Wideband Full-Duplex MIMO Relays with Blind Adaptive Self-interference Cancellation

Emilio Antonio-Rodríguez[†], Stefan Werner[†], Roberto López-Valcarce[*],
Taneli Riihonen[†], and Risto Wichman[†]
[*]Department of Signal Theory and Communications, University of Vigo, Vigo, Spain
[†]Department of Signal Processing and Acoustics, Aalto University, Helsinki, Finland

Abstract

We develop adaptive self-interference cancellation algorithms for both filter-and-forward and decode-and-forward multiple-input multiple-output relays. The algorithms are blind in the sense that they only exploit the spectral properties of the transmitted signal to identify the self-interference channel, while dealing with frequency-selective channels and arbitrary signal spectra. Our approach is non-intrusive in the sense that the algorithms can successfully identify, track, and cancel the self-interference distortion while the relay is operating in its normal mode. We study the stationary points of the algorithms and analyze under which conditions they achieve perfect cancellation of the self-interference. Simulation results show that the algorithms provide residual self-interference levels below the noise floor by using the time samples of only a few OFDM symbols.

I. Introduction

MIMO techniques, such as diversity schemes or spatial multiplexing, can provide high system throughput in a bandwidth-efficient manner [1]. However, extending network coverage while preserving consistent performance can be resource-demanding and, therefore, *in-band relaying* offers an attractive alternative for coverage extension due to the relaxed requirements for extra bandwidth usage [2], [3]. This paper concerns MIMO relays, which, as part of a MIMO link, are equipped with antenna arrays at both the transmit and receive sides to support end-to-end multistream communication.

Based on the employed transmission protocol, relays are usually classified as either filter-and-forward (FF) relays, of which amplify-and-forward (AF) is particular case, or decode-and-forward (DF) relays [4]. While FF relays simply forward the message, after some basic processing [5], [6], to the next network element without awareness of the data content, DF relays process the message and regenerate/re-encode the source data streams [7], [8]. As a consequence, DF relays generally yield better performance than FF relays [9], [10]. Relays can be further grouped into half-duplex (HD) or full-duplex (FD) relays. In particular, HD relays use orthogonal resources for transmission and reception, whereas FD relays share a common resource for both processes. For example, when in-band relaying is used, HD relays must designate two non-overlapping time slots: one for reception and one for transmission. On

the other hand, FD relays can receive and transmit at the same time, thus achieving higher data rates (approximately double) or higher spectral efficiency than HD relays [11], [12].

Full-duplex transmission introduces a new form of distortion caused by the relay itself, so-called *self-interference* (SI), wherein the transmitted signal loops back into the receive side. In particular, when both antenna arrays are closely located, the power imbalance between the transmitted signal and the signal received from the source can be high. As a consequence, we may end up with a scenario where the SI power at the relay input can be up to 100 dB higher than the information-bearing signal power,

rendering the relay useless [13]–[16]. Therefore, to realize maximum performance of the FD relay network, efficient SI mitigation techniques must be employed at the relay.

A. Self-interference mitigation methods

The different SI mitigation techniques found in the literature can be classified into the following three categories [17]:

1. *Isolation methods* [13], [18], [19] reduce SI power by optimizing the radiation pattern and improving the passive electromagnetic isolation between antenna arrays. Although isolation can reduce the interference by tens of decibels, the remaining residual SI in the system is likely high enough to affect the performance. Therefore, isolation techniques are typically used together with active analog and/or digital signal processing techniques that further reduce the residual interference.
2. *Spatial-domain suppression methods* [17], [20]–[23] exploit the degrees of freedom offered by multiple antennas to reduce the SI signal. For example, the signal may be transmitted in the nullspace of the SI channel, which ideally results in an interference-free signal at the receive side of the relay.
3. *Time-domain interference cancellation methods* [16], [17], [23]–[33] create a replica of the interference signal and subtract it from the relay input. Ideally, through an estimate of the SI channel, the residual SI will be negligible. Time-domain cancellation can be applied both in the analog domain to avoid glitches during analog-to-digital (A/D) conversion and in the digital domain to effectively cancel the remaining interference in the presence of multipath propagation.

The spatial suppression methods mentioned above, [17], [20]–[23] assume the explicit knowledge of the SI path. This is normally solved by means of an external estimation method or taking advantage of a pilot scheme embedded in the signal, which requires some processing in the frequency domain. Such approach only makes sense in a DF relay, where the signal is being regenerated. This not only intensify the computing requirements of the system, because each subcarrier should be treated independently, but also introduces an unavoidable delay due to the required demodulation/modulation process. Any additional delay can be harmful at the destination if the relative delay between the signal coming from the relay and the signal coming from the source exceeds the cyclic prefix length. Therefore, a shorter delay in the relay may result in better performance. In this work, we take a time-domain approach to reduce the processing delay by avoiding the demodulation/modulation to/from the frequency domain.

The time-domain cancellation methods presented in [25]–[28] are restricted to work in a single-stream transmission. In [26]–[28], an additional delay is introduced to perform the cancellation. As explained before, when the network uses OFDM any additional delay can be harmful. In [29], [30], a training-based cancellation method is proposed that disrupts the relay transmission, with the consequent data rate loss. If the relay is used for coverage extension, interrupting the transmission to redesign the interference canceller is not feasible. The use of a training sequence of a pilot embedded scheme introduces another implementation problem: the relay needs to be synchronized with the network, which, specially in the FF case, might not be possible. The use of blind techniques is convenient and in some cases the viable only option. In this work, we propose mitigation blind algorithms that do not introduce delay and do not require the use of any training sequence.

The implementation of the mitigation scheme follows one of the following categories: on-line or adaptive solutions and off-line or closed-form solutions. Whereas in the DF case an off-line solution is feasible (we provide a recursive implementation of an off-line solution), the nature of the problem does not allow to use a closed-form expression in the FF case. This limits the possible FF solutions to iterative processes, from which an adaptive solution is a natural choice. An adaptive solution has some advantages when compared to its off-line counterpart, e.g., reduced computational load per sample, zero processing delay (the adaptation is performed as samples arrive with no buffering required), and the ability to track time variations of the environment.

Finally, the mitigation scheme should not affect the relay normal operation. The link design is significantly simpler when assuming no self-interference at the relay and, therefore, it is important that self-interference is removed before any further processing of the received signal [17]. Our mitigation techniques are transparent to the relay operation and delivers a below-noise-level interference signal within the relay.

B. Contributions of the paper

This paper focuses on *low latency adaptive SI mitigation solutions for multistream relay networks*.[1] For this purpose, blind time-domain adaptive cancellation algorithms are proposed that effectively cope with SI. Our algorithms can be classified into stochastic gradient descent based algorithms and recursive least-squares based algorithms, and they are tailored separately for both FF and DF relays. The proposed algorithms can deal with arbitrary signal spectra and frequency-selective channels by exploiting statistical information about the source signal.

When considering OFDM relays, the proposed algorithms serve as attractive alternatives to frequency-domain methods, e.g., [17], [20]–[23], since the computational load is independent of the number of subcarriers. The algorithms also feature spectrally efficient implementation, since the estimation of the SI path requires only statistical information of the source signal. That is, no external estimation method or dedicated pilot scheme is required. This feature allows the algorithms to perform the interference mitigation without disrupting the relay's normal operation. Furthermore, the algorithms can efficiently cope with spatio-temporal inter-symbol interference (ISI), while previously mentioned methods only consider spatial ISI.

[1]By latency we mean any additional delay introduced by the self-interference mitigation scheme in the source-destination link.

Preliminary results pertaining to stochastic gradient implementations have appeared in [32] (FF case) and [33] (DF case). The associated material for the current paper has been rewritten and significantly extended. In addition, we present and analyze the stationary points of two new recursive least-squares based algorithms that features faster convergence and improved performance when compared to their stochastic gradient counterparts. The analysis presented herein is more thorough and the bias problem associated with MIMO FF relays is discussed in more detail.

C. Organization of the paper

The paper is organized as follows. Section II describes the system model of the considered relay network, discusses the impact of the SI in the system, and introduces the MIMO SI cancellation architecture. Section III addresses the problem of SI for a MIMO DF relay and derives gradient descent based and recursive least-squares based algorithms that effectively cancels the SI by using a power minimization approach. In this section we also provide details on algorithm convergence and stationary points. Section IV describes the bias problem originating from using a power minimization approach in a MIMO FF relay. Bias-corrected adaptive algorithms for SI cancellation and channel compensation are presented and stationary points are analyzed. Section V illustrates the performance of the proposed algorithms for an OFDM relay system, whereas Section VI draws the conclusions.

D. Notation

Let $\mathbf{H}[n] = \sum_{k=0}^{L_{\rm H}} \mathbf{H}(k)\delta[n-k]$ denote an $L_{\rm H}$th-order linear time-invariant causal filter of size $M \times N$ and coefficients $\{\mathbf{H}(k)\}_{k=0}^{L_{\rm H}}$. Let $\mathbf{x}[n]$ denote a signal vector of size $N \times 1$. The result of filtering $\mathbf{x}[n]$ with $\mathbf{H}[n]$, denoted by signal $\mathbf{y}[n]$ of size $M \times 1$, is given by the convolution operation ($\star$)

$$\mathbf{y}[n] = \mathbf{H}[n] \star \mathbf{x}[n] = \sum_{k=0}^{L_{\rm H}} \mathbf{H}(k)\mathbf{x}[n-k] = \boldsymbol{\mathcal{H}}\boldsymbol{x}[n] \tag{1}$$

where $\boldsymbol{\mathcal{H}} = [\mathbf{H}^T(0) \dots \mathbf{H}^T(L_{\rm H})]^T$ collects the taps of $\mathbf{H}[n]$ into a single matrix of size $M \times N(L_{\rm H}+1)$, and the vector $\boldsymbol{x}[n] = [\mathbf{x}^T[n] \dots \mathbf{x}^T[n-L_{\rm H}]]^T$, of compatible dimension, contains $L_{\rm H}+1$ past samples of $\mathbf{x}[n]$.

II. System model

The considered full-duplex MIMO relay link consists of a source node ($\mathcal{S}$), a destination node ($\mathcal{D}$), and an FD relay node ($\mathcal{R}$). The relay link supports the transmission of $M \geq 1$ data streams and each node is equipped with M receive/transmit antennas. Figure 1 illustrates the discrete-time equivalent baseband representation of the MIMO relay link where $\mathcal{S}$ transmits signal $\mathbf{s}_t[n] \in \mathbb{C}^M$, $\mathcal{D}$ receives signal $\mathbf{d}_r[n] \in \mathbb{C}^M$ and $\mathcal{R}$ receives $\mathbf{r}_r[n] \in \mathbb{C}^M$ while simultaneously transmitting $\mathbf{r}_t[n] \in \mathbb{C}^M$. In the following we assume that all channels are block fading. In other words, channels are assumed invariant for a few OFDM symbols or, equivalently, several hundreds or thousands of

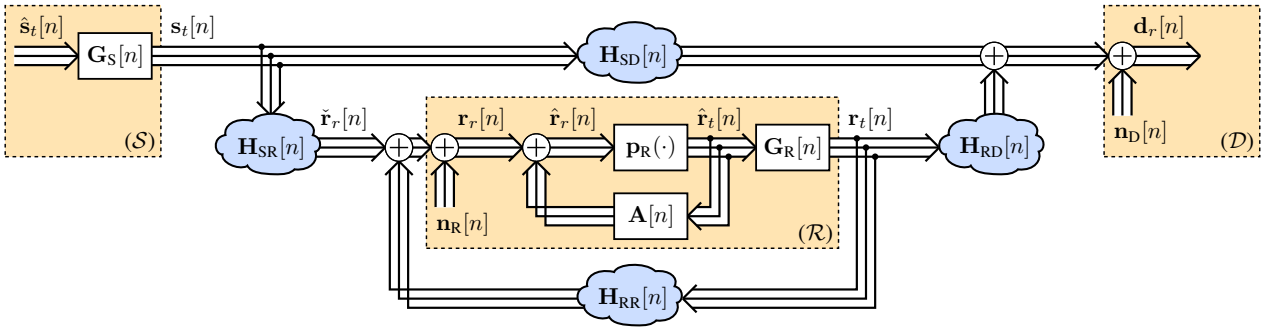

Fig. 1. System model of a single-frequency relay network with self-interference mitigation.

time-domain samples. The received signals at $\mathcal{R}$ and $\mathcal{D}$ are given by

$$\mathbf{r}_r[n] = \mathbf{H}_{\mathrm{SR}}[n] \star \mathbf{s}_t[n] + \mathbf{H}_{\mathrm{RR}}[n] \star \mathbf{r}_t[n] + \mathbf{n}_{\mathrm{R}}[n] \tag{2}$$

$$\mathbf{d}_r[n] = \mathbf{H}_{\mathrm{SD}}[n] \star \mathbf{s}_t[n] + \mathbf{H}_{\mathrm{RD}}[n] \star \mathbf{r}_t[n] + \mathbf{n}_{\mathrm{D}}[n] \tag{3}$$

where $\mathbf{H}_{ij}[n] \in \mathbb{C}^{M\times M}$, $i \in \{\mathrm{S},\mathrm{R}\}$ and $j \in \{\mathrm{R},\mathrm{D}\}$, is the L_{ij}th-order channel impulse response matrix between nodes i and j. Filter taps are drawn from a circularly-symmetric Gaussian distribution, i.e., $\mathrm{vec}\{\mathbf{H}_{ij}(k)\} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ up to a normalization factor. In (2) and (3), all channels are causal whereas $\mathbf{H}_{\mathrm{RR}}[n]$ is strictly causal, i.e., $\mathbf{H}_{\mathrm{RR}}[n] = \sum_{k=1}^{L_{\mathrm{RR}}} \mathbf{H}_{\mathrm{RR}}(k)\delta[n-k]$, $\mathbf{H}_{\mathrm{RR}}(0) = \mathbf{0}$. Finally, $\mathbf{n}_{\mathrm{D}}[n] \sim \mathcal{CN}(\mathbf{0},\sigma_d^2\mathbf{I}) \in \mathbb{C}^M$ is the noise at $\mathcal{D}$, while $\mathbf{n}_{\mathrm{R}}[n] \in \mathbb{C}^M$ represents all noise sources at $\mathcal{R}$. Specifically, $\mathbf{n}_{\mathrm{R}}[n]$ is composed of two different noise sources: one originating from the receive side of $\mathcal{R}$ while the other originates from the transmit side of $\mathcal{R}$ and couples back into the relay through $\mathbf{H}_{\mathrm{RR}}[n]$, i.e.,

$$\mathbf{n}_{\mathrm{R}}[n] = \mathbf{H}_{\mathrm{RR}}[n] \star \mathbf{n}_{\mathrm{R},t}[n] + \mathbf{n}_{\mathrm{R},r}[n] \tag{4}$$

where $\mathbf{n}_{\mathrm{R},r}[n] \sim \mathcal{CN}(\mathbf{0},\sigma_r^2\mathbf{I}) \in \mathbb{C}^M$ represents the relay reception noise and $\mathbf{n}_{\mathrm{R},t}[n] \sim \mathcal{CN}(\mathbf{0},\sigma_t^2\mathbf{I}) \in \mathbb{C}^M$ represents the relay transmit noise, which accounts for impairments in the transmission process [17], [34]–[39].

The relay is normally designed together with the other nodes of the link, therefore linear precoding might be present at the transmit side of $\mathcal{S}$ and $\mathcal{R}$, e.g., the wideband precoding techniques in [40], [41]. To account for precoding, let $\mathbf{G}_{\mathrm{S}}[n]$ denote the causal L_{S}th-order $M \times M$ precoding filter in $\mathcal{S}$ and $\mathbf{G}_{\mathrm{R}}[n]$ denote the causal L_{R}th-order $M \times M$ precoding filter in $\mathcal{R}$. Then, precoded signals at $\mathcal{S}$ and $\mathcal{R}$ are given by

$$\mathbf{s}_t[n] = \mathbf{G}_{\mathrm{S}}[n] \star \hat{\mathbf{s}}_t[n] \tag{5}$$

$$\mathbf{r}_t[n] = \mathbf{G}_{\mathrm{R}}[n] \star \hat{\mathbf{r}}_t[n] \tag{6}$$

where $\hat{\mathbf{s}}_t[n] \in \mathbb{C}^M$ and $\hat{\mathbf{r}}_t[n] \in \mathbb{C}^M$ are the respective signals before precoding in $\mathcal{S}$ and $\mathcal{R}$. If linear precoding is not present, we assume that $\mathbf{G}_{\mathrm{S}}[n] = \mathbf{I}\delta[n]$ and $\mathbf{G}_{\mathrm{R}}[n] = \mathbf{I}\delta[n]$. Note that mitigation should work independently of

the presence of $\mathbf{G}_{\mathrm{R}}[n]$.

The relay protocol, represented by $\hat{\mathbf{r}}_t[n] = \mathbf{p}_{\mathrm{R}}(\hat{\mathbf{r}}_r[n])$, controls the information forwarding process in accordance to some criterion or rule set. In particular, we distinguish between the following protocols:

- *Filter-and-forward (FF) protocol*: The relay simply forwards the received signal to the next element of the network. Consequently, we can model the family of FF protocols with $\mathbf{p}_{\mathrm{R}}(\hat{\mathbf{r}}_r[n]) = \mathbf{B}\hat{\mathbf{r}}_r[n]$, where $\mathbf{B}$ is an $M \times M$ full-rank matrix that spatially filters the different data streams. Note that the amplify-and-forward (AF) protocol is the particular case $\mathbf{G}_{\mathrm{R}}[n] = \mathbf{I}\delta[n]$.
- *Decode-and-forward (DF) protocol*: The relay regenerates the original M data streams from $\mathcal{S}$ before transmission. Ideally we have $\hat{\mathbf{r}}_t[n] = \hat{\mathbf{s}}_t[n-\delta]$, for some delay $\delta > 0$. The DF protocol includes various operations such as frame alignment, time- and frequency-synchronization, signal demodulation, equalization, and data decoding.

At $\mathcal{R}$, we denote the information-bearing signal coming from $\mathcal{S}$ by $\check{\mathbf{r}}_r[n] = \mathbf{H}_{\mathrm{SR}}[n] \star \mathbf{s}_t[n]$ whereas the SI signal is denoted by $\mathbf{i}_r[n] = \mathbf{H}_{\mathrm{RR}}[n] \star \mathbf{r}_t[n]$. The received signal at $\mathcal{R}$, $\mathbf{r}_r[n]$, can be expressed as the sum of the information signal, SI and noise:

$$\mathbf{r}_r[n] = \check{\mathbf{r}}_r[n] + \mathbf{i}_r[n] + \mathbf{n}_{\mathrm{R}}[n] \tag{7}$$

The SI term $\mathbf{i}_r[n]$ in (7) can be seen as an additional noise source whose power can be significantly stronger than that of $\check{\mathbf{r}}_r[n]$ and, since $\mathbf{r}_t[n]$ is a general function of $\check{\mathbf{r}}_r[n]$, it may be correlated in time with $\check{\mathbf{r}}_r[n]$. To fully enjoy the benefits of full-duplex relaying when compared to half-duplex relaying, the self interference must be reduced as much as possible, see, e.g., [11]. For this purpose, we employ the cancellation architecture depicted in Fig. 1, which consists of the strictly causal L_{A}th-order $M \times M$ MIMO filter $\mathbf{A}[n]$. The input of the relay protocol $\mathbf{p}_{\mathrm{R}}(\cdot)$ is now given by

$$\begin{aligned}
\hat{\mathbf{r}}_r[n] &= \check{\mathbf{r}}_r[n] + \mathbf{i}_r[n] + \mathbf{A}[n] \star \hat{\mathbf{r}}_t[n] + \mathbf{n}_{\mathrm{R}}[n] \\
&= \check{\mathbf{r}}_r[n] + \mathbf{H}_{\mathrm{RR}}[n] \star \mathbf{G}_r[n] \star \hat{\mathbf{r}}_t[n] + \mathbf{A}[n] \star \hat{\mathbf{r}}_t[n] + \mathbf{n}_{\mathrm{R}}[n] \\
&= \check{\mathbf{r}}_r[n] + \underbrace{(\mathbf{A}[n] + \mathbf{H}_{\mathrm{RR}}[n] \star \mathbf{G}_r[n]) \star \hat{\mathbf{r}}_t[n]}_{\hat{\mathbf{i}}_r[n]} + \mathbf{n}_{\mathrm{R}}[n]
\end{aligned} \tag{8}$$

where $\hat{\mathbf{i}}_r[n]$ denotes the residual SI at the input of $\mathbf{p}_{\mathrm{R}}(\cdot)$. If we let $\hat{\mathbf{H}}_{\mathrm{RR}}[n]$ denote the $L_{\hat{\mathrm{RR}}}$th-order impulse response of the equivalent SI channel, i.e.,

$$\hat{\mathbf{H}}_{\mathrm{RR}}[n] = -\mathbf{H}_{\mathrm{RR}}[n] \star \mathbf{G}_{\mathrm{R}}[n] \tag{9}$$

then we see from (8) that perfect cancellation is obtained when $\mathbf{A}[n] = \hat{\mathbf{H}}_{\mathrm{RR}}[n]$. This implies that filter order L_{A} must be chosen as $L_{\mathrm{A}} \geq L_{\hat{\mathrm{RR}}} = L_{\mathrm{RR}} + L_{\mathrm{R}}$.

In practice, perfect knowledge of $\hat{\mathbf{H}}_{\mathrm{RR}}[n]$ is not available at the relay and an estimate must be acquired. Furthermore, the estimation of $\hat{\mathbf{H}}_{\mathrm{RR}}[n]$ must be done while the relay is operating to avoid any interruption in the transmission. This justifies the use of a blind adaptive scheme that is also able to track temporal variations of

the SI path.

The setup of Fig. 1 resembles an identification scenario, since $\mathbf{A}[n]$ is placed in parallel with the effective SI path $\hat{\mathbf{H}}_{\text{RR}}[n]$. However, in contrast to a standard identification setup both $\mathbf{A}[n]$ and $\hat{\mathbf{H}}_{\text{RR}}[n]$ form part of a feedback loop, i.e., their common input signals depends of their respective outputs. Before proceeding with the derivation of the SI cancellation algorithms, we make two important remarks about the employed cancellation structure.

Remark 1: The self interference architecture in Fig. 1 does not increase the processing delay at the relay, i.e., the delay between $\check{\mathbf{r}}_r[n]$ and $\mathbf{r}_t[n]$ remains the same as without self-interference mitigation. As a consequence, the propagation delay between $\mathcal{S}$ and $\mathcal{R}$ is unchanged. This is a desirable feature in systems that exploit cyclic prefix for multipath compensation. This is in contrast with schemes which introduce an intentional delay in the relay processing path, see, e.g., [27], [28].

Remark 2: The sampling rate at which the cancellation algorithm operates, F_s, is related to the baseband signal bandwidth, B, as $F_s = 2k_{up}B$, where $k_{up} \geq 1$ is the oversampling factor. Therefore, the cancellation algorithm can operate directly with the waveform samples in an independent manner, without synchronization with the data. This is of major importance in FF relays where timing recovery is not necessary.

III. Adaptive interference cancellation for MIMO DF relays

In this section, we discuss the power minimization approach to SI cancellation and derive adaptive algorithms based on gradient descent and recursive least-squares principles. Analysis of stationary points establishes conditions for perfect cancellation.

A. Derivation of the algorithms

In the case of a MIMO DF relay, the associated decoding and regeneration process will introduce a significant delay. As a consequence we can adopt the following assumption.

$$\textbf{A1: } \textit{For the DF protocol, it holds that } \mathbb{E}\{\check{\mathbf{r}}_r[n]\hat{\mathbf{r}}_t^H[n-k]\} = \mathbb{E}\{\mathbf{n}_{\text{R}}[n]\hat{\mathbf{r}}_t^H[n-k]\} = \mathbf{0},\ k \geq 0. \tag{10}$$

The decorrelation property in (10) allows us, as we shall see below, the use of a recursive least-squares algorithm for updating the coefficients of the cancellation filter. We see from (8) that the residual interference $\hat{\mathbf{i}}_r[n]$ is the linear combination of past samples $\hat{\mathbf{r}}_t[n-k]$, with $k > 0$. Therefore, in view of **A1**, we have

$$\mathbb{E}\{\check{\mathbf{r}}_r[n]\hat{\mathbf{i}}_r^H[n]\} = \mathbb{E}\{\mathbf{n}_{\text{R}}[n]\hat{\mathbf{i}}_r^H[n]\} = 0 \tag{11}$$

Consequently, the power of $\hat{\mathbf{r}}_r[n]$ in (8) can be expressed as

$$P_{\hat{\mathbf{r}}_r} = \mathbb{E}\{\|\hat{\mathbf{r}}_r[n]\|^2\} = \mathbb{E}\{\|\hat{\mathbf{i}}_r[n]\|^2\} + \mathbb{E}\{\|\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]\|^2\} \tag{12}$$

Whereas the second term on the right-hand side of (12) does not depend on $\mathbf{A}[n]$, the first term clearly does. In particular, $\mathbf{E}\{\|\hat{\mathbf{i}}_r[n]\|^2\} = \mathbf{0}$ whenever $\mathbf{A}[n] = \hat{\mathbf{H}}_{\text{RR}}[n]$, which corresponds to perfect SI cancellation. Thus,

minimization of (an estimate of) $P_{\hat{\mathbf{r}}_r}$ with respect to the coefficients $\mathbf{A}[n]$ by means of an adaptive algorithm is a sensible approach to obtain an unbiased estimate of the SI path.

Using the previously defined notation, let rewrite signal $\hat{\mathbf{r}}_r[n]$ as

$$\hat{\mathbf{r}}_r[n] = \mathbf{r}_r[n] + \sum_{k=1}^{L_\mathrm{A}} \mathbf{A}[k]\hat{\mathbf{r}}_t[n-k] = \mathbf{r}_r[n] + \boldsymbol{\mathcal{A}}\hat{\boldsymbol{r}}_t[n-1] \tag{13}$$

where $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M\times ML_\mathrm{A}}$ collects all the L_A coefficients of $\mathbf{A}[n]$, and vector $\hat{\boldsymbol{r}}_t[n] \in \mathbb{C}^{ML_\mathrm{A}}$ collects the L_A past samples of the regenerated signal, $\hat{\mathbf{r}}_t[n]$, i.e.,

$$\boldsymbol{\mathcal{A}} = [\mathbf{A}[1]\ \mathbf{A}[2]\ \ldots\ \mathbf{A}[L_\mathrm{A}]] \tag{14}$$

$$\hat{\boldsymbol{r}}_t[n] = \left[\hat{\mathbf{r}}_t^H[n]\ \hat{\mathbf{r}}_t^H[n-1]\ \ldots\ \hat{\mathbf{r}}_t^H[n-L_\mathrm{A}+1]\right]^H \tag{15}$$

We can solve the minimization of $\mathbb{E}\{\|\hat{\mathbf{r}}_r[n]\|^2\}$ with respect to $\boldsymbol{\mathcal{A}}$ by using the following stochastic gradient descent algorithm, hereafter referred to as the decode-and-forward stochastic gradient descent (DF-SGD) algorithm

$$\begin{aligned}
\boldsymbol{\mathcal{A}}[n+1] &= \boldsymbol{\mathcal{A}}[n] - \mu_a \nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\|\hat{\mathbf{r}}_r[n]\|^2\right\} \\
&= \boldsymbol{\mathcal{A}}[n] - \mu_a \nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\left(\mathbf{r}_r[n] + \boldsymbol{\mathcal{A}}\hat{\boldsymbol{r}}_t[n-1]\right)^H \hat{\mathbf{r}}_r[n]\right\} \\
&= \boldsymbol{\mathcal{A}}[n] - \mu_a \nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\hat{\boldsymbol{r}}_t^H[n-1]\boldsymbol{\mathcal{A}}^H\hat{\mathbf{r}}_r[n]\right\} \\
&= \boldsymbol{\mathcal{A}}[n] - \mu_a \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1]
\end{aligned} \tag{16}$$

where $\mu_a > 0$ is the adaptation step size that controls the stability, convergence speed and misadjustment.[2] Note that $\boldsymbol{\mathcal{A}}[n]$ is updated for every new time-domain sample, so n denotes both sample and algorithm iteration index.

A recursive least-squares algorithm exhibiting better performance and faster convergence than DF-SGD algorithm is obtained by solving the following optimization problem [42], [43], i.e.,

$$\boldsymbol{\mathcal{A}}[n] = \arg\min_{\boldsymbol{\mathcal{A}}} \sum_{k=0}^{n} \lambda^{n-k}\|\hat{\mathbf{r}}_r[k]\|^2 \tag{17}$$

where $0 \ll \lambda \leq 1$ is the *forgetting factor*. Differentiating the cost function in (17) with respect to $\boldsymbol{\mathcal{A}}$ and solving for the minimum yields the following normal equation

$$\boldsymbol{\mathcal{A}}[n]\mathbf{P}^{-1}[n] = -\mathbf{T}[n] \tag{18}$$

where

$$\mathbf{P}^{-1}[n] = \sum_{k=0}^{n} \lambda^{n-k}\hat{\boldsymbol{r}}_t[k-1]\hat{\boldsymbol{r}}_t^H[k-1] \tag{19}$$

[2]When computing the gradient in (16) we treated $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{A}}^H$ as independent quantities and applied the gradient rule $\nabla_{\mathbf{X}^*}\left\{\mathbf{a}^H\mathbf{X}^H\mathbf{b}\right\} = \mathbf{b}\mathbf{a}^H$ ($\mathbf{X}$ being a matrix, and $\mathbf{a}$ and $\mathbf{b}$ being vectors).

TABLE I
STOCHASTIC GRADIENT DESCENT AND RECURSIVE LEAST-SQUARES POWER MINIMIZATION BASED ALGORITHMS FOR SI CANCELLATION IN DF RELAYS.

DF-SGD algorithm
$\boldsymbol{\mathcal{A}}[n+1] = \boldsymbol{\mathcal{A}}[n] - \mu_a \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1]$
DF-RLS algorithm
$\mathbf{P}[0] = \epsilon^{-1}\mathbf{I},\ 0 < \epsilon \ll 1$ $\mathbf{K}[n] = \mathbf{P}[n-1]\hat{\boldsymbol{r}}_t[n-1]$ $\mathbf{P}[n] = \frac{1}{\lambda}\left(\mathbf{P}[n-1] - \frac{\mathbf{K}[n]\mathbf{K}^H[n]}{\lambda + \mathbf{K}^H[n]\hat{\boldsymbol{r}}_t[n-1]}\right)$ $\boldsymbol{\mathcal{A}}[n] = \boldsymbol{\mathcal{A}}[n-1] - \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1]\mathbf{P}[n]$

$$\mathbf{T}[n] = \sum_{k=0}^{n} \lambda^{n-k} \mathbf{r}_r[k]\hat{\boldsymbol{r}}_t^H[k-1] \tag{20}$$

Substituting (13) into (18) yields the relation

$$\begin{aligned}\boldsymbol{\mathcal{A}}[n]\mathbf{P}^{-1}[n] &= -\lambda\mathbf{T}[n-1] - \mathbf{r}_r[n]\hat{\boldsymbol{r}}_t^H[n-1] \\ &= \lambda\boldsymbol{\mathcal{A}}[n-1]\mathbf{P}^{-1}[n-1] - \mathbf{r}_r[n]\hat{\boldsymbol{r}}_t^H[n-1] \\ &= \boldsymbol{\mathcal{A}}[n-1]\mathbf{P}^{-1}[n] - \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1]\end{aligned} \tag{21}$$

from which we obtain the following recursive update

$$\boldsymbol{\mathcal{A}}[n] = \boldsymbol{\mathcal{A}}[n-1] - \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1]\mathbf{P}[n] \tag{22}$$

The adaptation rule in (22) constitutes the decode-and-forward recursive least-squares based algorithm (referred to as DF-RLS) for SI cancellation. Matrix $\mathbf{P}[n] \in \mathbb{C}^{ML_{\text{A}} \times ML_{\text{A}}}$ is efficiently updated by applying the matrix inversion lemma to the recursive update of the auto-correlation matrix in (19), i.e., $\mathbf{P}^{-1}[n] = \lambda\mathbf{P}^{-1}[n-1] + \hat{\boldsymbol{r}}_t[n-1]\hat{\boldsymbol{r}}_t^H[n-1]$.

The DF-SGD and the DF-RLS algorithms are summarized in Table I. We note the similarities between the two algorithms wherein step size μ_a in the DF-SGD algorithm is replaced with matrix $\mathbf{P}[n]$ in the DF-RLS algorithm.

B. Algorithm analysis

For the purpose of the analysis, a stationary point is defined as follows.

Definition 1. *Let $\boldsymbol{\mathcal{A}}[n+1] = \boldsymbol{\mathcal{A}}[n] + \mathbf{X}[n]$ be a generic adaptive algorithm with $\mathbf{X}[n]$ being the driving term. Then, a stationary point of the algorithm, $\boldsymbol{\mathcal{A}}_\star$, results in a vanishing driving term, i.e.,*

$$\mathbb{E}\{\mathbf{X}[n]\}|_{\boldsymbol{\mathcal{A}}[n]=\boldsymbol{\mathcal{A}}_\star} = \mathbf{0} \tag{23}$$

when the coefficients of $\boldsymbol{\mathcal{A}}[n]$ are fixed to $\boldsymbol{\mathcal{A}}_\star$.

The concept of stationary points is useful for analyzing the algorithm performance upon convergence and the reader is referred to [44] and [45] for a thorough technical discussion on the validity of (23). In addition to **A1**, the analysis will use of the following assumption

A2: The process $\hat{\mathbf{r}}_t[n]$ is wide-sense stationary, i.e, we assume the *autocorrelation matrices*

$$\mathbf{R}_{\hat{\mathbf{r}}_t}[k] = \mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n-k]\}, \forall\, k, \text{ to be constant in time and independent of the relay state.}$$

That is, the statistical properties of the regenerated signal $\hat{\mathbf{r}}_t[n]$ only depend on the modulation format, and not on the particular realization of the data sequence embedded in $\hat{\mathbf{r}}_t[n]$ or internal parameters of the relay (relay state).

Assuming that $\mu_a \ll 1$ and $\lambda \approx 1$, i.e., (23) is a valid approximation, the conditions under which the SI distortion can be perfectly cancelled are stated in the following theorem.

Theorem 1. *If the L_Ath-order autocorrelation matrix of $\hat{\mathbf{r}}_t[n]$ is non-singular, then the DF-SGD and DF-RLS algorithms have a unique stationary point, $\boldsymbol{\mathcal{A}}_\star$. If $L_A \geq L_{\hat{RR}}$ holds, then this stationary point renders perfect cancellation.*

Proof. The SI $\mathbf{i}_r[n]$ in (7) can be expressed, using (9), as

$$\mathbf{i}_r[n] = \boldsymbol{\mathcal{H}}_{\text{RR}}\hat{\boldsymbol{r}}_i[n-1] \tag{24}$$

According to Definition 1, any stationary point, denoted by $\boldsymbol{\mathcal{A}}_\star$, must satisfy the following condition

$$\mathbb{E}\{\hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1]\mathbf{Q}[n]\} = \mathbf{0} \tag{25}$$

where $\mathbf{Q}[n] = \mathbf{I}$ for the DF-SGD algorithm and $\mathbf{Q}[n] = \mathbf{P}[n]$ for the DF-RLS algorithm. Combining (8)–(11),(13), (24), and (25), the above condition translates into

$$\boldsymbol{\mathcal{A}}_\star\mathbb{E}\{\hat{\boldsymbol{r}}_t[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\mathbf{Q}[n]\} = \boldsymbol{\mathcal{H}}_{\text{RR}}\mathbb{E}\{\hat{\boldsymbol{r}}_i[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\mathbf{Q}[n]\} \tag{26}$$

For sufficiently large n, we may replace $\mathbf{P}[n]$ in (26) with its expected value $\mathbb{E}\{\mathbf{P}[n]\}$ [46]. Factoring out $\mathbb{E}\{\mathbf{P}[n]\}$ we obtain

$$\mathbb{E}\{\hat{\boldsymbol{r}}_t[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\mathbf{P}[n]\} \approx \mathbb{E}\{\hat{\boldsymbol{r}}_t[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\}\mathbb{E}\{\mathbf{P}[n]\} \tag{27}$$

Thus, in view of **A2**, (26) becomes linear in $\boldsymbol{\mathcal{A}}_\star$. Assuming the L_{A}-th order autocorrelation matrix of $\hat{\mathbf{r}}_t[n]$ invertible, i.e., $\mathbb{E}\{\hat{\boldsymbol{r}}_t[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\}^{-1}$ exists, then $\mathbb{E}\{\mathbf{P}[n]\}$ is non-singular and the unique solution of (26) is given by

$$\boldsymbol{\mathcal{A}}_\star = \boldsymbol{\mathcal{H}}_{\text{RR}}\mathbb{E}\{\hat{\boldsymbol{r}}_i[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\}\mathbb{E}\{\hat{\boldsymbol{r}}_t[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\}^{-1} \tag{28}$$

Since $L_{\text{A}} \geq L_{\hat{\text{RR}}}$, $\hat{\boldsymbol{r}}_t[n-1]$ can be partitioned as

$$\hat{\boldsymbol{r}}_t[n-1] = \left[\hat{\boldsymbol{r}}_i^H[n-1]\ \hat{\mathbf{r}}_t^H[n-L_{\hat{\text{RR}}}-1]\ \ldots\ \hat{\mathbf{r}}_t^H[n-L_{\text{A}}]\right]^H \tag{29}$$

After inserting (29) into (28) and invoking **A2**, we can apply the block matrix inversion lemma to the partitioned

matrix $\mathbb{E}\{\hat{\boldsymbol{r}}_t[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\}$. Consequently, (28) reduces to

$$\mathbb{E}\{\hat{\boldsymbol{r}}_i[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\}\mathbb{E}\{\hat{\boldsymbol{r}}_t[n-1]\hat{\boldsymbol{r}}_t^H[n-1]\}^{-1} = \left[\mathbf{I}\ \mathbf{0}_{M\times M(L_{\text{A}}-L_{\text{RR}})}\right] \tag{30}$$

which gives $\boldsymbol{\mathcal{A}}_\star = [\boldsymbol{\mathcal{H}}_{\text{RR}}\ \mathbf{0}\ \ldots\ \mathbf{0}]$. □

Remark 3: Theorem 1 ensures that the stationary points of the algorithms provide perfect cancellation of the SI signal, whenever $\boldsymbol{\mathcal{A}}$ has enough degrees of freedom and $\hat{\mathbf{r}}_t[n]$ is persistently exciting.[3]

The convergence properties of DF-SGD follow from [43]. Global convergence is ensured whenever $\mu_a < 1/\rho_{max}$, where ρ_{max} is the largest eigenvalue of the autocorrelation matrix of $\hat{\mathbf{r}}_t[n]$. On the contrary, DF-RLS exhibits global convergence almost independently of the statistical properties of $\hat{\mathbf{r}}_t[n]$. The level of adaptation noise depends on the chosen values for μ_a and λ: higher values of μ_a and lower values of λ yield noisier steady-state responses, see, e.g., [42]. Finally, the differences in computational load between the DF-SGD and DF-RLS algorithms, is in the calculation of matrix $\mathbf{P}[n]$ for the DF-RLS algorithm.

IV. Adaptive interference cancellation for MIMO FF relays

In this section, we discuss the bias correction problem for proper SI cancellation in full-duplex MIMO FF relays. We derive and analyze stochastic gradient descent and recursive least-squares based algorithms that not only cancel the SI but also equalize the $\mathcal{S}$-$\mathcal{R}$ channel.

A. Derivation of the algorithms

For the case of MIMO FF relays, the minimization of (an estimate of) $P_{\hat{\mathbf{r}}_r}$ does not guarantee proper SI mitigation. The reason is that **A1** and **A2** do not hold for FF relays. In particular, in the FF relay case, $P_{\hat{\mathbf{r}}_r}$ can be expressed as

$$P_{\hat{\mathbf{r}}_r} = \mathbb{E}\{\|\hat{\mathbf{i}}_r[n]\|^2\} + \mathbb{E}\{\|\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]\|^2\} + 2\Re\{\mathbb{E}\{\hat{\mathbf{i}}_r^H[n](\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n])\}\} \tag{31}$$

where the last term in (31) is due to correlation between $\mathbf{i}_r[n]$ and $\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]$. The signal after cancellation, $\hat{\mathbf{r}}_r[n]$, can be written in terms of the input as

$$\hat{\mathbf{r}}_r[n] = \check{\mathbf{H}}_{\text{RR}}[n] \star (\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]) \tag{32}$$

where $\check{\mathbf{H}}_{\text{RR}}[n]$ is the impulse response of the equivalent system with transfer function

$$\check{\mathbf{H}}_{\text{RR}}(e^{j\omega}) = (\mathbf{I} - (\mathbf{A}(e^{j\omega}) - \hat{\mathbf{H}}_{\text{RR}}(e^{j\omega}))\mathbf{B})^{-1} \tag{33}$$

[3]A stationary process is persistently exciting *iff* its power spectral density (PSD) $\mathbf{S}_{\hat{\mathbf{r}}_t}(e^{j\omega})$ is full rank at all frequencies, or, equivalently, the autocorrelation matrix of $\hat{\mathbf{r}}_t[n]$ is positive definite.

In view of (33), perfect cancellation yields $\check{\mathbf{H}}_{\text{RR}}(e^{j\omega}) = \mathbf{I}$ and $\hat{\mathbf{r}}_r[n] = \check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]$, which gives $\mathbb{E}\{\|\hat{\mathbf{r}}_r[n]\|^2\} = \mathbb{E}\{\|\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]\|^2\}$. However, on a closer inspection it becomes clear that the solution $\mathbf{A}(e^{j\omega}) = \hat{\mathbf{H}}_{\text{RR}}(e^{j\omega})$ does not in general correspond to a minimum of (31) with respect to $\boldsymbol{\mathcal{A}}$. In particular, the following relation holds true

$$\min_{\boldsymbol{\mathcal{A}}} P_{\hat{\mathbf{r}}_r} \leq \mathbb{E}\{\|\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]\|^2\} \tag{34}$$

where, for L_{A} sufficiently large, the minimum of (34) is attained whenever $\check{\mathbf{H}}_{\text{RR}}(e^{j\omega}) = \boldsymbol{\Lambda}(e^{j\omega})$, $\boldsymbol{\Lambda}(e^{j\omega})$ being the optimum linear prediction error filter of signal $\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]$. Equality holds in (34) iff $\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]$ is a white signal, which corresponds to $\boldsymbol{\Lambda}(e^{j\omega}) = \mathbf{I}$. As a consequence, perfect SI cancellation through the minimization of $P_{\hat{\mathbf{r}}_r}$ with respect to $\boldsymbol{\mathcal{A}}$ is only possible for the case when $\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]$ is temporally white. If this condition is not met, the solution to (34) will be biased, i.e., $\mathbf{A}[n] + \hat{\mathbf{H}}_{\text{RR}}[n] \neq \mathbf{0}$. Even under ideal conditions, i.e., a nondispersive channel $\mathbf{H}_{\text{SR}}[n]$ and $\mathbf{n}_{\text{R}}[n] = \mathbf{0}$, the received signal cannot be assumed white [31]. To ease filtering requirements at the analog-to-digital and digital-to-analog conversion stages, the sampling rate at the relay is likely to exceed the Nyquist rate, which results in signals having correlated samples or non-flat PSDs. Also note that, in general, problem (34) has not a closed-form solution that allows us to use an off-line approach. An iterative process, such as an adaptive filter, is therefore required for solving the self-interference problem.

Even though the minimization of (31) with respect to $\boldsymbol{\mathcal{A}}$ does not directly lead to proper SI mitigation, we can still develop a recursive least-squares based algorithm by introducing a bias-correcting term that results in the desired interference cancellation properties [31]. For the sake of clarity, let first consider a bias-corrected minimization of (31), which was introduced in [33].

In particular, consider the following recursive update for $\boldsymbol{\mathcal{A}}$

$$\boldsymbol{\mathcal{A}}[n+1] = \boldsymbol{\mathcal{A}}[n] + \mu_a \left(\boldsymbol{\mathcal{R}}' - \nabla'_{\boldsymbol{\mathcal{A}}^*} \left\{\|\hat{\mathbf{r}}_r[n]\|^2\right\}\right) \tag{35}$$

where the $M \times ML_{\text{A}}$ constant matrix $\boldsymbol{\mathcal{R}}'$ represents the bias-correction term and $\nabla'_{\boldsymbol{\mathcal{A}}^*} \left\{\|\hat{\mathbf{r}}_r[n]\|^2\right\}$ is the gradient matrix, with size $M \times ML_{\text{A}}$, of $\|\hat{\mathbf{r}}_r[n]\|^2$ with respect to $\boldsymbol{\mathcal{A}}$. The stationary points of algorithm (35) satisfy

$$\mathbb{E}\{\nabla'_{\boldsymbol{\mathcal{A}}^*} \left\{\|\hat{\mathbf{r}}_t[n]\|^2\right\}\} = \boldsymbol{\mathcal{R}}' \tag{36}$$

consequently, adaptation rule (35) shifts the stationary points to the new location given by $\boldsymbol{\mathcal{R}}'$. In order to obtain an expression for the adaptation rule in (35), let us define the gradient of a matrix function $\mathbf{F}$ with respect to a matrix $\boldsymbol{\mathcal{A}}$ as [47]

$$\nabla_{\boldsymbol{\mathcal{A}}^*}\{\mathbf{F}(\boldsymbol{\mathcal{A}}^*)\} = \frac{\partial \operatorname{vec}\{\mathbf{F}\}}{\partial \operatorname{vec}^T\{\boldsymbol{\mathcal{A}}^*\}} \tag{37}$$

Note that when $\mathbf{F}$ is a scalar function, denoted by f, expression (37) has dimension $1 \times M^2 L_{\text{A}}$, whereas in (35) the gradient has dimension $M \times ML_{\text{A}}$. The relation between these two definitions is a simple component reordering,

$$\nabla_{\boldsymbol{\mathcal{A}}^*}\{f(\boldsymbol{\mathcal{A}}^*)\} = \operatorname{vec}^T\{\nabla'_{\boldsymbol{\mathcal{A}}^*}\{f(\boldsymbol{\mathcal{A}}^*)\}\} \tag{38}$$

where operator vec stacks together the columns of a matrix into a single column vector. Using (37) and (38), the gradient in (35) can be expressed as

$$\begin{aligned}\nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\|\hat{\mathbf{r}}_r[n]\|^2\right\} &= \nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\mathbf{r}_r^H[n]\hat{\mathbf{r}}_r[n]\right\} + \nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\hat{\boldsymbol{r}}_t^H[n-1]\boldsymbol{\mathcal{A}}^H\hat{\mathbf{r}}_r[n]\right\} \\ &= \hat{\mathbf{r}}_r^T[n]\boldsymbol{\mathcal{H}}_{\text{RR}}^*\nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\hat{\boldsymbol{r}}_i^H[n-1]\right\} + \mathbf{r}_t^T[n]\boldsymbol{\mathcal{A}}^*\nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\hat{\boldsymbol{r}}_t^H[n-1]\right\} + (\hat{\mathbf{r}}_r^T[n]\otimes\hat{\boldsymbol{r}}_t^H[n-1])\nabla_{\boldsymbol{\mathcal{A}}^*}\left\{\boldsymbol{\mathcal{A}}^H\right\}\end{aligned} \tag{39}$$

where $\otimes$ denotes the Kronecker product. We see that the gradient in (39) depends on both $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{H}}_{\text{RR}}$, which makes online gradient computation impossible in practice. To circumvent this issue, we assume that $\hat{\mathbf{r}}_t[n]$ is independent of $\boldsymbol{\mathcal{A}}$ in a similar fashion to linear regressions. This approximation is known as the *pseudo-linear regression* and allows us to neglect the nonlinear dependence induced by the two first terms of (39), see, e.g., [48]. Consequently, (39) can be approximated as

$$\nabla'_{\boldsymbol{\mathcal{A}}^*}\left\{\|\hat{\mathbf{r}}_t[n]\|^2\right\} \approx \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1] \tag{40}$$

and can be interpreted as the instantaneous crosscorrelation between signals $\hat{\mathbf{r}}_r[n]$ and $\hat{\mathbf{r}}_t[n]$. As discussed in Section II, $\hat{\mathbf{r}}_t[n]$ and $\hat{\mathbf{r}}_r[n]$ are related as

$$\hat{\mathbf{r}}_t[n] = \mathbf{B}\hat{\mathbf{r}}_r[n] \tag{41}$$

From Definition 1, we see that pre-multiplying the driving term of an adaptive algorithm by a full column-rank matrix does not alter the locations of the stationary points. Consequently, by first substituting (40) in (35) and then multiplying the driving term by $\mathbf{B}$, we obtain the following recursions

$$\begin{aligned}\boldsymbol{\mathcal{A}}[n+1] &= \boldsymbol{\mathcal{A}}[n] + \mu_a\mathbf{B}\left(\boldsymbol{\mathcal{R}}' - \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1]\right) \\ &= \boldsymbol{\mathcal{A}}[n] + \mu_a\left(\boldsymbol{\mathcal{R}} - \hat{\mathbf{r}}_t[n]\hat{\boldsymbol{r}}_t^H[n-1]\right)\end{aligned} \tag{42}$$

where $\boldsymbol{\mathcal{R}} = \mathbf{B}\boldsymbol{\mathcal{R}}'$. Equation (42) now satisfies the following stationary point condition

$$\mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\boldsymbol{r}}_t^H[n-1]\} = \boldsymbol{\mathcal{R}} \tag{43}$$

In other words, the algorithm in (42) adjusts the autocorrelation of $\hat{\mathbf{r}}_t[n]$, at lags 1 through L_{A}, to match the predetermined template matrix $\boldsymbol{\mathcal{R}} = [\boldsymbol{\mathcal{R}}(1)\ \cdots\ \boldsymbol{\mathcal{R}}(L_{\text{A}})]$. That is, at a stationary point we will have $\mathbf{R}_{\hat{\mathbf{r}}_t}(k) = \boldsymbol{\mathcal{R}}(k)$, $1 \leq k \leq L_{\text{A}}$, where the choice of $\boldsymbol{\mathcal{R}}(k)$ is at our disposal. To unambiguously adjust the autocorrelation of $\hat{\mathbf{r}}_t[n]$ we should extend the conditions in (43) to include the autocorrelation at lag zero $\mathbf{R}_{\hat{\mathbf{r}}_t}(0)$. This can be achieved by means of matrix $\mathbf{B}$ in the feedforward path, which, following the same procedure as with $\boldsymbol{\mathcal{A}}$, is updated according to

$$\mathbf{B}[n+1] = \mathbf{B}[n] + \mu_b\left(\boldsymbol{\mathcal{R}}(0) - \hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n]\right) \tag{44}$$

The adaptation rules in (42) and (44) constitute the filter-and-forward instantaneous autocorrelation shaping (FF-

TABLE II
INSTANTANEOUS AUTOCORRELATION SHAPING AND WEIGHTED INSTANTANEOUS AUTCORRELATION SHAPING ALGORITHMS FOR SI CANCELLATION IN FF RELAYS.

The FF-IAS algorithm
$\mathbf{B}[n+1] = \mathbf{B}[n] + \mu_b(\boldsymbol{\mathcal{R}}(0) - \hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n])$
$\boldsymbol{\mathcal{A}}[n+1] = \boldsymbol{\mathcal{A}}[n] + \mu_a(\boldsymbol{\mathcal{R}} - \hat{\mathbf{r}}_t[n]\hat{\boldsymbol{r}}_t^H[n])$
The FF-WIAS algorithm
$\mathbf{P}[0] = \epsilon^{-1}\mathbf{I},\ 0 < \epsilon \ll 1$
$\mathbf{K}[n] = \mathbf{P}[n-1]\hat{\boldsymbol{r}}_t[n-1]$
$\mathbf{P}[n] = \frac{1}{\lambda}\left(\mathbf{P}[n-1] - \frac{\mathbf{K}[n]\mathbf{K}^H[n]}{\lambda + \mathbf{K}^H[n]\hat{\boldsymbol{r}}_t[n-1]}\right)$
$\mathbf{B}[n] = \mathbf{B}[n-1] + \mu_b(\boldsymbol{\mathcal{R}}(0) - \hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n])$
$\boldsymbol{\mathcal{A}}[n] = \boldsymbol{\mathcal{A}}[n-1] + (\boldsymbol{\mathcal{R}} - \hat{\mathbf{r}}_t[n]\hat{\boldsymbol{r}}_t^H[n])\mathbf{P}[n]$

IAS) algorithm for SI cancellation. If L_{A} is chosen large enough, the FF-IAS algorithm can be seen as a spectrum shaping algorithm that forces the PSD of the output signal $\hat{\mathbf{r}}_t[n]$ to match the particular spectrum template given by the Fourier transform of $\boldsymbol{\mathcal{R}}(k)$. The proper selection of template matrices $\boldsymbol{\mathcal{R}}(k)$ is addressed in Section IV-B.

With the previous discussion in mind, we now derive a recursive least-squares based algorithm that exhibits better performance and faster convergence than the FF-IAS algorithm. The least-squares cost function with respect to $\boldsymbol{\mathcal{A}}$ is

$$\boldsymbol{\mathcal{A}}[n] = \arg\min_{\boldsymbol{\mathcal{A}}} \sum_{k=0}^{n} \lambda^{n-k} \|\hat{\mathbf{r}}_r[k]\|^2 \tag{45}$$

As was the case in (39), the gradient of (45) with respect to $\boldsymbol{\mathcal{A}}$ will also depend on $\boldsymbol{\mathcal{H}}_{\mathrm{RR}}$ and $\boldsymbol{\mathcal{A}}$. By invoking the pseudo-linear approximation, as in (40), the adaptation rule for $\boldsymbol{\mathcal{A}}$, including a bias correction term $\boldsymbol{\mathcal{R}}'$, is given by

$$\boldsymbol{\mathcal{A}}[n] = \boldsymbol{\mathcal{A}}[n-1] + (\boldsymbol{\mathcal{R}}' - \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1])\mathbf{P}[n] \tag{46}$$

where matrix $\mathbf{P}[n]$ is defined in (19). As explained before, we may pre-multiply the driving term in (46) by matrix $\mathbf{B}$ without altering the locations of the stationary points, i.e.,

$$\begin{aligned} \boldsymbol{\mathcal{A}}[n] &= \boldsymbol{\mathcal{A}}[n-1] + \mathbf{B}(\boldsymbol{\mathcal{R}}' - \hat{\mathbf{r}}_r[n]\hat{\boldsymbol{r}}_t^H[n-1])\mathbf{P}[n] \\ &= \boldsymbol{\mathcal{A}}[n-1] + (\boldsymbol{\mathcal{R}} - \hat{\mathbf{r}}_t[n]\hat{\boldsymbol{r}}_t^H[n-1])\mathbf{P}[n] \end{aligned} \tag{47}$$

Again, the particular choice of template matrix $\boldsymbol{\mathcal{R}}$ is discussed in Section IV-B. Note that, though based on a recursive least-squares principle, (44) and (47) do not correspond to a minimizer of (45). Nevertheless, as shown in Section IV-B, the algorithm results in proper cancellation of the SI. Equations (44) and (47) constitute the filter-and-forward weighted instantaneous autocorrelation shaping (FF-WIAS) algorithm for SI cancellation. The recursions of the FF-IAS and FF-WIAS algorithms are summarized in Table II.

B. Algorithm analysis

The algorithms in Table II require proper specification of matrices $\boldsymbol{\mathcal{R}}(0)$ and $\boldsymbol{\mathcal{R}} = [\boldsymbol{\mathcal{R}}(1) \cdots \boldsymbol{\mathcal{R}}(L_{\mathrm{A}})]$. These matrices are related to the autocorrelation of the transmitted signal $\hat{\mathbf{r}}_t[n]$ as summarized by the following proposition.

Proposition 1. *At any stationary point of the FF-IAS and FF-WIAS algorithms, the autocorrelation of* $\hat{\mathbf{r}}_t[n]$ *is given by the template matrices* $\boldsymbol{\mathcal{R}}(k)$ *in* (42)*,* (44)*, and* (47)*, i.e.,* $\mathbf{R}_{\hat{\mathbf{r}}_t}(k) := \mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n-k]\} = \boldsymbol{\mathcal{R}}(k)$ *for* $k = 0, \ldots, L_A$.

Proof. From (42), (44), and (47) it can be seen that any stationary point of the FF-IAS and FF-WIAS algorithms, denoted by $\mathbf{B}_\star$ and $\mathbf{A}_\star$, must satisfy:

I. $\mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n]\} = \boldsymbol{\mathcal{R}}[0]$, and;

II. $\mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\boldsymbol{r}}_t^H[n-1]\mathbf{Q}[n]\} = \boldsymbol{\mathcal{R}}\mathbb{E}\{\mathbf{Q}[n]\}$, where $\mathbf{Q}[n] = \mathbf{I}$ for the FF-IAS algorithm and $\mathbf{Q}[n] = \mathbf{P}[n]$ for the FF-WIAS algorithm.

It follows directly from Item I above that $\mathbf{R}_{\hat{\mathbf{r}}_t}(0) = \boldsymbol{\mathcal{R}}(0)$. To verify Item 2, we first assume that n is sufficiently large, which allows us to approximate $\mathbf{P}[n]$ with its expected value $\mathbb{E}\{\mathbf{P}[n]\}$ [46]. Therefore, $\mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\boldsymbol{r}}_t^H[n-1]\mathbf{P}[n]\} \approx \mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\boldsymbol{r}}_t^H[n-1]\}\mathbb{E}\{\mathbf{P}[n]\}$. Assuming that $\mathbb{E}\{\mathbf{P}[n]\}$ is invertible, it follows that the stationary points of the FF-IAS and FF-WIAS algorithms also satisfy $\mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n-k]\} = \boldsymbol{\mathcal{R}}(k)$, for $k = 1, \ldots, L_{\mathrm{A}}$. □

In view of Proposition 1, the FF-IAS and FF-WIAS algorithms shape the PSD of $\hat{\mathbf{r}}_t[n]$ to match the PSD associated with the predefined autocorrelation template. Designing $\boldsymbol{\mathcal{R}}(k)$ by exploiting knowledge of the (temporally correlated) signals observed at the relay is not feasible due to their dependency on the unknown channels $\mathbf{H}_{\mathrm{RR}}[n]$ and $\mathbf{H}_{\mathrm{SR}}[n]$. A better option is to exploit prior knowledge of the second-order statistics of the source signal $\hat{\mathbf{s}}_t[n]$. We propose here to choose $\boldsymbol{\mathcal{R}}(k) = \mathbf{R}_{\hat{\mathbf{s}}_t}(k) := \mathbb{E}\{\hat{\mathbf{s}}_t[n]\hat{\mathbf{s}}_t^H[n-k]\}$. It is reasonable to assume that during the design stage, the spectral characteristics at $\mathcal{S}$ are fixed and can be made available to every node. A relevant question is whether this choice of matrices $\boldsymbol{\mathcal{R}}(k)$ leads to SI cancellation upon algorithm convergence. We shall see that under certain conditions this particular choice will not only cancel the SI at the relay but also compensate for (i.e., equalize) the effects of the source-relay channel $\boldsymbol{\mathcal{H}}_{\mathrm{SR}}[n]$. In particular, we obtain the following result for $\hat{\mathbf{r}}_r[n]$.

Theorem 2. *Assume that the entries of* $\hat{\mathbf{s}}_t[n]$ *are mutually uncorrelated and with the same PSD,* L_A *is sufficiently large,* $P_{\mathbf{n}_R} \to 0$*, and the equivalent source-relay channel* $\mathbf{H}_{SR}(e^{j\omega})\mathbf{G}_S(e^{j\omega})$ *is of minimum-phase. Furthermore, let* $\boldsymbol{\mathcal{R}}(k) = \mathbf{R}_{\hat{\mathbf{s}}_t}(k) := \mathbb{E}\{\hat{\mathbf{s}}_t[n]\hat{\mathbf{s}}_t^H[n-k]\}$*. Then, any stationary point of the FF-IAS and FF-WIAS algorithms corresponding to a stable transfer function and with an invertible matrix* $\mathbf{B}$ *results in* $\hat{\mathbf{r}}_r[n] = \mathbf{V}\hat{\mathbf{s}}_t[n]$*, where* $\mathbf{V}$ *is a unitary matrix.*

Proof: See Appendix A.

Under the conditions of the theorem, we have $\mathbf{B}\check{\mathbf{H}}_{\mathrm{RR}}(e^{j\omega})\mathbf{H}_{\mathrm{SR}}(e^{j\omega})\mathbf{G}_{\mathrm{S}}(e^{j\omega}) = \mathbf{V}$ at any stationary point. This result has two implications: firstly, the SI is perfectly cancelled, and secondly, the $\mathcal{S}$-$\mathcal{R}$ channel is compensated. Simultaneous channel compensation and interference mitigation is a consequence of the restoration principle used

by the algorithm, which tries to restore the PSD of the source signal at the relay output. We note that the residual spatial rotation $\mathbf{V}$ has no impact on the system performance at $\mathcal{D}$. This is because the combined $\mathcal{R}$-$\mathcal{D}$ channel, $\mathbf{H}_{\text{RD}}(e^{j\omega})\mathbf{G}_{\text{R}}(e^{j\omega})\mathbf{V}$, yields the same SNR as the combined $\mathcal{R}$-$\mathcal{D}$ channel of a relay without SI, $\mathbf{H}_{\text{RD}}(e^{j\omega})\mathbf{G}_{\text{R}}(e^{j\omega})$. In case $\mathbf{H}_{\text{SR}}(e^{j\omega})\mathbf{G}_{\text{S}}(e^{j\omega})$ is a mixed-phase channel, only its minimum-phase part can be compensated and the overall system relating $\hat{\mathbf{s}}_t[n]$ and $\hat{\mathbf{r}}_t[n]$ will be a paraunitary function, i.e., $\mathbf{V}(e^{j\omega})$.

Remark 4: The resulting relay transfer function $\mathbf{B}\check{\mathbf{H}}_{\text{RR}}(e^{j\omega})$ can be interpreted as a three-stage filter: a first filter that whitens the signal $\check{\mathbf{r}}_r[n] + \mathbf{n}_{\text{R}}[n]$, followed by a unitary spatial rotation and, finally, a conformation filter that shapes the PSD conforming to $\{\boldsymbol{\mathcal{R}}(k)\}$, see (52) in Appendix A. Overall, from the perspective of $\mathcal{D}$, this approach may be suboptimal but without information about $\mathbf{H}_{\text{SD}}[n]$ and $\mathbf{H}_{\text{SR}}[n]$ at $\mathcal{R}$, it can be considered a "best-effort" solution.

Remark 5: When a noisy case is considered, the locations of the stationary points will shift and cause bias. Even in the case when $\mathbf{H}_{\text{SR}}(e^{j\omega})\mathbf{G}_{\text{S}}(e^{j\omega}) = \mathbf{I}$, $\check{\mathbf{H}}_{\text{RR}}(e^{j\omega})$ will be frequency dependent, resulting in residual SI. Unfortunately, the bias is related to the spectral factorization of both the information signal and noise, see (57) in Appendix A, and it becomes difficult to analytically characterize the resulting stationary points due to the nonlinear nature of the equations involved. Nevertheless, extensive simulations, provided in Section V, in noisy conditions confirm that the algorithms yield residual SI below noise level.

V. SIMULATIONS AND RESULTS

We consider a MIMO-OFDM system transmitting $M = 2$ independent streams of 64-QAM modulated data. The number of subcarriers is $N_s = 8192$ and the cyclic prefix length is $N_c = N_s/4$, which aims to overcome the relative propagation delay between the $\mathcal{S}$-$\mathcal{D}$ path and the $\mathcal{S}$-$\mathcal{R}$-$\mathcal{D}$ path. The sampling frequency is $F_s = k_{up}(N_c+N_s)F_{symb}$, where F_{symb} is an arbitrary OFDM symbol rate and $k_{up} = 1.15$ is the upsampling factor. With this choice of parameters, the signal spans approximately 87% of the available bandwidth. As we will see, the upsampling factor has an impact on the performance of the algorithms. The equivalent $\mathcal{S}$-$\mathcal{R}$ channel has order $L_{\text{SR}} = 2$ with a bulk delay of 5 samples accounting for propagation delay, i.e., $\mathbf{H}_{\text{SR}}(e^{j\omega}) = \mathbf{H}_{\text{SR}}(0)e^{-j5\omega}+\mathbf{H}_{\text{SR}}(1)e^{-j6\omega}+\mathbf{H}_{\text{SR}}(2)e^{-j7\omega}$, and each element of $\mathbf{H}_{\text{SR}}(k)$ is independently drawn from a zero-mean unit-variance circular complex Gaussian distribution [17].

The delay introduced by the analog filters and RF processing [31] in the relay is 5 samples and the SI channel $\hat{\mathbf{H}}_{\text{RR}}[n]$ has order $L_{\hat{\text{RR}}} = 2$, i.e., $\hat{\mathbf{H}}_{\text{RR}}(e^{j\omega}) = \hat{\mathbf{H}}_{\text{RR}}(1)e^{-j6\omega} + \hat{\mathbf{H}}_{\text{RR}}(2)e^{-j7\omega}$, and each element of $\hat{\mathbf{H}}_{\text{RR}}[k]$ is independently drawn from a zero-mean unit-variance circular complex Gaussian distribution up to a scaling factor, for $k = 1, \ldots, L_{\hat{\text{RR}}}$.

We respectively define the signal-to-interference ratio at the relay, the signal-to-interference ratio after cancellation and the signal-to-noise ratio as

$$\text{SIR}_{\mathcal{R}} = P_{\check{\mathbf{r}}_r}/P_{\check{\mathbf{i}}_r} \tag{48}$$

$$\hat{\text{SIR}}_{\mathcal{R}} = P_{\hat{\check{\mathbf{r}}}_r}/P_{\hat{\check{\mathbf{i}}}_r} \tag{49}$$

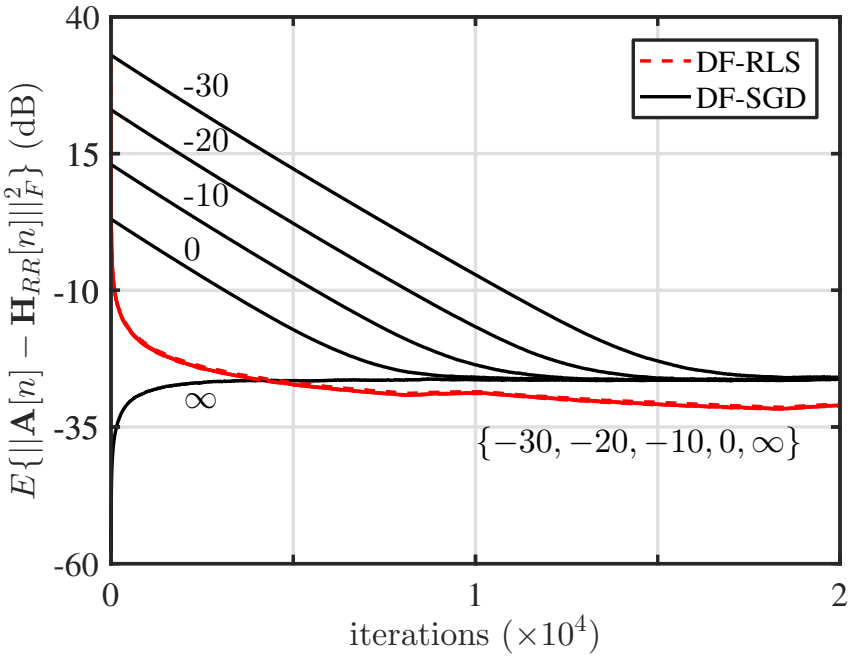

Fig. 2. Learning curves for different values of SI power. Vertical axis depicts the residual self-interference after cancellation.

$$\text{SNR}_{\mathcal{R}} = P_{\tilde{\mathbf{r}}_r}/P_{\mathbf{n}_\text{R}} \tag{50}$$

where $P_{\boldsymbol{\alpha}} \doteq \text{tr}\left(\mathbb{E}\{\boldsymbol{\alpha}[n]\boldsymbol{\alpha}^H[n]\}\right)$ denotes the power of a generic signal vector $\boldsymbol{\alpha}[n]$.

The receiver noise $\mathbf{n}_\text{R}[n]$ has PSD $\mathbf{S}_{\mathbf{n}_\text{R}}(e^{j\omega}) = \sigma_r^2\mathbf{I} + \sigma_t^2\mathbf{H}_{\text{RR}}(e^{j\omega})\mathbf{H}_{\text{RR}}^H(e^{j\omega})$, where $\sigma_r^2\mathbf{I}$ is the noise PSD at reception and $\sigma_t^2\mathbf{H}_{\text{RR}}(e^{j\omega})\mathbf{H}_{\text{RR}}^H(e^{j\omega})$ is the PSD of the transmit noise that couples back into $\mathcal{R}$. We consider a fixed SNR_t at the transmit side of $\text{SNR}_t = P_{\mathbf{r}_t}/(N_t\sigma_t^2) = 37$ dB, while $\text{SNR}_r = P_{\tilde{\mathbf{r}}_r}/(N_r\sigma_r^2)$ is variable. We evaluate the algorithm performance by measuring the residual interference after cancellation, $\hat{\text{SIR}}_{\mathcal{R}}$. The cancellation filter has order $L_\text{A} = L_{\hat{\text{RR}}}$ and the adaptive filter taps are all initialized with zeros. We also evaluate the algorithm convergence time, τ, which is defined as the number of iterations needed from initialization to a point satisfying $\|\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{H}}_{\text{RR}}\|_F^2/\|\boldsymbol{\mathcal{H}}_{\text{RR}}\|_F^2 < -25$ dB, i.e., the residual self-interference after cancellation is reduced more than 25 dB.

A. Results and discussion for the DF relay algorithms:

Figure 2 shows the learning curves of the algorithms when $\text{SNR}_r = 15$ dB and $\text{SIR}_{\mathcal{R}} = \{-30, -20, -10, 0, \infty\}$ dB. We see that the DF-RLS algorithm with $\lambda = 0.9999$ converges to the same steady-state residual SI independently of $\text{SIR}_{\mathcal{R}}$. On the other hand, the convergence of the DF-SGD algorithm with $\mu_a = 0.001$ slows down as $\text{SIR}_{\mathcal{R}}$ decreases, although the final steady-state residual SI does not depend on $\text{SIR}_{\mathcal{R}}$. Extensive simulations, for a wide range of values of SNR_r, confirm that both algorithms successfully reduces the SI well below noise level so that it can be considered negligible. Additionally, as pointed out in Section III-B, the performance of the DF-SGD and

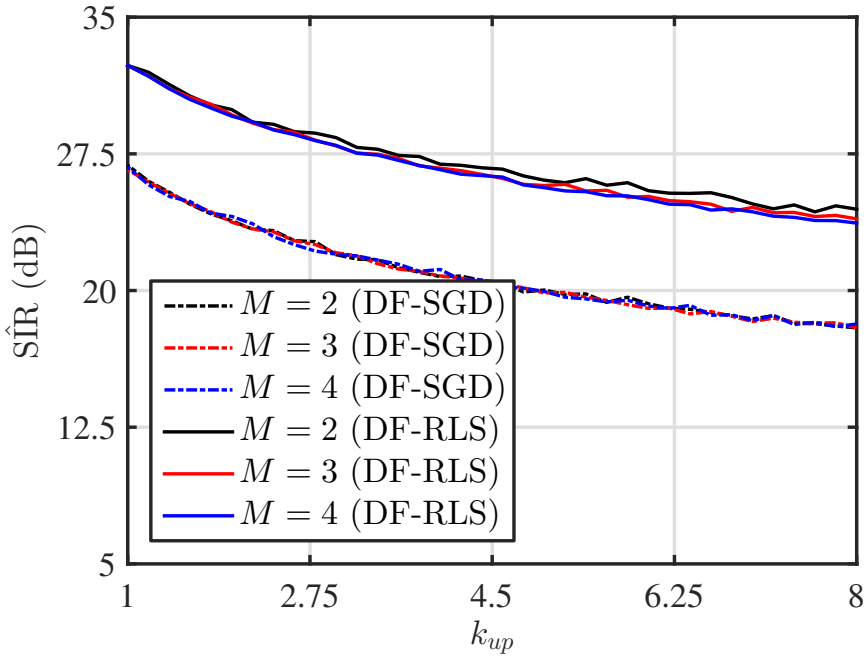

Fig. 3. Signal-to-interference ratio after cancellation as a function of the upsampling factor or the bandwidth of $\mathbf{S}_{\hat{\mathbf{r}}_t}(e^{j\omega})$.

DF-RLS algorithms will depend on other parameters, such as the spectral characteristics of $\hat{\mathbf{r}}_t[n]$. Figure 3 shows the SIR after cancellation versus the upsampling factor k_{up}, i.e., the bandwidth of $\mathbf{S}_{\mathbf{r}_t}(e^{j\omega})$, when $\text{SIR}_{\mathcal{R}} = -25$ dB. Note that $k_{up} = 1$ means that $\mathbf{r}_t[n]$ spans 100% of the available bandwidth, whereas $k_{up} = 8$ means that $\mathbf{r}_t[n]$ spans 12.5% of the available bandwidth. We see that signals of smaller bandwidths yield lower SIR values than wideband signals: approximately 8 dB for the DF-SGD algorithm, and 5 dB for the DF-RLS algorithm, which implies that μ_a and λ should be chosen proportional to the bandwidth of $\hat{\mathbf{r}}_t[n]$.

Figure 5 shows the contour lines of the convergence time, in terms of number of iterations, as a function of SNR_r and $\text{SIR}_{\mathcal{R}}$ for the DF-SGD algorithm when $\mu_a = 0.01$. We observe that the convergence time is approximately independent of SNR_r and primarily depends on $\text{SIR}_{\mathcal{R}}$. This is a consequence of $\check{\mathbf{r}}_r[n]$ being the major "noise source", from the algorithm's point-of-view. Typical values of convergence time are around 600 samples, which lies within the cyclic prefix of the OFDM symbol. Roughly speaking, the algorithm mitigates the SI signal by 25 dB every 600 samples, until the steady state is reached. The DF-RLS algorithm outperforms the DF-SGD algorithm in all cases. In general, the algorithms are fast enough to mitigate the SI within the duration of the cyclic prefix and, consequently, convergence time is not a limiting factor of the mitigation scheme.

As in every adaptive scheme, there is a tradeoff between the convergence time and the residual SI level. The value of μ_a, similarly λ, will determine the operation point in the tradeoff curve. Figure 4 shows the tradeoff curve for the DF-SGD algorithm (fixed μ_a) and the DF-RLS algorithm (fixed λ), when $\text{SIR}_{\mathcal{R}} = -25$ dB and $\text{SNR}_r = 15$. We see that a lower residual interference level is reached at the expense of slower convergence. In the case of (22),

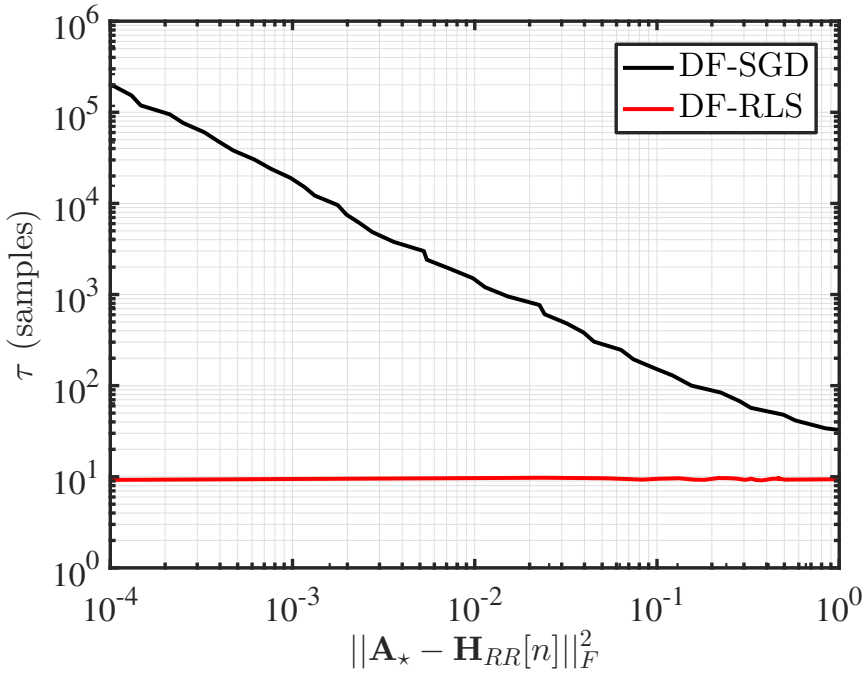

Fig. 4. Convergence speed versus steady-state performance for the DF-SGD and DF-RLS algorithms.

when $\lambda \approx 1$ lower residual interference is attained, while the convergence rate remains constant for any value of λ.

B. Results and discussion for the FF relay algorithms:

Figures 6 and 7 depict the contour lines of the residual SI after cancellation for the FF-IAS algorithm ($\mu_a = 10^{-5}$ and $\mu_b = 10\mu_a$) and the FF-WIAS algorithm ($\lambda = 0.9999$), respectively. In both cases, the residual interference only depends on SNR_r, i.e., noise power will bound the algorithm performance. As explained in Section IV-B, noise will introduce bias in the estimation of the SI channel. Despite this fact, the residual SI is at least 5 dB lower than the noise level, and even further below as SNR_r decreases. In other words, the residual SI is only noticeable for very high SNR_r values, which have little impact on the link performance. Therefore, under normal operation, the residual SI can be neglected. Also, the observed SNR is not changed after cancellation. We see that for this choice of parameters, the performance of the FF-WIAS algorithm is consistently better than that of the FF-IAS algorithm, especially for high SNR values.

Figures 8 and 9 show the convergence times for the FF-IAS and FF-WIAS algorithms, respectively. As was the case with the DF relay, convergence time is roughly independent of $\text{SNR}_{\mathcal{R}}$. In contrast, the convergence time increases for lower $\text{SIR}_{\mathcal{R}}$. Since the FF-IAS and FF-WIAS algorithms shape the PSD of the input signal, convergence time will depend on $\hat{\mathbf{H}}_{\text{RR}}[n]$. Lower values of $\text{SIR}_{\mathcal{R}}$, i.e., greater values of $\|\hat{\mathbf{H}}_{\text{RR}}[n]\|^2$ will result in a slower convergence. From Fig. 8, we see that when $\text{SIR}_{\mathcal{R}} = -20$ dB, the FF-IAS algorithm requires approximately 3 OFDM symbols to mitigate the interference by 35 dB. Comparing the two figures, we see that the FF-WIAS algorithm

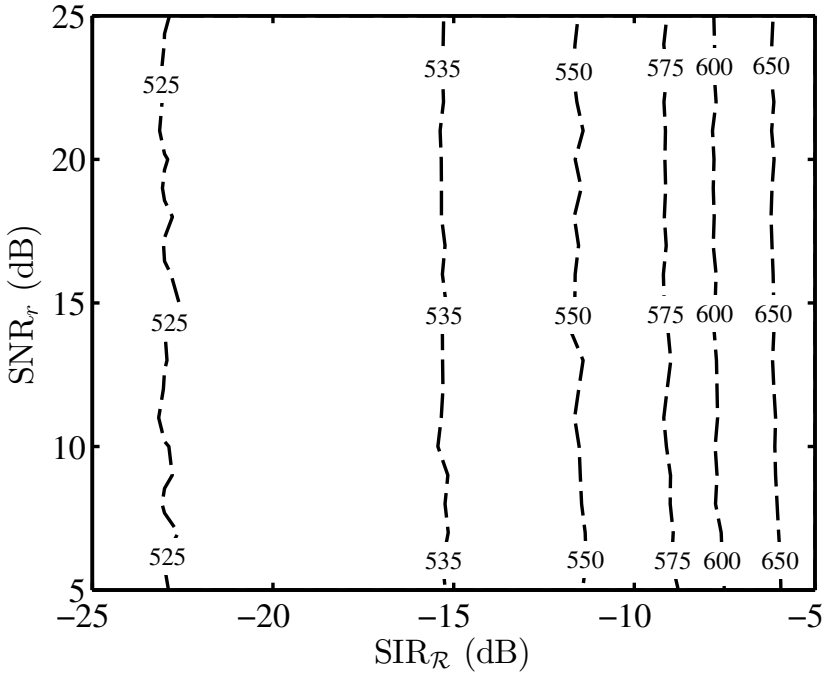

Fig. 5. Convergence time of the DF-SGD algorithm in number of iterations as a function of the signal-to-noise ratio (vertical axis) and the signal-to-interference ratio at the relay (horizontal axis).

convergences approximately seven times faster than the FF-IAS algorithm.

By Theorem 2 in Section III-B, the FF-IAS and FF-WIAS algorithms feature channel compensation capabilities. For purpose of illustration, FF-IAS algorithm is tested in a scenario where $\mathbf{H}_{\text{SR}}(e^{j\omega})\mathbf{G}_{\text{R}}(e^{j\omega})$ is a realization of a minimum-phase channel drawn from the following general channel expression: $\mathbf{H}_{\text{SR}}(e^{j\omega})\mathbf{G}_{\text{S}}(e^{j\omega}) = h_0\mathbf{H}_0e^{-j5\omega} + h_1\mathbf{H}_1e^{-j6\omega} + h_2\mathbf{H}_2e^{-j7\omega}$, with coefficients $h_0 = 1$, $h_1 = 0.4$ and $h_2 = 0.1$. The SI channel is randomly generated at each simulation run and the elements of $\hat{\mathbf{H}}_{\text{RR}}[k]$ and $\mathbf{H}_i$ are independently drawn from a zero-mean unit-variance circular complex Gaussian distribution. The coefficients of the FF-IAS algorithm are initialized to zero, step sizes are chosen as $\mu_a = 8 \cdot 10^{-5}$ and $\mu_b = 10\mu_a$, and filter order is set to $L_{\text{A}} = 2$.

Figure 10 shows the learning curves obtained for the FF-IAS algorithm (solid lines). For comparison, we also plot the curves obtained with a power minimization algorithm (dashed lines), i.e., by setting $\boldsymbol{\mathcal{R}}(k) = \mathbf{0}$ in the FF-IAS algorithm. We know from Section IV-B that such algorithm will result in a biased solution (since $\hat{\mathbf{s}}_t[n]$ is non-white), and Fig. 10 compares both algorithms. Note that $\mathbf{A}_\star$ denotes a point in which the overall system is a spatial rotation (see Theorem 2), so Fig. 10 can be understood as the residual multipath channel at $\hat{\mathbf{r}}_t[n]$. From Fig. 10 we can conclude that, firstly, convergence time depends on SIR$_\mathcal{R}$ (as already noted above) and, secondly, the bias resulting from direct power minimization leads to poor performance. The gap between the two algorithms is nearly 30 dB, and a bigger gap is expected for narrowband signals. Finally, from the obtained results, where the residual channel has gain of around -38 dB, we can conclude that the FF-IAS algorithm effectively compensates the $\mathcal{S}$-$\mathcal{R}$ channel while mitigating the SI.

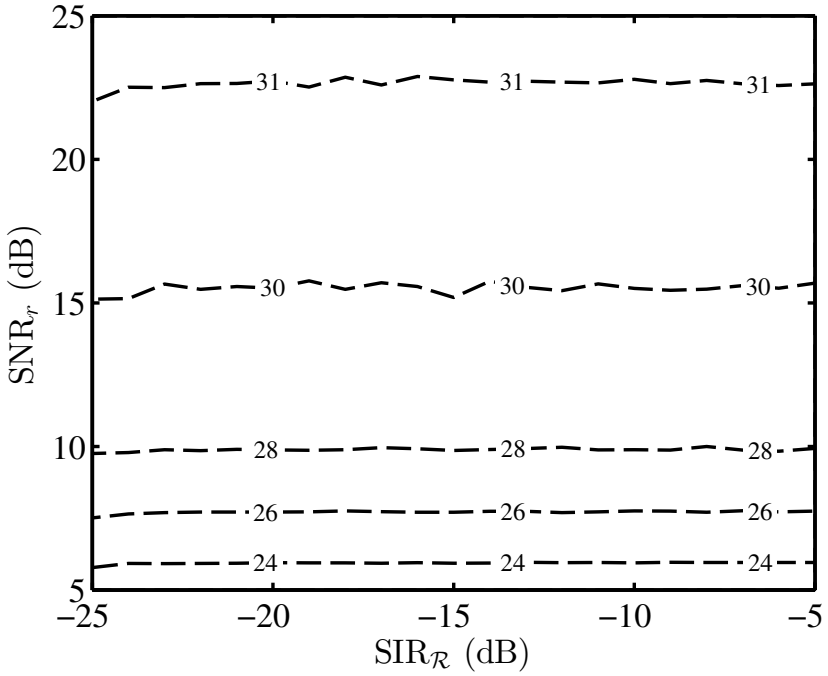

Fig. 6. SIR after cancellation for the FF-IAS algorithm as a function of the signal-to-noise ratio (vertical axis) and the signal-to-interference ratio at the relay (horizontal axis).

TABLE III
RESIDUAL SELF-INTERFERENCE IN DB AS A FUNCTION OF $L_{\hat{\text{RR}}}$.

$\hat{\text{SIR}}_{\mathcal{R}} \setminus L_{\hat{\text{RR}}}$	2	4	6	8	10
FF-WIAS	30.87	25.76	22.38	20.37	18.71
FF-IAS	28.41	23.51	20.18	17.69	15.98

The achievable $\hat{\text{SIR}}_{\mathcal{R}}$ depends on several factors related to the adaptive implementation, and we expect to see the usual performance trade-off between algorithm parameters such as step size (or memory) and filter order [42]. In particular, if we consider the step size of the algorithm fixed and sufficiently small (equivalently, $\lambda \to 1$), then increasing $L_{\hat{\text{RR}}}$ (equivalently, increasing L_{A}) will reduce the steady-state $\hat{\text{SIR}}_{\mathcal{R}}$ and increase convergence time. Table III shows the obtained $\hat{\text{SIR}}_{\mathcal{R}}$ versus $L_{\hat{\text{RR}}}$ for the case when $\text{SNR}_{\mathcal{R}} = 10$ dB and $\text{SIR}_{\mathcal{R}} = -10$ dB. We see that as $L_{\hat{\text{RR}}}$ increases (L_{A} increases), the achievable $\hat{\text{SIR}}_{\mathcal{R}}$ reduces due to, primarily, the misadjustment of the algorithm. In practice, the adaptation step size (or forgetting factor) is tuned to account for the existing trade-off between the convergence time and $\hat{\text{SIR}}_{\mathcal{R}}$ (see Fig. 4 for a trade-off curve).

VI. CONCLUSIONS

We have proposed adaptive SI mitigation algorithms for filter-and-forward and decode-and-forward full-duplex MIMO relays. In the decode-and-forward case, we derived and analyzed two power minimization based algorithms

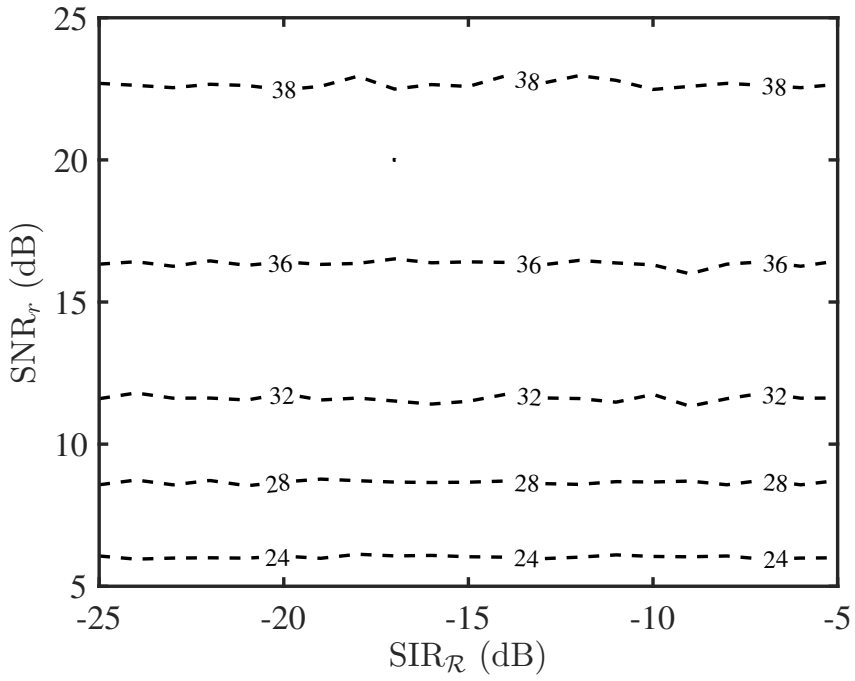

Fig. 7. SIR after cancellation for the FF-WIAS algorithm as a function of the signal-to-noise ratio (vertical axis) and the signal-to-interference ratio at the relay (horizontal axis).

that effectively mitigate the SI signal. On the other hand, we showed that for the case of filter-and-forward relays, a power minimization approach results in biased estimation of the SI channel. To overcome the bias problem, we derived and analyzed two algorithms that estimate the SI channel without bias as well as equalize the source-to-relay channel by only exploiting the second-order statistics of the source signal. Simulations confirmed that the proposed algorithms reduce the residual SI signal to levels below that of the noise.

APPENDIX A
STATIONARY POINTS OF FF-IAS AND FF-WIAS

Let $\boldsymbol{\alpha}[n]$ be an arbitrary vector of scalar wide-sense stationary sequences, with $\mathbf{R}_{\boldsymbol{\alpha}}(k) = \mathbb{E}\{\boldsymbol{\alpha}[n]\boldsymbol{\alpha}^H[n-k]\}$ denoting its autocorrelation matrix. The power spectral density (PSD) of $\boldsymbol{\alpha}[n]$ is defined as the Z-transform $\mathbf{S}_{\boldsymbol{\alpha}}(z)$, or Fourier transform $\mathbf{S}_{\boldsymbol{\alpha}}(e^{j\omega})$, of its autocorrelation matrix. The PSD can be decomposed as $\mathbf{S}_{\boldsymbol{\alpha}}(z) = \boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(z)\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}^H(1/z^*)$, where $\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(z)$ is a causal minimum-phase filter for which all poles are inside the unit circle ($|z| < 1$) and has full rank for $|z| \geq 1$. Evaluating $\mathbf{S}_{\boldsymbol{\alpha}}(z)$ on the unit-circle, i.e., for $z = e^{j\omega}$, we obtain $\mathbf{S}_{\boldsymbol{\alpha}}(e^{j\omega}) = \boldsymbol{\Gamma}_{\boldsymbol{\alpha}}(e^{j\omega})\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}^H(e^{j\omega})$. Finally, let $\mathbf{V}(z)$ denote a paraunitary matrix [49], i.e., $\mathbf{V}(z)\mathbf{V}^H(1/z^*) = \mathbf{V}^H(1/z^*)\mathbf{V}(z) = \mathbf{I}$. In case $\mathbf{V}(e^{j\omega}) = \mathbf{V}$ is constant for all frequencies, $\mathbf{V}$ is a unitary matrix.

Proof of Theorem 2. At a stationary point, if the resulting system is stable, we have from Proposition 1 that $\mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n-k]\} = \boldsymbol{\mathcal{R}}[k]$, for $k = 0, \ldots, L_{\mathrm{A}}$. Consequently, by choosing $\boldsymbol{\mathcal{R}}[k] = \mathbb{E}\{\hat{\mathbf{s}}_t[n]\hat{\mathbf{s}}_t^H[n-k]\}$ the

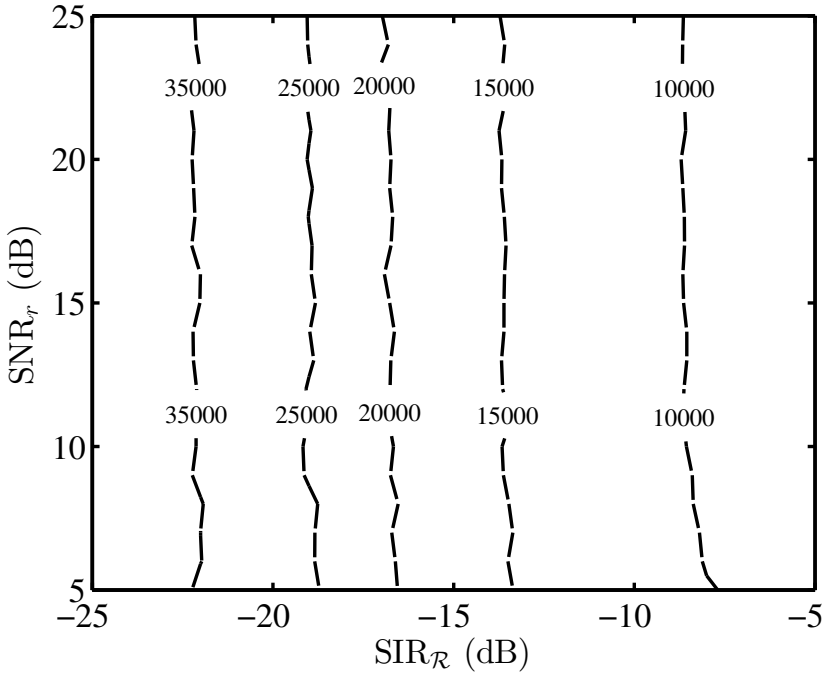

Fig. 8. Convergence time of the FF-IAS algorithms in number of iterations as a function of the signal-to-noise ratio (vertical axis) and the signal-to-interference ratio at the relay (horizontal axis).

resulting PSD of $\hat{\mathbf{r}}_t[n]$ equals $\mathbf{S}_{\hat{\mathbf{s}}_t}(e^{j\omega})$, i.e.,

$$\mathbf{S}_{\hat{\mathbf{s}}_t}(e^{j\omega}) = \mathbf{B}_\star \check{\mathbf{H}}_\star(e^{j\omega}) \mathbf{S}_{\check{\mathbf{r}}_r}(e^{j\omega}) \check{\mathbf{H}}_\star^H(e^{j\omega}) \mathbf{B}_\star^H \tag{51}$$

where $\check{\mathbf{H}}_\star(e^{j\omega}) = (\mathbf{I} - (\mathbf{A}_\star(e^{j\omega}) - \hat{\mathbf{H}}_{\text{RR}}(e^{j\omega}))\mathbf{B}_\star)^{-1}$, see (33). The overall relay response $\mathbf{B}_\star \check{\mathbf{H}}_\star(e^{j\omega})$ is necessarily causal by construction, and its inverse is given by $\mathbf{B}_\star^{-1} - (\mathbf{A}_\star(e^{j\omega}) - \hat{\mathbf{H}}_{\text{RR}}(e^{j\omega}))$, which is clearly causal as well. Thus $\mathbf{B}_\star \check{\mathbf{H}}_\star(e^{j\omega})$ is minimum phase. By assumption, $\mathbf{B}_\star$ is invertible at any stationary point. Any solution of (51) can be expressed in the Z-domain as [50]

$$\mathbf{B}_\star \check{\mathbf{H}}_\star(z) = \mathbf{\Gamma}_{\hat{\mathbf{s}}_t}(z) \mathbf{V}(z) \mathbf{\Gamma}_{\check{\mathbf{r}}_r}^{-1}(z) \tag{52}$$

for some paraunitary $\mathbf{V}(z)$, and with $\mathbf{\Gamma}_{\hat{\mathbf{s}}_t}(z)$, $\mathbf{\Gamma}_{\check{\mathbf{r}}_r}(z)$ the minimum phase spectral factors of $\mathbf{S}_{\hat{\mathbf{s}}_t}(z)$ and $\mathbf{S}_{\check{\mathbf{r}}_r}(z)$, respectively. Since $\mathbf{B}_\star \check{\mathbf{H}}_\star(z)$, $\mathbf{\Gamma}_{\hat{\mathbf{s}}_t}(z)$ and $\mathbf{\Gamma}_{\check{\mathbf{r}}_r}^{-1}(z)$ are causal transfer functions, it follows from (52) that $\mathbf{V}(z)$ must be a *causal* paraunitary matrix. On the other hand, the inverse system $(\mathbf{B}_\star \check{\mathbf{H}}_\star(z))^{-1} = \mathbf{\Gamma}_{\check{\mathbf{r}}_r}(z) \mathbf{V}^H(1/z*) \mathbf{\Gamma}_{\hat{\mathbf{s}}_t}^{-1}(z)$, is also causal and stable, and thus, by the same argument, $\mathbf{V}^H(1/z*)$ must be causal as well. It follows that $\mathbf{V}(z) = \mathbf{V}$, i.e., a unitary matrix independent of z. Using $\mathbf{\Gamma}_{\check{\mathbf{r}}_r}(z) = \gamma(z) \mathbf{H}_{\text{SR}}(z) \mathbf{G}_{\text{S}}(z)$, where $\gamma(z)$ is the minimum-phase factor of $\mathbf{S}_{\hat{\mathbf{s}}_t}(e^{j\omega})$, i.e., $\mathbf{S}_{\hat{\mathbf{s}}_t}(z) = \gamma(z)\gamma^*(1/z^*)\mathbf{I}$, then (52) is in the frequency domain

$$\mathbf{B}_\star \check{\mathbf{H}}_\star(e^{j\omega}) = \mathbf{V}(\mathbf{H}_{\text{SR}}(e^{j\omega}) \mathbf{G}_{\text{S}}(e^{j\omega}))^{-1} \tag{53}$$

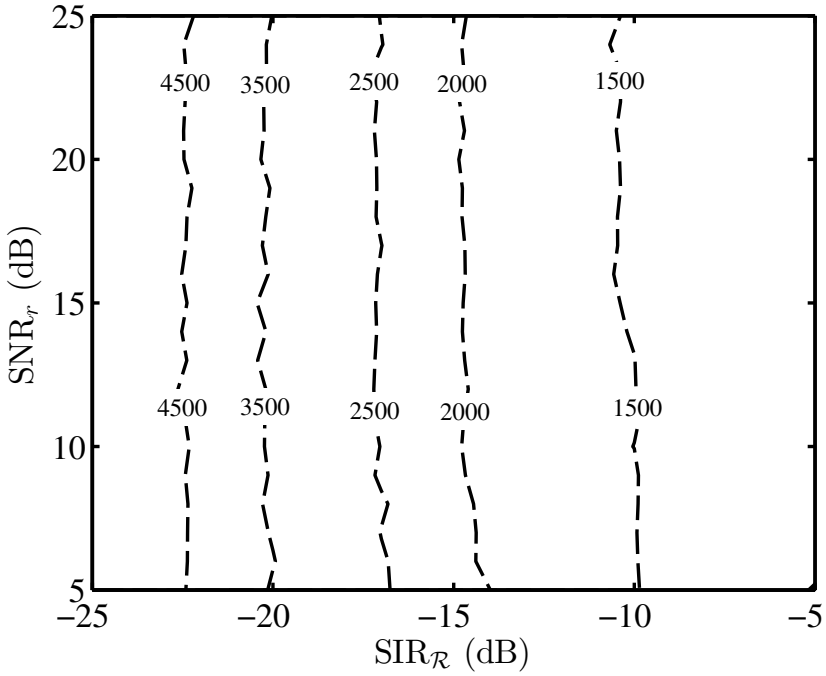

Fig. 9. Convergence time of the FF-WIAS algorithm in number of iterations as a function of the signal-to-noise ratio (vertical axis) and the signal-to-interference ratio at the relay (horizontal axis).

From (53) it follows that $\hat{\mathbf{r}}_t[n] = \mathbf{V}\hat{\mathbf{s}}_t[n]$.

□

When $\mathbf{H}_{\text{SR}}(e^{j\omega})\mathbf{G}_{\text{S}}(e^{j\omega})$ is not a minimum-phase channel, then only its minimum-phase component can be compensated, and $\mathbf{S}_{\hat{\mathbf{r}}_t}(e^{j\omega}) = \mathbf{V}(e^{j\omega})\mathbf{S}_{\hat{\mathbf{s}}_t}(e^{j\omega})$ for some causal paraunitary system. Also, if L_{A} is not sufficiently large, the autocorrelation of $\hat{\mathbf{r}}_t[n]$ equals that of $\hat{\mathbf{s}}_t[n]$ only for the first $L_{\text{A}}+1$ lags and (51) transforms into

$$\mathbf{S}_{\hat{\mathbf{s}}_t}(e^{j\omega})|_{L_{\text{A}}} + \mathbf{S}_e(e^{j\omega}) + \mathbf{S}_e^H(e^{j\omega}) = \mathbf{B}_\star\check{\mathbf{H}}_\star(e^{j\omega})\mathbf{S}_{\check{\mathbf{r}}_r}(e^{j\omega})\check{\mathbf{H}}_\star^H(e^{j\omega})\mathbf{B}_\star^H \tag{54}$$

where $\mathbf{S}_{\hat{\mathbf{s}}_t}(e^{j\omega})|_{L_{\text{A}}} = \sum_{k=-L_{\text{A}}}^{L_{\text{A}}} \mathbb{E}\{\hat{\mathbf{s}}_t[n]\hat{\mathbf{s}}_t^H[n-k]\}e^{-j\omega k}$ and $\mathbf{S}_e(e^{j\omega}) = \sum_{k=L_{\text{A}}+1} \mathbb{E}\{\hat{\mathbf{r}}_t[n]\hat{\mathbf{r}}_t^H[n-k]\}e^{-j\omega k}$ depends on $\mathbf{B}_\star$ and $\mathbf{A}_\star$. In the case that $\mathbf{H}_{\text{SR}}(e^{j\omega})\mathbf{G}_{\text{S}}(e^{j\omega})$ is a minimum-phase channel, then

$$\mathbf{B}_\star\check{\mathbf{H}}_\star(e^{j\omega}) = (\mathbf{H}_{\text{SR}}(e^{j\omega})\mathbf{G}_{\text{S}}(e^{j\omega}))^{-1} \tag{55}$$

is a solution of (54), for any $\mathbf{\Gamma}_{\hat{\mathbf{s}}_t}(e^{j\omega})$. Equation (55) requires that $L_{\text{A}} \geq \max\{L_{\hat{\text{RR}}}, L_{\text{SR}}+L_{\text{S}}\}$. Thus, the minimum required order for SI cancellation and channel compensation is $L_{\text{A}} = \max\{L_{\hat{\text{RR}}}, L_{\text{SR}}+L_{\text{S}}\}$. Finally, when $\text{SNR}_{\mathcal{R}} < \infty$, the stationary points are shifted to a new location given by

$$\mathbf{B}_\star\check{\mathbf{H}}_\star(e^{j\omega}) = \mathbf{\Gamma}_{\hat{\mathbf{s}}_t}(e^{j\omega})\mathbf{V}\mathbf{\Gamma}^{-1}_{\check{\mathbf{r}}_r+\mathbf{n}_{\text{R}}}(e^{j\omega}) \tag{56}$$

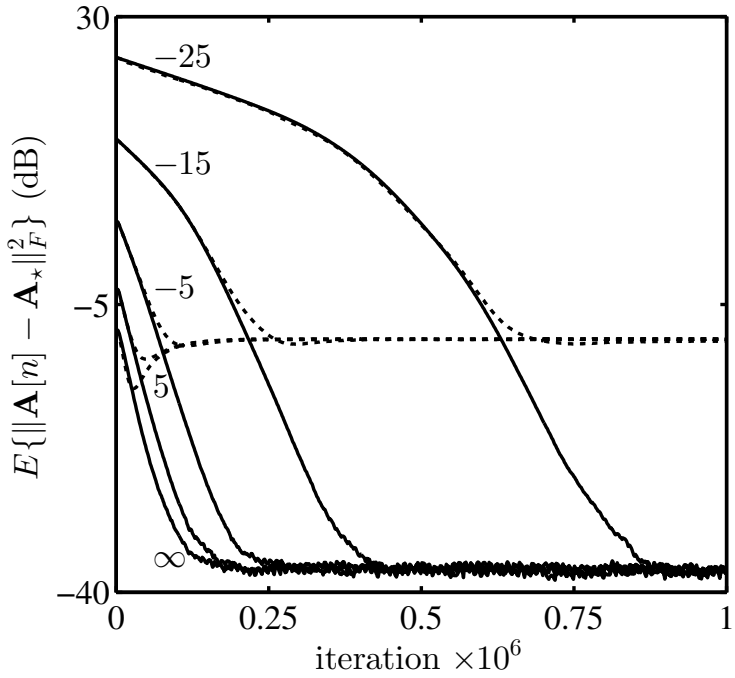

Fig. 10. Compensation of the $\mathcal{S}$-$\mathcal{R}$ channel and SI cancellation of algorithm FF-IAS. Vertical axis depicts the steady-state performance.

which, in general, results in a biased solution. The bias, $\boldsymbol{\Delta}$, is

$$\boldsymbol{\Delta} = \boldsymbol{\Gamma}_{\hat{\mathbf{s}}_t}(e^{j\omega})\mathbf{V}(\boldsymbol{\Gamma}^{-1}_{\check{\mathbf{r}}_r+\mathbf{n}_\mathrm{R}}(e^{j\omega}) - \boldsymbol{\Gamma}^{-1}_{\check{\mathbf{r}}_r}(e^{j\omega})) \tag{57}$$

Unfortunately, there is no closed-form relation between $\boldsymbol{\Gamma}_{\check{\mathbf{r}}_r+\mathbf{n}_\mathrm{R}}(e^{j\omega})$ and $\boldsymbol{\Gamma}_{\check{\mathbf{r}}_r}(e^{j\omega})$ that could provide deeper knowledge of the properties of (57).

ACKNOWLEDGMENT

Emilio Antonio-Rodríguez, Stefan Werner, Taneli Riihonen and Risto Wichman were funded by the Academy of Finland under project #258364 "In-band full-duplex MIMO transmission: A breakthrough to high-speed low-latency mobile networks". R. López-Valcarce was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) under projects COMPASS (TEC2013-47020-C2-1-R) and COMONSENS (TEC2015-69648-REDC), and by the Galician Regional Government and ERDF under projects "Consolidation of Research Units" (GRC2013/009), REdTEIC (R2014/037) and AtlantTIC.

REFERENCES

[1] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[2] A. Mattsson, "Single frequency networks in DTV," *IEEE Trans. Broadcast.*, vol. 51, no. 4, pp. 413–422, Dec. 2005.

[3] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[4] M. Iwamura, H. Takahashi, and S. Nagata, "Relay technology in LTE-advanced," *NTT DoCoMo Technical Journal*, vol. 12, no. 2, pp. 29–36, Sep. 2010.

[5] C. Xing, S. Ma, M. Xia, and Y.-C. Wu, "Cooperative beamforming for dual-hop amplify-and-forward multi-antenna relaying cellular networks," *Signal Processing*, vol. 92, no. 11, pp. 2689–2699, Nov. 2012.

[6] N. Bornhorst and M. Pesavento, "Filter-and-forward beamforming with adaptive decoding delays in asynchronous multi-user relay networks," *Signal Processing*, vol. 109, pp. 132–147, Apr. 2015.

[7] H. V. Khuong and T. Le-Ngoc, "A bandwidth-efficient cooperative relaying scheme with space-time block coding and iterative decoding," *Signal Processing*, vol. 89, no. 10, pp. 20 006–2012, Oct. 2009.

[8] S. Alabed, M. Pesavento, and A. Klein, "Non-coherent distributed spacetime coding techniques for two-way wireless relay networks," *Signal Processing*, vol. 93, no. 12, pp. 3371–3381, Dec. 2013.

[9] M. Yu and J. Li, "Is amplify-and-forward practically better than decode-and-forward or vice versa?" in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, vol. 3, Mar. 2005, pp. 365–368.

[10] M. Souryal and B. Vojcic, "Performance of amplify-and-forward and decode-and-forward relaying in Rayleigh fading with turbo codes," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, vol. 4, May 2006.

[11] T. Riihonen, S. Werner, and R. Wichman, "Hybrid full-duplex/half-duplex relaying with transmit power adaptation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 3074–3085, Sep. 2011.

[12] D. W. Bliss, P. A. Parker, and A. R. Margetts, "Simultaneous transmission and reception for improved wireless network performance," in *Proc. IEEE/SP Workshop on Statistical Signal Process. (SSP)*, Aug. 2007, pp. 478–482.

[13] K. Haneda, E. Kahra, S. Wyne, C. Icheln, and P. Vainikainen, "Measurement of loop-back interference channels for outdoor-to-indoor full-duplex radio relays," in *Proc. European Conf. on Antennas and Propagation (EuCAP)*, Apr. 2010.

[14] B. Day, A. Margetts, D. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1541–1553, Sep. 2012.

[15] B. Radunovic, D. Gunawardena, P. Key, A. Proutiere, N. Singh, V. Balan, and G. DeJean, "Rethinking indoor wireless mesh design: Low power, low frequency, full-duplex," in *Proc. IEEE Workshop on Wireless Mesh Networks (WIMESH)*, Jun. 2010.

[16] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. Conf. on Signals, Syst. and Comput. (ASILOMAR)*, Nov. 2010, pp. 1558–1562.

[17] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Dec. 2011.

[18] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, 2013, accepted for publication.

[19] N. Phungamngern, P. Uthansakul, and M. Uthansakul, "Digital and RF interference cancellation for single-channel full-duplex transceiver using a single antenna," in *Proc. Int. Conf. on Elect. Eng./Electron., Comput., Telecommun. and Inf. Technology (ECTI-CON)*, May 2013.

[20] P. Lioliou, M. Viberg, M. Coldrey, and F. Athley, "Self-interference suppression in full-duplex MIMO relays," in *Proc. Conf. on Signals, Syst. and Comput. (ASILOMAR)*, Nov. 2010, pp. 658–662.

[21] D. Senaratne and C. Tellambura, "Beamforming for space division duplexing," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, June 2011.

[22] B. Chun, E.-R. Jeong, J. Joung, Y. Oh, and Y. H. Lee, "Pre-nulling for self-interference suppression in full-duplex relays," in *Proc. Asia-Pacific Signal & Inform. Process. Assoc. Annu. Summit and Conf. (APSIPA ASC)*, Oct. 2009, pp. 91–97.

[23] P. Larsson and M. Prytz, "MIMO on-frequency repeater with self-interference cancellation and mitigation," in *Proc. IEEE Veh. Technology Conf. (VTC)*, Apr. 2009.

[24] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[25] H. Hamazumi, K. Imamura, N. Iai, K. Shibuya, and M. Sasaki, "A study of a loop interference canceller for the relay stations in an SFN for digital terrestrial broadcasting," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, vol. 1, 2000, pp. 167–171.

[26] N. Li, W. Zhu, and H. Han, "Digital interference cancellation in single channel, full duplex wireless communication," in *Proc. Int. Conf. on Wireless Commun., Networking and Mobile Computing (WiCOM)*, Sep. 2012.

[27] H. Sakai, T. Oka, and K. Hayashi, "A simple adaptive filter method for cancellation of coupling wave in OFDM signals at SFN relay station," in *Proc. European Signal Proces. Conf. (EUSIPCO)*, Sep. 2006.

[28] K. Hayashi, Y. Fujishima, M. Kaneko, H. Sakai, R. Kudo, and T. Murakami, "Self-interference canceller for full-duplex radio relay station using virtual coupling wave paths," in *Proc. Asia-Pacific Signal & Inform. Process. Assoc. Annu. Summit and Conf. (APSIPA ASC)*, Dec. 2012.

[29] D. Morgan and Z. Ma, "A same-frequency cellular repeater using adaptive feedback cancellation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 3825–3830.

[30] D. R. Morgan, M. G. Zierdt, D. A. Gudovskiy, J. Z. Pastalan, and Z. Ma, "FPGA implementation of a same-frequency cellular repeater using adaptive feedback cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, May 2013.

[31] R. Lopez-Valcarce, E. Antonio-Rodriguez, C. Mosquera, and F. Perez-Gonzalez, "An adaptive feedback canceller for full-duplex relays based on spectrum shaping," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1566–1577, Sep. 2012.

[32] E. Antonio-Rodríguez, R. López-Valcarce, T. Riihonen, S. Werner, and R. Wichman, "Autocorrelation-based adaptation rule for feedback equalization in wideband full-duplex amplify-and-forward MIMO relays," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, May 2013.

[33] E. Antonio-Rodríguez, R. López-Valcarce, T. Riihonen, S. Werner, and R. Wichman, "Adaptive self-interference cancellation in wideband full-duplex decode-and-forward MIMO relays," in *Proc. IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, Jun. 2013, pp. 370–374.

[34] C. Studer, M. Wenk, and A. Burg, "Mimo transmission with residual transmit-RF impairments," in *International ITG Workshop on Smart Antennas (WSA)*, Feb 2010, pp. 189–196.

[35] F. Gregorio, S. Werner, T. Laakso, and J. Cousseau, "Receiver cancellation technique for nonlinear power amplifier distortion in SDMA-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2499–2516, September 2007.

[36] H. Rowe, "Memoryless nonlinearities with Gaussian inputs: Elementary results," *The Bell System Technical Journal*, vol. 61, no. 7, pp. 1519–1525, Sept 1982.

[37] D. Dardari, V. Tralli, and A. Vaccari, "A theoretical characterization of nonlinear distortion effects in OFDM systems," *IEEE Transactions on Communications*, vol. 48, no. 10, pp. 1755–1764, October 2000.

[38] D. Bliss, T. Hancock, and P. Schniter, "Hardware phenomenological effects on cochannel full-duplex mimo relay performance," in *Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov 2012, pp. 34–39.

[39] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 375–386, Aug. 2013.

[40] C. H. Ta and S. Weiss, "A design of precoding and equalisation for broadband MIMO systems," in *Proc. Conf. on Signals, Syst. and Comput. (ASILOMAR)*, Nov. 2007, pp. 1616–1620.

[41] W. Al-Hanafy and S. Weiss, "Comparison of precoding methods for broadband MIMO systems," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Process. (CAMSAP)*, Dec. 2009, pp. 388–391.

[42] P. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 4th ed., ser. The International Series in Engineering and Computer Science. Springer, Aug. 2012, vol. 694.

[43] S. Haykin, *Adaptive filter theory*, 4th ed., ser. Prentice-Hall information and system sciences series. Prentice Hall, 2002.

[44] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes," *IEEE Trans. Autom. Control*, vol. 22, no. 4, pp. 539–551, Aug 1977.

[45] ——, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. 22, no. 4, pp. 551–575, Aug. 1977.

[46] A. H. Sayed, *Adaptive Filters*, 1st ed. John Wiley & Sons, Apr. 2011.

[47] A. Hjørungnes, *Complex-valued matrix derivatives: With Applications in Signal Processing and Communications*. Cambridge University Press, Mar. 2011.

[48] L. Ljung, *System Identification: Theory for the User*, 2nd ed., ser. Prentice-Hall information and system sciences series. Prentice-Hall PTR, Dec. 1999.

[49] J. Foster, J. McWhirter, M. Davies, and J. Chambers, "An algorithm for calculating the QR and singular value decompositions of polynomial matrices," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1263–1274, Mar. 2010.

[50] L. Crone, "Second order adjoint matrix equations," *Linear Algebra and its Applications*, vol. 39, pp. 61–71, Aug. 1981.