g|$, set $a_{opt}=a_{up}$ and $b_{opt}=b_{up}$ and stop.
(53) Set $a_{opt}=a_{low}$ and $b_{opt}=b_{low}$ and stop.

Appendix B. Linearity under real-valued model

In this appendix, we use a simple example to illustrate the fact that using 2×1 real-valued waveforms preserves the linearity in the presence of IQ imbalances.

Consider a radio transmitter with IQ imbalances. Its RF output can be written as [38]

$$\tilde{y}(t)=x_r(t)(1+\delta)\cos(2\pi f_c t+\phi)-x_i(t)(1-\delta)\sin(2\pi f_c t-\phi)$$

where f_c is the carrier frequency, δ and ϕ are the amplitude and phase imbalances, respectively. This RF output is governed by the complex-valued baseband waveform $x(t)\doteq x_r(t)+jx_i(t)\doteq\mathrm{Re}\{x(t)\}+j\,\mathrm{Im}\{x(t)\}$.

The relationship between the RF waveform $\tilde{y}(t)$ and its complex-valued baseband equivalent $y(t)$ is defined as $\tilde{y}(t)=\mathrm{Re}\{y(t)e^{j2\pi f_c t}\}$. Hence,

$$y(t)=x_r(t)(1+\delta)e^{j\phi}+jx_i(t)(1-\delta)e^{-j\phi}. \tag{84}$$

In terms of the 1×1 complex-valued waveforms $x(t)$ and $y(t)$, there is no linearity between them unless both δ and ϕ are zero.

However, if we replace each complex-valued baseband waveform by its corresponding 2×1 real-valued vector, i.e., $x(t)$ represented by $\overline{x}(t)=\begin{bmatrix}x_r(t)\\x_i(t)\end{bmatrix}$ and $y(t)$ represented by $\overline{y}(t)=\begin{bmatrix}y_r(t)\\y_i(t)\end{bmatrix}$, then it follows that

$$\overline{y}(t)=\overline{C}(\delta,\phi)\overline{x}(t) \tag{85}$$

where

$$\overline{C}(\delta,\phi)=\begin{bmatrix}(1+\delta)\cos(\phi) & (1-\delta)\sin(\phi)\\(1+\delta)\sin(\phi) & (1-\delta)\cos(\phi)\end{bmatrix}.$$

Clearly, there is a perfect linearity between the 2×1 real-valued waveforms $\overline{x}(t)$ and $\overline{y}(t)$ even in the presence of IQ imbalances.

On the other hand, a real-valued linear system model can be always used to represent a complex-valued linear system model. For example, consider a discrete-time complex-valued convolution $y[n]=h[n]*x[n]=\sum_{l=0}^{L}h[l]x[n-l]$, which can be equivalently written as $\overline{y}[n]=\overline{H}(h[n])*\overline{x}[n]$ with

$$\overline{H}(h[n])=\begin{bmatrix}\mathrm{Re}\{h[n]\} & -\mathrm{Im}\{h[n]\}\\\mathrm{Im}\{h[n]\} & \mathrm{Re}\{h[n]\}\end{bmatrix}. \tag{86}$$

Appendix C. Widely linearity

The real-valued linear vector equation (85) can be rewritten as the following complex-valued scalar equation:

$$y(t)=ax(t)+bx^*(t) \tag{87}$$

where $a \doteq a_r + ja_i$ and $b \doteq b_r + jb_i$ are complex numbers and

$$\begin{bmatrix} a_r \\ b_r \\ a_i \\ b_i \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \\ 1 & -1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} (1+\delta)\cos(\phi) \\ (1+\delta)\sin(\phi) \\ (1-\delta)\sin(\phi) \\ (1-\delta)\cos(\phi) \end{bmatrix} = \begin{bmatrix} \cos(\phi) \\ \delta\cos(\phi) \\ \delta\sin(\phi) \\ \sin(\phi) \end{bmatrix} \tag{88}$$

The above equation also suggests that even if the matrix $C(\delta, \phi)$ in (85) is replaced by an arbitrary 2×2 real matrix, Eq. (87) also holds. Conversely, for every given (87), there is a corresponding (85). An equation like (87) is a so-called widely linear equation [27].

Following the same procedure above, one can verify that the real-valued linear equation (6) can be rewritten into the following complex-valued widely linear equation:

$$\mathbf{y}[n] = \mathbf{A}[n] * \mathbf{x}[n] + \mathbf{B}[n] * \mathbf{x}^*[n] + \mathbf{w}[n] \tag{89}$$

where $\mathbf{A}[n]$ and $\mathbf{B}[n]$ for each n are $n_r \times n_t$ complex matrices corresponding to the $2n_r \times 2n_t$ real matrix $\overline{\mathbf{H}}[n]$ in (6).

However, in order to determine all independent transmit beamforming prefilters, using the widely linear model (89) such that $\mathbf{A}[n] * \mathbf{x}[n] + \mathbf{B}[n] * \mathbf{x}^*[n] = 0$ is not as straightforward as using the real-valued linear model (6) such that $\overline{\mathbf{H}}[n] * \overline{\mathbf{x}}[n] = 0$. The beamforming solution to the latter is simply given by (9)–(12).

Let us next discuss how to derive the complex-valued beamformer of the complex-valued widely linear model from the real-valued beamformer of the real-valued linear model.

Consider the complex matrix equation

$$\mathbf{A}[n] * \mathbf{x}[n] + \mathbf{B}[n] * \mathbf{x}^*[n] = 0 \tag{90}$$

with $(\mathbf{A}[n])_{i,j} = a_{i,j}[n]$, $(\mathbf{B}[n])_{i,j} = b_{i,j}[n]$ and $(\mathbf{x}[n])_j = x_j[n]$. We also let $\mathbf{A}[n] \in C^{p \times q}$, $\mathbf{B}[n] \in C^{p \times q}$ and $\mathbf{x}[n] \in C^{q \times 1}$. The real-valued equivalent of this equation can be written as

$$\overline{\mathbf{C}}_1[n] * \overline{\mathbf{x}}_1[n] = 0 \tag{91}$$

or

$$\overline{\mathbf{C}}_2[n] * \overline{\mathbf{x}}_2[n] = 0 \tag{92}$$

where

$$\mathbf{C}_1[n] = \begin{bmatrix} \mathrm{Re}\{\mathbf{A}[n]\} + \mathrm{Re}\{\mathbf{B}[n]\} & -\mathrm{Im}\{\mathbf{A}[n]\} + \mathrm{Im}\{\mathbf{B}[n]\} \\ \mathrm{Im}\{\mathbf{A}[n]\} + \mathrm{Im}\{\mathbf{B}[n]\} & \mathrm{Re}\{\mathbf{A}[n]\} - \mathrm{Re}\{\mathbf{B}[n]\} \end{bmatrix} \in R^{2p \times 2q} \tag{93}$$

$$\overline{\mathbf{x}}_1[n] = \begin{bmatrix} \mathrm{Re}\{\mathbf{x}[n]\} \\ \mathrm{Im}\{\mathbf{x}[n]\} \end{bmatrix} \in R^{2q \times 1} \tag{94}$$

$$(\overline{\mathbf{C}}_2[n])_{(i,j)-block} = \begin{bmatrix} \mathrm{Re}\{a_{i,j}[n]\} + \mathrm{Re}\{b_{i,j}[n]\} & -\mathrm{Im}\{a_{i,j}[n]\} + \mathrm{Im}\{b_{i,j}[n]\} \\ \mathrm{Im}\{a_{i,j}[n]\} + \mathrm{Im}\{b_{i,j}[n]\} & \mathrm{Re}\{a_{i,j}[n]\} - \mathrm{Re}\{b_{i,j}[n]\} \end{bmatrix} \in R^{2 \times 2} \tag{95}$$

$$(\overline{\mathbf{x}}_2[n])_{j-block} = \begin{bmatrix} \mathrm{Re}\{x_j[n]\} \\ \mathrm{Im}\{x_j[n]\} \end{bmatrix} \in R^{2 \times 1} \tag{96}$$

It is easy to verify that there are permutation matrices $\mathbf{P}_{2p \times 2p}$ and $\mathbf{P}_{2q \times 2q}$ such that

$$\overline{\mathbf{C}}_1[n] = \mathbf{P}_{2p \times 2p} \overline{\mathbf{C}}_2[n] \mathbf{P}^T_{2q \times 2q} \tag{97}$$

$$\overline{\mathbf{x}}_1[n] = \mathbf{P}_{2q \times 2q} \overline{\mathbf{x}}_2[n] \tag{98}$$

Now we let $\overline{\mathbf{D}}_2[n] \in R^{2q \times t}$ be a matrix containing t real-valued beam-vectors such that

$$\overline{\mathbf{C}}_2[n] * \overline{\mathbf{D}}_2[n] = 0 \tag{99}$$

the process of which is similar to (9)–(12). It follows that $\overline{\mathbf{D}}_1[n] \doteq \mathbf{P}_{2q \times 2q} \overline{\mathbf{D}}_2[n]$ satisfies

$$\overline{\mathbf{C}}_1[n] * \overline{\mathbf{D}}_1[n] = 0 \tag{100}$$

Furthermore, let us perform the partition

$$\overline{\mathbf{D}}_1[n] = \begin{bmatrix} \overline{\mathbf{D}}_{upper}[n] \\ \overline{\mathbf{D}}_{lower}[n] \end{bmatrix} \in R^{2q \times t} \tag{101}$$

with $\overline{\mathbf{D}}_{upper}[n] \in R^{q \times t}$ and $\overline{\mathbf{D}}_{lower}[n] \in R^{q \times t}$. Then such a matrix $\mathbf{D}[n] \doteq \overline{\mathbf{D}}_{upper}[n] + j\overline{\mathbf{D}}_{lower}[n] \in C^{q \times t}$ satisfies

$$\mathbf{A}[n] * \mathbf{D}[n] + \mathbf{B}[n] * \mathbf{D}^*[n] = 0 \tag{102}$$

Hence, the matrix $\mathbf{D}[n]$ contains the t complex-valued beam-vectors corresponding to the t real-valued beam-vectors in $\overline{\mathbf{D}}_2[n]$.

References

[1] N. Verhoeckx, H. van den Elzen, F. Snijders, P. van Gerwen, Digital echo cancellation for baseband data transmission, IEEE Trans. Acoust. Speech Signal Process. 27 (1979) 768–781.

[2] S. Chen, M.A. Beach, J.P. McGeehan, Division-free duplex for wireless applications, Electron. Lett. 34 (2) (1998).

[3] H. Hamazumi, K. Imamura, N. Iai, K. Shibuya, M. Sasaki, A study of a loop interference canceller for the relay stations in an SFN for digital terrestrial broadcasting, in: Proceedings of 2000 IEEE Global Communications Conference, 2000, pp. 167–171.

[4] H. Sakai, T. Oka, K. Hayashi, Simple adaptive filter method for cancellation of coupling wave in OFDM signals at SFN relay station, in: Proceedings of EUSIPCO 2006, September 2006.

[5] K.M. Nasr, J.P. Cosmas, M. Bard, J. Gledhill, Performance of an echo canceller and channel estimator for on-channel repeaters in DVBT/H networks, IEEE Trans. Broadcast. 53 (September) (2007) 609–618.

[6] D.W. Bliss, P.A. Parker, A.R. Margetts, Simultaneous transmission and reception for improved wireless network performance, in: Proceedings of IEEE 14th Workshop on Statistical Signal Processing, August 2007, pp. 26–29.

[7] H. Ju, E. Oh, D. Hong, Improving efficiency of resource usage in two-hop full-duplex relay systems based on resource sharing and interference cancellation, IEEE Trans. Wirel. Commun. 8 (August) (2009) 3933–3938.

[8] B. Chun, E. Jeong, J. Joung, Y. Oh, Y.H. Lee, Pre-nulling for self-interference suppression in full-duplex relays, in: Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit Conference, Sapporo, October 2009.

[9] Y. Hua, An overview of beamforming and power allocation for MIMO relays, in: Proceedings of IEEE Military Communications Conference, San Jose, CA, November 2010, pp. 375–380.

[10] P. Lioliou, M. Viberg, M. Coldrey, F. Athley, Self-interference suppression in full-duplex MIMO relays, in: Proceedings of Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, October 2010, pp. 658–662.

[11] T. Riihonen, S. Werner, R. Wichman, Residual self-interference in full-duplex MIMO relays after null-space projection and cancelation, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, November 2010.

[12] M. Duarte, A. Sabharwal, Full-duplex wireless communications using off-the shelf radios: feasibility and first results, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2010.

[13] A. Sahai, G. Patel, A. Sabharwal, Pushing the limits of full-duplex: design and real-time implementation, Online at arXiv:1107.0607.

[14] M. Duarte, C. Dick, A. Sabharwal, Experiment-driven characterization of full-duplex wireless systems, IEEE Trans. Wirel. Commun. 11 (December (12)) (2012) 4296–4307.

[15] T. Riihonen, S. Werner, R. Wichman, Transmit power optimization for multiantenna decode-and-forward relays with loopback

self-interference from full-duplex operation, in: Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2011, pp. 1408–1412.
[16] T. Riihonen, S. Werner, R. Wichman, Mitigation of loopback self-interference in full-duplex MIMO relays, IEEE Trans. Signal Process. 59 (December (12)) (2011) 5983–5993.
[17] D. Senaratne, C. Tellambura, Beamforming for space division duplexing, in: Proceedings of IEEE International Conference on Communications, Kyoto, Japan, June 2011, pp. 1–5.
[18] B. Chun, H. Park, A spatial-domain joint-nulling method of self-interference in full-duplex relays, IEEE Commun. Lett. 16 (April (4)) (2012) 436–438.
[19] Y. Hua, P. Liang, Y. Ma, A. Cirik, Q. Gao, A method for broadband full-duplex radio, IEEE Signal Process. Lett. (December) .
[20] S.H. Li, R.D. Murch, Full-duplex wireless communication using transmitter output based echo cancellation, in: Proceedings of 2011 IEEE Global Communications Conference, 2011, pp. 1–5.
[21] Y. Hua, Y. Ma, P. Liang, A. Cirik, Breaking the barrier of transmission noise in full-duplex radio, in: MILCOM 2013, San Diego, CA, November 2013.
[22] J.I. Choi, M. Jain, K. Srinivasan, P. Levis, S. Katti, Achieving single channel, full duplex wireless communication, in: Proceedings of MobiCom 2010, New York, NY, USA, 2010.
[23] S. Hong, J. Mehlman, S. Katti, Picasso: Flexible RF and Spectrum Slicing, in: Proc. Sigcomm 2012.
[24] D. Bharadia, E. McMilin, S. Katti, Full duplex radios, in: Sigcomm, 2013.
[25] J.G. McMichael, K.E. Kolodziej, Optimal tuning of analog self-interference cancellers for full-duplex wireless communication, in: Proceedings of Allerton Conference, 2012.
[26] Y.-S. Choi, H. Shirani-Mehr, Simultaneous transmission and reception: algorithm, design and system level performance, IEEE Trans. Wirel. Commun. (2014), to appear (posted at arXiv:1309.5546).
[27] B. Picinbono, P. Chevalier, Widely linear estimation with complex data, IEEE Trans. Signal Process. 43 (August (8)) (1995).
[28] D. Korpi, L. Anttila, V. Syrjala, M. Valkama, Widely-linear digital self-interference cancellation in direct-conversion full-duplex transceiver, arxiv-web3.library.cornell.edu/pdf/1402.6083v2.
[29] L. Anttila, D. Korpi, V. Syrjala, M. Valkama, Cancellation of power amplifier induced nonlinear self-interference in full-duplex transceivers, in: Asilomar Conference on Signals, Systems and Computers, November 2013.
[30] E. Ahmed, A.M. Eltawil, A. Sabharwal, Self-interference cancellation with nonlinear distortion suppression for full-duplex systems, in: Asilomar Conference on Signals, Systems and Computers, November 2013.
[31] A. Gholian, Y. Ma, Y. Hua, A numerical investigation of all-analog radio self-interference cancellation, in: IEEE Workshop on SPAWC, Toronto, Canada, June 2014.
[32] I. Cha, S.A. Kassam, Interference cancellation using radial basis function networks, Signal Process. 47 (December (3)) (1995) 247–268.
[33] K.A. Meraim, W. Qiu, Y. Hua, Blind system identification, Proc. IEEE 85 (8) (1997).
[34] X.-F. Xu, D.-Z. Feng, W.X. Zheng, H. Zhang, Convolutive blind source separation based on joint block Toeplitzation and block-inner diagonalization, Signal Process. 90 (1) (2010) 119–133.
[35] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge Press, 1985.
[36] G.H. Golub, C.F. Van Loan, Matrix Computations, John Hopkins University Press, 1996.
[37] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[38] D.M. Pozar, Microwave Engineering, 3rd ed. Wiley, 2005.
[39] C. Barnbaum, R.F. Bradley, A new approach to interference excision in radio astronomy: real-time adaptive cancellation, Astron. J. 155 (1998) 2598–2614.