1-24-2018

# Superimposed Signaling Inspired Channel Estimation in Full-Duplex Systems

Abbas Koohian
*Australian National University*

Hani Mehrpouyan
*Boise State University*

Ali A. Nasir
*King Fahd University of Petroleum and Minerals*

Salman Durrani
*Australian National University*

Steven D. Blostein
*Queen's University*

Koohian *et al. EURASIP Journal on Advances in Signal Processing* (2018) 2018:8
DOI 10.1186/s13634-018-0529-9

EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

# Superimposed signaling inspired channel estimation in full-duplex systems

Abbas Koohian[1*], Hani Mehrpouyan[2], Ali A. Nasir[3], Salman Durrani[1] and Steven D. Blostein[4]

**Abstract**

Residual self-interference (SI) cancellation in the digital baseband is an important problem in full-duplex (FD) communication systems. In this paper, we propose a new technique for estimating the SI and communication channels in a FD communication system, which is inspired from superimposed signaling. In our proposed technique, we add a constant real number to each constellation point of a conventional modulation constellation to yield asymmetric shifted modulation constellations with respect to the origin. We show mathematically that such constellations can be used for bandwidth efficient channel estimation without ambiguity. We propose an expectation maximization (EM) estimator for use with the asymmetric shifted modulation constellations. We derive a closed-form lower bound for the mean square error (MSE) of the channel estimation error, which allows us to find the minimum shift energy needed for accurate channel estimation in a given FD communication system. The simulation results show that the proposed technique outperforms the data-aided channel estimation method, under the condition that the pilots use the same extra energy as the shift, both in terms of MSE of channel estimation error and bit error rate. The proposed technique is also robust to an increasing power of the SI signal.

**Keywords:** Full-duplex systems, Channel estimation, Self-interference cancellation, Superimposed signaling

## 1 Introduction

### 1.1 Background

Full-duplex (FD) communication, allowing devices to transmit and receive over the same temporal and spectral resources, is a promising mechanism to potentially double the spectral efficiency of future wireless communication systems [1]. The main challenge in implementing a FD communication system is the cancellation of the strong self-interference (SI) signal, which is caused by transmission and reception in a single frequency band [2, 3]. This strong self-interference signal has to be suppressed to the receiver noise floor in order to ensure that it does not degrade the system performance. For instance, for small-cells in Long Term Evolution (LTE), the maximum transmit power is typically 23 dBm (200 mW) and the typical noise floor is − 90 dBm [4]. Ideally, this requires a total of 113 dB SI cancellation for realizing the full potential of FD systems [4].

*Correspondence: abbas.koohian@anu.edu.au
[1]Research School of Engineering, Australian National University, Canberra, Australia
Full list of author information is available at the end of the article

Recently, there has been a lot of interest in SI cancellation techniques for FD systems [5–11]. The SI cancellation techniques in the literature can be divided into two main categories [5]: (i) *passive* suppression in which the SI signal is suppressed by suitably isolating the transmit and receive antennas [5, 6] and (ii) *active* cancellation which uses knowledge of the SI signal to cancel the interference in either the analog domain (i.e., before the signal passes through the analog-to-digital converter) [5, 7] and/or the digital domain [8–10]. Depending upon the design, passive suppression and analog cancellation can provide about 40–60 dB cancellation in total [11]. Hence, in practice, the SI is cancelled in multiple stages, beginning with passive suppression and followed by cancellation in the analog and digital domains. In this paper, we focus on the SI after the passive suppression and analog cancellation, termed *residual SI*.

### 1.2 Motivation and related work

The residual SI can still be relatively strong in the baseband digital signal, e.g., for the LTE small-cell example, it can be as high as 50 dB assuming state-of-the-art passive suppression and analog cancellation provide 60 dB of the

total required SI cancellation of 113 dB. Thus, accurate digital SI cancellation is required to bring the SI as close to the noise floor as possible. The effectiveness of any digital interference cancellation technique depends strongly on the quality of the available channel estimates for both the SI and desired communication channels [12–14]. Typically, the baseband channels are estimated by using *data-aided channel estimation techniques*, where a portion of the data frame is allocated for known training sequences or pilot symbols [13, 15–17]. In this regard, a maximum-likelihood (ML) approach was proposed in [13] to jointly estimate the residual SI and communication channels by exploiting the known transmitted symbols and both the known pilot and unknown data symbols from the other intended transceiver. Another approach was proposed in [18] where a sub-space-based algorithm was developed to jointly estimate the residual SI and communication channels.

Compared to half-duplex (HD) systems, data-aided channel estimation in FD systems can be bandwidth inefficient. This is because, firstly, two channels need to be estimated and, secondly, accurate channel estimation requires a larger number of pilots [19, 20]. A bandwidth efficient channel estimation technique in HD systems is *superimposed training*, where no explicit time slots are allocated for channel estimation. Instead, a periodic low power training sequence is superimposed with the data symbols at the transmitter before modulation and transmission [21, 22]. The downside of this approach is that some power is consumed in superimposed training which could have otherwise been allocated to the data transmission. This lowers the effective signal-to-noise ratio (SNR) for the data symbols and affects the bit error rate (BER) at the receiver. In contrast to data-aided and superimposed training-based channel estimation techniques, *blind techniques* avoid the use of pilots altogether by exploiting statistical and other properties of the transmitted signal [23–27]. However, blind estimators can only estimate the channel up to a scaling factor and cannot recover the channel phases [23]. The necessary and sufficient conditions for ambiguity-free blind estimation can be determined using identifiability analysis, which determines whether a parameter can be uniquely estimated without any ambiguity [19, 23, 28–30]. To the best of our knowledge, bandwidth efficient and accurate channel estimation methods for FD systems are still an important open area of research.

### 1.3 Paper contributions

In this paper, we consider the problem of bandwidth efficient channel estimation in a single-input single-output (SISO) FD communication system. We propose a new technique for channel estimation and residual SI cancellation in FD systems. Our approach draws inspiration from (i) blind channel estimation techniques in that we examine the condition for identifiability of channel parameters in FD systems and (ii) superimposed signaling in that we superimpose (i.e., add) a constant real number to each constellation point of the modulation constellation. However, our proposed technique is distinct from superimposed signaling. In superimposed signaling, the superimposed signal is typically a periodic training sequence that is added to the data signal after the data symbols are modulated. Hence, the additional power of the superimposed signal is only used for channel estimation. In our proposed technique, the superimposed signal is a constant (non-random) signal and the objective is to shift the modulation constellation away from the origin, which we exploit for estimating the SI and communication channels without ambiguity. In addition, the additional power of the superimposed signal is used for both modulating the data symbols and channel estimation, which does not reduce the effective SNR as in superimposed signaling. The novel contributions are as follows:

- We derive the condition for identifiability of channel parameters in a FD system (cf. Theorem 1) and show that symmetric modulation constellations with respect to the origin cannot be used for ambiguity-free channel estimation in a FD system. Based on Theorem 1, our proposed technique is able to resolve the inherent ambiguity of blind channel estimation in FD communication via shifting the modulation constellation away from origin.
- Using the proposed technique, we derive a computationally efficient expectation maximization (EM) estimator for simultaneous estimation of both SI and communication channels. We derive a lower bound for the channel estimation error, which depends on the energy used for shifting the modulation constellations, and use it to find the minimum signal energy needed for accurate channel estimation in a given FD communication system.
- We use simulations to compare the performance of the proposed technique against that of the data-aided channel estimation method, under the condition that the pilots use the same extra power as the shift. Our results show that the proposed technique performs better than the data-aided channel estimation method both in terms of the mean square error (MSE) of channel estimation and BER. In addition, the proposed technique is robust to an increasing SI power.

### 1.4 Notation and paper organization

The following notation is used in this paper. Capital letters are used for random variables, and lower case letters are used for their realizations. $f_Y(y)$ denotes the

probability density function (PDF) of random variable $Y$. $\mathbb{E}_Y[\cdot]$ denotes the expectation with respect to the random variable $Y$. $p_X(x)$ denotes the probability mass function (PMF) of a discrete random variable $X$, and $P(X = a)$ is the probability of the discrete random variable $X$ taking the value $a$. $\mathcal{CN}\left(\mu, \sigma^2\right)$ denotes a complex Gaussian distribution with mean $\mu$ and variance $\sigma^2$. Bold face capital letters, e.g., $\mathbf{Y}$, are used for random vectors, and bold face lower case letters, e.g., $\mathbf{y}$, are used for their realizations. Capital letters in upright Roman font, e.g., G, are used for matrices. Lower case letters in upright Roman font, e.g., g, are used for functions. $\mathrm{I}_N$ represents the $N \times N$ identity matrix. $[\cdot]^T$ denotes vector and matrix transpose. $\mathrm{j} \triangleq \sqrt{-1}$ and the real and imaginary parts of a complex quantity are represented by $\Re\{\cdot\}$ and $\Im\{\cdot\}$, respectively. $z^*$ and $|z|$ indicate scalar complex conjugate and the absolute value of complex number $z$, respectively. Finally, $\det(\cdot)$ is the determinant operator.

This paper is organized as follows. The system model is presented in Section 2. The channel estimation problem and the proposed technique are formulated in Section 3. The EM estimator and the lower bound on the channel estimator error are derived in Section 4. The performance of the proposed technique is assessed in Section 5. Finally, conclusions are presented in Section 6.

## 2 System model

Consider the channel estimation problem for a SISO FD communication system between two nodes $a$ and $b$, as illustrated in Fig. 1. Transceiver nodes $a$ and $b$ are assumed to have passive suppression and analog cancellation stages, and we only consider the digital cancellation to remove the residual SI, i.e., the SI, which is still present after the passive suppression and analog cancellation. We consider the received signal available at the output of the analog-to-digital converter (ADC). The received signal at node $a$ is given by[1]

$$\mathbf{y_a} = h_{aa}\mathbf{x_a} + h_{ba}\mathbf{x_b} + \mathbf{w_a}, \tag{1}$$

where $\mathbf{x_a} \triangleq \left[x_{a_1}, \cdots, x_{a_N}\right]^T$ and $\mathbf{x_b} \triangleq \left[x_{b_1}, \cdots, x_{b_N}\right]^T$ are the $N \times 1$ vectors of transmitted symbols from nodes $a$ and $b$, respectively, $\mathbf{y_a} \triangleq \left[y_{a_1}, \cdots, y_{a_N}\right]^T$ is the $N \times 1$ vector of observations, $\mathbf{w_a}$ is the noise vector, which is modeled by $N$ independent Gaussian random variables, i.e., $f_{\mathbf{W_a}}(\mathbf{w_a}) = \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathrm{I}_N\right)$, and $h_{aa}$ and $h_{ba}$ are the residual SI and communication channel gains, respectively. Furthermore, we model $h_{aa}$ and $h_{ba}$ as independent random variables that are constant over one frame of data and change independently from frame to frame [12].

**Remark 1** *Including all the hardware impairments and unknown parameters in mathematical modeling of parameter estimation problem in FD communication results in a highly non-linear system model, which may not have a tractable solution. The current approach is to separate the estimation of the linear and non-linear parameters* [13, 18]. *In this paper, we focus on the estimation of linear parameters, while the estimation of non-linear parameters can be the topic of future works.*

### 2.1 Modulation assumptions and definitions

In this paper, we assume that the transmitted symbols are all equiprobable and call the set $\mathcal{A} \triangleq \{x_1, x_2, ..., x_M\}$, which contains an alphabet of $M$ constellation points, *a modulation set*. Let $\mathcal{K} \triangleq \{1, \cdots, M\}$ denote set of indices of the constellation points.

We define $E$ as the average symbol energy of a given constellation, i.e.,

$$E \triangleq \mathbb{E}_{X_k}[\,|x_k|^2] = \frac{\sum_{k=1}^{M} |x_k|^2}{M}, \tag{2}$$

where $x_k \in \mathcal{A}$. Note that the average symbol energy can be related to the average bit energy as $E_b \triangleq E/\log_2(M)$.

## 3 Channel estimation for FD systems

In this section, we first formulate the blind channel estimation problem for the FD system considered in Section 2. Based on this formulation, we present a

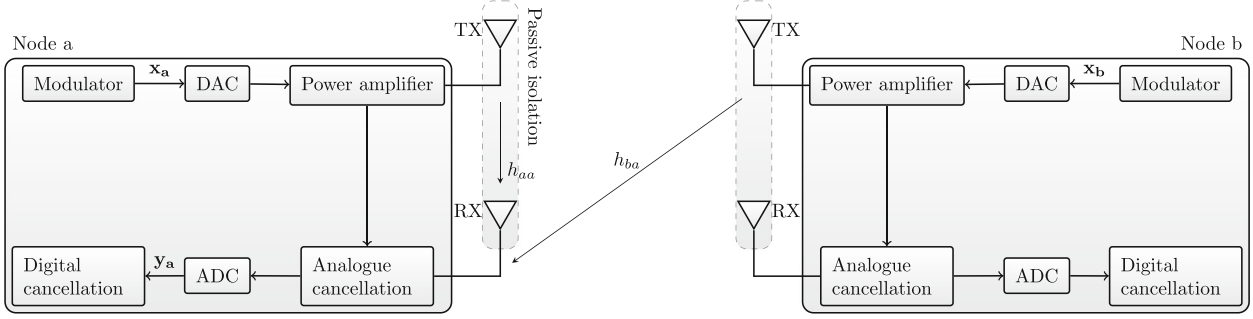


**Fig. 1** Illustration of full-duplex communication between two transceivers, each with a single transmit and a single receive antenna. ADC analog-to-digital converter, DAC digital-to-analog converter, TX transmit, RX receive

theorem which provides the necessary and sufficient condition for ambiguity-free channel estimation. Finally, we discuss the proposed technique to resolve the ambiguity problem.

### 3.1 Problem formulation

Without loss of generality, we consider the problem of baseband channel estimation at node $a$ only (similar results apply at node $b$). In formulating the problem, we make the following assumptions: (i) the transmitter is aware of its own signal, i.e., $\mathbf{x_a}$ is known at node $a$, which is a commonly adopted assumption in the literature [3, 5], (ii) the interference channel $h_{aa}$ and the communication channel $h_{ba}$ are unknown deterministic parameters, (iii) the transmit symbol from node $b$ is modeled using a discrete random distribution, and (iv) we observe $N$ independent received symbols.

The blind channel estimation problem requires the knowledge of the joint probability density function (PDF) of all observations, which is derived from the conditional PDF of a single observation. Given the system model in (1), the conditional PDF of a single observation is given by

$$f_{Y_{a_i}}(y_{a_i}|x_{b_i};h_{aa},h_{ba}) = \frac{1}{\pi\sigma^2}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-h_{ba}x_{b_i}-h_{aa}x_{a_i}|^2\right), \tag{3}$$

where $i \in \mathcal{I} \triangleq \{1,\cdots,N\}$, $y_{a_i}$ is the $i^{th}$ received symbol, and $x_{a_i}$ and $x_{b_i}$ are the $i^{th}$ transmitted symbols from nodes $a$ and $b$, respectively.

The marginal PDF of a single observation is then found by multiplying (3) by the uniform distribution $p_{X_{b_i}}(x_{b_i}) = \frac{1}{M}\mathbb{I}_{\{\mathcal{A}\}}(x_{b_i})$, and summing the results over all the possible values of $x_{b_i}$, where, $\mathbb{I}_{\{\mathcal{A}\}}(x) = 1$ if $x \in \mathcal{A}$ and 0 otherwise. Therefore, we have

$$\begin{aligned} f_{Y_{a_i}}(y_{a_i};h_{aa},h_{ba}) &= \sum_{\forall x_{b_i}} f_{Y_{a_i}}(y_{a_i}|x_{b_i};h_{aa},h_{ba})p_{X_{b_i}}(x_{b_i}) \\ &= \frac{1}{M\pi\sigma^2}\sum_{\forall x_{b_i}}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-h_{ba}x_{b_i}-h_{aa}x_{a_i}|^2\right)\mathbb{I}_{\{\mathcal{A}\}}(x_{b_i}) \\ &= \frac{1}{M\pi\sigma^2}\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-h_{ba}x_k-h_{aa}x_{a_i}|^2\right), \end{aligned} \tag{4}$$

where the last step follows from the fact that $\mathbb{I}_{\{\mathcal{A}\}}(x_{b_i}) = 1$ if and only if $x_{b_i} = x_k$, where $x_k \in \mathcal{A}$. Finally, since the transmitted symbols are assumed independent, and we observe $N$ independent observations, the joint PDF of all the observations is given by

$$\begin{aligned} f_{\mathbf{Y_a}}(\mathbf{y_a};h_{aa},h_{ba}) &= \prod_{i=1}^{N} f_{Y_{a_i}}(y_{a_i};h_{aa},h_{ba}) \\ &= \left(\frac{1}{M\pi\sigma^2}\right)^N \prod_{i=1}^{N}\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-h_{ba}x_k-h_{aa}x_{a_i}|^2\right). \end{aligned} \tag{5}$$

where we substitute the value of $f_{Y_{a_i}}(y_{a_i};h_{aa},h_{ba})$ from (4).

Using (5), we can state the channel estimation problem as shown in the proposition below.

**Proposition 1** *The blind maximum likelihood (ML) channel estimation problem in a SISO FD system is given by*

$$\arg\max_{h_{aa},h_{ba}} f_{\mathbf{Y_a}}(\mathbf{y_a};h_{aa},h_{ba}), \tag{6}$$

*where $f_{\mathbf{Y_a}}(\mathbf{y_a};h_{aa},h_{ba})$ is given by (5).*

In the next subsection, we show that (6) does not have a unique solution if modulation sets which are symmetric around the origin are used.

### 3.2 Identifiability analysis

In this subsection, we present the identifiability analysis for the blind channel estimation problem in (6), which allows us to determine when ambiguity-free channel estimation is possible. For ease of analysis, we first define $\boldsymbol{\theta} \triangleq [h_{aa}, h_{ba}]$ and rewrite (5) as

$$f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}) = \left(\frac{1}{M\pi\sigma^2}\right)^N \prod_{i=1}^{N}\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-\boldsymbol{\theta}(1)x_{a_i}-\boldsymbol{\theta}(2)x_k|^2\right), \tag{7}$$

where $\boldsymbol{\theta}(1)$ and $\boldsymbol{\theta}(2)$ represent the first and second elements of $\boldsymbol{\theta}$.

We start the identifiability analysis by presenting the following definition and remark:

**Definition 1** ([28], Definition 5.2) *If $\mathbf{Y}$ is a random vector distributed according to $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$, then $\boldsymbol{\theta}$ is said to be unidentifiable on the basis of $\mathbf{y}$, if $\forall\mathbf{y}$ there exists $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ for which $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}')$.*

**Remark 2** *Definition 1 states that $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ ($\boldsymbol{\theta} \neq \boldsymbol{\theta}'$) cannot be distinguished from a given set of observations if they both result in the same probability density function for the observations. This implies that if $\boldsymbol{\theta}$ is unidentifiable, then it is impossible for any estimator to uniquely determine the value of $\boldsymbol{\theta}$.*

In order to present the main result in this subsection, we first give the definitions of a symmetric modulation constellation [31] and a bijective function [32].

**Definition 2** *We mathematically define modulation constellation as the graph of the function* $\mathrm{f}(x_k) = x_k$*, where* $x_k \in \mathcal{A}\ \forall k \in \mathcal{K}$*. Then a modulation constellation is symmetric with respect to the origin if and only if* $\mathrm{f}(-x_k) = -\mathrm{f}(x_k)\ \forall x_k \in \mathcal{A}$ [31].

**Definition 3** *Let* $\mathcal{C}$ *and* $\mathcal{D}$ *be two sets. A function from* $\mathcal{C}$ *to* $\mathcal{D}$ *denoted* $\mathrm{t} : \mathcal{C} \rightarrow \mathcal{D}$ *is a bijective function if and only if it is both one-to-one and onto.*

The above definition states that a bijective function is a function between the elements of two sets, where each element of one set is paired with exactly one element of the other set and there are no unpaired elements. Note that a bijective function from a set to itself is also called a permutation [32].

In this work, we define and use the bijective function g: $\mathcal{K} \rightarrow \mathcal{K}$, i.e., g is a one-to-one and onto function on $\mathcal{K} \rightarrow \mathcal{K}$. Using this bijective function, we present the main result as below.

**Theorem 1** *There exists a* $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ *for which the joint probability density* $f_{\mathbf{Y_a}}(\mathbf{y_a}; \boldsymbol{\theta})$ *given by* (7) *is equal to* $f_{\mathbf{Y_a}}(\mathbf{y_a}; \boldsymbol{\theta}')\ \forall \mathbf{y}_a$*, if and only if there exists a bijective function* g: $\mathcal{K} \rightarrow \mathcal{K}$*, such that* $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$*, where* $c \neq 1$ *is a constant and* $|c| = 1$*, i.e., the modulation constellation is symmetric about the origin.*

*Proof* We prove the result in Theorem 1 in three steps. First, we assume $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ for which $f_{\mathbf{Y_a}}(\mathbf{y_a}; \boldsymbol{\theta}) = f_{\mathbf{Y_a}}(\mathbf{y_a}; \boldsymbol{\theta}')\ \forall \mathbf{y_a}$ exists and show that it leads to a bijective function g satisfying $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$. Then, we assume that a bijective function g satisfying $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$ exists and show that there exists a $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ for which $f_{\mathbf{Y_a}}(\mathbf{y_a}; \boldsymbol{\theta}) = f_{\mathbf{Y_a}}(\mathbf{y_a}; \boldsymbol{\theta}')\ \forall \mathbf{y_a}$. Finally, using Definition 2, we show that the condition $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$ is equivalent to the modulation constellation being symmetric with respect to the origin. The details are in Appendix 1. □

**Remark 3** *From Theorem 1, we can see that since the modulation constellations, such as M-ary quadrature amplitude modulation (M-QAM), satisfy the definition of symmetric modulation constellations in Definition 2, the blind channel estimation problem in* (6) *does not have a unique solution and suffers from an ambiguity problem.*

### 3.3 Proposed technique

In this subsection, we present our proposed technique to resolve the ambiguity problem in (6).

The rationale behind the proposed technique comes from the fact that Theorem 1 shows that symmetry of the modulation constellation with respect to the origin is the cause of the ambiguity. A simple way to achieve constellation asymmetry[2] is to add a constant $s$ to each element of $\mathcal{A}$. The resultant asymmetric shifted modulation constellation is formally defined as follows:

**Definition 4** *The asymmetric shifted modulation constellation,* $\bar{\mathcal{A}}$*, is defined as*

$$\bar{\mathcal{A}} \triangleq \{x_k + s, \quad \forall\, x_k \in \mathcal{A}, \quad s \in \mathbb{R}^+\}, \tag{8}$$

*where* $\mathbb{R}^+$ *is the set of positive real numbers.*

In the rest of the paper, we also use $\bar{x}_k = x_k + s$ to denote the $k$th element of $\bar{\mathcal{A}}$.

For illustration, Fig. 2 shows the effect of the proposed technique on the 16-QAM constellation. We can see that the resulting modulation constellation is shifted along the horizontal axis, which increases the average energy per symbol of the modulation constellation. This increase in the average energy per symbol can be justified as follows: in reality, it is inevitable to use some extra energy to estimate the unknown channels, whether it is done by pilots or by the proposed technique. In this regard, it is important to note that the smaller the energy used for shifting the modulation constellation, the closer the average energy of the proposed technique is to the ideal scenario where the channels are perfectly known at the receiver and no extra energy is needed for channel estimation.

**Remark 4** *The addition of the DC component lowers power efficiency similar to the use of superimposed training* [22][3]*. However, the proposed shifted modulation technique has the offsetting advantages that (i) bandwidth efficiency is not reduced and (ii) the DC offset can be used to reduce the peak-to-average power of the signal envelope during transmissions resulting in lowered cost/complexity power amplifiers. Moreover, the proposed scheme is well-suited to MQAM as investigated here. Since, this is a spectrally efficient modulation scheme used where power efficiency is not critical*

For the sake of numerically investigating the problem of smallest possible shift energy, we define $\beta$ as the portion of the average energy per symbol that is allocated to the shift and use the real constant $s \triangleq \sqrt{\beta E}$, where $0 < \beta < 1$ to shift the symmetric modulation constellation. In this case, the problem of smallest shift energy corresponds to the problem of finding the minimum value of $\beta$. The minimum value of $\beta$ is an indication of how much extra energy is needed compared to the perfect channel knowledge scenario.

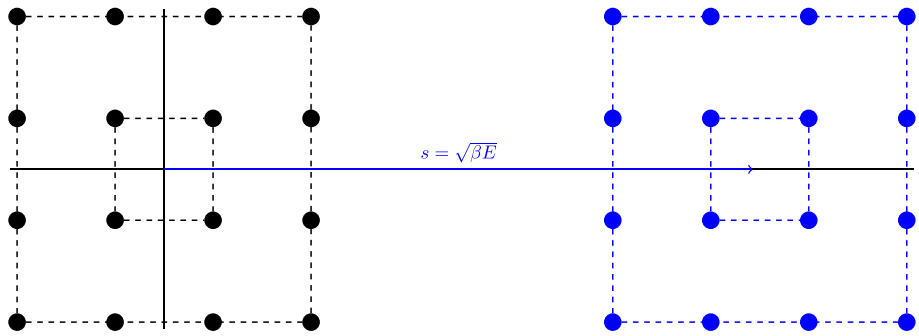


**Fig. 2** Effect of the proposed technique on the constellation of 16-QAM. The resulting constellation is shifted along the horizontal axis, i.e., it is asymmetric around the origin

In Section 4.1, we derive a lower bound on the estimation error, which allows us to numerically find the minimum value of $\beta$.

## 4 EM-based estimator

In this section, we derive an EM estimator to obtain channel estimates in a FD system with asymmetric shifted modulation constellation defined in Definition 4. We derive a lower bound on the estimation error of the estimator. Finally, we investigate the complexity of the proposed estimator.

For the sake of notational brevity, we first define

$$\boldsymbol{\phi} \triangleq [\Re(h_{aa}), \Im(h_{aa}), \Re(h_{ba}), \Im(h_{ba})]\,. \tag{9}$$

We can then reformulate the ML problem in (6) as follows

$$[\Re(\hat{h}_{aa}), \Im(\hat{h}_{aa}), \Re(\hat{h}_{ba}), \Im(\hat{h}_{ba})] \triangleq \arg\max_{\boldsymbol{\phi}}(\ln f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\phi})), \tag{10}$$

where $f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\phi})$ is given by (7) and $\ln f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\phi})$ is known as the log-likelihood function.

In formulating the channel estimation problem in (6) (and hence in (10)), we assumed unknown transmitted symbols. These unknown transmitted symbols can be treated as hidden data. A common approach to solving the maximization problem in (10) in the presence of hidden data is the EM algorithm [33], which is adopted in this work. The main steps of EM algorithm are

1. *Expectation step:* In the $E$-step, the expectation of the log-likelihood is taken over all the values of the hidden variable, conditioned on the vector of observations, and the $n$th estimate of $\boldsymbol{\phi}$ ($\boldsymbol{\phi}^{(n)}$). In (1), the hidden variable is $\bar{\mathbf{x}}_\mathbf{b}$ and consequently, we need to evaluate $Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)}) \triangleq \mathbb{E}_{\bar{\mathbf{X}}_\mathbf{b}|\mathbf{Y_a},\boldsymbol{\phi}^{(n)}}[\ln f_{\mathbf{Y_a}}(\mathbf{y_a}, \bar{\mathbf{x}}_\mathbf{b}|\boldsymbol{\phi})]$.
2. *Maximization step:* In the $M$-step, the function $Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)})$ obtained from the $E$-step is maximized with respect to $\boldsymbol{\phi}$.
3. *Iterations:* We iterate between the $E$- and $M$-steps until convergence is achieved.

The equations needed for the $E$- and $M$-steps are summarized in the propositions below.

**Proposition 2** *The E-step during nth iteration of the algorithm is given by*

$$Q\left(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)}\right) = -N\ln\left(M\pi\sigma^2\right) - \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{k=1}^{M} T_{k,i}^{(n)}|y_{a_i} - h_{ba}\bar{x}_k - h_{aa}\bar{x}_{a_i}|^2, \tag{11}$$

*where* $\boldsymbol{\phi}^{(n)} \triangleq \left[\hat{h}_{aa}^{(n)}, \hat{h}_{ba}^{(n)}\right]$ *are the estimates of the channels obtained from* $\boldsymbol{\phi}^{(n)}$ *during the nth iteration of the algorithm, and* $T_{k,i}^{(n)}$ *is defined as*

$$T_{k,i}^{(n)} \triangleq \frac{\exp\left(\frac{-1}{\sigma^2}|y_{a_i} - \hat{h}_{ba}^{(n)}\bar{x}_k - \hat{h}_{aa}^{(n)}\bar{x}_{a_i}|^2\right)}{\sum_{\bar{k}=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i} - \hat{h}_{ba}^{(n)}\bar{x}_{\bar{k}} - \hat{h}_{aa}^{(n)}\bar{x}_{a_i}|^2\right)}, \tag{12}$$

*where* $k \in \mathcal{K}$, $i \in \mathcal{I} \triangleq [1, 2, \cdots, N]$, $\bar{x}_k \in \bar{\mathcal{A}}$, $\bar{x}_{a_i} \in \bar{\mathcal{A}}$ *and* $\bar{x}_{\bar{k}} \in \bar{\mathcal{A}}$.

*Proof* See Appendix 2. □

**Proposition 3** *The M-step during the nth iteration of the algorithm is given by*

$$\boldsymbol{\phi}^{(n+1)} = \frac{1}{s_1 s_4 - s_2^2 - s_3^2}\begin{bmatrix} -s_2\nu_3 - s_3\nu_4 + s_4\nu_1 \\ -s_2\nu_4 + s_3\nu_3 + s_4\nu_2 \\ s_1\nu_3 - s_2\nu_1 + s_3\nu_2 \\ s_1\nu_4 - s_2\nu_2 - s_3\nu_1 \end{bmatrix}, \tag{13}$$

*where*

$$s_1 \triangleq \sum_{i=1}^{N} |\bar{x}_{a_i}|^2,\ s_2 \triangleq \sum_{i=1}^{N}\sum_{k=1}^{M} T_{k,i}^{(n)} \Re(\bar{x}_{a_i}\bar{x}_k^*), \tag{14a}$$

$$s_3 \triangleq \sum_{i=1}^{N}\sum_{k=1}^{M} T_{k,i}^{(n)} \Im(\bar{x}_{a_i}\bar{x}_k^*),\ s_4 \triangleq \sum_{i=1}^{N}\sum_{k=1}^{M} T_{k,i}^{(n)} |\bar{x}_k|^2, \tag{14b}$$

$$v_1 \triangleq \sum_{i=1}^{N} \Re(\bar{x}_{a_i}^* y_{a_i}),\quad v_2 \triangleq \sum_{i=1}^{N} \Im(\bar{x}_{a_i}^* y_{a_i}), \tag{14c}$$

$$v_3 \triangleq \sum_{i=1}^{N}\sum_{k=1}^{M} T_{k,i}^{(n)} \Re(y_{a_i}\bar{x}_k^*),\quad v_4 \triangleq \sum_{i=1}^{N}\sum_{k=1}^{M} T_{k,i}^{(n)} \Im(y_{a_i}\bar{x}_k^*). \tag{14d}$$

*Proof* See Appendix 2. □

**Remark 5** *It is well-known that the EM algorithm may be very sensitive to initialization* [34]. *Although different methods exist for EM initialization, generally they are not computationally efficient* [34, 35]. *For the given channel assumptions in Section 5, our empirical results showed that initializing the EM algorithm by* $\boldsymbol{\phi}^{(0)} \triangleq [0,0,0,0]$ *resulted in the lowest estimation error. Hence, this initialization is used in this work.*

### 4.1 Lower bound on the estimation error

In this section, we derive a closed-form lower bound on the estimation error of the proposed estimator. The derived lower bound directly links the channel estimation error to the parameter $\beta$, defined in Section 3.3.

The EM algorithm, defined in Propositions 2 and 3, is a ML estimator for the parameter vector $\boldsymbol{\phi}$ in (9). Hence, we aim to derive the lower bound for the variance of the proposed ML estimator. The ML estimator is asymptotically efficient [36] and its MSE is lower bounded by the inverse of the Fisher information matrix (FIM) [36]. This result is known as Cramer Rao lower bound (CRLB) and is given by

$$\mathbb{E}_{\hat{\Phi}_l}[\,|\hat{\phi}_l - \phi_l|^2] \geq \left[\mathrm{I}^{-1}\left[f_{\mathbf{Y_a}}(\mathbf{y}_a;\boldsymbol{\phi})\right]\right]_{l,l}, \tag{15}$$

where $\phi_l$ is the $l$th element of the parameter vector $\boldsymbol{\phi}$, $\hat{\phi}_l$ is an estimate of $\phi_l$, for $l \in \{1,2,3,4\}$, $[\cdot]_{l,l}$ is the $l$th diagonal element of a square matrix, and $\mathrm{I}^{-1}\left[f_{\mathbf{Y_a}}(\mathbf{y}_a;\boldsymbol{\phi})\right]$ is the inverse of FIM. Since the inverse of FIM in (15) cannot be found in closed-form [24, 37], we derive a lower bound on the MSE of the proposed estimator, which is in closed-from. The result is presented in the proposition below.

**Proposition 4** *The variance of the proposed estimator is lower bounded by*

$$\mathbb{E}_{\hat{\Phi}_l}[\,|\hat{\phi}_l - \phi_l|^2] \geq \left(\frac{\sigma^2}{2NE}\right)\frac{1+\beta}{(1+2\beta)}, \tag{16}$$

*where* $l \in \{1,\cdots,4\}$, $N$ *is the number of observations,* $E$ *is the average symbol energy of the modulation constellation before the shift, and* $\beta$ *is the portion of* $E$ *that is allocated to the shift.*

*Proof* See Appendix 3. □

**Remark 6** *The result in* (16) *links the closed-form lower bound of the estimation error to the average energy of the modulation constellation before the shift and the portion of this average energy allocated to the shift. This is important because in Section 5.1, we will use* (16) *to find the minimum shift energy needed for the proposed technique.*

### 4.2 Complexity analysis

To evaluate the feasibility in implementing the proposed estimator, we investigate the computational complexity of the estimator in terms of required floating point multiplications and additions (flops) [38].

Table 1 shows the number of multiplications and additions needed for the EM estimator for $h_{ba}$. Although we only present the complexity analysis of $h_{ba}$, similar complexity is also observed for estimating $h_{aa}$. In each row of the table, the number of required additions and multiplications to implement a given equation is presented and are then summed to obtain overall complexity.

It is clear from Table 1 that the complexity of EM estimator per iteration is proportional to $NM^2$. This analysis shows that the EM algorithm is computationally very efficient since, for a given modulation constellation with size $M$, the computational complexity of the EM estimator only grows linearly with the number of observations, $N$.

**Remark 7** *In data-aided approaches to channel estimation, both* $\mathbf{x}_a$ *and* $\mathbf{x}_b$ *in (1) are assumed to be known. Consequently, linear channel estimation can be performed*

**Table 1** Complexity analysis of the EM estimator

| EM—complexity per iteration | | |
|---|---|---|
| (Eq. no.) | Additions | Multiplications |
| (12) | $3M+2$ | $6M+6$ |
| (11) | $NM(3M+4)$ | $NM(6M+9)+5$ |
| (13) | $4NM(3M+2)+3N+14$ | $4NM(6M+8)+3N+33$ |
| $\hat{h}_{ba}$ | $15NM^2+12NM+3N+3M+16$ | $30NM^2+41NM+3N+6M+44$ |

*for linear Gaussian models (LGMs) as explained in* [36]. *It is a well-known fact that linear estimator complexity estimators for LGM is independent of the modulation size M, and only grows linearly with the number of observations* [36]. *The extra complexity of the proposed algorithm compared to linear estimators is expected. This is because as opposed to linear estimators, the proposed estimator requires no data-aided piloting, and hence, it allows for efficient use of the bandwidth for channel estimation.*

## 5 Simulation results

In this section, we present numerical and simulation results to investigate the performance of the proposed estimator with asymmetric shifted modulation constellation. We consider a FD communication system as illustrated in Fig. 1. The analysis in Section 4.1 shows an identical lower bound for the estimation error of both $h_{aa}$ and $h_{ba}$. Hence, in this section, we only present the results for the communication channel $h_{ba}$ since identical results are obtained for the SI channel $h_{aa}$.

For each simulation run, $N$ data and interfering symbols are randomly generated assuming shifted 16-QAM modulation constellation is used ($M = 16$). The channels are constant for the transmission of $N$ symbols, i.e., the quasi static assumption. We assume that there is no line-of-sight (LOS) communication link between the transmitter of node $b$ and the receiver of node $a$. Hence, the communication channel $h_{ba}$ can be modeled as a Rayleigh fading channel, i.e., $h_{ba} \sim \mathcal{CN}\left(0, \sigma^2_{h_{ba}}\right)$. For the SI channel, experimental results have shown that before passive and active cancellation, the SI channel has a strong LOS component and can be modeled as a Rician distribution with a large $K$ factor (approximately 20–25 dB). After passive suppression and analog cancellation, the strong LOS component is significantly reduced but still present and can be modeled as a Ricean distribution with $K = 0$ dB [5]. Hence, we generate the SI channel as

$$h_{aa} = \sqrt{\frac{K}{K+1}}\sigma_{h_{aa}} e^{\mathrm{j}\zeta} + \sqrt{\frac{1}{K+1}}\mathcal{CN}\left(0, \sigma^2_{h_{aa}}\right), \quad (17)$$

where $\zeta$ is uniformly distributed angle of arrival of the LOS component of the SI channel [39].

For the simulations, the signal-to-interference-noise ratio (SINR) is given by [5]

$$\mathrm{SINR} = \frac{1}{\frac{1}{\mathrm{SIR}} + \frac{1}{\mathrm{SNR}}}, \quad (18)$$

where the signal-to-interference ratio $\mathrm{SIR} = \frac{\sigma^2_{h_{ba}}}{\sigma^2_{h_{aa}}}$ assuming both nodes use constellations with the same average energy, the desired signal-to-noise ratio $\mathrm{SNR} = \frac{\sigma^2_{h_{ba}} \log_2(M) E_b}{N_0}$, $E_b$ is the average bit energy which is defined below (2) and $N_0$ is the noise power spectral density.

As discussed in Section 1, even with state-of-the-art passive suppression and analog cancellation, the SIR can still be around $-5$ dB [5, 40]. Hence, we adopt this value of the SIR in the simulations while assuming that the communication channel has average energy of unity, i.e., $\mathbb{E}\left[|h_{ba}|^2\right] = \sigma^2_{h_{ba}} = 1$. Furthermore, in order to investigate the performance of the proposed estimator over a range of SINR, we fix $N_0 = 1$ and run the simulations for different values of $E_b/N_0$ (in dB). The figures of merit used are the average mean square error (MSE) and the BER, which are obtained by averaging over 5000 Monte Carlo simulation runs.

### 5.1 Minimum energy needed for channel estimation

In this subsection, we are interested in finding the minimum value of $\beta$, for a given $E_b/N_0$ and $N$. As discussed in Section 3.3, we use $s \triangleq \sqrt{\beta E}$, where $0 < \beta < 1$, to shift the symmetric modulation constellation. Hence, a lower value of $\beta$ is desirable since it means less energy is used to shift the modulation constellation.

In order to find a minimum value of $\beta$ suitable for a practical range of $E_b/N_0$ and $N$, we use the average MSE lower bound in (16) to observe the behavior of the proposed estimator as a function of $\beta$ at low $N$ and low $E_b/N_0$. This is motivated by the fact that the minimum value of $\beta$ found for low $N$ and low $E_b/N_0$ will ensure that the desired estimation error will also be achieved for high $E_b/N_0$ and/or when the number of observations $N$ is large. This intuition is confirmed from (16), which indicates that higher values of $\beta$ are needed at low $E_b/N_0$ to reach a given estimation error. Furthermore, since the lower bound on the estimation error also decreases with $N$, the minimum value of $\beta$ found for smaller $N$ can also serve for larger $N$. Since the experimental results of [5, 41] show that the FD communication channel is normally constant for more than $N > 128$ symbols, we propose to find the minimum $\beta$ at $N = 128$ and $E_b/N_0 = 0$ dB.

Figure 3 shows the MSE performance of the proposed technique versus $\beta$ for $E_b/N_0 = 0$ dB, $N = 128$, and SIR $= -50$ dB. If the desired estimation error is taken to be within 10% of the lower bound error, then we can see from the figure that for $\beta < 0.2$, the MSE of the proposed estimator is within 10% of the lower bound. Consequently, the minimum value of $\beta$ is 0.2.

Figure 4 shows the MSE performance of the proposed estimator with $\beta = 0.2$ (the selected minimum value of $\beta$) vs. $E_b/N_0$ (dB) for $N = 128$ and SIR $= -50$ dB. The lower bound in (16) is plotted as a reference. The figure shows that as $E_b/N_0$ increases, the gap between the performance of the proposed estimator and the lower bound decreases. The gap is less than 2 dB after $E_b/N_0 = 20$ dB.

In the following sections, we set $\beta = 0.2$ and $N = 128$ to study the performance of the FD communication system.

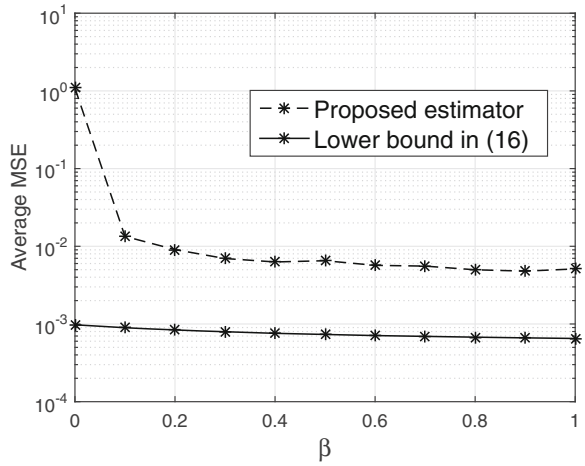


**Fig. 3** MSE performance of the proposed channel estimator for different values of $\beta$ for $E_b/N_0 = 0$ dB, $N = 128$, and SIR $= -50$ dB

### 5.2 Comparison with data-aided channel estimation

In this section, we compare the MSE and BER performance of the proposed estimator against a data-aided channel estimator for the case that the average energy per transmitted frame is the same for both methods[4]. For the proposed technique, we assume that (i) all the transmitted symbols are data symbols and (ii) shifting the modulation constellation increases the average energy by 20% compared to the ideal scenario when no channel estimation is needed (corresponds to $\beta = 0.2$). For the data-aided channel estimation, we assume that (i) 64 pilot symbols are used in a frame of 128 symbols and (ii) these pilots also require an extra 20% energy.

#### 5.2.1 MSE performance

The average MSE reveals the accuracy of the channel estimation. Figure 5 plots the average MSE vs. $E_b/N_0$ with $\beta = 0.2$, $N = 128$, and SIR $= -50$ dB. The lower bound from (16) is plotted as a reference. We also plot the MSE for data-aided channel estimation with (i) 64 pilot symbols in a frame of 128 symbols and (ii) 128 pilot symbols in a frame of 128 symbols. Figure 5 shows that the proposed technique outperforms data-aided channel estimation when both methods use the same extra amount of energy for channel estimation. At high $E_b/N_0$, the MSE performance of the proposed technique is within $3-4$ dB of the lower bound.

It has been shown in [9] that the effect of quantization error in FD communication system can be modeled as an additive Gaussian noise. This means the system model given by (1) implicitly includes the effect of quantization noise as well as the effect of thermal noise in the Gaussian noise term $\mathbf{w}_a$. Consequently, the effect of quantization noise on the performance of the proposed estimator can be studied by observing the MSE results of the proposed estimator in the low SNR region as shown by Fig. 5. The results of Fig. 5 show that at low SNR region, a noticeable MSE gain is not obtained by using the proposed estimator instead of the data-aided technique that uses 128 pilots. However, the proposed technique is still more attractive compared to the data-aided technique because of the bandwidth efficiency.

#### 5.2.2 BER performance

Figure 6 shows the average BER vs. $E_b/N_0$ (dB) with $\beta = 0.2$, $N = 128$, and SIR $= -50$ dB. The BER performance with perfect channel knowledge is plotted as a reference. We also plot the BER for data-aided channel estimation with 64 pilot symbols in a frame of 128 symbols. Figure 6

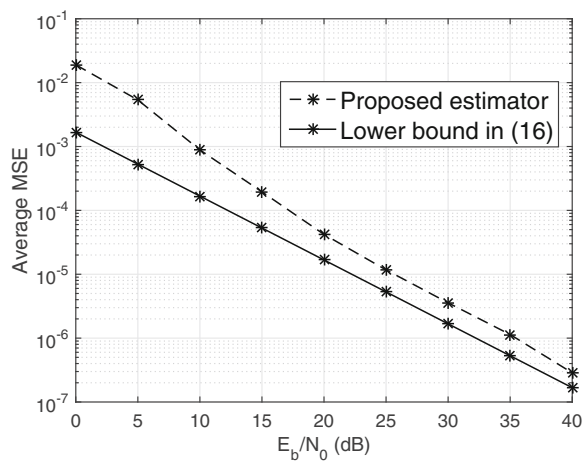


**Fig. 4** MSE performance of the proposed channel estimator vs. $E_b/N_0$ for $\beta = 0.2$, $N = 128$, and SIR $= -50$ dB

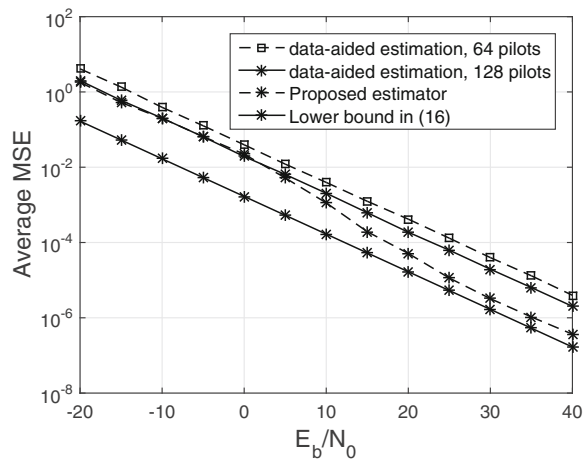


**Fig. 5** MSE performance of the proposed technique

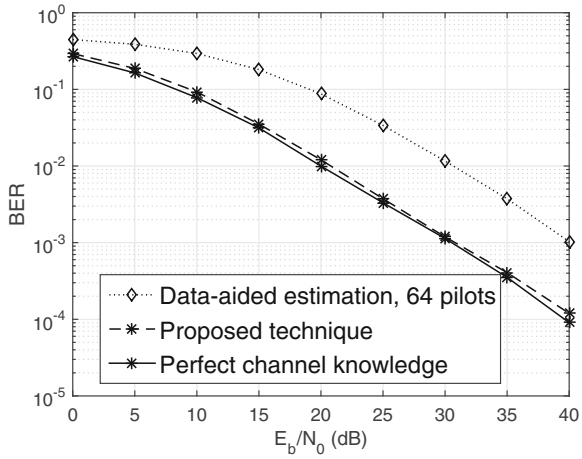


**Fig. 6** BER performance of the proposed technique

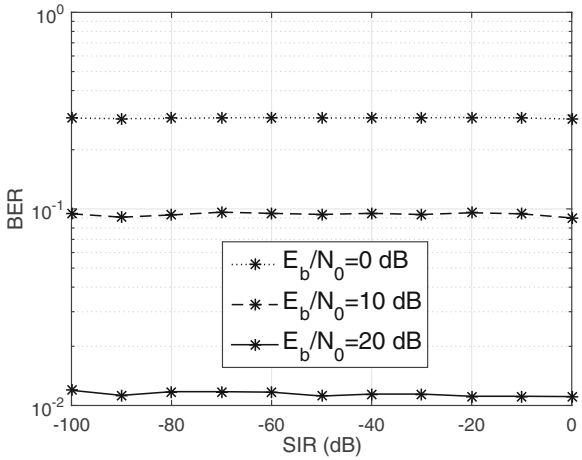


**Fig. 7** Effect of SI power level on the BER performance of the proposed technique

shows that the proposed technique outperforms the data-aided channel estimation in terms of the BER. This is to be expected since, as shown in Fig. 5, for the same extra amount of energy for channel estimation the proposed technique has much lower MSE compared to data-aided channel estimation. In addition, at high $E_b/N_0$, the BER performance of the proposed technique is within 1 dB of the ideal performance obtained with perfect channel knowledge.

As shown in this sub-section, in comparison to the data-aided algorithms, the proposed algorithm is more bandwidth efficient. Also compared to the existing blind algorithms, which suffer from phase ambiguity [23], the proposed algorithm can estimate the channel with no phase ambiguity. These advantages have been obtained by an increase in complexity as explained in Remark 7. However, the superior performance of the proposed algorithm as shown in Figs. 5 and 6 produces attractive tradeoffs compared to the existing data-aided and blind algorithms.

### 5.3 Effect of power of SI signal

In the results so far, we have set the SIR to $-50$ dB. In this section, we assess the impact of the SI power level on the performance of the proposed technique.

Figure 7 plots the BER versus the SIR (dB) for $E_b/N_0 = 0, 10, 20$ dB, with $\beta = 0.2$ and $N = 128$. We can see that as the SI power increases, the BER performance of the proposed technique remains nearly constant. This is because in FD communication the self-interference signal is completely known to the receiver [3]. Consequently, for relatively small channel estimation error of the proposed estimator, the SI can be cancelled regardless of its power. Figure 7 illustrates the robustness of the proposed technique, i.e., even with weak passive suppression and analog cancellation requiring digital SI cancellation to handle a very large SIR (e.g., $-100$ dB), the BER is not significantly altered.

## 6 Conclusions

In this paper, we have proposed a new technique to estimate the SI and communication channels in a FD communication systems for residual SI cancellation. In the proposed technique, we add a real constant number to each constellation point of a modulation constellation to yield asymmetric shifted modulation constellations with respect to the origin. Using identifiability analysis, we show mathematically that such a modulation constellation can be used for ambiguity-free channel estimation in FD communication systems. We proposed a computationally efficient EM-based estimator to estimate the SI and communication channels simultaneously using the proposed technique. We also derived a lower bound for the estimation error of the proposed estimator. The results showed that the proposed technique is robust to the level of SI power.

## Endnotes

[1] Note that the system model for FD communication in (1) is applicable to per subcarrier communication in orthogonal frequency division multiplexing (OFDM) FD communication [3, 5, 41].

[2] Note that it may be possible to achieve constellation asymmetry through other means, such as design of irregular modulation constellations. The optimum design of such modulation constellations is outside the scope of this work.

[3] Note that the existing hardware implementations for superimposed training [42] can also be used here for shifting the modulation constellation.

[4] Note that the simulation results have been obtained by normalizing the shifted modulation so the extra power needed to shift the modulation does not push the power amplifiers into saturation and hence, power amplifiers do not experience any non-linearities.

## Appendix 1

### Proof of Theorem 1

The proof consists of three main steps.

Step 1: We show that if $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ exists such that $f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}') = f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta})\ \forall \mathbf{y}_a$, then a bijective function g : $\mathcal{K} \rightarrow \mathcal{K}$ exists, such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$, where $c \neq 1$ is a constant and $|c| = 1$. This is done as follows.

If the two joint probability densities $f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}')$ and $f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta})$ are equal $\forall \mathbf{y}_a$, then it easily follows that the marginal densities $f_{Y_{a_i}}(y_{a_i};\boldsymbol{\theta}')$ and $f_{Y_{a_i}}(y_{a_i};\boldsymbol{\theta})$ are also equal $\forall y_{a_i}$. From (4), $f_{Y_{a_i}}(y_{a_i};\boldsymbol{\theta}')$ and $f_{Y_{a_i}}(y_{a_i};\boldsymbol{\theta})$ are given by

$$f_{Y_{a_i}}(y_{a_i};\boldsymbol{\theta}) = \left(\frac{1}{M\pi\sigma^2}\right)\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i} - \boldsymbol{\theta}(1)x_{a_i} - \boldsymbol{\theta}(2)x_k|^2\right), \tag{19}$$

$$f_{Y_{a_i}}(y_{a_i};\boldsymbol{\theta}') = \left(\frac{1}{M\pi\sigma^2}\right)\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i} - \boldsymbol{\theta}'(1)x_{a_i} - \boldsymbol{\theta}'(2)x_k|^2\right). \tag{20}$$

If $f_{Y_{a_i}}(y_{a_i};\boldsymbol{\theta}')$ and $f_{Y_{a_i}}(y_{a_i};\boldsymbol{\theta})$ are equal $\forall y_{a_i}$, they should also be equal for $y_{a_i} = \boldsymbol{\theta}(1)x_{a_i} + \boldsymbol{\theta}(2)x_1$. In this case, we have

$$\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|\boldsymbol{\theta}(2)(x_1 - x_k)|^2\right) = \sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}\left|(\boldsymbol{\theta}(1) - \boldsymbol{\theta}'(1))x_{a_i} + \boldsymbol{\theta}(2)x_1 - \boldsymbol{\theta}'(2)x_k\right|^2\right). \tag{21}$$

The left hand side (LHS) of (21) is independent of $i$, while the right hand side (RHS) of (21) depends on $i$ through $x_{a_i}$. Consequently, for (21) to hold for $\forall y_{a_i}$, the coefficient of $x_{a_i}$ should be zero, i.e., $\boldsymbol{\theta}'(1) = \boldsymbol{\theta}(1)$. Knowing that $\theta'(1) = \theta(1)$ and equating (19) and (20), we have

$$\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i} - \boldsymbol{\theta}(1)x_{a_i} - \boldsymbol{\theta}(2)x_k|^2\right) = \sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i} - \boldsymbol{\theta}(1)x_{a_i} - \boldsymbol{\theta}'(2)x_k|^2\right). \tag{22}$$

By taking the first and second order derivatives of both sides of (22) with respect to $y_{a_i}$, it can be shown that $\forall k \in \mathcal{K}$, the points $y_{a_i} = \boldsymbol{\theta}(1)x_{a_i} + \boldsymbol{\theta}(2)x_k$ and $y_{a_i} = \boldsymbol{\theta}(1)x_{a_i} + \boldsymbol{\theta}'(2)x_k$ maximize the summations of the $M$ exponential functions on the LHS and RHS of (22), respectively. Consequently, since (22) holds $\forall y_{a_i}$, the points that maximize the summation of $M$ exponential on the LHS of (22) are the same as the points that maximize the summation of $M$ exponentials on the RHS of (22). Hence, for a bijective function g : $\mathcal{K} \rightarrow \mathcal{K}$

$$\boldsymbol{\theta}(1)x_{a_i} + \boldsymbol{\theta}(2)x_k = \boldsymbol{\theta}(1)x_{a_i} + \boldsymbol{\theta}'(2)x_{\mathrm{g}(k)} \tag{23}$$

It can easily be verified that if (23) holds $\forall y_{a_i}$, then,

$$\boldsymbol{\theta}(2)x_k = \boldsymbol{\theta}'(2)x_{\mathrm{g}(k)}, \tag{24}$$

or

$$\frac{\boldsymbol{\theta}'(2)}{\boldsymbol{\theta}(2)} = \frac{x_k}{x_{\mathrm{g}(k)}}. \tag{25}$$

The LHS of (25) does not depend on $k$; consequently, the RHS of (25) should also be independent of $k$ and should be a constant. Hence, for bijective function g, $\frac{x_k}{x_{\mathrm{g}(k)}} = c$, where $c \neq 1$ is a constant. We note that if $c = 1$ then $\theta(2) = \theta'(2)$ and hence $\boldsymbol{\theta} = \boldsymbol{\theta}'$, which violates the assumption that $f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}') = f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta})$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$.

Let us now define permutation $\Pi$ on the ordered set $\mathcal{A} = \{x_1, \cdots, x_M\}$ as

$$\Pi \triangleq \begin{pmatrix} x_1 & x_2 & \cdots & x_M \\ \Pi(x_1) = x_{\mathrm{g}(1)} & \Pi(x_2) = x_{\mathrm{g}(2)} & \cdots & \Pi(x_M) = x_{\mathrm{g(M)}} \end{pmatrix}. \tag{26}$$

The sequence $(x_k, \Pi(x_k), \Pi(\Pi(x_k)), \cdots, x_k)$ forms an orbit of the permutation $\Pi$ [32]. If $x_k = cx_{\mathrm{g(k)}}\ \forall k \in \mathcal{K}$, then from the definition of the orbit it is clear that $x_k = c^m x_k\ \forall k \in \mathcal{K}$, where $m$ is the length of the orbit sequence and $c \neq 1$ is a constant. Since, $x_k = c^m x_k\ \forall k \in \mathcal{K}$, we can conclude that $c^m = 1$ and $|c| = 1$.

Step 2: We show that if the bijective function g: $\mathcal{K} \rightarrow \mathcal{K}$ exists, such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$, then there exists a $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ for which $f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}') = f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta})\ \forall \mathbf{y}_a$. This is done as follows.

Firstly, in Section 3.2 we showed that the joint PDF of all the observations is given by

$$f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}) = \left(\frac{1}{M\pi\sigma^2}\right)^N \prod_{i=1}^{N}\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-\boldsymbol{\theta}(1)x_{a_i}-\boldsymbol{\theta}(2)x_k|^2\right), \tag{27}$$

We can see that for a fixed $i$ in (27), $M$ different exponential functions corresponding to different values of $x_k \in \mathcal{A}$ are summed together. Since, $\mathrm{g} : \mathcal{K} \rightarrow \mathcal{K}$ is a bijective function and consequently replacing $x_k$ by $x_{\mathrm{g}(x)}$ only affects the order of the exponential functions, we can rewrite (27) as

$$f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}) = \left(\frac{1}{M\pi\sigma^2}\right)^N \prod_{i=1}^{N}\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-\boldsymbol{\theta}(1)x_{a_i}-\boldsymbol{\theta}(2)x_{\mathrm{g}(k)}|^2\right), \tag{28}$$

Secondly, for $\boldsymbol{\theta}' \triangleq [\boldsymbol{\theta}(1), \frac{\boldsymbol{\theta}(2)}{c}], f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}')$ is given by

$$f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}') = \left(\frac{1}{M\pi\sigma^2}\right)^N \prod_{i=1}^{N}\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-\boldsymbol{\theta}(1)x_{a_i}-\frac{\boldsymbol{\theta}(2)}{c}x_k|^2\right), \tag{29}$$

and since $\frac{x_k}{c} = x_{\mathrm{g}(k)}\ \forall k \in \mathcal{K}$, (29) can be written as

$$f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}') = \left(\frac{1}{M\pi\sigma^2}\right)^N \prod_{i=1}^{N}\sum_{k=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-\boldsymbol{\theta}(1)x_{a_i}-\boldsymbol{\theta}(2)x_{\mathrm{g}(k)}|^2\right), \tag{30}$$

Comparing (30) with (28) reveals that for $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$, $f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}') = f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta})\ \forall \mathbf{y}_a$. Consequently, if bijective function g: $\mathcal{K} \rightarrow \mathcal{K}$ exits, such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$, then there exists a $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ for which $f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta}') = f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\theta})\ \forall \mathbf{y}_a$.

Step 3: Thirdly, we show that the condition $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$ is equivalent to the modulation constellation being symmetric around the origin. To prove this equivalency, we need to consider the following two sub-cases:

(i) Firstly, we need to show that if a bijective function $\mathrm{g} : \mathcal{K} \rightarrow \mathcal{K}$ exists such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c$, then the modulation constellation is symmetric with respect to origin. Equivalently, we can show that if a bijective function $\mathrm{g} : \mathcal{K} \rightarrow \mathcal{K}$ does not exist such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c$, then the modulation constellation is not symmetric with respect to origin. To prove this equivalent statement, we use proof by contradiction. We assume $\mathrm{g} : \mathcal{K} \rightarrow \mathcal{K}$ does not exist such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c$, but the modulation constellation is symmetric with respect to origin. If the modulation is symmetric with respect to the origin, then it satisfies the condition of Definition 2 and hence,

$$\mathrm{f}(-x_k) = -\mathrm{f}(x_k). \tag{31}$$

However, since the function $\mathrm{f}(x_k)$ is defined on set $\mathcal{A}$, (31) holds if and only if both $x_k$ and $-x_k$ are in the set $\mathcal{A}$. Consequently, set $\mathcal{A}$ can be represented by

$$\mathcal{A} = \{x_1, x_2, \cdots, x_{\frac{M}{2}}, -x_1, -x_2, \cdots, -x_{\frac{M}{2}}\}. \tag{32}$$

Now, if the bijective function g is defined as $\mathrm{g}(k) = k + \frac{M}{2}$, then $\frac{x_k}{x_{\mathrm{g}(k)}} = -1$. However, this contradicts the assumption that bijective function $\mathrm{g} : \mathcal{K} \rightarrow \mathcal{K}$ does not exist, such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c \forall k \in \mathcal{K}$. Hence, if $\mathrm{g} : \mathcal{K} \rightarrow \mathcal{K}$ does not exist such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c$, then the modulation constellation cannot be symmetric with respect to origin.

(ii) Secondly, we need to show that if the modulation is symmetric then a bijective function $\mathrm{g} : \mathcal{K} \rightarrow \mathcal{K}$ exists such that $\frac{x_k}{x_{\mathrm{g}(k)}} = c\ \forall k \in \mathcal{K}$. This easily follows from the proof of previous step, where we showed that if the modulation constellation is symmetric then $\mathcal{A}$ can be represented by (32). Consequently, a bijective function $\mathrm{g} : \mathcal{K} \rightarrow \mathcal{K}$ exists such that $\frac{x_k}{x_{\mathrm{g}(k)}} = -1$ $\forall k \in \mathcal{K}$, i.e., $\mathrm{g}(k) = k + \frac{M}{2}\ \forall k \in \mathcal{K}$.

Combining the proofs of the three steps, Theorem 1 is proved.

## Appendix 2

### Proof of Propositions 2 and 3

Propositions 2 and 3 correspond to the $E$ and $M$ steps of the EM algorithm. We assume that both transmitters at nodes $a$ and $b$ use the asymmetric shifted modulation constellation $\bar{\mathcal{A}}$ defined in Definition 4, i.e., $\bar{x}_{a_i}, \bar{x}_{b_i} \in \bar{\mathcal{A}}$ and assume a uniform discrete distribution for the transmitted symbols.

#### *2.1 Proof of E-step*

In the $E$-step of the algorithm function $Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)})$ is given by

$$Q\left(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)}\right) = \mathbb{E}_{\bar{\mathbf{X}}_\mathbf{b}|\mathbf{y_a},\boldsymbol{\phi}^{(n)}}\left[\ln f_{\mathbf{Y_a}}(\mathbf{y_a}, \bar{\mathbf{x}}_\mathbf{b}|\boldsymbol{\phi})\right]. \tag{33}$$

To calculate (33), we require $\ln f_{\mathbf{Y_a}}(\mathbf{y}_a, \bar{\mathbf{x}}_\mathbf{b}|\boldsymbol{\phi})$. Hence, we start with the following joint PDF

$$f_{Y_{a_i}}(y_{a_i}, \bar{x}_{b_i};\boldsymbol{\phi}) = f_{Y_{a_i}}(y_{a_i}|\bar{X}_{b_i} = \bar{x}_{b_i};\boldsymbol{\phi})p_{\bar{X}_{b_i}}(\bar{x}_{b_i}) = \frac{1}{M\pi\sigma^2}\sum_{k=1}^{M}\delta_{\bar{x}_k,\bar{x}_{b_i}}\exp\left(\frac{-1}{\sigma^2}|y_{a_i}-h_{ba}\bar{x}_k - h_{aa}\bar{x}_{a_i}|^2\right), \tag{34}$$

where, $\delta_{\bar{x}_k,\bar{x}_{b_i}}$ is the Kronecker delta function and $\delta_{\bar{x}_k,\bar{x}_{b_i}} = 1$ if $\bar{x}_{b_i} = \bar{x}_k$ and 0 otherwise [31]. Consequently, the log-likelihood of all the observations is given by

$$\ln(f_{\mathbf{Y_a}}(\mathbf{y_a},\mathbf{x_b};\boldsymbol{\phi})) = \sum_{i=1}^{N} \ln(f_{Y_{a_i}}(y_{a_i},\bar{x}_{b_i};\boldsymbol{\phi})) = -N\ln(M\pi\sigma^2) - \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{k=1}^{M}\delta_{\bar{x}_k,\bar{x}_{b_i}}|y_{a_i} - h_{ba}\bar{x}_{b_i} - h_{aa}\bar{x}_{a_i}|^2. \tag{35}$$

The expectation in (33) is conditioned on knowing $\boldsymbol{\phi}^{(n)}$ during the $n$th iteration of the algorithm, which is obtained from the $M$-step. Substituting (35) in (33), we have

$$Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)}) = \mathbb{E}_{\tilde{\mathbf{x}}_\mathbf{b}|\mathbf{y_a},\boldsymbol{\phi}^{(n)}}[\ln f_{\mathbf{Y_a}}(\mathbf{y_a},\tilde{\mathbf{x}}_\mathbf{b};\boldsymbol{\phi})] = -N\ln(M\pi\sigma^2) - \frac{1}{\sigma^2}\mathbb{E}_{\tilde{\mathbf{x}}_\mathbf{b}|\mathbf{y_a},\boldsymbol{\phi}^{(n)}}\left[\sum_{i=1}^{N}\sum_{k=1}^{M}\delta_{\bar{x}_k,\bar{x}_{b_i}}|y_{a_i} - h_{ba}\bar{x}_{b_i} - h_{aa}\bar{x}_{a_i}|^2\right]. \tag{36}$$

The assumption of independent transmitted symbols allows to rewrite (36) as follows

$$\begin{aligned} Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)}) &= -N\ln\left(M\pi\sigma^2\right) - \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{k=1}^{M}\mathbb{E}_{\bar{x}_{b_i}|\mathbf{y_a},\boldsymbol{\phi}^{(n)}}\left[\delta_{\bar{x}_k,\bar{x}_{b_i}}|y_{a_i} - h_{ba}\bar{x}_{b_i} - h_{aa}\bar{x}_{a_i}|^2\right], \\ &= -N\ln(M\pi\sigma^2) - \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{k=1}^{M}P(\bar{x}_{b_i} = \bar{x}_k|\mathbf{y}_a,\boldsymbol{\phi}^{(n)})|y_{a_i} - h_{ba}\bar{x}_k - h_{aa}\bar{x}_{a_i}|^2. \end{aligned} \tag{37}$$

We define

$$T_{k,i}^{(n)} \triangleq P(\bar{x}_{b_i} = \bar{x}_k|\mathbf{y}_a,\boldsymbol{\phi}^{(n)}). \tag{38}$$

Then, it can easily be shown that

$$T_{k,i}^{(n)} = \frac{\exp\left(\frac{-1}{\sigma^2}|y_{a_i} - \hat{h}_{ba}^{(n)}\bar{x}_k - \hat{h}_{aa}^{(n)}\bar{x}_{a_i}|^2\right)}{\sum_{\bar{k}=1}^{M}\exp\left(\frac{-1}{\sigma^2}|y_{a_i} - \hat{h}_{ba}^{(n)}\bar{x}_{\bar{k}} - \hat{h}_{aa}^{(n)}\bar{x}_{a_i}|^2\right)}. \tag{39}$$

Finally, substituting (39) into (37), $Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)})$ can be found as in (33). This concludes the proof of Proposition 2.

### *2.2 Proof of M-step*

The maximization-step of the EM algorithm is given by

$$\boldsymbol{\phi}^{(n+1)} = \arg\max_{\boldsymbol{\phi}} Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)}) = \arg\min_{\boldsymbol{\phi}} \sum_{i=1}^{N}\sum_{k=1}^{M} T_{k,i}^{(n)}|y_{a_i} - h_{ba}\bar{x}_k - h_{aa}\bar{x}_{a_i}|^2. \tag{40}$$

We define the following function

$$\mathrm{r}(\boldsymbol{\phi}) \triangleq \sum_{i=1}^{N}\sum_{k=1}^{M} T_{k,i}^{(n)}|y_{a_i} - h_{ba}\bar{x}_k - h_{aa}\bar{x}_{a_i}|^2, \tag{41}$$

The minimum of function $\boldsymbol{\phi}$ (the maximum of the likelihood function), which corresponds to the solution of the $M$-step of the EM algorithm during the $n$th iteration, happens at the critical point $\boldsymbol{\phi}^{(n+1)}$ for which the Jacobian is zero, i.e., $\mathrm{J} = 0$ [43]. To find this critical point, the Jacobian matrix should be constructed and set equal to zero. This is done by taking the derivative of $\mathrm{r}(\boldsymbol{\phi})$ with respect to the four elements of vector $\boldsymbol{\phi}$, as defined by (9), to construct the Jacobian matrix and then set it equal to zero. Then, it can be easily shown that the critical point $\boldsymbol{\phi}^{(n+1)}$ is given by

$$\boldsymbol{\phi}^{(n+1)} = \mathrm{S}^{-1}\boldsymbol{\nu}, \tag{42}$$

where

$$\mathrm{S} \triangleq \begin{bmatrix} s_1 & 0 & s_2 & s_3 \\ 0 & s_1 & -s_3 & s_2 \\ s_2 & -s_3 & s_4 & 0 \\ s_3 & s_2 & 0 & s_4 \end{bmatrix}, \quad \boldsymbol{\nu} \triangleq \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{bmatrix}, \tag{43}$$

where the elements of S and $\boldsymbol{\nu}$ are given by (14b)–(14d).

However, to ensure that the critical point $\boldsymbol{\phi}^{(n+1)}$ is the minimum of function $\mathrm{r}(\boldsymbol{\phi})$, the Hessian matrix H should be positive semi-definite [43]. By taking the second derivatives of $\mathrm{r}(\boldsymbol{\phi})$ with respect to the four elements of vector $\boldsymbol{\phi}$, we can show that $\mathrm{H} = 2\mathrm{S}$. Then, according to Sylvester's criterion [43], H is positive semi-definite if and only if all the following are positive

$$s_1, \quad \det\left(\begin{bmatrix} s_1 & 0 \\ 0 & s_1 \end{bmatrix}\right), \quad \det\left(\begin{bmatrix} s_1 & 0 & s_2 \\ 0 & s_1 & -s_3 \\ s_2 & -s_3 & s_4 \end{bmatrix}\right), \quad \det(\mathrm{S}). \tag{44}$$

It can easily be shown that $\det(\mathrm{S}) = \left(s_1 s_4 - s_2^2 - s_3^2\right)^2$ and is always positive. According to (14b), $s_1$ is always positive, and it is clear that the second determinant is always positive. However, the positivity of the third determinant, i.e., $s_1 s_4 - s_2^2 - s_3^2$, directly depends on the initialization. This is evident from definitions in (14b)–(14d), which link $s_1$, $s_2$, $s_3$, and $s_4$ to the function $T_{k,i}^{(n)}$ and the derivation of function $T_{k,i}^{(n)}$ in (39), which is a function of $\hat{h}_{aa}^{(n)}$ and $\hat{h}_{ba}^{(n)}$, i.e., the estimates from the $n$th iteration. Our numerical investigation shows that for the initialization vector $\boldsymbol{\phi}^{(0)} \triangleq [0,0,0,0]$, the Hessian matrix H is always positive and hence the critical point $\boldsymbol{\phi}^{(n+1)}$ is indeed the minimum of function $\mathrm{r}(\boldsymbol{\phi})$. Consequently, the EM algorithm with initialization vector $\boldsymbol{\phi}^{(0)} \triangleq [0,0,0,0]$ converges to the maximum of the likelihood function.

This concludes the proof of Proposition 3.

## Appendix 3

### Proof of Proposition 4

It is shown in [44] that for any random variables $X$ and $Y$ and any parameter $\theta$, if the probability distribution of $X$ is independent of the parameter $\theta$, then $\mathrm{I}[f_Y(y;\theta)] < \mathrm{I}[f_Y(y|x;\theta)]$, where $\mathrm{I}[\cdot]$ is the FIM and $f(\cdot;\theta)$ is the probability density function parameterized by $\theta$. Consequently, the performance of the proposed estimator is lower bounded by the inverse of $\mathrm{I}[f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};\boldsymbol{\phi})]$, i.e.,

$$\mathbb{E}_{\hat{\Phi}_l}[\,|\hat{\phi}_l-\phi_l|^2] \geq \left[\mathrm{I}^{-1}\left[f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};\boldsymbol{\phi})\right]\right]_{l,l} \quad \forall l\in\{1,2,3,4\}, \tag{45}$$

where $\phi_l$ is the $l$th element of the parameter vector $\boldsymbol{\phi}$, $\hat{\phi}_l$ is an estimate of $\phi_l$, and $[\,\cdot]_{l,l}$ is the $l$th diagonal element of matrix. This is because (i) the performance of the proposed estimator is lower bounded by $\mathrm{I}[f_{\mathbf{Y_a}}(\mathbf{y_a};\boldsymbol{\phi})]$ according to (15) and (ii) $p_{\bar{\mathbf{X}}_\mathbf{b}}(\bar{\mathbf{x}}_\mathbf{b})$ is independent of $\boldsymbol{\phi}$. Furthermore, since (45) holds $\forall \bar{\mathbf{x}}_\mathbf{a}, \bar{\mathbf{x}}_\mathbf{b}$, then the variance is also lower-bounded by

$$\mathbb{E}_{\hat{\Phi}_l}\left[|\hat{\phi}_l-\phi_l|^2\right] \geq \left[\mathrm{I}_{avg}^{-1}\left[f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};\boldsymbol{\phi})\right]\right]_{l,l} \quad \forall l\in\{1,2,3,4\}, \tag{46}$$

where $\mathrm{I}_{avg} = \mathbb{E}_{\bar{\mathbf{X}}_\mathbf{b},\bar{\mathbf{X}}_\mathbf{a}}\left[\mathrm{I}[f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};\boldsymbol{\phi})]\right]$. The value of $\mathrm{I}[f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};\boldsymbol{\phi})]$, needed to evaluate $\mathrm{I}_{avg}$, is presented in the Lemma below.

**Lemma 1** *$\mathrm{I}[f(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};\boldsymbol{\phi})]$ is a $4\times 4$ matrix with its elements given by*

$$i_{1,1}=i_{2,2}=\frac{2}{\sigma^2}\sum_{i=1}^{N}|\bar{x}_{a_i}|^2, \qquad \mathbb{I}_{3,3}=\mathbb{I}_{4,4}=\frac{2}{\sigma^2}\sum_{i=1}^{N}|\bar{x}_{b_i}|^2, \tag{47a}$$

$$i_{2,4}=i_{4,2}=i_{1,3}=i_{3,1} = \frac{2}{\sigma^2}\sum_{i=1}^{N}\left(\Re\{\bar{x}_{a_i}\}\Re\{\bar{x}_{b_i}\}+\Im\{\bar{x}_{a_i}\}\Im\{\bar{x}_{b_i}\}\right), \tag{47b}$$

$$i_{2,3}=i_{3,2}=-i_{1,4}=-i_{4,1} = \frac{2}{\sigma^2}\sum_{i=1}^{N}\left(\Re\{\bar{x}_{b_i}\}\Im\{\bar{x}_{a_i}\}-\Re\{\bar{x}_{a_i}\}\Im\{\bar{x}_{b_i}\}\right). \tag{47c}$$

*Proof* It can be easily seen from (5) that $f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};h_{aa},h_{ba})$ is given by

$$f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};h_{aa},h_{ba}) = \left(\frac{1}{\pi\sigma^2}\right)^N \exp\sum_{i=1}^{N}\left(\frac{-|y_{a_i}-h_{aa}\bar{x}_{a_i}-h_{ba}\bar{x}_{b_i}|^2}{\sigma^2}\right). \tag{48}$$

Then, for $l,l'\in\{1,2,3,4\}$, $\mathrm{I}[f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};h_{aa},h_{ba})]$ is [36]

$$\mathrm{I}_{l,l'} = -\mathbb{E}_{\mathbf{Y_a}}\left[\frac{\partial^2}{\partial_{\phi_m}\partial_{\phi_n}}\ln f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};h_{aa},h_{ba})\right], \tag{49}$$

where $m,n\in\{1,2,3,4\}$, $\phi_m$ and $\phi_n$ are the $m$th and $n$th elements of $\boldsymbol{\phi}=[\Re\{h_{aa}\},\Im\{h_{aa}\},\Re\{h_{ba}\},\Im\{h_{ba}\}]$. By evaluating (49), using the joint PDF given by (48), the non-zero elements of $\mathrm{I}[f_{\mathbf{Y_a}}(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};h_{aa},h_{ba})]$ can be found and are given by (47b)–(47c). □

Using the value of $\mathrm{I}[f(\mathbf{y_a}|\bar{\mathbf{x}}_\mathbf{b};\boldsymbol{\phi})]$ given by the above lemma, we need to evaluate the expectations in order to find $\mathrm{I}_{avg}$. As discussed in Section 3.3, we assume that both nodes $a$ and $b$ use a real constant $s \triangleq \sqrt{\beta E}$ to shift the modulation constellation. Since all the constellation points are equally likely to be transmitted, before shifting the modulation constellation we have $\mathbb{E}_{\Im\{\bar{X}_{a_i}\}}[\Im\{x_{a_i}\}] = \mathbb{E}_{\Im\{\bar{x}_{b_i}\}}[\Im\{x_{b_i}\}] = \mathbb{E}_{\Re\{\bar{X}_{a_i}\}}[\Re\{x_{a_i}\}] = \mathbb{E}_{\Re\{\bar{X}_{b_i}\}}[\Re\{x_{b_i}\}] = 0$. However, after the shift, $\mathbb{E}_{\Re\{\bar{X}_{a_i}\}}[\Re\{\bar{x}_{a_i}\}] = \mathbb{E}_{\Re\{\bar{X}_{b_i}\}}[\Re\{\bar{x}_{b_i}\}] = \sqrt{\beta E}$ and $\mathbb{E}_{\Im\{\bar{X}_{a_i}\}}[\Im\{\bar{x}_{a_i}\}] = \mathbb{E}_{\Im\{\bar{X}_{b_i}\}}[\Im\{\bar{x}_{b_i}\}] = 0$. Furthermore, $\mathbb{E}_{X_{a_i}}[\,|\bar{x}_{a_i}|^2] = \mathbb{E}_{X_{b_i}}[\,|\bar{x}_{b_i}|^2] = E+\beta E$ since the average energy of the constellation after the shift is increased by the shift energy ($|s|^2=\beta E$). Consequently, the average FIM with respect to $\bar{\mathbf{x}}_\mathbf{a}$ and $\bar{\mathbf{x}}_\mathbf{b}$ is given by

$$\mathrm{I}_{avg} = \frac{2NE}{\sigma^2}\begin{pmatrix} 1+\beta & 0 & \beta & 0\\ 0 & 1+\beta & 0 & \beta\\ \beta & 0 & 1+\beta & 0\\ 0 & \beta & 0 & 1+\beta\end{pmatrix}, \tag{50}$$

and $\mathrm{I}_{avg}^{-1}$ is given by

$$\mathrm{I}_{avg}^{-1} = \frac{\sigma^2}{2NE}\begin{pmatrix} \frac{\beta+1}{(2\beta+1)} & 0 & -\frac{\beta}{(2\beta+1)} & 0\\ 0 & \frac{\beta+1}{(2\beta+1)} & 0 & -\frac{\beta}{(2\beta+1)}\\ -\frac{\beta}{(2\beta+1)} & 0 & \frac{\beta+1}{(2\beta+1)} & 0\\ 0 & -\frac{\beta}{(2\beta+1)} & 0 & \frac{\beta+1}{(2\beta+1)}\end{pmatrix}. \tag{51}$$

Using (46) and considering the diagonal elements of (51), we arrive at the result in (16).

#### Acknowledgements

The authors would like to thank Mohammad Azarbad for insightful discussions and comments.

#### Funding

This work was supported by the Australian Research Councils Discovery Project Funding Scheme (Project number DP140101133). The work of Abbas Koohian was supported by an Australian Government Research Training Program (RTP) Scholarship. The work of Hani Mehrpouyan was partially funded by the NSF ERAS grand award number 1642865. The work of Steven Blostein was partially funded by the Natural Sciences and Engineering Research Council of Canada 05061-2014.

#### Authors' contribution

All authors contributed equally to this work. The final manuscript has been read and approved by all authors for submission.

#### Competing interests

The authors declare that they have no competing interests.

## Publisher's Note



**Author details**

[1]Research School of Engineering, Australian National University, Canberra, Australia. [2]Department of Electrical and Computer Engineering, Boise State University, Idaho, USA. [3]Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia. [4]Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada.

Received: 3 July 2017 Accepted: 11 January 2018
Published online: 24 January 2018

**References**

1. D Korpi, J Tamminen, M Turunen, T Huusari, YS Choi, L Anttila, S Talwar, M Valkama, Full-duplex mobile device: pushing the limits. IEEE Commun. Mag. **54**(9), 80–87 (2016)
2. Z Zhang, X Chai, K Long, AV Vasilakos, L Hanzo, Full duplex techniques for 5G networks: self-interference cancellation, protocol design, and relay selection. IEEE Commun. Mag. **53**(5), 128–137 (2015)
3. A Sabharwal, P Schniter, D Guo, DW Bliss, S Rangarajan, R Wichman, In-band full-duplex wireless: challenges and opportunities. IEEE J. Sel. Areas Commun. **32**(9), 1637–1652 (2014)
4. MS Sim, M Chung, D Kim, J Chung, DK Kim, CB Chae, Nonlinear self-interference cancellation for full-duplex radios: from link-level and system-level performance perspectives. IEEE Commun. Mag. **55**(9), 158–167 (2017)
5. M Duarte, C Dick, A Sabharwal, Experiment-driven characterization of full-duplex wireless systems. IEEE Trans. Wireless Commun. **11**(12), 4296–4307 (2012)
6. E Everett, A Sahai, A Sabharwal, Passive self-interference suppression for full-duplex infrastructure nodes. IEEE Trans. Wireless Commun. **13**(2), 680–694 (2014)
7. B Kaufman, J Lilleberg, B Aazhang, in *Proc. Asilomar Conf. on Signals, Syst. and Computers*. An analog baseband approach for designing full-duplex radios, (2013)
8. D Korpi, T Riihonen, V Syrjala, L Anttila, M Valkama, R Wichman, Full-duplex transceiver system calculations: analysis of ADC and linearity challenges. IEEE Trans. Wireless Commun. **13**(7), 3821–3836 (2014)
9. D Korpi, L Anttila, V Syrjala, M Valkama, Widely-linear digital self-interference cancellation in direct-conversion full duplex transceiver. IEEE J. Sel. Areas Commun. **32**(9), 1674–1687 (2014)
10. E Ahmed, AM Eltawil, All-digital self-interference cancellation technique for full-duplex systems. IEEE Trans. Wireless Commun. **14**(7), 3519–3532 (2015)
11. A Nadh, J Samuel, A Sharma, S Aniruddhan, RK Ganti, A linearization technique for self-interference cancellation in full-duplex radios. (2017). arXiv Technical Report. [Online]. Available: https://arxiv.org/abs/1605.01345
12. D Kim, H Ju, S Park, D Hong, Effects of channel estimation error on full-duplex two-way networks. IEEE Trans. Veh. Technol. **62**(9), 4666–4672 (2013)
13. A Masmoudi, T Le-Ngoc, A maximum-likelihood channel estimator for self-interference cancelation in full-duplex systems. IEEE Trans. Veh. Technol. **65**(7), 5122–5132 (2016)
14. X Li, C Tepedelenliolu, H Şenol, Channel estimation for residual self-interference in full-duplex amplify-and-forward two-way relays. IEEE Trans. Wirel. Commun. **16**(8), 4970–4983 (2017)
15. S Li, RD Murch, in *Proc. IEEE GLOBECOM*. Full-duplex wireless communication using transmitter output based echo cancellation, (2011)
16. M Sohaib, H Nawaz, K Ozsoy, O Gurbuz, I Tekin, A low complexity full-duplex radio implementation with a single antenna (accepted to appear.) IEEE Trans. Veh. Technol (2017)
17. F Shu, J Wang, J Li, R Chen, W Chen, Pilot optimization, channel estimation, and optimal detection for full-duplex OFDM systems with IQ imbalances. IEEE Trans. Veh. Technol. **66**(8), 6993–7009 (2017)
18. A Masmoudi, T Le-Ngoc, Channel estimation and self-interference cancellation in full-duplex communication systems. IEEE Trans. Veh. Technol. **66**(1), 321–334 (2017)
19. A Koohian, H Mehrpouyan, M Ahmadian, M Azarbad, in *Proc. IEEE ICC*. Bandwidth efficient channel estimation for full duplex communication systems, (2015)
20. A Koohian, H Mehrpouyan, AA Nasir, S Durrani, SD Blostein, in *Proc. IEEE ICC*. Residual self-interference cancellation and data detection in full-duplex communication systems, (2017)
21. F Mazzenga, Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence. IEEE Trans. Broadcast. **46**(2), 170–176 (2000)
22. JK Tugnait, W Luo, On channel estimation using superimposed training and first-order statistics. IEEE Commun. Lett. **7**(9), 413–415 (2003)
23. E de Carvalho, DTM Slock, Blind and semi-blind FIR multichannel estimation: (global) identifiability conditions. IEEE Trans. Signal Process. **52**(4), 1053–1064 (2004)
24. S Abdallah, IN Psaromiligkos, EM-based semi-blind channel estimation in amplify-and-forward two-way relay networks. IEEE Wirel. Commun. Lett. **2**(5), 527–530 (2013)
25. S Abdallah, IN Psaromiligkos, Blind channel estimation for amplify-and-forward two-way relay networks employing $M$-PSK modulation. IEEE Trans. Signal Process. **60**(7), 3604–3615 (2012)
26. S Zhang, S-C Liew, H Wang, Blind known interference cancellation. IEEE J. Sel. Areas Commun. **31**(8), 1572–1582 (2013)
27. Y Zhu, D Guo, ML Honig, A message-passing approach for joint channel estimation, interference mitigation, and decoding. IEEE Trans. Wirel. Commun. **8**(12), 6008–6018 (2009)
28. E Lehmann, G Casella, *Theory of Point Estimation*. (Springer Verlag, New York, 1998)
29. L Tong, S Perreau, Multichannel blind identification: from subspace to maximum likelihood methods. Proc. IEEE. **86**(10), 1951–1968 (1998)
30. K Abed-Meraim, W Qiu, Y Hua, Blind system identification. Proc. IEEE. **85**(8), 1310–1322 (1997)
31. PJ Pahl, R Damrath, *Mathematical Foundations of Computational Engineering: A Handbook*. (Springer, Berlin Heidelberg, 2001)
32. N Loehr, *Bijective Combinatorics,* 1st ed. (Chapman & Hall/CRC, Boca Raton, 2011)
33. AP Dempster, NM Laird, DB Rubin, Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B. **39**, 1–38 (1977)
34. V Melnykov, I Melnykov, Initializing the EM algorithm in Gaussian mixture models with an unknown number of components. Comput. Stat. Data Anal. **56**(6), 1381–1395 (2012)
35. C Biernacki, G Celeux, G Govaert, Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. Comput. Stat. Data Anal. **41**(3-4), 561–575 (2003)
36. SM Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. (Prentice-Hall, Inc., Upper Saddle River, 1993)
37. S Abdallah, IN Psaromiligkos, Exact Cramer Rao bounds for semiblind channel estimation in amplify-and-forward two-way relay networks employing square QAM. IEEE Trans. Wirel. Commun. **13**(12), 6955–6967 (2014)
38. H Mehrpouyan, SD Blostein, Bounds and algorithms for multiple frequency offset estimation in cooperative networks. IEEE Trans. Wirel. Commun. **10**(4), 1300–1311 (2011)
39. D Tse, P Viswanath, *Fundamentals of Wireless Communication*. (Cambridge University Press, New York, 2005)
40. E Ahmed, AM Eltawil, All-digital self-interference cancellation technique for full-duplex systems. IEEE Trans. Wirel. Commun. **14**(7), 3519–3532 (2015)
41. M Duarte, A Sabharwal, in *Proc. Asilomar Conf. on Signals, Syst. and Computers*. Full-duplex wireless communications using off-the-shelf radios: feasibility and first results, (2010)
42. E Romero-Aguirre, R Parra, AG Orozco, R Carrasco-Alvarez, in *Workshop on Signal Processing Systems (SiPS)*. Full-hardware architectures for data-dependent superimposed training channel estimation, (2011)
43. G Strang, *Linear Algebra and Its Applications*. (Harcourt Brace Jovanovich Publishers, Essex England, 1988)
44. R Zamir, A proof of the Fisher information inequality via a data processing argument. IEEE Trans. Inf. Theory. **44**(3), 1246–1250 (1998)