Breaking the Barrier of Transmission Noise in Full-Duplex Radio

Yingbo Hua, Yiming Ma, Ping Liang, and Ali Cirik
Department of Electrical Engineering, University of California, Riverside, CA 92521
yhua@ee.ucr.edu, yma@ee.ucr.edu, liang@ee.ucr.edu, aciri002@student.ucr.edu

***Abstract*— The key technical challenge in making a full-duplex radio is self-interference cancelation (SIC). The self-interference received by a full-duplex radio has two major components: one corresponds to an information-carrying waveform meant for a remote radio and the other corresponds to the noise generated from the transmit chain of the full-duplex radio. The transmission signal-to-noise ratio (SNR) of a typical radio is only about 30dB. This severely limits the performance of any SIC methods which ignore the transmission noise. None of the previously known digital or hybrid methods for SIC has any built-in mechanism to handle the transmission noise. In this paper, we present a new analog-digital hybrid method whose performance is no longer limited by the transmission noise. This method also involves a blind system identification and equalization algorithm for finding the optimal parameters of the cancelation filter.**

***Index Terms*— interference cancellation, full-duplex radio**

I. Introduction

As the radio spectrum becomes more crowded than ever, the need for spectrally efficient radio technologies increases. One such technology is known as full-duplex radio which can transmit and receive at the same time and same frequency. While the full-duplex radio technology appears on the verge to be proven feasible for many commercial and military applications, a key technical challenge still remains which is self-interference cancelation (SIC). A full-duplex radio must have at least one radio transmit chain and one radio receive chain. The signal emitted out of the transmit chain is also picked up by the receive chain, which is the self-interference.

The self-interference can be first reduced by increasing the attenuation between the transmit chain and the receive chain. This can be achieved by using various antenna technologies (and even possibly using some radio blocker/absorber in between transmit and receive antennas in some situations). This approach is also called passive cancelation in the literature [1].

The remaining self-interference has to be (actively) canceled by one or more SIC methods. For any SIC method, a cancelation waveform must be first generated based on a source signal from the transmit chain and then used for cancelation somewhere in the receive chain. Given the RF nature of the interference, it is natural to think of an analog cancelation path between the transmitter and the receiver at the RF frontend. This is exactly what was proposed in [2] and [3] where a tunable analog circuit is used for interference cancelation. We will refer to these analog methods as all-analog where the cancelation path has analog input interface, analog filter and analog output interface. See Fig. 1. An advantage of all-analog passive circuits is that virtually no noise is introduced and the remaining interference can be further canceled at a later stage.

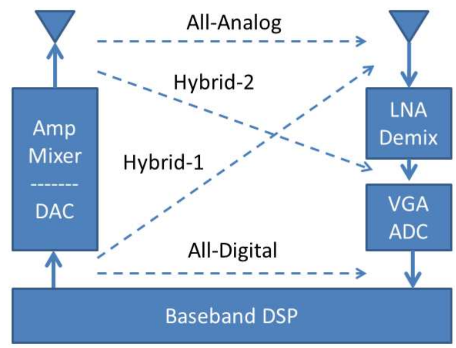

Fig. 1. Four approaches for SIC critical for full-duplex radio. The dash lines denote the cancelation paths. The all-analog path denotes a path with analog input interface, analog filter and analog output interface. The all-digital path denotes a path with digital input interface, digital filter and digital output interface. The hybrid-1 path denotes a path with digital input interface, digital filter and analog output interface. The hybrid-2 path denotes a path with analog input interface, digital filter and analog output interface. The two antennas can be merged into one by using a circulator [8]. The concepts shown in this paper can also be applied to MIMO full-duplex radio.

An alternative to all-analog is all-digital. There are well established theories for adaptive filters [4] that can be readily implemented in baseband DSP circuits. An all-digital cancelation path has digital input interface, digital filter and digital output interface. But this method works only if the interference (or residue interference after an initial cancelation) is not much stronger than the desired signal from a remote radio or otherwise the desired signal suffers from a large quantization noise. Furthermore, this method also suffers from the transmission noise. The interference caused by the noise originated from the transmit chain cannot be regenerated in the baseband for cancelation.

The alternative to all-analog and all-digital is hybrid. In order to preserve the desired (weak) signal in the receive chain, the strong interference should be canceled at the RF frontend of the receiver. For this purpose, several authors have proposed various forms of transmit beamforming based methods (see for example [5], [6], [7], [8]) where the transmitters are prefiltered such that the waveform from a primary transmit chain and the waveform from a secondary (cancelation) chain cancels each other at the receiver's RF frontend. We view this group of methods as hybrid-1 as in Fig. 1.The cancelation path is

This work was supported in part by DoD, ARL and UCOP.

978-0-7695-5124-1/13 $31.00 © 2013 IEEE
DOI 10.1109/MILCOM.2013.263

Fig. 2. A programmable radio board with 2.4 GHz carrier where the FPGA firmware and software have been programmed (substantially changed from original WARP http://warp.rice.edu/trac/wiki/cite) for real-time SIC using the time-domain transmit beamforming method [8], which differs from the programming at the data packet level shown in [5]. The latter is a frequency-domain method and has a prefix problem as discussed in [8].

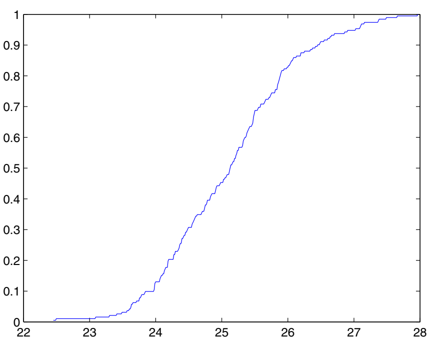

Fig. 3. A cumulative distribution function of the amount (dB) of interference cancelation using the radio shown in Fig. 2. The source waveforms are Hamming-windowed sinc functions of 15MHz bandwidth. The average amount of the cancelation is slightly above 25dB. With about 80% probability, the cancelation is within 24-27dB.

driven by a digital source waveform and also filtered digitally, but however the output of the cancelation path cancels the interference in an analog fashion at the RF frontend of the receiver. Compared to all-digital, the hybrid-1 reduces the burden of potential saturation of the receiver's frontend. But it still suffers from the transmission noise as for all-digital.

In fact, we have recently implemented the method shown in [8] on a programmable radio board. See Fig. 2. This radio has a lower transmission SNR (higher transmit chain noise figure) than the Agilent vector generators which we used for the data reported in [8]. On this radio, we have achieved 25dB interference cancelation (see Fig. 3.) of waveforms with 15 MHz bandwidth although 50dB cancelation was previously achieved on the Agilent equipment.

In this paper, we present a new hybrid approach labeled as hybrid-2 in Fig. 1. In this approach, the cancelation path taps a source waveform directly from the RF end of the transmit chain, converts it into baseband for digital filtering and then yields a baseband analog cancelation waveform for cancelation just before VGA (variable gain amplifier) in the receive chain. In other words, this cancelation path has analog input interface, digital filter and analog output interface. The hybrid-2 is robust to the transmission noise as discussed later in detail. The hybrid-2 can also be integrated with the hybrid-1 for improved benefits. (Shortly before submitting this paper, we realized that our blind system algorithm shown in sections V and VI for computing the parameters of the digital filter in the hybrid-2 cancelation path can also be used for computing the parameters of the analog filter in the all-analog cancelation path shown in [3], which we will address in detail in a future paper.)

In order for a full-duplex radio to be used in an LTE cellular network, for example, the required total interference suppression is about 160dB. With the best possible passive cancelation [1], we need to achieve about 90dB active cancelation. With the state-of-art all-analog method [3], we still have about 40dB interference left. By breaking the performance barrier caused by 30dB transmission SNR of a typical radio transmitter, the realization of a full-duplex radio being used for real world applications should not be far in the future.

In section II, we highlight the impact of the transmission noise on a hybrid-1 method. In section III, we present the basic configuration of the new (hybrid-2) method. In section IV, we show how a hybrid-2 method can be integrated with a hybrid-1 method. After that, in sections V and VI, we outline some of the fundamental results necessary for implementing the hybrid-2 method. At the conference, we will present the latest simulation and experimental findings.

II. The Barrier of Transmission Noise

Let us consider a hybrid-1 method for a full-duplex radio with one primary transmit chain and one secondary (cancelation) transmit chain. Let SNR_T be the SNR of the transmitted signal from each transmit chain. Before cancelation, the receiver receives $g_1\sqrt{P_T}x_T(t) + g_1\sqrt{\frac{P_T}{SNR_T}}n_T(t) + g_1\sqrt{P_R}x_R(t) + n_R(t)$ where g_1 is the receiver gain (normally affected by both LNA and VGA), $x_T(t)$ is the normalized (unit variance) interfering waveform, $n_T(t)$ is the normalized transmission noise, $x_R(t)$ is the normalized desired signal from another node, $n_R(t)$ is the normalized receiver noise, and P_T and P_R are the powers of $\sqrt{P_T}x_T(t)$ and $\sqrt{P_R}x_R(t)$. After cancelation, the best possible result for what the receiver receives is $g_2\sqrt{\frac{P_T}{SNR_T}}n_T(t)+g_2\sqrt{\frac{P_T}{SNR_T}}n'_T(t)+g_2\sqrt{P_R}x_R(t)+n_R(t)$ where g_2 is the receiver gain after cancelation and the 2nd term is the additional noise from the cancelation path.

Clearly, if $P_R < \frac{P_T}{SNR_T}$, the desired signal is buried under the transmission noise. Unfortunately, this condition holds for most practical situations where P_T/P_R is larger than 50dB and SNR_T is only about 30dB.

Also, if $g_1 = g_2$ is forced, the ratio of interference powers before and after cancelation is less than SNR_T (or equals $\frac{SNR_T+1}{2}$ from the above analysis). This is the performance barrier caused by the transmission noise.

III. Basic Configuration of the New Method

To break the barrier of the transmission noise, we propose a new method whose basic configuration is shown in Fig. 4.

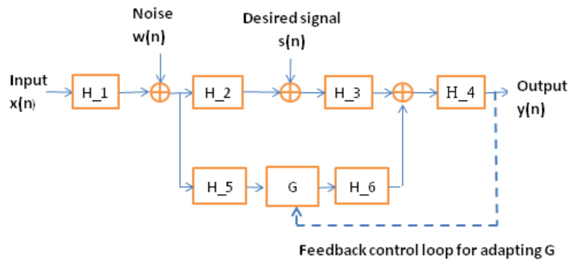

Fig. 4. Basic configuration of the new method. The path of H_5GH_6 is the cancelation path where G is adaptive. All H functions represent analog-interfaced channels or components and should be treated as unknown linear transfer functions. (This configuration also applies to the all-analog approach shown in [3].)

Here, the input $x(n)$ represents the digital source interference signal before DAC (digital-analog converter) in the transmit chain. The noise $w(n)$ is the unknown transmission noise from the entire transmit chain. The output $y(n)$ represents the received digital signal after ADC (analog-digital converter) in the receive chain. This observable has two components: one is due to the desired signal $s(n)$ from a remote radio and the other is due to both $x(n)$ and $w(n)$. The cancelation path is represented by H_5, G and H_6 where G is an adaptive filter. If the transfer function G is such that $H_6GH_5 = -H_3H_2$, then neither $x(n)$ nor $w(n)$ affects $y(n)$. It is important to note that none of the H transfer functions (H_1 to H_6) is known precisely enough and they should be treated as unknown. This is because the exact knowledge of the transfer function of an analog-interfaced component is difficult to obtain. Given unknown $H_1, \cdots, H_6$ and unknown $w(n)$, finding G is a blind system identification and equalization problem (which however differs from the conventional blind equalization problems in the literature). More detailed descriptions of the H functions are as follow:

H_1 represents the equivalent baseband channel transfer function (or simply channel) between a digitally generated baseband waveform $x(n)$ and the output of the RF power amplifier in the transmit chain. The noise $w(n)$ represents a combination of all noises generated in the transmit chain, which include the quantization noise in generating the transmitted baseband waveform at DAC, the noise from the up-conversion RF mixer, and the noise from the transmit power amplifier.

H_2 represents the channel between the transmit antenna and the receive antenna when two separate antennas are used for transmitting and receiving. H_2 may also represent the isolation path of an RF circulator when a single antenna is used along with the RF circulator for both transmitting and receiving.

H_3 represents the channel between the receive antenna and an analog baseband signal combiner just before VGA. (For all-analog, the signal combiner should be RF and before LNA.)

H_4 is the channel between the analog signal combiner and the output $y(n)$, which includes VGA and ADC. (For all-analog, H_4 is the entire receive chain after the RF combiner.)

H_5 is the channel between the output of the transmit power amplifier in the transmit chain and the input of the digital filter G, which may include a baseband-frequency sampler, a low pass filter and an ADC. This block does not need any carrier-frequency oscillator which tends to have a large phase noise.

H_6 is the channel between the output of the digital filter G and the signal combiner, which includes a DAC for analog signal combining before VGA. (For all-analog, G is an analog filter, H_5 models the input interface of G, and H_6 models the output interface of G.)

The cancelation path comprising H_5, G and H_6 can be made relatively noise free compared to the transmission noise. Note that the power of the quantization noise from a 14-bits ADC, for example, is over 70dB weaker than the signal power, which is insignificant compared to a typical transmission noise. For most applications, all H functions may appear all-pass with some delays while H_2 tends to be highly frequency-selective depending on the environment surrounding the transmit antenna and the receive antenna. However, for high-quality interference cancelation (such as 50dB or more), all H functions need to be treated as unknown when the parameters of G are optimized in minimizing the interference in $y(n)$. To find the optimal G online, the output signal $y(n)$ is the only observable we have to rely on, which we will discuss the detail shortly.

IV. A Cascade Form of Hybrid-2

To reduce the noise caused by LNA, it is desirable to reduce the interference at the RF frontend of the receiver. With a reduced interference at the RF front-end, the gain of LNA can be increased. The noise figure of LNA generally decreases with the gain, which is typically as small as 3dB at the highest gain. Although the hybrid-2 alone cannot achieve that, it can be used in tandem after the all-analog. Furthermore, hybrid-2 can be used in cascade with hybrid-1 (and both hybrids can be used after the all-analog cancelation).

Shown in Fig. 5 is a cascade form of the hybrid-2 with a special form of the time-domain transmit beamforming method (hybrid-1). Here, C_a and C_b are the waveform prefilters which should be chosen to reduce the self-interference at the RF frontend (at the sum immediately after the desired signal $s(n)$. C_a is the prefilter for the primary transmit chain, and C_b for the secondary (cancelation) transmit chain. The hardware associated with H_{ia} and H_{ib} is similar to that of H_i in Fig. 4 where $i = 1, \cdots, 6$. If there were no transmission noises $w_a(n)$ and $w_b(n)$, C_a and C_b could be chosen to yield a zero net contribution at the sum after $s(n)$. But with $w_a(n)$ and $w_b(n)$ (due to RF oscillator, RF mixers and power amplifiers embedded in H_1a and H_1b), there can be a significant amount of residue interference even if C_a and C_b are perfectly chosen. Due to noisy channel estimates, the choice of C_a and C_b can not be perfect, which introduces additional residue interference. To reduce the residue interference left from C_a and C_b, the adaptive filter G with two inputs can be used as shown.

V. The System Model and Solution for G

In order to find the optimal transfer function G of the adaptive filter during training, we must have a system model

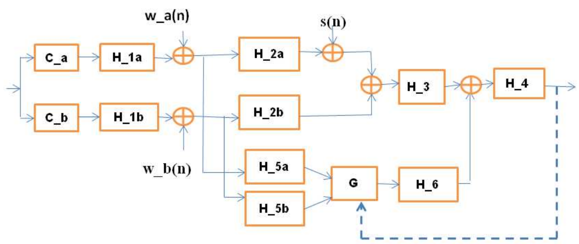

Fig. 5. A cascade of hybrid-1 and hybrid-2 methods. Ideally, the prefilters C_a and C_b should be such that $C_a = H_{1b}H_{2b}$ and $C_b = -H_{1a}H_{2a}$. Due to transmission noises $w_a(n)$ and $w_b(n)$, even the ideal C_a and C_b would still leave a large residue interference. However, regardless of the cancelation performance by C_a and C_b, if G is such that $G = [G_a, G_b]$, $H_3H_{2a} = -H_6G_aH_{5a}$, $H_3H_{2b} = -H_6G_bH_{5b}$, then all interferences are canceled at the output of this system.

in terms of G. This model must take into account the unknown nature of the H functions. In the following, we will only consider the basic configuration shown in Fig. 4. (Although useful for finding the parameters of the all-analog cancelation path [3], the following discussions assume that G is digital.)

During training, we assume the absence of $s(n)$, and hence the output $y(n)$ is simply the self-interference. We model $y(n)$ as a linear (but otherwise unknown) function of $x(n)$ and $w(n)$, and also as an affine (but otherwise unknown) function of the impulse response of $G(z)$. The unknown nature here is due to the unknown H functions. The optimal solution for $G(z)$ follows directly from the system model as shown next.

A. Complex Linear System

We will denote the impulse response of $G(z)$ by $g(0), \cdots, g(L)$. If the complex linear system model holds for the entire circuit (where each pair of I/Q components forms a complex number), which is a reasonable assumption, then for any given training pulse $x(n)$, $n = 0, 1, \cdots, N-1$ and a proper guard interval, we can write

$$\mathbf{y}_i = (\mathbf{F} + \mathbf{W}_i)\,\mathbf{g} + \mathbf{f} + \mathbf{w}_i \tag{1}$$

where $\mathbf{y}_i = [y(0), \cdots, y(N-1)]^T$ is the vector of the corresponding (complex) outputs in the ith realization, $\mathbf{g} = [g(0), \cdots, g(L)]^T$, $\mathbf{F}$ and $\mathbf{f}$ are unknown linear functions of $x(n)$, $n = 0, 1, \cdots, N-1$, and $\mathbf{W}_i$ and $\mathbf{w}_i$ are unknown linear functions of the noise $w(n)$, $n = 0, 1, \cdots, N-1$ in the ith realization. Here, we assume that $x(n)$, $n = 0, 1, \cdots, N-1$ is used repeatedly for all realizations.

Denote the energy of $\mathbf{y}_i$ by $\|\mathbf{y}_i\|^2$ with $i = 1, \cdots, M$. Then, it follows that for a large M, the average energy of the output (the self-interference) is given by

$$\begin{aligned} e = \frac{1}{M}\sum_{i=1}^{M}\|\mathbf{y}_i\|^2 &= \mathbf{g}^H\left(\mathbf{R}_F + \mathbf{R}_W\right)\mathbf{g} + \\ &+2Re\{(\mathbf{r}_F + \mathbf{r}_W)^H\mathbf{g}\} + \mathbf{r}_f + \mathbf{r}_w \end{aligned} \tag{2}$$

where $\mathbf{R}_F = \mathbf{F}^H\mathbf{F}$, $\mathbf{R}_W = \frac{1}{M}\sum_{i=1}^{M}\mathbf{W}_i^H\mathbf{W}_i$, $\mathbf{r}_F^H = \mathbf{f}^H\mathbf{F}$, $\mathbf{r}_W^H = \frac{1}{M}\sum_{i=1}^{M}\mathbf{w}_i^H\mathbf{W}_i$, $\mathbf{r}_f = \mathbf{f}^H\mathbf{f}$, and $\mathbf{r}_w = \frac{1}{M}\sum_{i=1}^{M}\mathbf{w}_i^H\mathbf{w}_i$. Equivalently, we can write

$$e = \mathbf{g}^H\mathbf{A}\mathbf{g} + Re\{\mathbf{g}^H\mathbf{b}\} + c \tag{3}$$

where $\mathbf{A} = \mathbf{A}^H$, $\mathbf{b}$ and c are the unknowns to be determined before we can find the optimal $\mathbf{g}$ to minimize e.

Now, define

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_r & -\mathbf{A}_i \\ \mathbf{A}_i & \mathbf{A}_r \end{bmatrix} \tag{4}$$

$$\bar{\mathbf{g}} = [\mathbf{g}_r^T, \mathbf{g}_i^T]^T \tag{5}$$

$$\bar{\mathbf{b}} = [\mathbf{b}_r^T, \mathbf{b}_i^T]^T \tag{6}$$

where $\mathbf{A}_r = \mathbf{A}_r^T$ and $\mathbf{A}_i = -\mathbf{A}_i^T$. Then, (3) is equivalent to

$$e = \bar{\mathbf{g}}^T\bar{\mathbf{A}}\bar{\mathbf{g}} + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \tag{7}$$

Furthermore, we can write

$$\begin{aligned} e &= \bar{\mathbf{g}}^T \otimes \bar{\mathbf{g}}^T vec(\bar{\mathbf{A}}) + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \\ &= [\mathbf{g}_r^T, \mathbf{g}_i^T] \otimes [\mathbf{g}_r^T, \mathbf{g}_i^T] \begin{bmatrix} vec\begin{pmatrix} \mathbf{A}_r \\ \mathbf{A}_i \end{pmatrix} \\ vec\begin{pmatrix} -\mathbf{A}_i \\ \mathbf{A}_r \end{pmatrix} \end{bmatrix} + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \\ &= \mathbf{g}_r^T \otimes [\mathbf{g}_r^T, \mathbf{g}_i^T] vec\begin{pmatrix} \mathbf{A}_r \\ \mathbf{A}_i \end{pmatrix} \\ &\quad + \mathbf{g}_i^T \otimes [\mathbf{g}_r^T, \mathbf{g}_i^T] vec\begin{pmatrix} -\mathbf{A}_i \\ \mathbf{A}_r \end{pmatrix} + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \\ &= \{\mathbf{g}_r^T \otimes [\mathbf{g}_r^T, \mathbf{g}_i^T] \\ &\quad + \mathbf{g}_i^T \otimes [\mathbf{g}_i^T, -\mathbf{g}_r^T]\}\, vec\begin{pmatrix} \mathbf{A}_r \\ \mathbf{A}_i \end{pmatrix} + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \end{aligned} \tag{8}$$

Let $\mathbf{P}$ be such that $vec\begin{pmatrix} \mathbf{A}_r \\ \mathbf{A}_i \end{pmatrix} = \mathbf{P}\begin{pmatrix} vec(\mathbf{A}_r) \\ vec(\mathbf{A}_i) \end{pmatrix}$. Then,

$$\begin{aligned} e &= \{\mathbf{g}_r^T \otimes [\mathbf{g}_r^T, \mathbf{g}_i^T] + \mathbf{g}_i^T \otimes [\mathbf{g}_i^T, -\mathbf{g}_r^T]\}\,\mathbf{P}\begin{pmatrix} vec(\mathbf{A}_r) \\ vec(\mathbf{A}_i) \end{pmatrix} \\ &\quad + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \\ &= \{[\mathbf{g}_r^T \otimes \mathbf{g}_r^T, \mathbf{g}_r^T \otimes \mathbf{g}_i^T] + [\mathbf{g}_i^T \otimes \mathbf{g}_i^T, -\mathbf{g}_i^T \otimes \mathbf{g}_r^T]\} \cdot \\ &\quad \cdot \begin{pmatrix} vec(\mathbf{A}_r) \\ vec(\mathbf{A}_i) \end{pmatrix} + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \\ &= [\mathbf{g}_r^T \otimes \mathbf{g}_r^T + \mathbf{g}_i^T \otimes \mathbf{g}_i^T, \mathbf{g}_r^T \otimes \mathbf{g}_i^T - \mathbf{g}_i^T \otimes \mathbf{g}_r^T] \cdot \\ &\quad \cdot \begin{pmatrix} vec(\mathbf{A}_r) \\ vec(\mathbf{A}_i) \end{pmatrix} + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \end{aligned} \tag{9}$$

Since $\mathbf{A}_r = \mathbf{A}_r^T$ and $\mathbf{A}_i = -\mathbf{A}_i^T$, we know that the only the lower triangular (including the diagonal) elements of $\mathbf{A}_r$ are independent and only the strictly lower triangular (excluding the diagonal) elements of $\mathbf{A}_i$ are independent.

We now define the selection matrices $\mathbf{S}_{L+1}$ and $\bar{\mathbf{S}}_{L+1}$ as such that $\mathbf{S}_{L+1}vec(\mathbf{A}_r)$ contains only the lower triangular elements of $\mathbf{A}_r$, and $\bar{\mathbf{S}}_{L+1}vec(\mathbf{A}_i)$ contains only the strictly lower triangular elements of $\mathbf{A}_i$.

Then, it follows that

$$e = \mathbf{u}^T\mathbf{v} \tag{10}$$

where

$$\mathbf{v} = \begin{bmatrix} \mathbf{S}_{L+1}vec(\mathbf{A}_r) \\ \bar{\mathbf{S}}_{L+1}vec(\mathbf{A}_i) \\ \bar{\mathbf{b}} \\ c \end{bmatrix} \tag{11}$$

$$\mathbf{u}^T = \left[\ \mathbf{u}_1^T \mid \mathbf{u}_2^T \mid \bar{\mathbf{g}}^T \mid 1\ \right] \tag{12}$$

where $\mathbf{u}_1^T = (\mathbf{g}_r^T \otimes \mathbf{g}_r^T + \mathbf{g}_i^T \otimes \mathbf{g}_i^T)\mathbf{S}_{L+1}^T \mathbf{D}_{L+1}$, $\mathbf{u}_2^T = 2(\mathbf{g}_r^T \otimes \mathbf{g}_i^T - \mathbf{g}_i^T \otimes \mathbf{g}_r^T)\bar{\mathbf{S}}_{L+1}^T$, and for example

$$\mathbf{D}_4 = diag\left[\begin{array}{cccc|ccc|cc|c} 1 & 2 & 2 & 2 & 1 & 2 & 2 & 1 & 2 & 1 \end{array}\right] \tag{13}$$

Now we assume that for every M realizations, we use a unique $\mathbf{g}$ from the set of $\mathbf{g}(k)$, $k = 1, 2, \cdots, K$. Then, corresponding to this set, we have a set of $e(k)$, $k = 1, 2, \cdots, K$. It follows that

$$\mathbf{e} = \mathbf{G}\mathbf{v} \tag{14}$$

where $\mathbf{e} = [e(1), \cdots, e(K)]^T$ and

$$\mathbf{G} = \left[\begin{array}{c} \mathbf{u}(1)^T \\ \cdots \\ \mathbf{u}(K)^T \end{array}\right] \in R^{K \times ((L+1)^2 + 2L + 3)} \tag{15}$$

and $\mathbf{u}(k)$ is defined in (12) with $\mathbf{g} = \mathbf{g}(k)$.

If $\mathbf{G}$ has a full-column-rank matrix (see next section), the least square solution of $\mathbf{v}$ is given by

$$\mathbf{v} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{e} \tag{16}$$

From this $\mathbf{v}$, we can then construct the estimates of $\bar{\mathbf{A}}$, $\bar{\mathbf{b}}$ and c.

With the model (7) and the estimates of $\bar{\mathbf{A}}$, $\bar{\mathbf{b}}$ and c, the optimal $\bar{\mathbf{g}}$ which minimizes e (the self-interference energy) is given by

$$\bar{\mathbf{g}} = -\frac{1}{2}\bar{\mathbf{A}}^{-1}\bar{\mathbf{b}} \tag{17}$$

B. Real Linear System

For the same system shown earlier, we can also model it as

$$e = \bar{\mathbf{g}}^T\hat{\mathbf{A}}\bar{\mathbf{g}} + \bar{\mathbf{g}}^T\bar{\mathbf{b}} + c \tag{18}$$

where the real matrix $\hat{\mathbf{A}}$ only has the symmetric property $\hat{\mathbf{A}}^T = \hat{\mathbf{A}}$, which is more relaxed than the structure shown in (4). If there is an I/Q imbalance in the circuit, the complex linear model fails but the real linear model still holds [8].

In this case, we can similarly show that

$$e = \hat{\mathbf{u}}^T\hat{\mathbf{v}} \tag{19}$$

where

$$\hat{\mathbf{v}} = \left[\begin{array}{c} \mathbf{S}_{2L+2}vec(\hat{\mathbf{A}}) \\ \bar{\mathbf{b}} \\ c \end{array}\right] \tag{20}$$

$$\hat{\mathbf{u}}^T = \left[\begin{array}{c|c|c} (\bar{\mathbf{g}}^T \otimes \bar{\mathbf{g}}^T)\mathbf{S}_{2L+2}^T\mathbf{D}_{2L+2} & \bar{\mathbf{g}}^T & 1 \end{array}\right] \tag{21}$$

VI. How to choose the training vectors $\mathbf{g}(k)$, $k = 1, 2, \cdots, K$

A necessary condition for the training vectors is that the matrix $\mathbf{G}$ defined in (15) is of full column rank. Furthermore, we want the columns of $\mathbf{G}$ to be "nearly" orthogonal with each other, which is important to reduce the noise sensitivity of the estimate given by (16). In this section, we introduce a choice of the training vectors that result in a sparse matrix $\mathbf{G}$ that meets the above requirement. The sparseness is also useful for reduced computation.

A. For real system

For the real system, we only treat $\hat{\mathbf{A}}$ as a symmetric matrix. Let $\bar{\mathbf{g}}$ have the dimension $m \times 1$. Then there are total $N_m = m(m+1)/2 + m + 1$ unknown real parameters in the system. We let the corresponding $\mathbf{G}$ be denoted by $\mathbf{G}_m$. We also use $\bar{\mathbf{g}}^T(n) = [g_1(n), \cdots, g_m(n)]^T$ with $n = 1, 2, \cdots, N$.

We first consider the case of $m = 2$ for which $N_2 = 6$. With $N = N_2 = 6$, we can write $\mathbf{G}_2$ as

$$\mathbf{G}_2 = \left[\begin{array}{cc|c|cc|c} g_1(1)^2 & 2g_1(1)g_2(1) & g_2(1)^2 & g_1(1) & g_2(1) & 1 \\ g_1(2)^2 & 2g_1(2)g_2(2) & g_2(2)^2 & g_1(2) & g_2(2) & 1 \\ g_1(3)^2 & 2g_1(3)g_2(3) & g_2(3)^2 & g_1(3) & g_2(3) & 1 \\ g_1(4)^2 & 2g_1(4)g_2(4) & g_2(4)^2 & g_1(4) & g_2(4) & 1 \\ g_1(5)^2 & 2g_1(5)g_2(5) & g_2(5)^2 & g_1(5) & g_2(5) & 1 \\ g_1(6)^2 & 2g_1(5)g_2(6) & g_2(6)^2 & g_1(6) & g_2(6) & 1 \end{array}\right] \tag{22}$$

To construct a nonsingular $\mathbf{G}_2$, we let

1) $\bar{\mathbf{g}}(1) = [0, 0]$
2) $\bar{\mathbf{g}}(2) = [1, 0]$
3) $\bar{\mathbf{g}}(3) = [-1, 0]$
4) $\bar{\mathbf{g}}(4) = [0, 1]$
5) $\bar{\mathbf{g}}(5) = [0, -1]$
6) $\bar{\mathbf{g}}(6) = [1, 1]$

Then,

$$\mathbf{G}_2 = \left[\begin{array}{cc|c|cc|c} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & -1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 \end{array}\right] \tag{23}$$

To show that this matrix has the full rank $N_2 = 6$, we can use eliminations by row combinations (starting from the top rows). Then, we can obtain $\mathbf{T}_2$ with $\det(\mathbf{T}_2) = 1$ such that

$$\mathbf{T}_2\mathbf{G}_2 = \left[\begin{array}{cc|c|cc|c} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \end{array}\right] \tag{24}$$

There is a column permutation matrix $\mathbf{P}_{c2}$ such that $\mathbf{T}_2\mathbf{G}_2\mathbf{P}_{c2} = diag[1, \mathbf{J}, \mathbf{J}, 2]$ where

$$\mathbf{J} = \left[\begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array}\right] \tag{25}$$

Hence, we have $\det(\mathbf{G}_2) = (-2)(-2)2 = 2^3$.

For the case of $m = 3$ where $N_3 = N_2 + 4 = 10$, we choose

1) $\bar{\mathbf{g}}(1) = [0, 0, 0]^T$
2) $\bar{\mathbf{g}}(2) = [1, 0, 0]^T$
3) $\bar{\mathbf{g}}(3) = [-1, 0, 0]^T$
4) $\bar{\mathbf{g}}(4) = [0, 1, 0]^T$
5) $\bar{\mathbf{g}}(5) = [0, -1, 0]^T$
6) $\bar{\mathbf{g}}(6) = [0, 0, 1]^T$
7) $\bar{\mathbf{g}}(7) = [0, 0, -1]^T$
8) $\bar{\mathbf{g}}(8) = [1, 1, 0]^T$
9) $\bar{\mathbf{g}}(9) = [1, 0, 1]^T$
10) $\bar{\mathbf{g}}(10) = [0, 1, 1]^T$

Then, it follows that

$$\mathbf{G}_3 = \left[\begin{array}{ccc|cc|c|ccc|c} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 \\ 1 & 2 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 1 & 1 & 1 \end{array}\right] \quad (26)$$

and there is an elimination-by-row matrix $\mathbf{T}_3$ with $\det(\mathbf{T}_3) = 1$ such that

$$\mathbf{T}_3\mathbf{G}_3 = \left[\begin{array}{ccc|cc|c|ccc|c} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad (27)$$

It is then easy to verify that there is a column permutation $\mathbf{P}_{c3}$ such that $\mathbf{T}_3\mathbf{G}_3\mathbf{P}_{c3} = diag[1, \mathbf{J}, \mathbf{J}, \mathbf{J}, 2, 2, 2]$. It follows that $\det(\mathbf{G}_3) = -2^6$.

For any given m, we can now show that the following generalization of the previous cases can make a full rank $\mathbf{G}_m$:

1) $g_i(1) = 0$ for $i = 1, \cdots, m$
2) $g_i(2i) = 1$ for $i = 1, \cdots, m$
3) $g_i(2i+1) = -1$ for $i = 1, \cdots, m$
4) $g_{i_k}(k+m+1) = g_{j_k}(k+m+1) = 1$, $k = 1, \cdots, N_m - m - 1$, where (i_k, j_k) is the kth pair of elements out of $\{1, \cdots, m\}$ and $i_k < j_k$.
5) $g_i(n) = 0$ otherwise

To prove this, we can let the $N_{m-1} \times N_{m-1}$ matrix $\mathbf{G}_{m-1}$ have the full rank N_{m-1}. Next, we prove that the $N_m \times N_m$ matrix $\mathbf{G}_m$ have the full rank N_m. Note that $N_m = N_{m-1} + m + 1$. It means that $\mathbf{G}_m$ have $m+1$ additional columns and rows than $\mathbf{G}_{m-1}$. One can then verify that there are a row permutation matrix $\mathbf{P}_1$, a column permutation matrix $\mathbf{P}_2$ and an elimination-by-row matrix $\mathbf{T}_m$ such that

$$\mathbf{P}_1\mathbf{T}_m\mathbf{G}_m\mathbf{P}_2 = \begin{bmatrix} \mathbf{G}_{m-1} & & \\ & \mathbf{J} & \\ & & 2\mathbf{I}_{m-1} \end{bmatrix} \quad (28)$$

where the two rows associated with the component $\mathbf{J}$ in the above are due to the two entries in $\mathbf{G}_m$: $\bar{\mathbf{g}}(2m) = [0, \cdots, 0, 1]^T$ and $\bar{\mathbf{g}}(2m+1) = [0, \cdots, 0, -1]^T$. The $m-1$ rows associated with the component $2\mathbf{I}_{m-1}$ in the above are due to the $m-1$ entries in $\mathbf{G}_m$: $\bar{\mathbf{g}}(N_m - m + 1 + i) = [\mathbf{e}_i^T, 1]^T$, $i = 1, \cdots, m-1$, where $\mathbf{e}_i$ is the $(m-1) \times 1$ vector of all zeros except the one value at its ith position. Hence, $\det(\mathbf{G}_m) = -2^m \det(\mathbf{G}_{m-1})$. Given $\det(\mathbf{G}_2) = 2^3$, we have $\det(\mathbf{G}_m) = (-1)^m 2^{\frac{m(m+1)}{2}}$.

B. For complex system

For the complex system, the matrix $\mathbf{G}$ has an additional structure. If we let $\bar{\mathbf{g}}$ have the dimension $k \times 1$ where $k = 2m$, then there are total $M_k = (k/2)^2 + k + 1$ unknown real parameters in the system. Note that $\bar{\mathbf{g}}(n) = [\mathbf{g}_r^T(n), \mathbf{g}_i^T(n)]^T$ where $\mathbf{g}_r(n) = [g_{r,1}(n), \cdots, g_{r,m}(n)]^T$ and $\mathbf{g}_i(n) = [g_{i,1}(n), \cdots, g_{i,m}(n)]^T$. Also note that $M_k = N_m + \frac{(m-1)m}{2} + m$. We will denote the corresponding $\mathbf{G}$ by $\mathbf{G}_{(k)}$. We will propose a choice of $\bar{\mathbf{g}}(n)$ for $n = 1, \cdots, M_k$ to construct a nonsingular $M_k \times M_k$ matrix $\mathbf{G}_{(k)}$.

There is a permutation matrix $\mathbf{P}_{(k)}$ such that as follows:

$$\mathbf{G}_{(k)}\mathbf{P}_{(k)} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{G}_{1,2} \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} \end{bmatrix} \quad (29)$$

where the first block column only depends on $\mathbf{g}_r(n)$ and the second block column depends on both $\mathbf{g}_r(n)$ and $\mathbf{g}_i(n)$. Furthermore, $\mathbf{G}_{1,1} \in R^{N_m \times N_m}$ has the exactly same structure as $\mathbf{G}_m$ in the real system with $\bar{\mathbf{g}}(n)$ replaced by $\mathbf{g}_r(n)$.

For $1 \leq n \leq N_m$, we propose to choose $\mathbf{g}_i(n) = 0$ but at the same time choose $\mathbf{g}_r(n)$ in the same way as for the real system. Then, we know that $\mathbf{G}_{1,1}$ is nonsingular and $\mathbf{G}_{1,2} = 0$. Furthermore, for any $\mathbf{G}_{2,2}$, there is an elimination-by-row matrix $\mathbf{T}_{(k)}$ such that

$$\mathbf{T}_{(k)}\mathbf{G}_{(k)}\mathbf{P}_{(k)} = \begin{bmatrix} \mathbf{G}_{1,1} & 0 \\ 0 & \mathbf{G}_{2,2} \end{bmatrix} \quad (30)$$

It is now sufficient to choose $M_k - N_m = \frac{(m-1)m}{2} + m$ additional $\bar{\mathbf{g}}(n)$ to make $\mathbf{G}_{2,2}$ full rank.

For $m = 3$, the structure of the nth row of $\mathbf{G}_{2,2}$ is

$$(\mathbf{G}_{2,2})_{n,:} = \left[\begin{array}{c|c|c|c} 2\delta g_{r,i}(1,2) & 2\delta g_{r,i}(1,3) & 2\delta g_{r,i}(2,3) & \hat{\mathbf{g}}_i^T \end{array}\right] \quad (31)$$

where $\delta g_{r,i}(k,l) = g_r(k)g_i(l) - g_i(k)g_r(l)$ and $\hat{\mathbf{g}}_i^T = [g_i(1), g_i(2), g_i(3)]$. If we choose

1) $\bar{\mathbf{g}}(N_3 + 1) = [0, 0, 0, 1, 0, 0]^T$
2) $\bar{\mathbf{g}}(N_3 + 2) = [0, 0, 0, 0, 1, 0]^T$
3) $\bar{\mathbf{g}}(N_3 + 3) = [0, 0, 0, 0, 0, 1]^T$
4) $\bar{\mathbf{g}}(N_3 + 4) = [1, 0, 0, 0, 1, 0]^T$
5) $\bar{\mathbf{g}}(N_3 + 5) = [1, 0, 0, 0, 0, 1]^T$
6) $\bar{\mathbf{g}}(N_3 + 6) = [0, 1, 0, 0, 0, 1]^T$

then we have (up to a column permutation)

$$\mathbf{G}_{2,2} = diag\,(\mathbf{I}_3, 2\mathbf{I}_3) \quad (32)$$

For any given m, we choose $\bar{\mathbf{g}}(n)$ for $N_m + 1 \leq n \leq M_k$ as follows:

1) $g_{i,n}(n + N_m) = 1$ for $n = 1, \cdots, m$
2) $g_{r,n_k}(N_m + m + k) = g_{i,m_k}(N_m + m + k) = 1$ for $1 \leq k \leq \frac{(m-1)m}{2}$ where (n_k, m_k) is the kth pair of elements from $\{1, \cdots, m\}$ satisfying $n_k < m_k$
3) $g_{r,j}(n) = 0$ and $g_{i,j}(n) = 0$ otherwise.

One can verify that up to a column permutation we have

$$\mathbf{G}_{2,2} = diag\left(\mathbf{I}_m, 2\mathbf{I}_{\frac{(m-1)m}{2}}\right) \quad (33)$$

which implies that $\det(\mathbf{G}_{2,2}) = 2^{\frac{(m-1)m}{2}}$.

Therefore, $\det(\mathbf{G}_{(k)}) = \det(\mathbf{G}_{1,1})\det(\mathbf{G}_{2,2}) = \det(\mathbf{G}_m)\det(\mathbf{G}_{2,2}) = 2^{m^2}$.

REFERENCES

[1] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," arXiv, 2013.

[2] M. Jain, J. I. Choi, T. M. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless", In Proc. Mobicom 2011.

[3] J. G. McMichael and K. E. Kolodziej, "Optimal tuning of analog self-interference cancellers for full-duplex wireless communication", 5th Annual Allerton Conference, Oct. 2012.

[4] S. Haykin, Adaptive Filter Theory, Prentice-Hall.

[5] A. Sahai, G. Patel, and A. Sabharwal, "Pushing the limits of full-duplex: Design and real-time implementation", Online at arXiv, 2011.

[6] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," IEEE Trans Signal Proc., Vol. 59, No. 12, Dec 2011.

[7] Y. Hua, "An overview of beamforming and power allocation for MIMO relays," Proc of MILCOM 2010, pp. 99-104, San Jose, CA, Nov 2010.

[8] Y. Hua, P. Liang, Y. Ma, A. Cirik and Q. Gao, "A method for broadband full-duplex MIMO radio," IEEE Signal Processing Letters, 2012.