# Rapidly Time-Varying Channel Estimation for Full-Duplex Amplify-and-Forward One-Way Relay Networks

Habib Şenol [ID], *Member, IEEE*, Xiaofeng Li [ID], and Cihan Tepedelenlioğlu [ID], *Member, IEEE*

*Abstract*—Estimation of both cascaded and residual self-interference (RSI) channels and a new training frame structure are considered for full-duplex (FD) amplify-and-forward (AF) one-way relay networks with rapidly time-varying individual channels. To estimate the RSI and the rapidly time-varying cascaded channels, we propose a new training frame structure in which orthogonal training blocks are sent by the source node and delivered to the destination over an FD-AF relay. Exploiting the orthogonality of the training blocks, we obtain two decoupled training signal models for the estimation of the RSI and the cascaded channels. We apply linear minimum mean square error (MMSE) based estimators to the cascaded channel as well as RSI channel. In order to investigate the mean square error (MSE) performance of the system, we also derive the Bayesian Cramer–Rao lower bound. As another performance benchmark, we also assess the symbol error rate (SER) performances corresponding to the estimated and the perfect channel state information available at the receiver side. Computer simulations exhibit the proposed training frame structure and the linear MMSE estimator MSE and SER performances are shown.

*Index Terms*—Full duplex, one way relay, self interference, time varying, channel estimation.

## I. INTRODUCTION

EFFICIENT use of the radio spectrum has become more important because of the ever-increasing demands on the limited wireless bandwidth. From this perspective, recently, in-band full-duplex communication radio has gained significant attention since its spectral efficiency, as a measure of number of information bits reliably communicated per second per Hz, doubles that of half-duplex communication radio [1]–[3]. Conventional communication systems operating in half-duplex radio, transmit and receive in different time slots or over different frequency bands whereas in-band full-duplex communication systems transceive simultaneously over the same frequency band. Due to the so-called self-interference (SI) that

is caused by the fact that a full-duplex transceiver will also receive its own transmit signal, in-band full-duplex communication systems have not been used in widespread practice so far. However, recent achievements in SI cancelation [4], [5] may allow widespread use of in-band full-duplex which is a suitable candidate to meet higher spectral efficiency needs of upcoming fifth generation (5G) radio communication systems [6]. SI cancelation/suppression in full duplex (FD) mode is achieved by estimating the SI channels [7]–[12]. In FD communication, residual self interference (RSI) still remains despite the SI cancelation [1], [5], [8], [13], and RSI channel needs to be estimated and canceled at the destination so as to improve the system performance. Relaying is a key solution especially in urban environments where strong shadowing effects appear and very useful for the coverage of the structures in which point-to-point communication is inadequate such as underground tunnels and subways. Transmitted signal can be delivered over the relay to the destination by analog or digital transmission methods. *Amplify-and-Forward* (AF) is one analog option, whereas in digital transmission, observations at relay are quantized, encoded, and transmitted via digital modulation. Accordingly, one of the most important reasons to prefer the AF relaying is the limited power consumption at the relay [14], [15].

As mentioned, FD relaying suffers from severe SI and the cancelation of RSI emerges as a challenging problem. Studies in the literature investigate the system performance in the presence of RSI channel with respect to different criteria [13], [16]–[18]. The study in [13], assuming imperfect channel state information (CSI), concentrates on spatial domain RSI power suppression using null-space projection by deploying spatially separated receive and transmit antenna arrays in the relay without considering the channel estimation problem. Authors in [17] investigate the error and diversity performances of full-duplex AF relaying under the effect of RSI. In [18], the authors analyze the optimal power allocation scheme and the corresponding capacity limit for the same system setup of [17]. Both works in [17] and [18] assume a perfect CSI scenario.

Apart from one-way relay network (OWRN) scheme as in [13], [16]–[18], an FD-AF relaying can be employed in also two-way relay networks (TWRN) [19]–[23]. In an FD-TWRN, two source nodes transmit their signals to each other over relay node. References [19], [20] analyze the effect of channel estimation error and suppression of SI in TWRN. In [21], authors investigate achievable rates of the FD-AF-TWRN comparing

with the half duplex (HD) AF TWRN. In [22], instead of purely suppressing the self-interference, end-to-end performance maximization is addressed by jointly optimizing the beamforming matrix at the MIMO relay as well as the power control at the sources. Reference [23] focuses on estimation of the individual channels as well as the RSI channel simultaneously and analyzes the effects of channel parameters and transmit powers on the Fisher information.

The FD-AF relay system can be also combined with MIMO [8], [24], [25]. In the MIMO case, the signal distortion caused by hardware impediments like the limited dynamic-range of non-ideal amplifiers, oscillators, ADCs, and DACs has to be considered [26]–[31] when modeling the RSI channel. However, with sufficient passive self-interference suppression and analog cancelation in RF [5], the distortion can be reduced [1] to a Gaussian model for the RSI channel [22], [32], [33].

In the context of the RSI channel estimation in an FD-AF relay network, [33] addresses two channel estimation problems for large scale antenna arrays consisting of one base station, one relay and a group of users around the relay. These are *cascaded estimation*, where only the base station estimates the cascaded two-hop channel, and *individual estimation*, where the base station estimates its respective channel, and the relay simultaneously estimates both the source-to-relay and SI channels. The authors conclude that the resulting interference affects the accuracy of channel estimation significantly. Reference [34] proposes a maximum likelihood RSI channel estimator and derives the Cramer Rao Lower Bound (CRLB) so as to provide benchmark for the MSE performance of the estimator. Nevertheless, works in [33], [34], [35] only consider time-invariant channel scenarios. However, the relay channel is more susceptible to the time-varying fading as compared to the conventional point-to-point communication systems, because of the motion of either the source, the relay or the destination. To the best of our knowledge, time-varying channel estimation in an FD-AF relay network has not yet been reported. This open problem motivates our current work.

Time-varying channels are usually modeled by using the Gauss-Markov model [36] or basis expansion model (BEM) [11], [12], [37]–[40]. In Gauss-Markov model, channel variation is tracked through symbol-by-symbol updating. In the BEM, time-varying channel variation is represented by a few basis functions scaled with corresponding basis coefficients. Thus, time-varying channel estimation becomes possible by means of estimating only unknown basis coefficients, which are significantly smaller than the total number of unknown channel coefficients. The BEMs proposed in [11], [12], [37]–[40] are Karhunen-Loeve, discrete-prolate spheroidal, complex exponential and discrete cosine BEMs, respectively. All of these BEMs except Karhunen-Loeve BEM have an error floor due to high Doppler spread environments. However, the Karhunen-Loeve BEM requires prior knowledge of channel statistics. In addition to these BEMs, [41] proposes the discrete Legendre polynomial (DLP) BEM which models the time-varying channel very accurately without the knowledge of channel statistics.

In this work, in contrast to our previous work in [23] and [34] in which the mobile channels on the source-relay and relay-destination links are assumed to be time-invariant, we address the estimation problem of both RSI and time-varying cascaded channels in an FD-AF-OWRN. We use the discrete Legendre polynomial [41] which captures the time-varying cascaded channel variation by a finite set of parameters to be estimated. The RSI channel at the relay results in an end-to-end inter symbol interference (ISI) channel [16], [23]. Thus, the resulting ISI channel includes high order powers of RSI channel. In other words, the observation is nonlinear with respect to the RSI channel. In order to obtain a linear training signal model with respect to RSI channel by way of eliminating the higher powers of RSI channel, we propose a new training frame structure in which one-sample consecutive orthogonal training blocks are sent by the source node. By exploiting the orthogonality of consecutive training blocks, two decoupled training signal models are obtained so as to estimate simultaneously the BEM coefficients of the cascaded channel and the RSI channel. Thus, after estimation and the equalization, the RSI is canceled at the destination node. In order to estimate the BEM coefficients of the cascaded channel as well as RSI channel, linear minimum mean square error (MMSE) based estimators are proposed.

In this paper, our contributions are five-fold:
1) We propose time-varying cascaded channel estimation for the FD-AF-OWRNs. To the best of our knowledge, time-varying channel estimation in an FD-AF relay network has not yet been reported in the literature.
2) Conventionally, BEM is used to represent point-to-point channels. In this work, we modified and apply the BEM to the rapidly time varying cascaded channel of the FD-AF-OWRNs.
3) We express the power scaling factor at the relay analytically for the FD-AF-OWRNs with time varying individual channels.
4) A new training structure suitable for FD-AF-OWRNs is proposed. In the training frame, one-sample consecutive orthogonal training blocks are sent by the source node. By exploiting the orthogonality of consecutive training blocks, two decoupled training signal models are obtained so as to estimate simultaneously the BEM coefficients of the cascaded channel and the RSI channel.
5) We derive the Bayesian CRLBs for both the RSI channel and BEM coefficients of the rapidly time-varying cascaded channel.

The rest of the paper is organized as follows: Section II describes the system model of an FD-AF-OWRN, introduces basis expansion model to represent the rapidly time-varying cascaded channel by a few basis functions. Section III provides a new training frame structure and two decoupled training signal models so as to estimate simultaneously the BEM coefficients of the cascaded channel and the RSI channel. Section IV evaluates the performances of the proposed estimators via computer simulations. Finally, Section V summarizes the main conclusions of the paper.
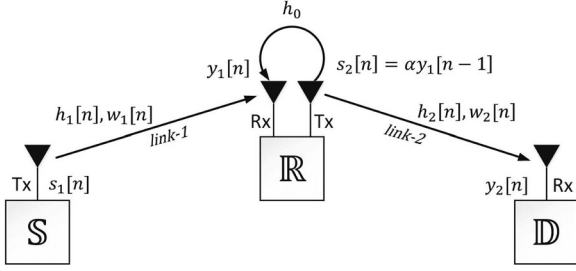
Fig. 1.  FD-AF-OWRN with rapidly time-varying channels

## II. System Model

We consider a wireless cooperative communication system in which the source node $\mathbb{S}$ transmits information to the destination node $\mathbb{D}$ with the assistance of the relay node $\mathbb{R}$. In our system, as seen in Fig. 1, the source and the destination nodes have only one HD antenna while the relay node has an FD antenna. Because of limited power consumption, we assume that the relay node does not perform channel estimation but just amplifies-and-forwards (AF) its received signal in order to keep the computational complexity as low as possible and to satisfy an average transmit power constraint. In the HD mode, since the relay node has only one HD antenna, the source node transmits during the first half of the transmit interval, and the relay node amplifies-and-forwards the received symbols during the second half [42]. In contrast, in our system operating in the FD mode, the relay keeps receiving the new symbols while amplifying-and-forwarding previously received symbols and, thus, both the source and the relay nodes transmit over the entire transmit interval. We assume that there is one sampling delay between the received and the transmitted signals at the relay node [16]. For notational simplicity, we call the communication links $\mathbb{S} \to \mathbb{R}$ and $\mathbb{R} \to \mathbb{D}$ as *link-1* and *link-2*, respectively. Let $\left\{ s_l[n], y_l[n], h_l[n], w_l[n] \right\}_{l=1}^{2}$ denote the $n$th discrete time samples of the transmitted signal, received signal, time-varying flat fading baseband channel and the additive white Gaussian noise (AWGN) on the $l$th link, respectively. A separate pre-stage is assumed to gather the information of the self-interference channel to perform analog and digital cancelation methods in the next transmission stage [5], [33]. During the transmission, the self-interference is reduced in RF with analog cancelation methods until the residual self-interference power falls in the ADC dynamic range, and then is further suppressed by digital methods. However, despite these suppression methods, the RSI is still present at the destination. We assume that RSI channel is a time-invariant flat fading channel and denote it by $h_0$.

The time-varying flat fading individual channels $h_1[n]$ and $h_2[n]$ are independent from each other and modeled as wide-sense stationary (WSS) zero-mean complex Gaussian random processes with variances $\Omega_1^2$ and $\Omega_2^2$. Time variations of the channels are caused by the mobility of the three nodes [43]. Assuming the Jakes model, the discrete autocorrelation functions of $h_l[n]$'s are given by

$$\rho_l[n - n'] = E\left\{ h_l[n] h_l^*[n'] \right\}$$
$$= \Omega_l^2 J_0\left( 2\pi f_{D_l} T_s(n - n') \right), \ l = 1, 2, \quad (1)$$

where $(\cdot)^*$ denotes the complex conjugate operator, $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, $f_{D_l}$ is the maximum (one-sided) Doppler shift and $T_s$ denotes the sampling period. The power of the signal transmitted by the source is represented by $\sigma_s^2$ and the variance of the complex valued AWGN on the $l$th link is denoted by $\sigma_l^2$. We also assume that the RSI channel $h_0$ is a zero-mean complex Gaussian with variance $\Omega_0^2$ [22], [33].

The received symbol at the $n$th discrete time at the relay and the destination nodes are given by

$$y_1[n] = h_1[n] s_1[n] + h_0 s_2[n] + w_1[n] \quad (2)$$

$$y_2[n] = h_2[n] s_2[n] + w_2[n], \ n = 0, 1, \dots, (N - 1), \quad (3)$$

where transmitted symbols by the source and relay nodes are $s_1[n]$ and $s_2[n] = \alpha y_1[n - 1]$, respectively, the final observation of the system at the destination node is $y_2[n]$ and $N$ stands for the number of samples within one observation interval. The relay gain $\alpha$ is used to satisfy the average transmit power constraint at the relay. We will derive a closed-form expression for $\alpha$ in terms of the average power $\mathcal{P}$, $\sigma_\theta^2$ and $\Omega_0^2$ in Appendix-B. Substituting these definitions into (2) and (3), the recursive solution of (2) is expressed as

$$y_1[n] = \sum_{\ell=0}^{n} \theta^\ell \left( h_1[n - \ell] s_1[n - \ell] + w_1[n - \ell] \right), \quad (4)$$

where $\theta = \alpha h_0$. The received symbol at the destination in (3) is found as follows

$$y_2[n] = \alpha h_2[n] \sum_{\ell=1}^{n} \theta^{(\ell-1)} h_1[n - \ell] s_1[n - \ell] + \eta[n], \quad (5)$$

where the overall additive colored noise term $\eta[n]$ is defined as

$$\eta[n] = \underbrace{\alpha h_2[n] \sum_{\ell=1}^{n} \theta^{(\ell-1)} w_1[n - \ell])}_{\text{colored noise}} + w_2[n]. \quad (6)$$

In (5), we assume distortion of the signal caused by hardware impediments is reduced due to sufficient passive self-interference suppression and analog cancelation in RF [1]. However, if the distortion has to be considered, (5) does not change because the distortion can be incorporated as part of the noise. To be specific, the distortion from the transmitter and the receiver are incorporated into the colored noise term and white noise term of $\eta[n]$, respectively [29] by modifying the variances of these terms.

### A. Receive Signal Over the Cascaded Channel

The higher powers of $\theta$ in (5) occur due to the RSI loop resulting in inter-block interference as well. Thus, we adopt a block based transmission with a guard time of the effective length of RSI loop in which the sources keep silent to prevent inter-block interference. We use a block-based transmission with a guard time to avoid inter-block interference [23]. Choosing guard time of $L$ as the effective length of the RSI loop where the most of the energy, (e.g., 99.9%), is contained, the received signal model in (5) after discarding the last $L$ symbols can be

rewritten as

$$y_2[n] = \alpha h_2[n] \sum_{\ell=1}^{L} \theta^{(\ell-1)} h_1[n-\ell] s_1[n-\ell] + \eta[n]. \quad (7)$$

where note that channel terms are $\left\{ \alpha h_2[n] \theta^{(\ell-1)} h_1[n-\ell] \right\}_{\ell=1}^{L}$. Basically, the estimation problem of the cascaded mobile radio channel is to estimate the channel terms $\left\{ h_2[n] h_1[n-\ell] \right\}_{\ell=1}^{L}$. However, comparing the speed of change in the terms of $\theta^{(\ell-1)}$ and $h_1[n-\ell]$ inside the summation, assuming $\left\{ h_1[n-\ell] \right\}_{\ell=1}^{L}$ terms are almost the same within short discrete time interval $L$, it can be possible to reduce the number of unknown channel terms that enables computationally efficient mobile radio channel estimator. If $|\theta| \ll 1$, it is straightforward to obtain $|\theta^{\ell}|/|\theta^{\ell-1}| \ll 1$, $\ell = 1, 2, \ldots, L$ that shows how fast $\left\{ \theta^{(\ell-1)} \right\}_{\ell=1}^{L}$ terms change. On the other hand, using (1), we can figure out how fast the channel terms $\left\{ h_1[n-\ell] \right\}_{\ell=1}^{L}$ change by the autocorrelation coefficient of $h_1[n-\ell]$, $\rho_1[1]\big|_{\Omega_1^2=1} = J_0\left(2\pi f_{D_1} T_s\right)$, for one sample discrete time shift, where we note that the zeroth-order Bessel function of the first kind $J_0$ is less than but very close to 1 since $f_{D_1} T_s$ is very small. Consequently, we conclude from $|\theta^{\ell}|/|\theta^{\ell-1}| \ll 1$ and $\rho_1[1]\big|_{\Omega_1^2=1} = J_0\left(2\pi f_{D_1} T_s\right)$ that $\theta^{\ell}$ goes to zero very quickly if $|\theta| \ll 1$, whereas $h_1[n-\ell]$ is almost constant within the range of $1 \leqslant \ell \leqslant L$ for small $L$ values. So, we conclude that very rapid changes of $\{\theta^{(\ell-1)}\}_{\ell=1}^{L}$ dominate slow changes of $\{h_1[n-\ell]\}_{\ell=1}^{L}$ and we can assume that $h_1[n]$ is almost constant within short time interval (e.g. $L = 2, 3, 4$) in the system setup. Besides, when the channel $h_1[n-\ell]$ is not constant, our simulation results in Section-IV also reflect that this assumption is accurate for the parameters considered. It follows from (7), the received signal model can be approximated as

$$\begin{aligned} y_2[n] =\ & \alpha h_2[n] \sum_{\ell=1}^{L} \theta^{(\ell-1)} h_1[n-\ell] s_1[n-\ell] + \eta[n] \\ =\ & \alpha h[n] \sum_{\ell=1}^{L} \theta^{(\ell-1)} s_1[n-\ell] + \eta[n], \quad (8) \end{aligned}$$

where $h[n] := h_2[n] \tilde{h}_1[n]$ is the so-called time-varying cascaded channel that represents the overall channel throughout source-relay-destination and $\tilde{h}_1[n] \cong h_1[n-\ell]$, $\ell = 1, 2, \ldots, L$. The autocorrelation of $h[n]$ is determined as

$$\begin{aligned} \rho[n-n'] :=\ & E\{h[n] h^*[n']\} \\ =\ & E\{\tilde{h}[n] \tilde{h}_1^*[n']\} E\{h_2[n] h_2^*[n']\} \\ =\ & \rho_1[n-n'] \rho_2[n-n']. \quad (9) \end{aligned}$$

Stacking the time samples, the observation model in (8) can be written in vector form as

$$\begin{aligned} \mathbf{y} =\ & \alpha \mathbf{H} \bar{\mathbf{H}}_0 \mathbf{s} + \boldsymbol{\eta} \\ =\ & \alpha \operatorname{diag}(\bar{\mathbf{H}}_0 \mathbf{s}) \mathbf{h} + \boldsymbol{\eta}, \quad (10) \end{aligned}$$

together with the following additive noise vector definition

$$\boldsymbol{\eta} = \alpha \mathbf{H}_2 \mathbf{H}_0 \mathbf{w}_1 + \mathbf{w}_2, \quad (11)$$

where

$$\begin{aligned} \mathbf{y} =\ & [y_2[0], y_2[1], \ldots, y_2[N-1]]^T \\ \mathbf{s} =\ & [s_1[0], s_1[1], \ldots, s_1[N-1]]^T, \\ \mathbf{H} =\ & \operatorname{diag}(\mathbf{h}), \ \mathbf{h} = [h[0], h[1], \ldots, h[N-1]]^T \\ \boldsymbol{\eta} =\ & [\eta[0], \eta[1], \ldots, \eta[N-1]]^T \\ \mathbf{w}_l =\ & [w_l[0], w_l[1], \ldots, w_l[N-1]]^T, \ l = 1, 2, \\ \mathbf{H}_l =\ & \operatorname{diag}(\mathbf{h}_l), \ \mathbf{h}_l = [h_l[0], h_l[1], \ldots, h_l[N-1]]^T, \ l = 1, 2. \end{aligned} \quad (12)$$

In (11), $\mathbf{H}_0$ matrix whose first row entries are all-zero is an $N \times N$ Toeplitz matrix with the first column vector of $[0, 1, \theta^1, \theta^2, \ldots, \theta^{(N-2)}]^T$ and $\bar{\mathbf{H}}_0$ in (10) is a banded version of $\mathbf{H}_0$ with the powers of $\theta$ greater than $L$ being set to zero. An entry of $\mathbf{H}_0$ at $p$th row and $q$th column can be given as

$$[\mathbf{H}_0]_{p,q} = \begin{cases} \theta^{(p-q-1)} & , q < p \\ 0 & , \text{otherwise} \end{cases} \quad (13)$$

where $p, q \in \{1, 2, \ldots, N\}$. From (9), it is straightforward that the autocorrelation matrix of the non-Gaussian cascaded channel vector $\mathbf{h}$ is given by

$$\mathbf{R}_{\mathbf{h}} = \mathbf{R}_{\mathbf{h}_1} \odot \mathbf{R}_{\mathbf{h}_2} \quad (14)$$

and the overall additive noise vector $\boldsymbol{\eta}$ in (10) and (11) is zero-mean non-Gaussian vector with uncorrelated entries. Detailed derivation of the autocorrelation matrix $\mathbf{R}_{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ is given in Appendix-A, where it is shown to be a diagonal matrix. Recalling the definitions in (12), it follows from (5) that the vector form of the observation model with respect to individual channel vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ can be written also as

$$\mathbf{y} = \mathbf{G}\mathbf{s} + \boldsymbol{\eta}, \quad (15)$$

where the channel matrix $\mathbf{G}$ is given by

$$\mathbf{G} = \alpha \mathbf{H}_2 \mathbf{H}_0 \mathbf{H}_1. \quad (16)$$

### B. Basis Expansion Model (BEM) of the Time-Varying Cascaded Channel

The performance of the receiver critically depends on the estimate of the time-varying cascaded channel vector $\mathbf{h} \in \mathcal{C}^N$ and the RSI channel $\theta$. It seems the estimation of the $N+1$ unknowns is impossible by means of $\mathbf{y} \in \mathcal{C}^N$ since there are more unknowns to be determined than known equations. In order to express the time variations of the cascaded channel by a finite number of parameters, basis expansion model (BEM) can be applied to approximate the time-varying cascaded channel $h[n]$ in (8). As the cascaded channel $h[n]$ is essentially a lowpass process whose bandwidth is determined by the Doppler frequency, it can be well approximated by the weighted sum of substantially fewer number $D(\ll N)$ of orthonormal basis functions $\{\psi_j[n]\}$ in the discrete time interval $[0, N-1]$ as follows

$$h[n] \cong \sum_{j=0}^{D-1} c_j \psi_j[n], \quad (17)$$

where $c_j$ are the expansion coefficients. Similarly, using the orthogonality property of the basis functions, the expansion coefficients can be evaluated by the inverse transformation as

$$c_j = \sum_{n=0}^{N-1} h[n]\psi_j[n]. \tag{18}$$

Thus, the channel and the expansion coefficients can be expressed in matrix form:

$$\mathbf{h} \cong \boldsymbol{\Psi}\mathbf{c}$$
$$\mathbf{c} = \boldsymbol{\Psi}^T\mathbf{h}, \tag{19}$$

where $\mathbf{h} = [h[0], h[1], \ldots, h[N-1]]^T$, $\mathbf{c} = [c_0, c_1, \ldots, c_{(D-1)}]^T$ and the transformation matrix $\boldsymbol{\Psi}$ contains the orthonormal basis vectors as

$$\boldsymbol{\Psi} = [\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{(D-1)}] \in \mathcal{R}^{N \times D}, \tag{20}$$

with $\boldsymbol{\psi}_j = [\psi_j[0], \psi_j[1], \ldots, \psi_j[N-1]]^T$. In (17), the number of basis functions $D$ is a function of the maximum Doppler frequency $f_d$ of the cascaded channel $h[n]$ and the interval length $NT_s$. From (19), it is straightforward that the autocorrelation matrix of the BEM coefficient vector of the cascaded channel is found as

$$\mathbf{R_c} = \boldsymbol{\Psi}^T \mathbf{R_h} \boldsymbol{\Psi} \in \mathcal{C}^{D \times D}. \tag{21}$$

The number of basis coefficients satisfies [41], [44]

$$2f_D T_s N \leq D \leq N, \tag{22}$$

where $f_D = f_{D1} + f_{D2}$, and $f_{D1}$ and $f_{D2}$ denote the maximum (one-sided) Doppler bandwidth of link-1 and link-2 due to relative motions between source-relay and relay-destination, respectively.

In our work, we make use of a BEM, based on the orthonormal discrete Legendre polynomial (DLP) basis expansion model (DLP-BEM), to represent the time variations of the channel in an observation interval. DLP-BEM is well suited to represent the low-pass equivalent of the Doppler channel by means of a small number of basis functions. Also, the DLP basis functions have the advantages of being independent of the channel statistics and having expansion coefficients that become uncorrelated as the number of observations $N$ gets larger, as proven in [41]. The Legendre polynomials are generated by carrying out Gram-Schmidt orthogonalization on the polynomials $\{1, n, n^2, \ldots\}$ with respect to the time-varying channels in a neighborhood of the middle point of the considered interval. The orthonormal Legendre polynomials can be defined as [41]

$$\psi_j[n] = \frac{\phi_j[n]}{\sqrt{\sum_{n=0}^{N-1} \phi_j^2[n]}}, \tag{23}$$

where $\phi_j[n]$ denotes discrete orthogonal Legendre polynomial that can be computed recursively as

$$\phi_j[n] = \frac{(2j-1)(N-1-2n)}{j(N-j)} \phi_{j-1}[n]$$
$$- \frac{(j-1)(N+j-1)}{j(N-j)} \phi_{j-2}[n], \, j = 2, 3, \ldots, (D-1) \tag{24}$$

with the following initial polynomials

$$\phi_0[n] = 1, \ \phi_1[n] = 1 - \frac{2n}{N-1}. \tag{25}$$

Substituting (19) into (10), the observation model can be written with respect to the symbol and the BEM coefficient vector as follows

$$\mathbf{y} = \alpha \operatorname{diag}(\boldsymbol{\Psi}\mathbf{c})\bar{\mathbf{H}}_0\mathbf{s} + \boldsymbol{\eta}$$
$$= \alpha \operatorname{diag}(\bar{\mathbf{H}}_0\mathbf{s})\boldsymbol{\Psi}\mathbf{c} + \boldsymbol{\eta}, \tag{26}$$

## III. TRAINING SIGNAL MODEL AND CHANNEL ESTIMATION

### A. Training Signal Model

In this section, in order to estimate the residual self interference (RSI) and the rapidly time-varying cascaded channels, we propose a new training frame structure in which one-sample consecutive orthogonal training blocks are sent by the source node and delivered to the destination over FD-AF relay. Exploiting the orthogonality of training blocks, we obtain two decoupled training signal models, each of them for the estimation of the RSI and the cascaded channels.

Collecting the observations $\{y_2[n_k - m]\}_{m=0}^{L-1}$ in (8), we obtain

$$\mathbf{y}_k = \alpha \operatorname{diag}(\mathbf{h}_k)\mathbf{S}_k\boldsymbol{\theta} + \boldsymbol{\eta}_k, \, k = 1, 2, \ldots, K, \tag{27}$$

where

$$\mathbf{y}_k = [y[n_k - (L-1)], y[n_k - (L-2)], \ldots, y[n_k]]^T$$
$$\mathbf{h}_k = [h[n_k - (L-1)], h[n_k - (L-2)], \ldots, h[n_k]]^T$$
$$\mathbf{S}_k = [\mathbf{s}_{k,1}, \mathbf{s}_{k,2}, \ldots, \mathbf{s}_{k,L}]$$
$$\mathbf{s}_{k,q} = [s[n_k - (q+L-1)],$$
$$\qquad s[n_k - (q+L-2)], \ldots, s[n_k - q]]^T$$
$$\boldsymbol{\theta} = [1, \theta, \ldots, \theta^{(L-1)}]^T$$
$$\boldsymbol{\eta}_k = [\eta[n_k - (L-1)], \eta[n_k - (L-2)], \ldots, \eta[n_k]]^T, \tag{28}$$

$k$ stands for the training block index and the smallest value of the training time index should satisfy $n_1 \geq 2L - 1$ since $n_k - (q + L - 2) \geq 0$ for $k = 1$ and $q = L$. Note that the signal matrix $\mathbf{S}_k$ is a Toeplitz matrix. In the training signal model in (27), choosing $\mathbf{S}_k = b\mathbf{A}$, we arrive at

$$\mathbf{y}_k = \alpha b \operatorname{diag}(\mathbf{h}_k)\mathbf{A}\boldsymbol{\theta} + \boldsymbol{\eta}_k, \tag{29}$$

where $b$ is the unique training symbol, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_L]$ is an $L \times L$ Toeplitz matrix having orthogonal column vectors of

Euclidean norm $\sqrt{L}$ and each entry has unity norm. It is possible to show many orthogonal matrices having such properties but in this we use the following orthogonal Toeplitz matrices for $L = 2, 3, 4$, respectively,

$$
\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} e^{j2\pi/3} & 1 & 1 \\ 1 & e^{j2\pi/3} & 1 \\ 1 & 1 & e^{j2\pi/3} \end{bmatrix}, \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}
$$

(30)

For matrices that have different values of $L > 4$, one can use the results in [45], which generates deterministic matrices that are orthogonal Toeplitz and symmetric. Note that

$$
\begin{aligned}
\mathbf{a}_\ell^\dagger \mathbf{y}_k &= \alpha b\, \mathbf{a}_\ell^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{A}\boldsymbol{\theta} + \mathbf{a}_\ell^\dagger \boldsymbol{\eta}_k, \\
&= \alpha b\, \mathbf{a}_\ell^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{A}_\ell^- \boldsymbol{\theta}_\ell^- \\
&\quad + \alpha b\, \mathbf{a}_\ell^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{A}_\ell^+ \boldsymbol{\theta}_\ell^+ + \mathbf{a}_\ell^\dagger \boldsymbol{\eta}_k,
\end{aligned}
$$

(31)

where $\boldsymbol{\theta}_\ell^- = [1, \theta, \ldots, \theta^{(\ell-1)}]^T$, $\mathbf{A}_\ell^- = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_\ell]^T$, $\boldsymbol{\theta}_\ell^+ = [\theta^\ell, \theta^{(\ell+1)}, \ldots, \theta^{(L-1)}]^T$ and $\mathbf{A}_\ell^+ = [\mathbf{a}_{(\ell+1)}, \mathbf{a}_{(\ell+2)}, \ldots, \mathbf{a}_L]^T$ and $(\cdot)^\dagger$ denotes the conjugate transpose operator. Because of orthonormality of the vector set $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_L\}$ and $|\theta| \ll 1$, it is clear that $|\mathbf{a}_\ell^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{A}_\ell^- \boldsymbol{\theta}_\ell^-| \gg |\mathbf{a}_\ell^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{A}_\ell^+ \boldsymbol{\theta}_\ell^+|$. Therefore, we may incorporate the second term on the right hand side of (31) or simply omit it. For simplicity, we prefer to omit this term. It follows from (31) that

$$
\mathbf{a}_\ell^\dagger \mathbf{y}_k = \alpha b\, \mathbf{a}_\ell^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{A}_\ell^- \boldsymbol{\theta}_\ell^- + \mathbf{a}_\ell^\dagger \boldsymbol{\eta}_k,
$$

(32)

Keeping in mind that $\mathbf{a}_\ell \odot \mathbf{a}_\ell^* = \mathbf{1}_L\ \forall \ell$ and using (32), we define

$$
\begin{aligned}
x_k &= \mathbf{a}_1^\dagger \mathbf{y}_k \\
&= \alpha b\, \mathbf{a}_1^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{a}_1 + \mathbf{a}_1^\dagger \boldsymbol{\eta}_k \\
&= \alpha b\, \mathbf{1}_L^T \mathbf{h}_k + \mathbf{a}_1^\dagger \boldsymbol{\eta}_k \\
&= \alpha b\, (\mathbf{1}_L^T \boldsymbol{\Psi}_k)\mathbf{c} + \mathbf{a}_1^\dagger \boldsymbol{\eta}_k
\end{aligned}
$$

(33)

and

$$
\begin{aligned}
z_k &= \mathbf{a}_2^\dagger \mathbf{y}_k \\
&= \alpha b\, \mathbf{a}_2^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{a}_1 + \alpha b\, \mathbf{a}_2^\dagger \operatorname{diag}(\mathbf{h}_k) \mathbf{a}_2 + \mathbf{a}_2^\dagger \boldsymbol{\eta}_k \\
&= \alpha b\, (\mathbf{a}_2 \odot \mathbf{a}_1^*)^\dagger \mathbf{h}_k + \alpha b\, \mathbf{1}_L^T \mathbf{h}_k\, \theta + \mathbf{a}_2^\dagger \boldsymbol{\eta}_k \\
&= \alpha b\, ((\mathbf{a}_2 \odot \mathbf{a}_1^*)^\dagger \boldsymbol{\Psi}_k)\mathbf{c} + \alpha b\, (\mathbf{1}_L^T \boldsymbol{\Psi}_k)\mathbf{c}\, \theta + \mathbf{a}_2^\dagger \boldsymbol{\eta}_k,
\end{aligned}
$$

(34)

where the transform matrix $\boldsymbol{\Psi}_k$ as in (20) is given within $k$th training block time interval $\{n_k - m\}_{m=0}^{(L-1)}$ by (35) shown at the bottom of this page.

Stacking the observation vectors in (33) and (34), we arrive at the following observation models

$$
\mathbf{x} = \alpha b \boldsymbol{\Phi}_1 \mathbf{c} + \boldsymbol{\omega} \quad \in \mathcal{C}^{K \times 1}
$$

(36)

and

$$
\mathbf{z} = \alpha b \boldsymbol{\Phi}_2 \mathbf{c} + \alpha b \boldsymbol{\Phi}_1 \mathbf{c}\theta + \boldsymbol{\xi} \quad \in \mathcal{C}^{K \times 1},
$$

(37)

where

$$
\begin{aligned}
\boldsymbol{\omega} &= (\mathbf{I}_K \otimes \mathbf{a}_1^\dagger)\boldsymbol{\eta}_{\text{tr}} \quad \in \mathcal{C}^{K \times 1} \\
\boldsymbol{\xi} &= (\mathbf{I}_K \otimes \mathbf{a}_2^\dagger)\boldsymbol{\eta}_{\text{tr}} \quad \in \mathcal{C}^{K \times 1} \\
\boldsymbol{\eta}_{\text{tr}} &= [\boldsymbol{\eta}_1^\dagger, \boldsymbol{\eta}_2^\dagger, \ldots, \boldsymbol{\eta}_K^\dagger]^\dagger \quad \in \mathcal{C}^{KL \times 1} \\
\boldsymbol{\Phi}_1 &= (\mathbf{I}_K \otimes \mathbf{1}_L^T)\boldsymbol{\Psi}_{\text{tr}} \quad \in \mathcal{C}^{K \times D} \\
\boldsymbol{\Phi}_2 &= (\mathbf{I}_K \otimes (\mathbf{a}_2 \odot \mathbf{a}_1^*)^\dagger)\boldsymbol{\Psi}_{\text{tr}} \quad \in \mathcal{C}^{K \times D} \\
\boldsymbol{\Psi}_{\text{tr}} &= [\boldsymbol{\Psi}_1^\dagger, \boldsymbol{\Psi}_2^\dagger, \ldots, \boldsymbol{\Psi}_K^\dagger]^\dagger \quad \in \mathcal{C}^{KL \times D}.
\end{aligned}
$$

(38)

Here note that the operation $\otimes$ denotes the Kronecker product. As seen from (28) and (38), the additive noise vector $\boldsymbol{\eta}_{\text{tr}}$ is obtained by stacking the discrete time samples $\{\eta[n_k - (L-1)], \eta[n_k - (L-2)], \ldots, \eta[n_k]\}_{k=1}^K$. So, the autocorrelation matrix of the noise vectors $\boldsymbol{\eta}_{\text{tr}}$ is defined as the sub-matrix of the diagonal autocorrelation matrix $\mathbf{R}_\eta$ in (56) as follows

$$
\mathbf{R}_{\boldsymbol{\eta}_{\text{tr}}} = [\mathbf{R}_\eta]_{\text{tr}},
$$

(39)

where, noting that time index starts from zero, $[\cdot]_{\text{tr}}$ represents all entries whose both row and column indices are in the index set of $\{n_k - (L-2), n_k - (L-3), \ldots, n_k + 1\}_{k=1}^K$. Toeplitz matrix $\mathbf{S}_k$ in (28) is constructed with $(2L-1)$ training signals transmitted within one training block. Accordingly, discrete time indices $\{n_k\}_{k=1}^K$ leading to the training blocks are chosen as equally spaced such that

$$
n_k = \lfloor (2k-1)\Delta \rfloor + k(2L-1), \; k = 1, 2, \ldots, K,
$$

(40)

where $\lfloor \cdot \rfloor$ rounds the number to its integer part. The number of training blocks $K$ should satisfy that

$$
\Delta = \frac{N - (2L-1)K}{2K} \geq 0.
$$

(41)

and $\Delta = 0$ means that all transmitted signals are considered as training signals. Accordingly, it is straightforward from (38) that

$$
\mathbf{R}_{\boldsymbol{\omega}} = (\mathbf{I}_K \otimes \mathbf{a}_1^\dagger)\mathbf{R}_{\boldsymbol{\eta}_{\text{tr}}}(\mathbf{I}_K \otimes \mathbf{a}_1^\dagger)^\dagger
$$

(42)

and

$$
\mathbf{R}_{\boldsymbol{\xi}} = (\mathbf{I}_K \otimes \mathbf{a}_2^\dagger)\mathbf{R}_{\boldsymbol{\eta}_{\text{tr}}}(\mathbf{I}_K \otimes \mathbf{a}_2^\dagger)^\dagger.
$$

(43)

$$
\boldsymbol{\Psi}_k = \begin{bmatrix} \psi_0(n_k - (L-1)) & \psi_1(n_k - (L-1)) & \cdots & \psi_{D-1}(n_k - (L-1)) \\ \psi_0(n_k - (L-2)) & \psi_1((n_k - (L-2)) & \cdots & \psi_{D-1}(n_k - (L-2)) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_0(n_k) & \psi_1(n_k) & \cdots & \psi_{D-1}(n_k) \end{bmatrix} \in \mathcal{R}^{L \times D}.
$$

(35)

## B. Estimation of Cascaded and RSI Channels

Signal term of the receive model in (36) does not depend on $\theta$ and this model is linear with respect to $\mathbf{c}$. However, signal term in (37) depends on both $\theta$ and $\mathbf{c}$. Hence, we estimate both parameters in a successive way: first we estimate $\mathbf{c}$ using the observation model in (36) and then using the estimate value of $\mathbf{c}$ and regarding estimation error of $\mathbf{c}$ in (37), we estimate $\theta$. So, using (36), we obtain the linear MMSE estimate of $\mathbf{c}$ as

$$\hat{\mathbf{c}} = \boldsymbol{\Gamma}^\dagger \mathbf{x}, \tag{44}$$

where the coefficient matrix $\boldsymbol{\Gamma}^\dagger$ that minimizes the average MSE, $\frac{1}{D} E\{(\mathbf{c} - \hat{\mathbf{c}})^\dagger (\mathbf{c} - \hat{\mathbf{c}})\}$, is obtained as

$$
\begin{aligned}
\boldsymbol{\Gamma}^\dagger &= E\{\mathbf{c}\mathbf{x}^\dagger\}\left(E\{\mathbf{x}\mathbf{x}^\dagger\}\right)^{-1} \quad \in \mathcal{C}^{D \times K} \\
&= \alpha b^* \mathbf{R_c} \boldsymbol{\Phi}_1^\dagger \left(\alpha^2 |b|^2 \boldsymbol{\Phi}_1 \mathbf{R_c} \boldsymbol{\Phi}_1^\dagger + \mathbf{R}_{\boldsymbol{\omega}}\right)^{-1} \\
&= \alpha b^* \left(\alpha^2 |b|^2 \boldsymbol{\Phi}_1^\dagger \mathbf{R}_{\boldsymbol{\omega}}^{-1} \boldsymbol{\Phi}_1 + \mathbf{R_c}^{-1}\right)^{-1} \boldsymbol{\Phi}_1^\dagger \mathbf{R}_{\boldsymbol{\omega}}^{-1} \quad (45)
\end{aligned}
$$

and the corresponding pre-computed estimation error covariance matrix becomes

$$
\begin{aligned}
\mathbf{R}_{\boldsymbol{\varepsilon}} &= E\{(\mathbf{c} - \hat{\mathbf{c}})(\mathbf{c} - \hat{\mathbf{c}})^\dagger\} \\
&= E\{(\mathbf{c} - \hat{\mathbf{c}})\mathbf{c}^\dagger\} - \underbrace{E\{(\mathbf{c} - \hat{\mathbf{c}})\hat{\mathbf{c}}^\dagger\}}_{0} \\
&= \left(\mathbf{I}_D - \alpha b \boldsymbol{\Gamma}^\dagger \boldsymbol{\Phi}_1\right)\mathbf{R_c}. \tag{46}
\end{aligned}
$$

After obtaining the linear MMSE estimate of $\mathbf{c}$, in order to take advantage of the statistical orthogonality principle of linear MMSE estimate $\hat{\mathbf{c}}$ and the estimation error $\boldsymbol{\varepsilon} = \mathbf{c} - \hat{\mathbf{c}}$, we rewrite (37) as follows

$$\underbrace{\mathbf{z} - \alpha b \boldsymbol{\Phi}_2 \hat{\mathbf{c}}}_{\tilde{\mathbf{z}}} = \alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}} \theta + \underbrace{\alpha b \boldsymbol{\Phi}_2 (\mathbf{c} - \hat{\mathbf{c}}) + \alpha b \boldsymbol{\Phi}_1 (\mathbf{c} - \hat{\mathbf{c}}) \theta + \boldsymbol{\xi}}_{\boldsymbol{\nu} : \text{ noise term}} \tag{47}$$

From (47), the linear MMSE estimate of $\theta$ is obtained as follows

$$
\begin{aligned}
\hat{\theta} &= E\{\theta \tilde{\mathbf{z}} | \hat{\mathbf{c}}\}\left(E\{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^\dagger | \hat{\mathbf{c}}\}\right)^{-1} \tilde{\mathbf{z}} \\
&= \sigma_\theta^2 (\alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}})^\dagger \left(\sigma_\theta^2 (\alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}})(\alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}})^\dagger + \mathbf{R}_{\boldsymbol{\nu}}\right)^{-1} (\mathbf{z} - \alpha b \boldsymbol{\Phi}_2 \hat{\mathbf{c}}) \\
&= \frac{\sigma_\theta^2 (\alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}})^\dagger \mathbf{R}_{\boldsymbol{\nu}}^{-1} (\mathbf{z} - \alpha b \boldsymbol{\Phi}_2 \hat{\mathbf{c}})}{1 + \sigma_\theta^2 (\alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}})^\dagger \mathbf{R}_{\boldsymbol{\nu}}^{-1} (\alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}})}, \tag{48}
\end{aligned}
$$

where the pre-computed autocovariance of the noise term vector $\boldsymbol{\nu}$ is given by

$$\mathbf{R}_{\boldsymbol{\nu}} = \mathbf{R}_{\boldsymbol{\xi}} + \alpha^2 |b|^2 \left(\sigma_\theta^2 \boldsymbol{\Phi}_1 \mathbf{R}_{\boldsymbol{\varepsilon}} \boldsymbol{\Phi}_1^\dagger + \boldsymbol{\Phi}_2 \mathbf{R}_{\boldsymbol{\varepsilon}} \boldsymbol{\Phi}_2^\dagger\right). \tag{49}$$

In order to evaluate the MSE performances, we present Bayesian CRLB for each estimator of $\mathbf{c}$ and $\theta$. The exact derivations of CRLBs are given in Appendix-C. After obtaining the $\hat{\mathbf{c}}$ and $\hat{\theta}$, using (19) and (10), transmitted unknown symbols are determined by first linear MMSE estimation and then rounding to nearest constellation point.

## TABLE I
### SIMULATION PARAMETERS

| | |
|---|---|
| Sampling Frequency ($F_s = 1/T_s$) | 600 KHz |
| Observation Time Length ($T = NT_s$) | $200 T_s$ |
| Number of BEM coefficients ($D$) | 2 |
| Number of Training Blocks ($K$) | 4 |
| Effective Length of RSI Loop ($L$) | 2, 3, 4 |
| Training Block Length ($2L - 1$) | 3, 5, 7 |
| Normalized Doppler Frequency ($f_{D_l} T, l = 1, 2$) | 0.01, 0.1 |
| RSI Channel Power ($\Omega_0^2$) | 0.01, 0.1 Watt |
| Channel Power on the $l$th link ($\Omega_l^2, l = 1, 2$) | 1 Watt |
| Average Relay Transmit Power ($P_r$) | 1 Watt |
| Modulation Formats | BPSK, QPSK, 16-QAM |
| Transmit Signal Power ($\sigma_s^2$) | 1 Watt |

## C. Computational Complexity

In this subsection, we evaluate the computational complexity of the proposed estimation algorithms in terms of complex multiplications (CMs) and complex additions (CAs). The computational complexity of the proposed estimation algorithm is determined by observation block length $N$, number of training blocks $K$, number of BEM coefficients $D$ and the adopted effective length of the RSI channel $L$.

Each of $x_k$ and $z_k$ in (33) and (34) is transformed from the observation vector $\mathbf{y}_k$ corresponding to $k$th training block and needs $DL$ CMs and $D(L - 1)$ CAs, and thus, by stacking them, the overall calculation of each of $\mathbf{x}$ and $\mathbf{z}$ in (36) and (37) requires $KDL$ CMs and $KD(L - 1)$ CAs. We should note that the matrices $\mathbf{R}_{\boldsymbol{\eta}_{\text{tr}}}, \mathbf{R}_{\boldsymbol{\omega}}$ and $\mathbf{R}_{\boldsymbol{\xi}}$ are pre-computed, and following in a sequential manner, the matrices $\boldsymbol{\Gamma}^\dagger$, $\mathbf{R}_{\boldsymbol{\varepsilon}}$ and $\mathbf{R}_{\boldsymbol{\nu}}$ are also pre-computed. So, we will not take these matrices into account in the complexity calculations.

It is straightforward that the estimator of $\mathbf{c}$ in (44) requires $DK$ CMs and $D(K - 1)$ CAs, and consequently according to the real valued orthogonal transformation matrix in (19), the estimator of the time-varying cascaded channel takes $ND/2$ CMs and $N(D - 1)$ CAs. In the estimation of $\theta$ in (48), calculation of the term $\alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}}$ needs $KD$ CMs and $K(D - 1)$ CAs. In addition, the calculation of $(\alpha b \boldsymbol{\Phi}_1 \hat{\mathbf{c}})^\dagger \mathbf{R}_{\boldsymbol{\nu}}^{-1}$ that is involved two times in (48) takes $K^2$ CMs and $K(K - 1)$ CAs and the calculation of the term $(\mathbf{z} - \alpha b \boldsymbol{\Phi}_2 \hat{\mathbf{c}})$ uses $K$ CAs. The rest of calculations in the estimation of $\theta$ need $2K + 1$ CMs and $2K - 1$ CAs. As a result, the overall computational complexity load of $\hat{\theta}$ in (48) is $K^2 + K(D + 2) + 1$ CMs and $K^2 + (K - 1)(D + 2) + 1$ CAs. Consequently, we calculate the total complexity cost of the proposed estimators is $ND/2 + K^2 + 2K(DL + D + 1) + 1$ CMs and $N(D - 1) + K^2 + K(2DL + D + 2) - (3D + 1)$ CAs which is linear in $N$, $D$, $L$, and quadratic in $K$. Finally, since $K$, $D$ and $L$ take small integer values as seen in Table I, we conclude that the total computational complexity load to implement our estimation algorithms is significantly low.

## D. Training Overhead

In our system setup, we use $K$ training blocks in order to estimate the cascaded and the RSI channels. Each training block consists of a training sequence with length of $2L - 1$ which means that there is total $K(2L - 1)$ training symbols among

the $N$ transmitted symbols. We can define the training overhead metric as a percentage of the total number of training symbol to the number of transmitted symbol as follows

$$\kappa = \frac{K(2L-1)}{N} \times 100. \quad (50)$$

In our system setup, training overhead measure $\kappa\%$ takes the values of 10%, 12%, and 14% for $L = 2, 3$ and 4, respectively. As we can see from (22), the number of BEM coefficients $D$ increases with more mobility between source, relay and destination nodes, or with larger observation time length. In order to estimate $D$ unknown BEM coefficients, we need $K \geq D$ observation equations that are obtained by deploying $K$ equally spaced training blocks within total observation time length to track the time-varying cascaded channel. On the other hand, as the power of the RSI channel increases, the adopted effective length of the RSI loop $L$ should be increased in order to cancel higher powers of the RSI channel. This results larger training blocks. As seen from (22), it is possible to reduce the training overhead by increasing the total observation time length which trades off computational complexity with training overhead.

## IV. SIMULATION RESULTS

In this section, we present computer simulation results to assess the performance of FD-AF-OWRN operating with the proposed channel estimation algorithm and the training model. While evaluating the system performance, the time-varying flat fading channels are generated directly by using singular value decomposition (SVD) method [46] that exploits the autocorrelations of the individual channels given in (1). However, we use DLP-BEM while estimating the cascaded channel. Simulation parameters are chosen as in Table I. The estimates of the RSI and the cascaded channels are performed by the low complexity linear MMSE estimation technique using two decoupled training signal models that are obtained by exploiting the orthogonality of one-sample consecutive orthogonal training blocks. In order to test the performance of the FD-AF-OWRN operating under various scenarios, SER and MSE simulation plots are taken in four categories such as for different values of RSI channel power, mobility, adopted effective length of RSI loop and modulation format. In these figures, SER and MSE results are compared with the perfect channel state knowledge (CSI) and the CRLB results, respectively.

The plots in Figs. 2, 3 and 4 show the system performance for different values of RSI channel power. Fig. 2 exhibits the MSE performance of the linear MMSE estimator of the RSI channel for different values of RSI channel power. In Fig. 2, we observe performance gaps between MSE and relevant CRLB curves increasing with the higher SNR levels. As we see from the curves, gaps between MSE and relevant CRLB curves get smaller as the RSI channel power becomes smaller. These gaps occur because of the fact that the overall additive noise vector $\boldsymbol{\xi}$ in (37) still nonlinearly depends on RSI channel to be estimated. On the other hand as seen in Fig. 3, the proposed linear MMSE estimator of the DLP-BEM coefficient vector almost attains the CRLB for the RSI channel power around $\Omega_0^2 = 0.01$ Watt. As
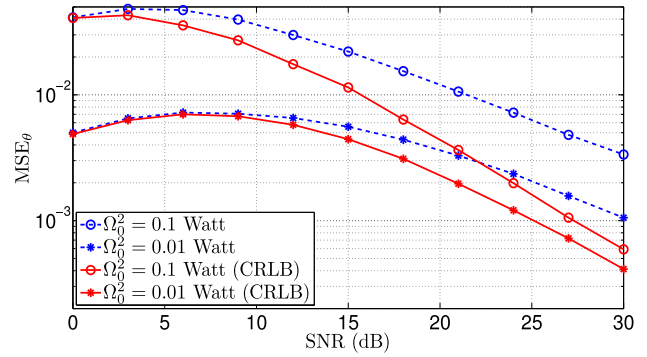


Fig. 2. MSE of $\theta$ vs. SNR results for different $\Omega_0^2$ values. ($N = 200, K = 4, L = 3, f_{D_1}T = f_{D_2}T = 0.01$, QPSK const.)
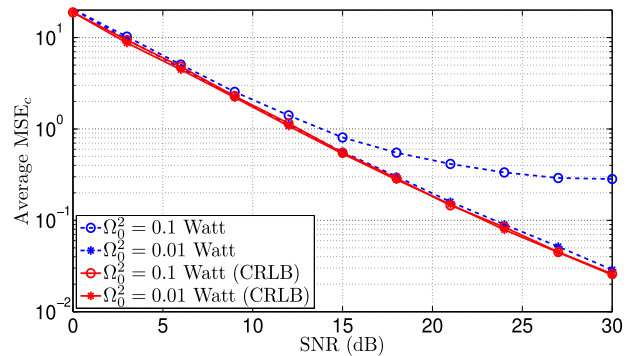


Fig. 3. Average MSE of **c** vs. SNR results for different $\Omega_0^2$ values. ($N = 200, K = 4, L = 3, f_{D_1}T = f_{D_2}T = 0.01$, QPSK const.)
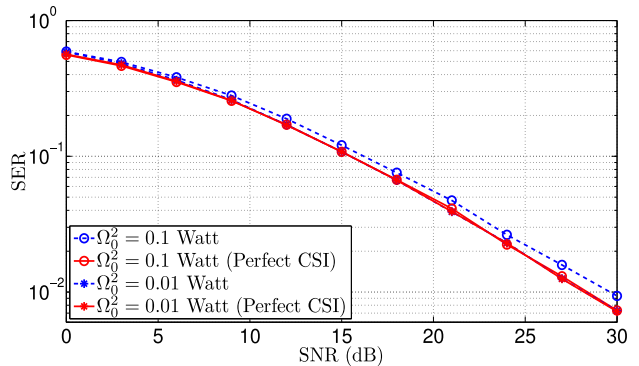


Fig. 4. SER vs. SNR results for different $\Omega_0^2$ values. ($N = 200, K = 4, L = 3, f_{D_1}T = f_{D_2}T = 0.01$, QPSK const.)

observed from the curve plotted for $\Omega_0^2 = 0.1$ Watt, the proposed estimator of the DLP-BEM coefficient vector gets away from CRLB for higher RSI channel powers. This is because the fact that the adopted effective length of RSI loop ($L$) is taken as 3 for both $\Omega_0^2 = 0.01$ Watt and $\Omega_0^2 = 0.1$ Watt values. In fact, the adopted effective length of RSI loop should be set up to higher values for higher RSI channel powers. On the other hand, the value of $\Omega_0^2 = 0.1$ as we consider as a benchmark means that 10% of the average transmit power at the relay ($Pr$) is dissipated as RSI channel power and is too high and not encountered in
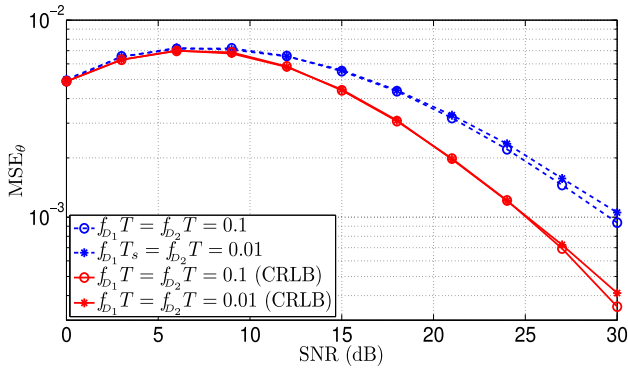
Fig. 5.    MSE of $\theta$ vs. SNR results for different normalized Doppler values. ($N = 200, K = 4, L = 3, \Omega_0^2 = 0.01$, QPSK const.)
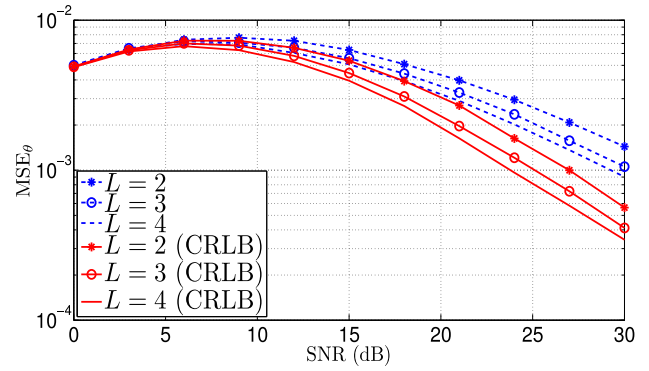


Fig. 8.    MSE of $\theta$ vs. SNR results for different $L$ values. ($N = 200, K = 4, \Omega_0^2 = 0.01, f_{D_1}T = f_{D_2}T = 0.01$, QPSK const.)
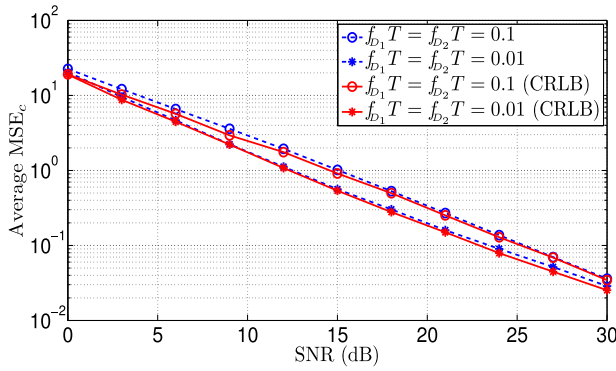


Fig. 6.    Average MSE of $\mathbf{c}$ vs. SNR results for different normalized Doppler values. ($N = 200, K = 4, L = 3, \Omega_0^2 = 0.01$, QPSK const.)



Fig. 9.    Average MSE of $\mathbf{c}$ vs. SNR results for different $L$ values. ($N = 200, K = 4, \Omega_0^2 = 0.01, f_{D_1}T = f_{D_2}T = 0.01$, QPSK const.)



Fig. 7.    SER vs. SNR results for different normalized Doppler values. ($N = 200, K = 4, L = 3, \Omega_0^2 = 0.01$, QPSK const.)
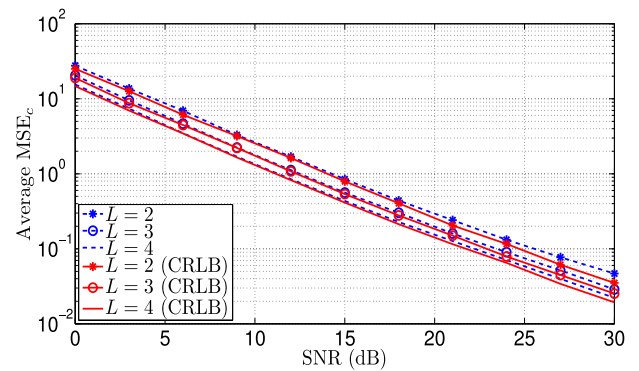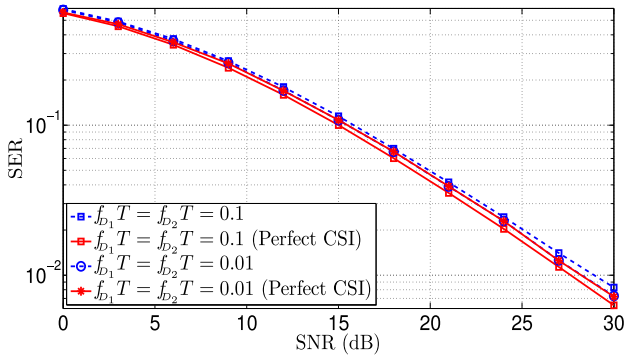
practice. As shown in Fig. 4, the resulting SER performance for $\Omega_0^2 = 0.01$ Watt achieves to SER performance that is plotted under perfect CSI scenario. However, a SER performance loss is observed for the case of $\Omega_0^2 = 0.1$.

Figs. 5, 6 and 7 exhibit the MSE and SER performances of the proposed estimators for the normalized Doppler frequencies of 0.01 and 0.1. In Fig. 5, we observe that higher Doppler does not degrade the MSE performance. This is not surprising, since our BEM captures fast variations of the channel, and larger Doppler does not cause a mismatch between our channel model and the

true channel. As seen from the curves in Fig. 6, the proposed estimator of the DLP-BEM coefficient vector almost attains the CRLB for both Doppler frequencies. In Fig. 7, the SER performance achieves the SER performance of the perfect CSI case for the normalized Doppler frequency of 0.01. Even if the MSE performance of the RSI channel estimator in Fig. 5 is better for the normalized Doppler frequency of 0.1, we still observe a SER performance loss in Fig. 7 for the same Doppler frequency. Thus, we conclude that the performance of the cascaded mobile channel dominates the performance of the RSI channel estimator while obtaining the resulting SER performance.

In Figs. 8 and 9, we investigate the MSE performances of the proposed estimators for three different values of the adopted effective length of RSI loop . It is clearly seen from these figures that the proposed estimators are having better SER and MSE performances as the value of the adopted effective length of RSI loop increases.

Figs. 10, 11 and 12 depict the MSE and SER performance curves of the proposed estimators for binary phase shift-keying (BPSK), quadrature phase shift-keying (QPSK) and 16-ary quadrature amplitude modulation (16QAM) signaling formats. As seen from the curves in Figs. 11 and 12, MSE performances for the constant envelope type modulation (i.e., BPSK, QPSK) formats are the same. Especially, the proposed estimator of the DLP-BEM coefficient vector almost attains the CRLB for each
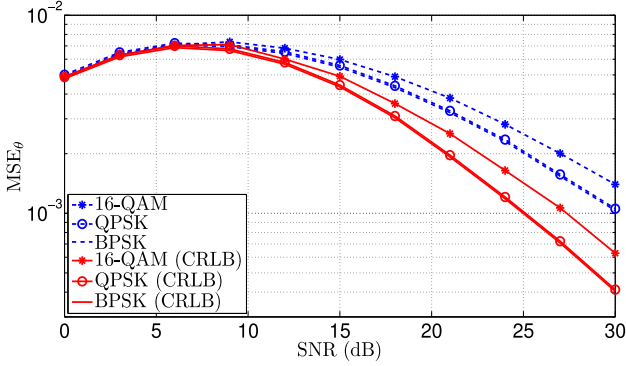
Fig. 10. MSE of $\theta$ vs. SNR results for different signal constellations. ($N = 200, K = 4, L = 3, \Omega_0^2 = 0.01, f_{Dl}T = f_{D2}T = 0.01$)
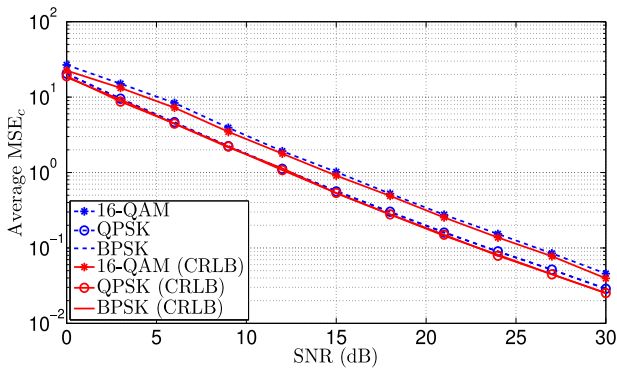


Fig. 11. Average MSE of **c** vs. SNR results for different signal constellations. ($N = 200, K = 4, L = 3, \Omega_0^2 = 0.01, f_{D1}T = f_{D2}T = 0.01$)
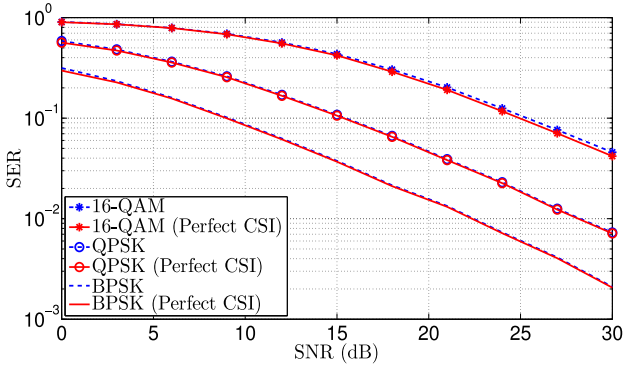


Fig. 12. SER vs. SNR results for different signal constellations. ($N = 200, K = 4, L = 3, \Omega_0^2 = 0.01, f_{D1}T = f_{D2}T = 0.01$)

signaling format and the resulting SER performances achieve to SER performances of the perfect CSI cases.

## V. CONCLUSIONS

In this work, we have employed the DLP-BEM to represents the time-varying cascaded channel by BEM coefficients that are a finite set of parameters to be estimated. We have proposed a new training frame structure in which one-sample consecutive orthogonal training blocks are sent by the source node and

delivered to the destination over FD-AF relay. Exploiting the orthogonality of training blocks, we have obtained two decoupled training signal models, each of them for the linear MMSE based estimation of the RSI channel and the BEM coefficient vector of the cascaded channel. The Bayesian CRBs for both estimators are derived. We have tested the SER and MSE performances of the FD-AF-OWRN operating under various scenarios such as for different values of RSI channel power, mobility, adopted effective length of RSI loop and modulation format. Simulations illustrate that the MSE performance of the proposed estimator for the DLP-BEM coefficient vector almost attains the CRLB and the FD-AF-OWRN system has excellent SER performance under these scenarios.

## APPENDIX A
## CALCULATION OF THE AUTOCORRELATION MATRIX OF $\boldsymbol{\eta}$

First, we need to derive the expectations $E_\theta\{\mathbf{H}_0\mathbf{H}_0^\dagger\}$ and $E_\theta\{\mathbf{H}_0^\dagger\mathbf{H}_0\}$ that are used while obtaining the autocorrelation matrix of the overall additive noise vector $\boldsymbol{\eta}$ and the power scaling factor in Appendix-B, respectively. Since $\theta \sim \mathcal{CN}(0, \sigma_\theta^2)$ with $\sigma_\theta^2 = \alpha^2\Omega_0^2$, it can be easily shown that $E_\theta\{|\theta|^{2m}\} = m!\,\sigma_\theta^{2m}$ and

$$
\begin{aligned}
E_\theta\{\theta^p(\theta^*)^q\} &= E_\theta\left\{|\theta|^{p+q}\right\}\delta[p-q] \\
&= E_\theta\left\{|\theta|^{2p}\right\}\delta[p-q] \\
&= p!\,\sigma_\theta^{2p}\,\delta[p-q], \quad (51)
\end{aligned}
$$

where $\delta[\cdot]$ is the Krönecker $\delta$ function. Using (13) and (51), we can show that

$$
\begin{aligned}
\left[E_\theta\{\mathbf{H}_0\mathbf{H}_0^\dagger\}\right]_{p,q} &= \sum_{m=1}^{N} E_\theta\left\{[\mathbf{H}_0]_{p,m}[\mathbf{H}_0]_{q,m}^*\right\} \\
&= \sum_{m=1}^{\min(p,q)-1} E_\theta\left\{\theta^{(p-m-1)}(\theta^*)^{(q-m-1)}\right\} \\
&= \sum_{m=1}^{p-1} E_\theta\left\{|\theta|^{2(p-1-m)}\right\}\delta[p-q] \\
&= \sum_{m=0}^{p-2} E_\theta\left\{|\theta|^{2m}\right\}\delta[p-q] \\
&= \vartheta[p]\,\delta[p-q], \quad (52)
\end{aligned}
$$

where

$$
\vartheta[p] = \sum_{m=0}^{p-2} m!\,\sigma_\theta^{2m}, \; p = 1, 2, \ldots, N, \quad (53)
$$

and $(\cdot)^\dagger$ denotes the Hermitian transpose operator. From (52), we conclude that $E_\theta\{\mathbf{H}_0\mathbf{H}_0^\dagger\}$ is an $N \times N$ diagonal matrix with $p$th diagonal entry given in (53). Note that the first entry on the main diagonal $\vartheta[1] = 0$ by the definition of summation since the final value of the summation index is less than initial value for $p = 1$ in (52). In other words,

$$
E_\theta\{\mathbf{H}_0\mathbf{H}_0^\dagger\} = \text{diag}\{\vartheta[1], \vartheta[2], \ldots, \vartheta[N]\}. \quad (54)
$$

Following similar steps in (52), we can obtain also

$$E_\theta\{\mathbf{H}_0^\dagger\mathbf{H}_0\} = \text{diag}\{\vartheta[N], \vartheta[N-1], \ldots, \vartheta[1]\}, \quad (55)$$

It is straightforward that the overall additive noise vector $\boldsymbol{\eta}$ in (11) is zero-mean non-Gaussian random vector whose autocorrelation matrix can be obtained as

$$
\begin{aligned}
\mathbf{R}_{\boldsymbol{\eta}} &= E\{\boldsymbol{\eta}\boldsymbol{\eta}^\dagger\} \\
&= E\{(\alpha\mathbf{H}_2\mathbf{H}_0\mathbf{w}_1 + \mathbf{w}_2)(\alpha\mathbf{H}_2\mathbf{H}_0\mathbf{w}_1 + \mathbf{w}_2)^\dagger\} \\
&= \alpha^2\sigma_1^2 E\{\mathbf{H}_2\mathbf{H}_0\mathbf{H}_0^\dagger\mathbf{H}_2^\dagger\} + \sigma_2^2\mathbf{I}_N \\
&= \alpha^2\sigma_1^2\left[E\{\mathbf{H}_0\mathbf{H}_0^\dagger\} \odot E\{\mathbf{h}_2\mathbf{h}_2^\dagger\}\right] + \sigma_2^2\mathbf{I}_N \\
&= \alpha^2\Omega_2^2\sigma_1^2\,\text{diag}\{\vartheta[1], \vartheta[2], \ldots, \vartheta[N]\} + \sigma_2^2\mathbf{I}_N, \quad (56)
\end{aligned}
$$

where $\odot$ is the Hadamard product and note that $\mathbf{R}_{\boldsymbol{\eta}}$ is a diagonal matrix which means that the samples of the time domain overall additive noise are uncorrelated.

## APPENDIX B
### CALCULATION OF POWER SCALING FACTOR

In the AF protocol, the relay amplifies the received signal by factor $\alpha > 0$ to satisfy the average transmit power constraint. From (2), (15), (16) and (11), the transmitted signals at the relay can be given by

$$\mathbf{s}_2 = \alpha\mathbf{H}_0(\mathbf{H}_1\mathbf{s} + \mathbf{w}_1), \quad (57)$$

and the average transmit power of the relay is defined as

$$
\begin{aligned}
P_r &= \frac{1}{N}\sum_{n=1}^{N-1} E\{|s_2[n]|^2\} \\
&= \frac{1}{N} E\{\mathbf{s}_2^\dagger\mathbf{s}_2\} \\
&= \frac{\alpha^2}{N} E\{(\mathbf{H}_1\mathbf{s} + \mathbf{w}_1)^\dagger E_\theta\{\mathbf{H}_0^\dagger\mathbf{H}_0\}(\mathbf{H}_1\mathbf{s} + \mathbf{w}_1)\}. \quad (58)
\end{aligned}
$$

In this calculation, we approximate $\mathbf{H}_0$ with its truncated version with the maximum 2nd power of $\sigma_\theta^2$ since $\sigma_\theta^2 \ll 1$. Under this approximation, the entries of the diagonal matrix $E_\theta\{\mathbf{H}_0^\dagger\mathbf{H}_0\}$ are given in (53) as

$$\vartheta[1] = 0,\ \vartheta[2] = 1,\ \vartheta[3] = 1 + \sigma_\theta^2,$$
$$\vartheta[p] \cong 1 + \sigma_\theta^2 + 2\sigma_\theta^4,\ p = 4, 5, \ldots, N. \quad (59)$$

Making the training signal power equal to the transmit signal power $\sigma_s^2$, assuming $N \gg 1$ and using (55), we obtain

$$
\begin{aligned}
P_r &= \alpha^2(\Omega_1^2\sigma_s^2 + \sigma_1^2)\frac{1}{N}\sum_{p=1}^{N}\vartheta[p] \\
&= \alpha^2(\Omega_1^2\sigma_s^2 + \sigma_1^2)\frac{(N-1) + (N-2)\sigma_\theta^2 + 2(N-3)\sigma_\theta^4}{N} \\
&\cong \alpha^2(\Omega_1^2\sigma_s^2 + \sigma_1^2)(1 + \sigma_\theta^2 + 2\sigma_\theta^4). \quad (60)
\end{aligned}
$$

Multiplying both sides of (60) with $\Omega_0^2/(\Omega_1^2\sigma_s^2 + \sigma_1^2)$ and recalling $\sigma_\theta^2 = \alpha^2\Omega_0^2$, (60) can be rearranged as he following cubic

equation of $\sigma_\theta^2$

$$2(\sigma_\theta^2)^3 + (\sigma_\theta^2)^2 + \sigma_\theta^2 - \mathcal{P} = 0, \quad (61)$$

where $\mathcal{P} = \Omega_0^2 P_r/(\Omega_1^2\sigma_s^2 + \sigma_1^2)$. One real root and a pair of complex conjugate roots of (61) are found as

$$
\begin{aligned}
\sigma_\theta^2 &= -\frac{1}{6} + E + F, \\
\sigma_\theta^2 &= -\frac{1}{6} - \frac{1}{2}(E + F) + \frac{i\sqrt{3}}{2}(E - F), \\
\sigma_\theta^2 &= -\frac{1}{6} - \frac{1}{2}(E + F) - \frac{i\sqrt{3}}{2}(E - F), \quad (62)
\end{aligned}
$$

respectively, where

$$E = \sqrt[3]{\frac{\mathcal{P}}{4} + \frac{1}{27} + \sqrt{\left(\frac{\mathcal{P}}{4} + \frac{1}{27}\right)^2 + \left(\frac{5}{36}\right)^3}}$$

$$F = -\sqrt[3]{\left|\frac{\mathcal{P}}{4} + \frac{1}{27} - \sqrt{\left(\frac{\mathcal{P}}{4} + \frac{1}{27}\right)^2 + \left(\frac{5}{36}\right)^3}\right|}. \quad (63)$$

Since $\alpha$ must be real, eventually it is calculated with respect to real valued root given in the first line of (62) as follows

$$\alpha = \sqrt{\frac{\sigma_\theta^2}{\Omega_0^2}} = \frac{1}{\Omega_0}\sqrt{-\frac{1}{6} + E + F}. \quad (64)$$

## APPENDIX C
### EVALUATIONS OF BAYESIAN CRÁMER RAO LOWER BOUNDS

In this appendix, Bayesian Crámer Rao Lower Bounds (CRLBs) on the MSEs for the estimation of the unknown parameter set $\boldsymbol{\gamma} = [\theta\ \mathbf{c}^T]^T$ are derived. There are various Bayesian CRLB on the error correlation matrix in the literature[47]. For CRLB evaluations we use the following observation model

$$\mathbf{r} = \boldsymbol{\mu} + \boldsymbol{\zeta}, \quad (65)$$

that is obtained by stacking the observations in (36) and (37), where

$$
\mathbf{r} = \begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix},\ \boldsymbol{\mu} = \alpha b\begin{bmatrix} \boldsymbol{\Phi}_1\mathbf{c} \\ \boldsymbol{\Phi}_2\mathbf{c} + \boldsymbol{\Phi}_1\mathbf{c}\,\theta \end{bmatrix},
$$

$$
\boldsymbol{\zeta} = \begin{bmatrix} \boldsymbol{\omega} \\ \boldsymbol{\xi} \end{bmatrix} = \boldsymbol{\Upsilon}\boldsymbol{\eta}_{\text{tr}},\ \boldsymbol{\Upsilon} = \begin{bmatrix} \mathbf{I}_K \otimes \mathbf{a}_1^\dagger \\ \mathbf{I}_K \otimes \mathbf{a}_2^\dagger \end{bmatrix}. \quad (66)
$$

The classical Bayesian CRLB for the estimation of the parameter set $\boldsymbol{\gamma}$ is given by

$$\mathcal{B}_{\boldsymbol{\gamma}} = \mathbf{J}_{\boldsymbol{\gamma}}^{-1}, \quad (67)$$

where $\mathbf{J}_{\boldsymbol{\gamma}}$ is the Bayesian Fisher Information Matrix (FIM) defined as

$$
\begin{aligned}
\mathbf{J}_{\boldsymbol{\gamma}} &= E_{\mathbf{r},\boldsymbol{\gamma}}\left\{\frac{\partial\log p(\mathbf{r},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^*}\frac{\partial\log p(\mathbf{r},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T}\right\} \\
&= -E_{\mathbf{r},\boldsymbol{\gamma}}\left\{\frac{\partial^2\log p(\mathbf{r},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^*\partial\boldsymbol{\gamma}^T}\right\}. \quad (68)
\end{aligned}
$$

However, in our model, Bayesian FIM is intractable. Instead of the Bayesian CRLB in (67), we use following tractable Bayesian

CRLB [47], [48]

$$\mathcal{B}_{\boldsymbol{\gamma}} = E_{\mathbf{h}_2}\{\mathbf{J}_{\boldsymbol{\gamma}|\mathbf{h}_2}^{-1}\}, \tag{69}$$

that is the average of the (67) over nuisance parameter $\mathbf{h}_2$. The Bayesian FIM in (69) is defined as

$$\begin{aligned}
\mathbf{J}_{\boldsymbol{\gamma}|\mathbf{h}_2} &= E_{\mathbf{r},\boldsymbol{\gamma}|\mathbf{h}_2}\left\{\frac{\partial \log p(\mathbf{r},\boldsymbol{\gamma}|\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^*}\frac{\partial \log p(\mathbf{r},\boldsymbol{\gamma}|\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^T}\right\} \\
&= -E_{\mathbf{r},\boldsymbol{\gamma}|\mathbf{h}_2}\left\{\frac{\partial^2 \log p(\mathbf{r},\boldsymbol{\gamma}|\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^*\partial \boldsymbol{\gamma}^T}\right\}. \tag{70}
\end{aligned}$$

In [48], it is stated that exact ordering between the bounds in (67) and (69) is not possible and, in some cases, (69) can be tightest bound. $\mathbf{J}_{\boldsymbol{\gamma}|\mathbf{h}_2}$ in (70) can be expressed as the sum of $\mathbf{J}_{\boldsymbol{\gamma}|\mathbf{h}_2}$ data and $\mathbf{J}_{\boldsymbol{\gamma}|\mathbf{h}_2}$ prior that are the Fisher Informations evaluated from data and prior information, respectively:

$$\mathbf{J}_{\boldsymbol{\gamma}|\mathbf{h}_2} = \mathbf{J}^{\text{data}} + \mathbf{J}^{\text{prior}}, \tag{71}$$

where

$$\begin{aligned}
\mathbf{J}^{\text{data}} &= E_{\mathbf{r},\boldsymbol{\gamma}|\mathbf{h}_2}\left\{\frac{\partial \log p(\mathbf{r}|\boldsymbol{\gamma},\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^*}\frac{\partial \log p(\mathbf{r}|\boldsymbol{\gamma},\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^T}\right\} \\
&= -E_{\mathbf{r},\boldsymbol{\gamma}|\mathbf{h}_2}\left\{\frac{\partial^2 \log p(\mathbf{r}|\boldsymbol{\gamma},\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^*\partial \boldsymbol{\gamma}^T}\right\} \\
\mathbf{J}^{\text{prior}} &= E_{\boldsymbol{\gamma}|\mathbf{h}_2}\left\{\frac{\partial \log p(\boldsymbol{\gamma}|\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^*}\frac{\partial \log p(\boldsymbol{\gamma}|\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^T}\right\} \\
&= -E_{\boldsymbol{\gamma}|\mathbf{h}_2}\left\{\frac{\partial^2 \log p(\boldsymbol{\gamma}|\mathbf{h}_2)}{\partial \boldsymbol{\gamma}^*\partial \boldsymbol{\gamma}^T}\right\}. \tag{72}
\end{aligned}$$

Taking $\boldsymbol{\gamma}_1 = \theta$ and $\boldsymbol{\gamma}_2 = \mathbf{c}$ with $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1^T\ \boldsymbol{\gamma}_2^T]^T$, each of $\mathbf{J}^{\text{data}}$ and $\mathbf{J}^{\text{prior}}$ can be partitioned into sub-matrices as follows

$$\mathbf{J}^{\text{data}} = \begin{bmatrix} \mathbf{J}_{11}^{\text{data}} & \mathbf{J}_{12}^{\text{data}} \\ \mathbf{J}_{21}^{\text{data}} & \mathbf{J}_{22}^{\text{data}} \end{bmatrix}, \quad \mathbf{J}^{\text{prior}} = \begin{bmatrix} \mathbf{J}_{11}^{\text{prior}} & \mathbf{J}_{12}^{\text{prior}} \\ \mathbf{J}_{21}^{\text{prior}} & \mathbf{J}_{22}^{\text{prior}} \end{bmatrix}, \tag{73}$$

and using rightmost expectations in (72), the sub-matrices of FIMs can also be defined as follows

$$\begin{aligned}
\mathbf{J}_{ij}^{\text{data}} &= -E_{\mathbf{r},\theta,\mathbf{c}|\mathbf{h}_2}\left\{\frac{\partial^2 \log p(\mathbf{r}|\theta,\mathbf{c},\mathbf{h}_2)}{\partial \boldsymbol{\gamma}_i^*\partial \boldsymbol{\gamma}_j^T}\right\}, \ i,j \in \{1,2\} \\
&= -E_\theta E_{\mathbf{c}|\mathbf{h}_2}E_{\mathbf{r}|\theta,\mathbf{c},\mathbf{h}_2}\left\{\frac{\partial^2 \log p(\mathbf{r}|\theta,\mathbf{c},\mathbf{h}_2)}{\partial \boldsymbol{\gamma}_i^*\partial \boldsymbol{\gamma}_j^T}\right\} \tag{74}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{J}_{ij}^{\text{prior}} &= -E_{\theta,\mathbf{c}|\mathbf{h}_2}\left\{\frac{\partial^2 \log p(\theta,\mathbf{c}|\mathbf{h}_2)}{\partial \boldsymbol{\gamma}_i^*\partial \boldsymbol{\gamma}_j^T}\right\}, \ i,j \in \{1,2\} \\
&= -E_\theta\left\{\frac{\partial^2 \log p(\theta)}{\partial \boldsymbol{\gamma}_i^*\partial \boldsymbol{\gamma}_j^T}\right\} - E_{\mathbf{c}|\mathbf{h}_2}\left\{\frac{\partial^2 \log p(\mathbf{c}|\mathbf{h}_2)}{\partial \boldsymbol{\gamma}_i^*\partial \boldsymbol{\gamma}_j^T}\right\}. \tag{75}
\end{aligned}$$

In (74) and (75), note that $\{\mathbf{r}|\theta,\mathbf{c},\mathbf{h}_2\} \sim \mathcal{CN}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $\{\mathbf{c}|\mathbf{h}_2\} \sim \mathcal{CN}(\mathbf{0}_D,\mathbf{R}_{\mathbf{c}|\mathbf{h}_2})$ together with

$$\begin{aligned}
\boldsymbol{\Sigma} &= E\{\boldsymbol{\zeta}\boldsymbol{\zeta}^\dagger|\theta,\mathbf{c},\mathbf{h}_2\} \\
&= \boldsymbol{\Upsilon}E\{\boldsymbol{\eta}_{\text{tr}}\boldsymbol{\eta}_{\text{tr}}^\dagger|\theta,\mathbf{c},\mathbf{h}_2\}\boldsymbol{\Upsilon}^\dagger \\
&= \boldsymbol{\Upsilon}\left[E\{\boldsymbol{\eta}\boldsymbol{\eta}^\dagger|\theta,\mathbf{c},\mathbf{h}_2\}\right]_{\text{tr}}\boldsymbol{\Upsilon}^\dagger \\
&= \boldsymbol{\Upsilon}\left[\alpha^2\sigma_1^2(\mathbf{H}_0\mathbf{H}_0^\dagger)\odot(\mathbf{h}_2\mathbf{h}_2^\dagger)+\sigma_2^2\mathbf{I}_N\right]_{\text{tr}}\boldsymbol{\Upsilon}^\dagger \in \mathcal{C}^{2K\times 2K}, \tag{76}
\end{aligned}$$

and

$$\frac{\partial\boldsymbol{\Sigma}}{\partial\theta} = \boldsymbol{\Upsilon}\left[\alpha^2\sigma_1^2(\mathbf{Q}_0\mathbf{H}_0^\dagger)\odot(\mathbf{h}_2\mathbf{h}_2^\dagger)\right]_{\text{tr}}\boldsymbol{\Upsilon}^\dagger \in \mathcal{C}^{2K\times 2K}, \tag{77}$$

$$\begin{aligned}
\mathbf{R}_{\mathbf{c}|\mathbf{h}_2} &= E\{\mathbf{c}\mathbf{c}^\dagger|\mathbf{h}_2\} \\
&= \boldsymbol{\Psi}^T E\{\mathbf{h}\mathbf{h}^\dagger|\mathbf{h}_2\}\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}^T\left[\mathbf{R}_{\mathbf{h}_1}\odot(\mathbf{h}_2\mathbf{h}_2^\dagger)\right]\boldsymbol{\Psi} \in \mathcal{C}^{D\times D}, \tag{78}
\end{aligned}$$

where $\mathbf{Q}_0 = \partial\mathbf{H}_0/\partial\theta$. Using the following matrix identities

$$\begin{aligned}
\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\theta} &= -\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta}\boldsymbol{\Sigma}^{-1} \\
\frac{\partial\log\det(\boldsymbol{\Sigma})}{\partial\theta} &= \text{trace}\left[\frac{\partial\boldsymbol{\Sigma}}{\partial\theta}\boldsymbol{\Sigma}^{-1}\right] \\
\text{trace}[\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2] &= \text{trace}[\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1], \tag{79}
\end{aligned}$$

we obtain the sub-matrices of FIMs as follows

$$J_{11}^{\text{data}}$$
$$= E_\theta E_{\mathbf{c}|\mathbf{h}_2}\left\{\text{trace}\left[\frac{\partial\boldsymbol{\Sigma}}{\partial\theta^*}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta}\boldsymbol{\Sigma}^{-1}\right]+\frac{\partial\boldsymbol{\mu}^\dagger}{\partial\theta^*}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta}\right\}$$
$$= E_\theta\left\{\text{trace}\left[\frac{\partial\boldsymbol{\Sigma}}{\partial\theta^*}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta}\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}^{-1}E_{\mathbf{c}|\mathbf{h}_2}\left\{\frac{\partial\boldsymbol{\mu}}{\partial\theta}\frac{\partial\boldsymbol{\mu}^\dagger}{\partial\theta^*}\right\}\right]\right\}$$

$$J_{22}^{\text{data}} = E_\theta E_{\mathbf{c}|\mathbf{h}_2}\left\{\frac{\partial\boldsymbol{\mu}^\dagger}{\partial\mathbf{c}^*}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{c}^T}\right\}$$
$$= E_\theta\left\{\frac{\partial\boldsymbol{\mu}^\dagger}{\partial\mathbf{c}^*}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{c}^T}\right\}$$

$$J_{12}^{\text{data}} = \mathbf{0}_D^T, \ J_{21}^{\text{data}} = \mathbf{0}_D \tag{80}$$

where

$$\frac{\partial\boldsymbol{\mu}}{\partial\theta} = \alpha b\begin{bmatrix}\mathbf{0}_K \\ \boldsymbol{\Phi}_1\mathbf{c}\end{bmatrix}, \quad \frac{\partial\boldsymbol{\mu}^\dagger}{\partial\theta^*} = \left(\frac{\partial\boldsymbol{\mu}}{\partial\theta}\right)^\dagger$$

$$\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{c}^T} = \alpha b\begin{bmatrix}\boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2+\boldsymbol{\Phi}_1\theta\end{bmatrix}, \quad \frac{\partial\boldsymbol{\mu}^\dagger}{\partial\mathbf{c}^*} = \left(\frac{\partial\boldsymbol{\mu}}{\partial\mathbf{c}^T}\right)^\dagger$$

$$E_{\mathbf{c}|\mathbf{h}_2}\left\{\frac{\partial\boldsymbol{\mu}}{\partial\theta}\frac{\partial\boldsymbol{\mu}^\dagger}{\partial\theta^*}\right\} = |\alpha b|^2\begin{bmatrix}0 & 0 \\ 0 & 1\end{bmatrix}\otimes(\boldsymbol{\Phi}_1\mathbf{R}_{\mathbf{c}|\mathbf{h}_2}\boldsymbol{\Phi}_1^\dagger) \tag{81}$$

$$\mathbf{J}_{11}^{\text{prior}} = \frac{1}{\sigma_\theta^2}, \quad \mathbf{J}_{22}^{\text{prior}} = \mathbf{R}_{\mathbf{c}|\mathbf{h}_2}^{-1},$$

$$\mathbf{J}_{12}^{\text{prior}} = \mathbf{0}_D^T, \quad \mathbf{J}_{21}^{\text{prior}} = \mathbf{0}_D \tag{82}$$

As seen (80) and (82), FIM matrices $\mathbf{J}^{\text{data}}$ and $\mathbf{J}^{\text{prior}}$ are block diagonal matrices, so the inverse of the FIM $\mathbf{J}_{\boldsymbol{\gamma}|\mathbf{h}_2}$ is a block diagonal matrix that can be partitioned as follows

$$
\begin{aligned}
\mathcal{B}_{\boldsymbol{\gamma}} &= E_{\mathbf{h}_2}\{\mathbf{J}_{\boldsymbol{\gamma}|\mathbf{h}_2}^{-1}\} \\
&= E_{\mathbf{h}_2}\{(\mathbf{J}^{\text{data}} + \mathbf{J}^{\text{prior}})^{-1}\} \\
&= \begin{bmatrix} \mathcal{B}_{\theta} & \mathbf{0}_D^T \\ \mathbf{0}_D & \mathcal{B}_{\mathbf{c}} \end{bmatrix},
\end{aligned} \tag{83}
$$

where $\mathcal{B}_{\theta}$ and $\mathcal{B}_{\mathbf{c}}$ represent Bayesian error covariance bound for the estimation of $\theta$ and $\mathbf{c}$, respectively, and they are calculated as

$$
\begin{aligned}
\mathcal{B}_{\theta} &= E_{\mathbf{h}_2}\{(\mathbf{J}_{11}^{\text{data}} + \mathbf{J}_{11}^{\text{prior}})^{-1}\}, \\
\mathcal{B}_{\mathbf{c}} &= E_{\mathbf{h}_2}\{(\mathbf{J}_{22}^{\text{data}} + \mathbf{J}_{22}^{\text{prior}})^{-1}\}.
\end{aligned} \tag{84}
$$

To compute the expectations in (80) and (84) with respect to $\theta$ and $\mathbf{c}$, respectively, we simply use the Gibbs sampling technique and generate samples from pdfs $\theta^{[i]} \sim \mathcal{CN}(0, \sigma_{\theta}^2)$, $i = 1, 2, \ldots, N_1$ and $\mathbf{h}_2^{[j]} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{\mathbf{h}_2})$, $j = 1, 2, \ldots, N_2$ then calculate Gibbs samples of the sample means of the expressions inside the expectations and replace sample means with the expectations. Eventually, Bayesian MSE bound for the estimation of $\theta$ is obtained as

$$
\text{MSE}_{\theta} = E_{\mathbf{r}, \theta}\{|\theta - \hat{\theta}|^2\} \geq \mathcal{B}_{\theta}, \tag{85}
$$

and consequently, average Bayesian MSE bound for the estimation of the vector $\mathbf{c}$ is found as

$$
\begin{aligned}
\text{Average MSE}_{\mathbf{c}} &= \frac{1}{D} E_{\mathbf{r}, \mathbf{c}}\{(\mathbf{c} - \hat{\mathbf{c}})^{\dagger}(\mathbf{c} - \hat{\mathbf{c}})\} \\
&= \frac{1}{D} \text{trace}[E_{\mathbf{r}, \mathbf{c}}\{(\mathbf{c} - \hat{\mathbf{c}})(\mathbf{c} - \hat{\mathbf{c}})^{\dagger}\}] \\
&\geq \frac{1}{D} \text{trace}[\mathcal{B}_{\mathbf{c}}].
\end{aligned} \tag{86}
$$

## References

[1] 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.

[2] B. Debaillie *et al.*, "Analog/RF solutions enabling compact full-duplex radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1662–1673, Sep. 2014.

[3] Z. Zhang, X. Chai, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 128–137, May 2015.

[4] M. Heino *et al.*, "Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 91–101, May 2015.

[5] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Jun. 2014.

[6] S.-K. Hong *et al.*, "Applications of self-interference cancellation in 5G and beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb. 2014.

[7] J. Ma, G. Y. Li, J. Zhang, T. Kuze, and H. Iura, "A new coupling channel estimator for cross-talk cancellation at wireless relay stations," in *Proc. IEEE Global Telecommun. Conf.*, Honolulu, HI, USA, Nov. 30–Dec. 4, 2009, pp. 1–6.

[8] 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.

[9] A. Masmoudi and T. Le-Ngoc, "A maximum-likelihood channel estimator for self-interference cancellation in full-duplex systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5122–5132, Jul. 2016.

[10] A. Koohian, H. Mehrpouyan, M. Ahmadian, and M. Azarbad, "Bandwidth efficient channel estimation for full duplex communication systems," in *Proc. IEEE Int. Conf. Commun.*, London, U.K., Jun. 8–12, 2015, pp. 4710–4714.

[11] R. Li, A. Masmoudi and T. Le-Ngoc, "Self-interference cancellation with phase-noise suppression in full-duplex systems," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Hong Kong, Aug. 30–Sep. 2, 2015, pp. 261–265.

[12] R. Li, A. Masmoudi, and T. Le-Ngoc, "Self-interference cancellation with nonlinearity and phase-noise suppression in full-duplex systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2118–2129, Mar. 2018.

[13] T. Riihonen, S. Werner, and R. Wichman, "Residual self-interference in full-duplex MIMO relays after null-space projection and cancellation," in *Proc. IEEE 44th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 7–10, 2010, pp. 653–657.

[14] H. Senol and C. Tepedelenlioglu, "Performance of distributed estimation over unknown parallel fading channels," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6057–6068, Dec. 2008.

[15] E. Panayirci, H. Senol, M. Uysal, and H. V. Poor, "Sparse channel estimation and equalization for OFDM-based underwater cooperative systems with amplify-and-forward relaying," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 214–228, Jan. 2016.

[16] T. M. Kim and A. Paulraj, "Outage probability of amplify-and-forward cooperation with full duplex relay," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Shanghai, China, Apr. 1–4, 2012, pp. 75–79.

[17] L. Jimenez Rodriguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex AF relaying in the presence of residual self-interference," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1752–1764, Sep. 2014.

[18] L. Jimenez Rodriguez, N. H. Tran, and T. Le-Ngoc, "Optimal power allocation and capacity of full-duplex AF relaying under residual self-interference," *IEEE Wireless Commun. Lett.*, vol. 3, no. 2, pp. 233–236, Apr. 2014.

[19] F. S. Tabataba, P. Sadeghi, C. Hucher, and M. R. Pakravan, "Impact of channel estimation errors and power allocation on analog network coding and routing in two-way relaying," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3223–3239, Sep. 2012.

[20] D. Kim, H. Ju, S. Park, and D. Hong, "Effects of channel estimation error on full-duplex two-way networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4666–4672, Nov. 2013.

[21] X. Cheng, B. Yu, X. Cheng, and L. Yang, "Two-way full-duplex amplify-and-forward relaying," in *Proc. IEEE Mil. Commun. Conf.*, San Diego, CA, USA, Nov. 18–20, 2013, pp. 1–6.

[22] G. Zheng, "Joint beamforming optimization and power control for full-duplex MIMO two-way relay channel," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 555–566, Feb. 2015.

[23] X. Li, C. Tepedelenlioglu, and H. Senol, "Channel estimation for residual self-interference in full duplex amplify-and-forward two-way relays," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4970–4983, Aug. 2017.

[24] J. S. Lemos, F. A. Monteiro, I. Sousa, and A. Rodrigues, "Full-duplex relaying in MIMO-OFDM frequency-selective channels with optimal adaptive filtering," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Orlando, FL, USA, Dec. 14–16, 2015, pp. 1081–1085.

[25] M. Mohammadi, B. K. Chalise, H. A. Suraweera, C. Zhong, G. Zheng, and I. Krikidis, "Throughput analysis and optimization of wireless-powered multiple antenna full-duplex relay systems," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1769–1785, Apr. 2016.

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

[27] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3702–3713, Jul. 2012.

[28] O. Taghizadeh, M. Rothe, A. C. Cirik, and R. Mathar, "Distortion-loop analysis for full-duplex amplify-and-forward relaying in cooperative multicast scenarios," in *Proc. 9th Int. Conf. Signal Process. Commun. Syst.*, Cairns, QLD, Australia, Dec. 14–16, 2015, pp. 1–9.

[29] A. C. Cirik, M. C. Filippou, and T. Ratnarajaht, "Transceiver design in full-duplex MIMO cognitive radios under channel uncertainties," *IEEE Trans. Cogn. Commun. Netw.*, vol. 2, no. 1, pp. 1–14, Mar. 2016.

[30] O. Taghizadeh, T. Yang, A. C. Cirik, and R. Mathar, "Distortion-loop-aware amplify-and-forward full-duplex relaying with multiple antennas," in *Proc. Int. Symp. Wireless Commun. Syst.*, Poznan, Poland, Sep. 20–23, 2016, pp. 54–58.

[31] A. Masmoudi and T. Le-Ngoc, "Self-interference mitigation using active signal injection for full-duplex MIMO-OFDM systems," in *Proc. IEEE 84th Veh. Technol. Conf.*, Montreal, QC, Canada, Sep. 18–21, 2016.

[32] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sep. 2014.

[33] X. Xiong, X. Wang, T. Riihonen, and X. You, "Channel estimation for full-duplex relay systems with large-scale antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6925–6938, Oct. 2016.

[34] X. Li, C. Tepedelenlioglu, and H. Senol, "Optimal training for residual self-interference for full duplex one-way relays," *IEEE Trans. Commun.*, to be published.

[35] X. Li and C. Tepedelenlioğlu, "Maximum likelihood channel estimation for residual self-interference cancellation in full duplex relay," in *Proc. IEEE 49th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 8–11, 2015, pp. 807–811.

[36] A. Gaston, W. Chriss, and E. Walker, "A multipath fading simulator for radio," *IEEE Trans. Veh. Technol.*, vol. VT-22, no. 4, pp. 241–244, Nov. 1973.

[37] K. A. D. Teo and S. Ohno, "Optimal MMSE finite parameter model for doubly-selective channels," in *Proc. IEEE Global Telecommun. Conf.*, St. Louis, MO, USA, Nov. 28–Dec. 2, 2005, pp. 3503–3507.

[38] T. Zemen and C. F. Mecklenbrauker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sep. 2005.

[39] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time varying channels," *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986, Oct. 1998.

[40] E. Panayirci, H. Senol, and H. V. Poor, "Joint channel estimation, equalization and data detection for OFDM systems in the presence of very high mobility," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4225–4238, Aug. 2010.

[41] H. Senol, E. Panayirci, and H. V. Poor, "Non-data-aided joint channel estimation and equalization for OFDM systems in very rapidly varying mobile channels," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4236–4253, Aug. 2012.

[42] 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.

[43] C. S. Patel and G. L. Stuber, "Channel estimation for amplify and forward relay based cooperation diversity systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 3348–3356, Jun. 2007.

[44] G. Wang, F. Gao, W. Chen, and C. Tellambura, "Channel estimation and training design for two-way relay networks in time-selective fading environments," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2681–2691, Aug. 2011.

[45] A. Böttcher, "Orthogonal symmetric Toeplitz matrices," *Complex Anal. Operator Theory*, vol. 2, no. 2, pp. 285–298, May 2008.

[46] O. Edfors, M. Sandell, J. J. van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[47] H. L. V. Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Hoboken, NJ, USA: Wiley, 2007.

[48] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramér-Rao bounds," *Ann. Statist.*, vol. 15, no. 4, pp. 1421–1438, Dec. 1987.

**Habib Şenol** (S'04–M'07) received the B.S. and M.S. degrees from Istanbul University, Istanbul, Turkey, in 1993 and 1999, respectively, and the Ph.D. degree from Işık University, Istanbul, Turkey, in 2006, all in electronics engineering.

He is currently an Associate Professor of Computer Engineering with Kadir Has University, Istanbul, Turkey. From 2007 to 2008, he was with the Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA, working on channel estimation and power optimization algorithms for *Wireless Sensor Networks*. His research interests include statistical signal processing techniques and their applications to wireless electrical/underwater acoustic/optical communication systems, estimation and equalization algorithms for wireless communications, orthogonal frequency division multiplexing, multicarrier communications for 5G and beyond wireless networks, and statistical machine learning.
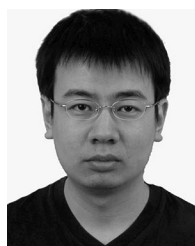
**Xiaofeng Li** received the B.S. and M.S. degrees in electronic engineering from Beijing Institute of Technology, Beijing, China, in 2009 and 2012, respectively. He is currently working toward the Ph.D. degree in electrical engineering at Arizona State University, Tempe, AZ, USA. His research interests include channel estimation, full-duplex relays, and mm-Wave MIMO.

**Cihan Tepedelenlioğlu** (S'97–M'01) was born in Ankara, Turkey, in 1973. He received the B.S. degree with highest Hons. from the Florida Institute of Technology, Melbourne, FL, USA, and the M.S. degree from the University of Virginia, Charlottesville, VA, USA, in 1995 and 1998, respectively, both in electrical engineering, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA.

From January 1999 to May 2001, he was a Research Assistant with the University of Minnesota. He is currently an Associate Professor of electrical engineering with Arizona State University, Tempe, AZ, USA. His research interests include statistical signal processing, system identification, wireless communications, estimation and equalization algorithms for wireless systems, multiantenna communications, OFDM, ultrawideband systems, distributed detection and estimation, and data mining for PV systems. He was the recipient of the NSF (early) Career grant in 2001, and was an Associate Editor for several IEEE Transactions including the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.