RESEARCH   **Open Access**

CrossMark

Subspace-based self-interference cancellation for full-duplex MIMO transceivers

Ahmed Masmoudi[*] and Tho Le-Ngoc

Abstract

This paper addresses the self-interference (SI) cancellation at baseband for full-duplex MIMO communication systems in consideration of practical transmitter imperfections. In particular, we develop a subspace-based algorithm to jointly estimate the SI and intended channels and the nonlinear distortions. By exploiting the covariance and pseudo-covariance of the received signal, we can increase the dimension of the received signal subspace while keeping the dimension of the signal subspace constant, and hence, the proposed algorithm can be applied to most of full-duplex MIMO configurations with arbitrary numbers of transmit and receive antennas. The channel coefficients are estimated, up to an ambiguity term, without any knowledge of the intended signal. A joint detection and ambiguity identification scheme is proposed. Simulation results show that the proposed algorithm can properly estimate the channel with only one pilot symbol and offers superior SI cancellation performance.

Keywords: Full-duplex communication, SI suppression, MIMO, parameter estimation, Subspace method, Second-order statistics

1 Introduction

Half-duplex transmission is commonly used in the current communication systems by transmitting and receiving over orthogonal channels. Full-duplex communication represents an attractive alternative to save channel resources or to increase the transmission efficiency. The main deterrent to employ full-duplex is the large self-interference (SI) from the simultaneous transmission and reception over the same frequency band. The SI is usually several orders of magnitude higher than the intended signal received from the other transmitter, because the later travels a longer distance than the former signal. Recent works have shown that, using different cancellation stages, the SI can be sufficiently suppressed to properly detect the intended signal [1, 2].

The SI is first cancelled at the radio-frequency (RF) level, prior to the low-noise amplifier (LNA) and the analog-to-digital converter (ADC), to avoid overloading/saturation of these devices [1–3]. In other words, the SI should be sufficiently suppressed at RF to maintain the receiver's limited dynamic range. Then, further SI suppression can be done after the ADC at the baseband [4, 5].

In the following, we assume that a cancellation stage at RF is available and we concentrate on the SI cancellation in the baseband.

To further reduce the SI, channel state information of the interference link should be available. Therefore, estimating the SI channel is a critical issue in full-duplex systems. In [6], the SI channel estimation is performed in the frequency domain using a least square (LS) technique. LS and minimum mean square error (MMSE) channel estimations are proposed in [7] to estimate the SI channel in the relay station. However, these approaches ignore the intended signal coming from the other transceiver and treat it as additive noise. An adaptive least mean square algorithm to estimate the SI channel is proposed in [8] where the large SI compared to the intended signal and additive noise is exploited to obtain an estimate of the SI channel. A more elaborate LS-based estimator was presented in [9] where a first estimate of the SI channel is obtained by considering the intended signal as additive noise. Then an iterative detection of the intended signal and channel estimation is performed to obtain a better estimate of the channel. On the other hand, spatial domain cancellation attempts to reduce the SI by precoding at the transmit chain and decoding at the receive chain. Spatial domain cancellation is formulated in the frequency

*Correspondence: ahmed.masmoudi@mail.mcgill.ca
Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec, Canada

domain [10–12]. An alternative time domain formulation was presented in [13] by precoding the transmitted SI to coincide with the null space of the SI channel. These techniques are based on the knowledge of both the SI and intended channels at the two transceivers, which further motivates the development of channel estimators for full-duplex systems. A novel cancellation method is proposed in [14] by adding a cancelling signal to the original signal.

In addition to the SI channel information for SI cancellation, intended channel knowledge is an important prerequisite for signal detection. Motivated by this fact, channel estimation has been the subject of intense research. In the case of data-aided transmissions, training-based techniques can be applied [15, 16]. However, the amount of training increases dramatically with the number of antennas and channel order. Blind approaches have been proposed as more bandwidth efficient techniques [17, 18] where subspace methods, initially presented in [19], have a great potential. By decomposing the covariance matrix of the received signal, subspace methods exploit the orthogonality between the noise and the signal subspaces in the observation space to express the channel coefficients as a linear combination of a basis of the signal subspace. Although previous researches have shown the potential of this procedure to give an accurate estimate of the channel, it remains of limited practical interest. Actually, considering that the noise subspace needs to be nondegenerated, it is legitimate to wonder how we can satisfy this condition. Previous works rely on oversampling of the received signal or using more receive antennas than transmit antennas [20, 21]. However, such solutions increase the receiver cost and need additional hardware. Moreover, they may result in correlated noise which makes the subspace technique inappropriate. A maximum likelihood estimator was presented in [22] by exploiting the pilots in the intended signal.

In the full-duplex context, the transmitter impairments, including power amplifier (PA) nonlinearity and IQ mixer imbalance, become limiting factors and need to be reduced to properly detect the intended signal. In practice, the inband image resulting from the IQ mixer in mobile user is about 28 dB lower than the direct signal [23]. In the presence of strong SI of about 50 dB higher than the intended signal, this IQ image represents additional interference for the intended signal. The effects of transceiver impairments are illustrated in detail in [3, 24]. Due to the importance of the nonlinearities, a digital cancellation procedure has been proposed to reduce the effects of the PA in [25] by estimating the nonlinear coefficients of the PA and another algorithm has been proposed to deal with the IQ mixer imbalance [26]. However, there is no discussion about the intended signal in the existing literature, which limits the estimation performance if it is considered as additive noise.

In this work, we incorporate the intended signal in the estimation process. We also take into account the transmitter impairments when modelling the SI signal. For realistic multipath propagation channels, we need to estimate the SI channel, the intended channel and the distorted SI. And noting that the intended signal is unknown, we propose to use a novel subspace method to efficiently estimate the different parameters. Since the received signal consists of the SI and intended signals, the dimension of the signal subspace in full-duplex operation is at least twice that in traditional half-duplex operation [5, 27]. Thus an essential shortcoming of the existing subspace-based technique is that it can be applied only when the number of receive antennas is larger than the number of transmit antennas. In the following, we circumvent this condition and develop a subspace-based algorithm suitable for MIMO full-duplex systems with larger or equal numbers of transmit and receive antennas. We exploit both the covariance and pseudo-covariance matrices of the received signal to effectively increase the dimension of the observation space while keeping the dimension of the signal subspace unchanged. The joint processing of the received signal and its complex conjugates has been used in many works to improve the detection performance on various systems [28, 29]. Also, in an entirely different context, the improper property of the received signal was first exploited for channel identification in [30] to obtain a virtual SIMO model from a SISO one. Preliminary results can be found in [31] for real-valued symbols to enable the application of widely linear processing techniques, but entail a loss in spectral efficiency compared to complex-valued symbols. We propose in this paper a method to use the widely linear processing to complex symbols by forcing the transmit signal to be improper. We justify the advocated time domain approach and compare its performances to a frequency domain approach and we generalize the PA model to any nonlinearity order. In practice, we cannot blindly recover the channel coefficients since an ambiguity term always appears in the final estimate [5]. This ambiguity is resolved using a sequence of pilot symbols, considerably shorter than needed in training-based techniques. In the following, we propose a joint data detection and estimation of the ambiguity term to considerably reduce the length of the pilot sequence. We show through simulation that just one pilot symbol is sufficient to perfectly estimate the channel.

The paper is organized as follows. In Section 2, the full-duplex system model is presented. The subspace-based channel estimation is described in Section 3. In Section 4, we describe the joint decoding and ambiguity removal procedure. Illustrative simulation results are given in Section 5 and Section 6 presents the conclusion.

Notations commonly used in this paper are presented. Subscripts $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ refer to conjugate, transpose

and conjugate transpose for matrices or vectors, respectively. For a given vector x, diag(x) returns a diagonal matrix whose diagonal elements are the entries of x. rank(M) returns the rank of a given matrix M, det(M) returns the determinant of M and vect(M) stacks the columns of M into one vector. The operator $\otimes$ refers to the Kronecker product of two matrices. $\Re(\cdot)$ and $\Im(\cdot)$ return the real and imaginary parts of complex numbers. $E(\cdot)$ denotes the mathematical expectation. $||\cdot||_2$ returns the Euclidean norm of a vector. I_p refers to the $p \times p$ identity matrix and $\mathbf{1}_p$ the $p \times 1$ vector with 1 at all elements. A term accented by a hat, $\widehat{x}$, means an estimate of x.

2   Full-duplex MIMO system model

Consider two transceivers communicating in a full-duplex fashion. The simultaneous transmission and reception creates self-interference (SI) to be cancelled before the demodulation process. The SI signal is first suppressed at RF, prior to the low-noise amplifier (LNA) and analog-to-digital converter (ADC) to avoid overloading/saturation of these components [2, 3, 32]. In [5], we proposed an efficient compressed-sensing (CS)-based algorithm for the RF SI cancellation stage. In this work, we concentrate on the development of subspace-based algorithm to jointly estimate the SI and intended channels and the nonlinear distortions for the baseband SI cancellation stage of a full-duplex MIMO transceiver with arbitrary numbers of transmit and receive antennas. The output signal of the RF SI cancellation stage consists of the residual SI, the intended signal received from the other transceiver and the additive thermal noise. Figure 1 shows a simplified block diagram of a MIMO transceiver. The residual SI can be further suppressed at the baseband after ADC using digital signal processing (DSP). The advantage of working in the digital domain, as compared to RF, is that sophisticated DSP methods can be handled. Both transceivers are equipped with N_t transmitting antennas and N_r receiving antennas. At transmitting antenna q, a group of N data symbols $X_q = [X_q(0), \ldots, X_q(N-1)]^T$ is first modulated by the IFFT matrix to form an OFDM block, then the time domain vector $x_q = [x_q(0), \ldots, x_q(N-1)]^T$ is extended by the cyclic prefix of length[1] N_{cp} and the resulting vector is sent sequentially. In the transmit stream q, the complex signal $x_q(t)$ after the digital-to-analog conversion (DAC), is passed through an imbalance IQ mixer whose output is as follows:

$$x_q^{IQ}(t) = k_{1,q}x_q(t) + k_{2,q}x_q^*(t), \qquad (1)$$

where $k_{1,q}$ and $k_{2,q}$ are the responses of the IQ mixer at antenna q to the direct signal and the image, respectively. Then, the signal is amplified with a nonlinear PA. In the
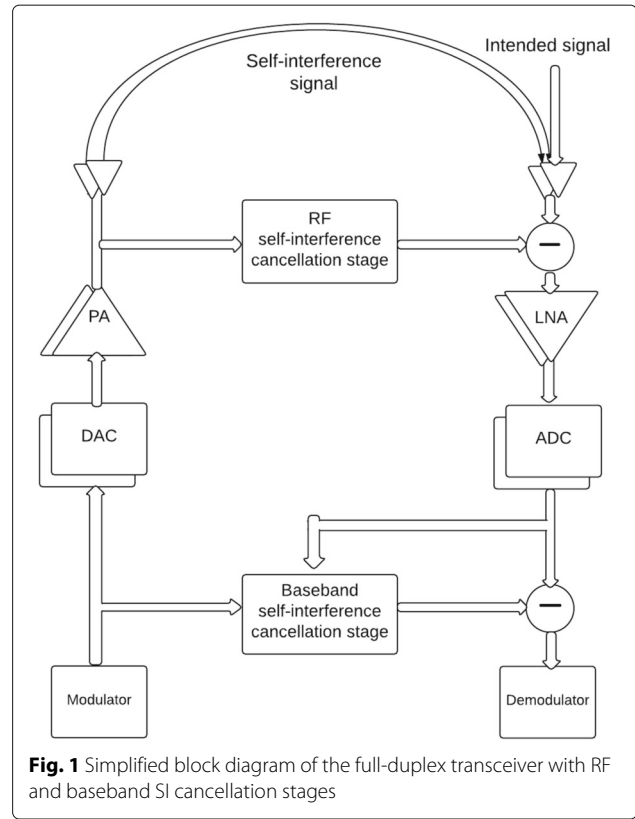

Fig. 1 Simplified block diagram of the full-duplex transceiver with RF and baseband SI cancellation stages

following, we model the PA response with a Hammerstein model whose response is:

$$x_q^{PA}(t) = \left(\sum_{p=0}^{P} \alpha_{2p+1,q} x_q^{IQ}(t) |x_q^{IQ}(t)|^{2p} \right) \star f(t), \qquad (2)$$

where $\alpha_{2p+1,q}$, for $p = 0, \ldots, P$, are the nonlinearity coefficients of the PA at transmit antenna q, P is the nonlinearity order and $f(t)$ is the memory of the PA. In (2), $\star$ denotes the convolution operator. The transmitted signal is coupled to produce SI in the receiver. Considering multipath channels, the received signal at antenna r is as follows:

$$y_r^{ant}(t) = \sum_{q=1}^{N_t} h_{r,q}^c(t) \star x_q^{PA}(t) + \sum_{q=1}^{N_t} h_{r,q}^s(t) \star s_q(t) + w_{th,r}(t),$$

$$(3)$$

where $s_q(t)$ is the transmitted signal from the qth antenna of the other intended transceiver. $h_{r,q}^c(t)$ is the response of the SI channel from transmitting antenna q to receiving antenna r of the same transceiver. $h_{r,q}^s(t)$ is the response of the intended channel from transmitting antenna q of the other intended transceiver to receiving antenna r of the same transceiver. $w_{th,r}(t)$ is the additive thermal noise in

Rx stream r. To reduce the SI before the LNA and ADC, the RF cancellation stage is performed as follows:

$$y_r^{RF}(t) = y_r^{ant}(t) - \sum_{q=1}^{N_t} \widehat{h}_{r,q}^c(t) \star x_q^{PA}(t), \qquad (4)$$

where $\widehat{h}_{r,q}^c(t)$ is a first estimate of the SI channel [1, 6]. $\widehat{h}_{r,q}^c(t)$ is used to adjust the phase, amplitude and delay of the SI to the main propagation path. To include the transmitter distortion in the RF cancellation process, the reference signal is taken from the output of the PA. This RF SI cancellation can attenuate the SI by 30 dB, as reported in practical experiments [6, 33]. Then, the received signal passes through the LNA:

$$y_r^{LNA}(t) = k_{LNA} y_r^{RF}(t) + w_{LNA}(t), \qquad (5)$$

where $w_{LNA}(t)$ is the additive noise caused by the LNA and k_{LNA} is the gain of the LNA. Finally, the received signal is adjusted by the variable gain amplifier (VGA) to match the dynamic range of the ADC. For simplicity, we suppose that the linear gains $k_{1,q}$ and $\alpha_{1,q}$ of the IQ mixer and PA are equal to 1. Combining (2), (3) and (5), the received samples are given by

$$y_r(n) = \sum_{q=1}^{N_t} \sum_{l=0}^{L} h_{r,q}^{(i)}(l) x_q^{IQ}(n-l) + \sum_{p=1}^{P} \alpha_{2p+1,q} h_{r,q}^{(i)}(l) x_{q,ip,p}(n-l)$$
$$+ h_{r,q}^{(s)}(l) s_q(n-l) + w_r(n), \qquad (6)$$

where $x_{q,ip,p}(n) = x_q^{IQ}(n) |x_q^{IQ}(n)|^{2p}$ resulting from the cascade of IQ mismatch and PA $(2p+1)^{rd}$ order nonlinearity and $w_r(n)$ collects the thermal noise, the LNA noise and the quantization noise. In (6), the global channel responses are given by

$$h_{r,q}^{(i)}(l) = k_{LNA}(h_{r,q}^c(l) \star f(l) - \widehat{h}_{r,q}^c(l)),$$
$$h_{r,q}^{(s)}(l) = k_{LNA} h_{r,q}^s(l). \qquad (7)$$

To have a homogeneous notation, all channels are supposed to have the same order L and the channels of order lower than L are zero-padded so that the different channels have the same order and L still satisfies $L < N_{cp}$. The received vector $\boldsymbol{y}(n) = [y_1(n), \ldots, y_{N_r}(n)]^T$ over the N_r antennas is given by

$$\boldsymbol{y}(n) = \sum_{q=1}^{N_t} \sum_{l=0}^{L} \boldsymbol{h}_q^{(i)}(l) x_q^{IQ}(n-l) + \sum_{p=1}^{P} \alpha_{2p+1,q} \boldsymbol{h}_q^{(i)}(l) x_{q,ip,p}(n-l)$$
$$+ \boldsymbol{h}_q^{(s)}(l) s_q(n-l) + \boldsymbol{w}(n), \qquad (8)$$

where

$$\boldsymbol{h}_q^{(i)}(l) = [h_{1,q}^{(i)}(l), h_{2,q}^{(i)}(l), \ldots, h_{N_r,q}^{(i)}(l)]^T,$$
$$\boldsymbol{h}_q^{(s)}(l) = [h_{1,q}^{(s)}(l), h_{2,q}^{(s)}(l), \ldots, h_{N_r,q}^{(s)}(l)]^T, \qquad (9)$$

for $l = 0, 1, \ldots, L$ and $\boldsymbol{w}(n) = [w_1(n), w_2(n), \ldots, w_{N_r}(n)]^T$. For a more compact representation, we gather the transmitted signals from the N_t antennas to obtain

$$\boldsymbol{y}(n) = \sum_{l=0}^{L} \boldsymbol{H}^{(i)}(l) \boldsymbol{x}(n-l) + \boldsymbol{H}^{(s)}(l) \boldsymbol{s}(n-l) + \boldsymbol{w}(n), \quad (10)$$

where the $N_r \times N_t$ matrices $\boldsymbol{H}^{(i)}(l)$ and $\boldsymbol{H}^{(s)}(l)$ are given by

$$\boldsymbol{H}^{(i)}(l) = [\boldsymbol{h}_1^{(i)}(l), \boldsymbol{h}_2^{(i)}(l), \ldots, \boldsymbol{h}_{N_t}^{(i)}(l)],$$
$$\boldsymbol{H}^{(s)}(l) = [\boldsymbol{h}_1^{(s)}(l), \boldsymbol{h}_2^{(s)}(l), \ldots, \boldsymbol{h}_{N_t}^{(s)}(l)], \qquad (11)$$

for $l = 0, \ldots, L$ and

$$\boldsymbol{x}_i(n) = [x_1(n), x_2(n), \ldots, x_{N_t}(n)]^T,$$

$$\boldsymbol{x}_{dist}(n) = \left[k_{2,1} x_1^*(n) + \sum_{p=1}^{P} \alpha_{2p+1,1} x_{1,ip,p}(n), \ldots, k_{2,N_t} x_{N_t}^*(n) \right.$$
$$\left. + \sum_{p=1}^{P} \alpha_{2p+1,N_t} x_{N_t,ip,p}(n) \right]^T,$$

$$\boldsymbol{x}(n) = \boldsymbol{x}_i(n) + \boldsymbol{x}_{dist}(n),$$
$$\boldsymbol{s}(n) = [s_1(n), s_2(n), \ldots, s_{N_t}(n)]^T. \qquad (12)$$

We then group the channel matrices $\boldsymbol{H}^{(i)}(l)$ and $\boldsymbol{H}^{(s)}(l)$ in one $N_r \times 2N_t$ matrix $\boldsymbol{H}(l) = [\boldsymbol{H}^{(i)}(l), \boldsymbol{H}^{(s)}(l)]$ and gather all the channel coefficients in the following $N_r M \times 2N_t N$ block Toeplitz matrix:

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}(0) & \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{H}(L) & \ldots & \boldsymbol{H}(1) \\ \boldsymbol{H}(1) & \boldsymbol{H}(0) & & & & \ddots & \vdots \\ \vdots & \boldsymbol{H}(1) & \ddots & & & & \boldsymbol{H}(L) \\ \boldsymbol{H}(L) & \vdots & & \ddots & & & \boldsymbol{0} \\ & \boldsymbol{H}(L) & & & & & \boldsymbol{H}(0) \\ \boldsymbol{0} & & \ddots & & & & \boldsymbol{H}(1) \\ \vdots & & & & \ddots & & \vdots \\ \boldsymbol{0} & & \ldots & & \boldsymbol{0} & & \boldsymbol{H}(L) \end{pmatrix}. \qquad (13)$$

The received OFDM block on the N_r antennas is:

$$\boldsymbol{y} = [\boldsymbol{y}^T(0), \boldsymbol{y}^T(1), \ldots, \boldsymbol{y}^T(M-1)]^T = \boldsymbol{H}\boldsymbol{u} + \boldsymbol{w}, \quad (14)$$

where $M = N + L$, the $2N_t N \times 1$ data vector $\boldsymbol{u}$ is given by

$$\boldsymbol{u} = [\boldsymbol{x}^T(0), \boldsymbol{s}^T(0), \ldots, \boldsymbol{x}^T(N-1), \boldsymbol{s}^T(N-1)]^T, \quad (15)$$

and

$$\boldsymbol{w} = [\boldsymbol{w}^T(0), \boldsymbol{w}^T(1), \ldots, \boldsymbol{w}^T(M-1)]^T. \qquad (16)$$

For multi-block transmission, the received vector in (14) is indexed by the block number t, i.e., $\boldsymbol{y}_t$. For convenience, we omit this indexation and we will consider later a given number of transmitted blocks to compute the covariance matrix of the received vector.

3 Subspace-based channel estimator

We propose to apply a subspace-based algorithm to jointly estimate the SI and intended channel coefficients along with the nonlinear coefficients. Subspace methods rely on the orthogonality property between the signal and noise subspaces. These two subspaces are obtained from eigendecomposition of the covariance matrix of the received signal y. Denoting by R_u, the covariance of u, the covariance matrix R_y of the received vector y is given by

$$R_y = HR_uH^H + \sigma^2 I_{MN_r}, \tag{17}$$

as long as the signal samples are uncorrelated from the noise samples[2].

The signal subspace is spanned by the columns of the matrix H. Noting that the columns of H are, by construction, linearly independent as soon as there exists an $l \in [0, L]$ such that $H(l)$ is full rank[3], the matrix H is a full-rank matrix. Therefore, the dimension of the signal subspace is $2NN_t$. It follows that, to obtain a nondegenerate noise subspace, its dimension $N_rM - 2N_tN$ should be larger than zero, and thus, the number of receiving antennas should be larger than the number of transmitting antennas to make the subspace method work, and in [5], we developed the linear subspace algorithm for this setting. In the following, we will develop the subspace-based algorithm for general numbers of transmit and receive antennas. When $N_t = N_r$, the matrix R_y cannot be directly used to find the noise subspace. As an alternative different approach, we consider the augmented received vector as

$$\widetilde{y} = \begin{pmatrix} y \\ y^* \end{pmatrix} = \begin{pmatrix} H & 0 \\ 0 & H^* \end{pmatrix} \begin{pmatrix} u \\ u^* \end{pmatrix} + \begin{pmatrix} w \\ w^* \end{pmatrix}. \tag{18}$$

The use of the augmented received vector is usually referred as widely linear processing. In this case, the augmented covariance matrix $R_{\widetilde{y}}$ of $\widetilde{y}$ has the following structure:

$$R_{\widetilde{y}} = \widetilde{H}R_{\widetilde{u}}\widetilde{H}^H + \sigma^2 I_{2MN_r}, \tag{19}$$

where $R_{\widetilde{u}}$ denotes the covariance matrix of the augmented transmit signal $\widetilde{u} = \begin{pmatrix} u \\ u^* \end{pmatrix}$ and

$$\widetilde{H} = \begin{pmatrix} H & 0 \\ 0 & H^* \end{pmatrix}. \tag{20}$$

It is worth mentioning that the proper noise has a vanishing pseudo-covariance [34]. The main purpose of using the extended received signal is to increase the dimension of the received signal and thus avoid the degenerate noise subspace. Hence, the subspace identification procedure can be derived only if the signal part covariance matrix, given by $\widetilde{H}R_{\widetilde{u}}\widetilde{H}^H$, of the covariance matrix $R_{\widetilde{y}}$ is singular. It results that $d_s = \text{rank}(\widetilde{H}R_{\widetilde{u}}\widetilde{H}^H) < 2MN_r$. In this case, the signal is confined in a d_s-dimensional subspace and the remaining noise subspace is with dimension $2MN_r - d_s$.

Singularity of $R_{\widetilde{u}}$ is a necessary condition to obtain a non-degenerate noise subspace. Actually, noting that $\widetilde{H}$ is full rank, nonsingular $R_{\widetilde{u}}$ results in rank$(\widetilde{H}R_{\widetilde{u}}\widetilde{H}^H) = 2MN_r$, and thus, the matrix $\widetilde{H}R_{\widetilde{u}}\widetilde{H}^H$ spans all the observation space. On the other hand, since the matrix $\widetilde{H}$ is a tall matrix, singularity of $R_{\widetilde{u}}$ is not a sufficient condition to guarantee the singularity of $\widetilde{H}R_{\widetilde{u}}\widetilde{H}^H$.

The matrix $R_{\widetilde{u}}$ can be expressed in a block form in terms of the covariance matrix of u, $R_u = E(uu^H)$, the pseudo-covariance matrix $C_u = E(uu^T)$ and their complex conjugates as

$$R_{\widetilde{u}} = \begin{pmatrix} R_u & C_u \\ C_u^* & R_u^* \end{pmatrix}. \tag{21}$$

In the following, we distinguish two cases of real and complex modulated symbols.

For real modulated symbols, it can be shown that $R_{\widetilde{u}} = \alpha^2 M \otimes I_{2N_t}$ with the $2N \times 2N$ matrix M having the following form:

$$M = \begin{pmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \ddots & & & 0 & & & 1 \\ \vdots & & \ddots & & \vdots & & \iddots & \\ 0 & \dots & 0 & 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & & & 1 & 0 & \ddots & & \\ \vdots & & \iddots & & \vdots & & \ddots & \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 & 1 \end{pmatrix}. \tag{22}$$

From (22), we note that each column of M appears exactly two times (the first column of M is the same as the $(N + 1)^{th}$ column, and the ith column of M is the same as the $(2N-i+2)^{th}$ column, for $i = 2, \dots, N$). Therefore, the matrix M has exactly N-independent columns and thus its rank is N. It follows that the rank of $R_{\widetilde{u}}$ is $2NN_t$. In Appendix 1, we show that $R_{\widetilde{u}}$ has zero eigenvalue with multiplicity $2NN_t$ and $2\alpha^2$ also with multiplicity $2NN_t$. Then, the matrix $R_{\widetilde{u}}$ is decomposed as UDU^H where D is the $4NN_t \times 4NN_t$ diagonal matrix with zeroes in the first $2NN_t$ diagonal elements and $2\alpha^2$ in the last $2NN_t$ diagonal elements and U is an orthogonal matrix whose columns are the corresponding eigenvectors of $R_{\widetilde{u}}$.

For complex symbols, the pseudo-covariance matrix C_u is generally equal to the zero matrix, which makes the matrix $R_{\widetilde{u}}$ of full rank. To avoid this problem, we apply a simple precoding at the input of the IFFT. It transforms the data symbol X_q to

$$\widetilde{X}_q = PX_q + QX_q^*. \tag{23}$$

where P and Q are two matrices. By combining the data symbol X_q and its complex conjugate, we force the pseudo-covariance matrix to be different from zero. Appendix 2 gives a detailed discussion about the choice of

the matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ so that the covariance matrix $\boldsymbol{R_{\widetilde{u}}}$ has rank $2NN_t$ and can be decomposed as $\boldsymbol{UDU}^H$ with $\boldsymbol{D}$ as the $4NN_t \times 4NN_t$ diagonal matrix with zeroes in the first $2NN_t$ diagonal elements.

The noise subspace is the span of the $p = 2MN_r - 2NN_t$ eigenvectors of $\boldsymbol{R_{\widetilde{y}}}$ corresponding to the smallest eigenvalue σ^2, and the columns of $\widetilde{\boldsymbol{H}}\boldsymbol{R_{\widetilde{u}}}\widetilde{\boldsymbol{H}}^H$ belong to the signal subspace. Due to the orthogonality between the signal and the noise subspaces, each column of $\widetilde{\boldsymbol{H}}\boldsymbol{R_{\widetilde{u}}}\widetilde{\boldsymbol{H}}^H$ is orthogonal to any vector in the noise subspace. Let $\{\boldsymbol{v}_i\}_{i=1}^p$ denote the p co-orthogonal eigenvectors corresponding to the smallest eigenvalue of $\boldsymbol{R_{\widetilde{y}}}$. Then we have the following set of equations:

$$\boldsymbol{v}_i^H \widetilde{\boldsymbol{H}}\boldsymbol{R_{\widetilde{u}}}\widetilde{\boldsymbol{H}}^H = \boldsymbol{0}, \ i = 1, 2, \ldots, p. \tag{24}$$

From (24), we conclude that $\boldsymbol{v}_i$ spans the left null space of $\widetilde{\boldsymbol{H}}\boldsymbol{R_{\widetilde{u}}}\widetilde{\boldsymbol{H}}^H$. For convenience, $\boldsymbol{U}$ is written as a block of 4 $2NN_t \times 2NN_t$ matrices:

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \\ \boldsymbol{U}_3 & \boldsymbol{U}_4 \end{pmatrix}, \tag{25}$$

where the columns of $[\boldsymbol{U}_1^T, \ \boldsymbol{U}_3^T]^T$ are the eigenvectors of $\boldsymbol{R_{\widetilde{u}}}$ corresponding to the eigenvalue zero and the columns of $[\boldsymbol{U}_2^T, \ \boldsymbol{U}_4^T]^T$ are the other eigenvectors. Then, taking into account the eigenvalue decomposition of $\boldsymbol{R_{\widetilde{u}}}$, the set of equations in (24) are equivalent to

$$\boldsymbol{v}_i^H \begin{pmatrix} \boldsymbol{H}\boldsymbol{U}_2 \\ \boldsymbol{H}^*\boldsymbol{U}_4 \end{pmatrix} = \boldsymbol{0}, \ i = 1, 2, \ldots, p. \tag{26}$$

By dividing $\boldsymbol{v}_i$ into two $MN_r \times 1$ vectors, i.e., $\boldsymbol{v}_i = [\boldsymbol{v}_{i,1}^T, \ \boldsymbol{v}_{i,2}^T]^T$, (26) is rewritten as

$$\boldsymbol{v}_{i,1}^H \boldsymbol{H}\boldsymbol{U}_2 + \boldsymbol{v}_{i,2}^H \boldsymbol{H}^*\boldsymbol{U}_4 = \boldsymbol{0}, \tag{27}$$

for $i = 1, 2, \ldots, p$. The matrix $\boldsymbol{H}$ is completely defined by the set of matrices $\boldsymbol{H}(l)$, for $l = 0, 1, \ldots, L$. Therefore, the specific structure of $\boldsymbol{H}$ should be taken into consideration when solving the equations in (27) to obtain a more accurate estimate of the channels. To that end, we divide the two vectors $\boldsymbol{v}_{i,1}$ and $\boldsymbol{v}_{i,2}$ as follows:

$$\boldsymbol{v}_{i,j} = \left[\boldsymbol{v}_{i,j}^T(M), \ \boldsymbol{v}_{i,j}^T(M-1), \ldots, \ \boldsymbol{v}_{i,j}^T(1) \right]^T,$$
$$j = 1, 2, \ i = 1, 2, \ldots, p, \tag{28}$$

where each $\boldsymbol{v}_{i,j}(n)$, for $n = 1, 2, \ldots, M$, is a $N_r \times 1$ vector. From (13) and (28), each term $\boldsymbol{v}_{i,1}^H \boldsymbol{H}$ in (27) is rewritten as

$$\sum_{l=0}^{L} \boldsymbol{v}_{i,1}^H(n+L-l)\boldsymbol{H}(l) + \sum_{l=n}^{L} \boldsymbol{v}_{i,1}^H(M-l+n)\boldsymbol{H}(l),$$
$$\text{for } n = 1, \ldots, L,$$

$$\sum_{l=0}^{L} \boldsymbol{v}_{i,1}^H(n+L-l)\boldsymbol{H}(l), \text{for } n = L+1, \ldots, M,$$
$$\tag{29}$$

and $\boldsymbol{v}_{i,2}^H \boldsymbol{H}^*$ can also be partitioned in the same manner. By introducing $\check{\boldsymbol{h}}(l) = \text{vect}(\boldsymbol{H}(l))$ and $\boldsymbol{V}_{i,j}(n) = \boldsymbol{I}_{2N_t} \otimes \boldsymbol{v}_{i,j}^H(n)$, for $i = 1, \ldots, p$ and $j = 1, 2$, it is easy to verify that $\boldsymbol{v}_{i,j}^H(n)\boldsymbol{H}(l) = \check{\boldsymbol{h}}^T(l)\boldsymbol{V}_{i,j}^T(n)$. Let us denote the $2NN_t \times 2N_tN_r(L+1)$ matrices $\boldsymbol{V}_{i,j}$, for $j = 1, 2$, as

$$\boldsymbol{V}_{i,j} = \begin{pmatrix} \boldsymbol{V}_{i,j}(L+1) & \boldsymbol{V}_{i,j}(L) & \ldots & \boldsymbol{V}_{i,j}(1) \\ \boldsymbol{V}_{i,j}(L+2) & \boldsymbol{V}_{i,j}(L+1) & \ldots & \boldsymbol{V}_{i,j}(2) \\ \boldsymbol{V}_{i,j}(L+3) & \boldsymbol{V}_{i,j}(L+2) & \ldots & \boldsymbol{V}_{i,j}(3) \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{V}_{i,j}(N+L) & \boldsymbol{V}_{i,j}(N+L-1) & \ldots & \boldsymbol{V}_{i,j}(N) \end{pmatrix}$$
$$+ \begin{pmatrix} \boldsymbol{0} & \boldsymbol{V}_{i,j}(N+L) & \ldots & \boldsymbol{V}_{i,j}(N+1) \\ & & \ddots & \vdots \\ \vdots & & & \boldsymbol{V}_{i,j}(N+L) \\ \vdots & & \boldsymbol{0} & \\ \vdots & & & \\ \boldsymbol{0} & & \boldsymbol{0} & \end{pmatrix}, \tag{30}$$

and $\check{\boldsymbol{h}} = [\check{\boldsymbol{h}}^T(0), \ \check{\boldsymbol{h}}^T(1), \ldots, \ \check{\boldsymbol{h}}^T(L)]^T$. Then, using the previous notations, (27) is rearranged to obtain

$$\check{\boldsymbol{h}}^T \boldsymbol{V}_{i,1}^T \boldsymbol{U}_2 + \check{\boldsymbol{h}}^H \boldsymbol{V}_{i,2}^T \boldsymbol{U}_4 = \boldsymbol{0}, \tag{31}$$

or, by taking the transpose of the previous equation:

$$\boldsymbol{U}_2^T \boldsymbol{V}_{i,1} \check{\boldsymbol{h}} + \boldsymbol{U}_4^T \boldsymbol{V}_{i,2} \check{\boldsymbol{h}}^* = \boldsymbol{0}, \tag{32}$$

for $i = 1, 2, \ldots, p$. Note that the difference between (27) and (32) is that (32) takes into account the Toeplitz blocks structure of $\boldsymbol{H}$. Now, collecting all the previous equations, we obtain

$$\boldsymbol{\Theta}_1 \check{\boldsymbol{h}} + \boldsymbol{\Theta}_2 \check{\boldsymbol{h}}^* = \boldsymbol{0}, \tag{33}$$

where

$$\boldsymbol{\Theta}_1 = \left[\left(\boldsymbol{U}_2^T \boldsymbol{V}_{1,1}\right)^T, \ \left(\boldsymbol{U}_2^T \boldsymbol{V}_{2,1}\right)^T, \ldots, \ \left(\boldsymbol{U}_2^T \boldsymbol{V}_{p,1}\right)^T \right]^T,$$
$$\boldsymbol{\Theta}_2 = \left[\left(\boldsymbol{U}_4^T \boldsymbol{V}_{1,2}\right)^T, \ \left(\boldsymbol{U}_4^T \boldsymbol{V}_{2,2}\right)^T, \ldots, \ \left(\boldsymbol{U}_4^T \boldsymbol{V}_{p,2}\right)^T \right]^T. \tag{34}$$

Separating the real and imaginary parts of (33), we have

$$\underbrace{\begin{pmatrix} \Re(\boldsymbol{\Theta}_1 + \boldsymbol{\Theta}_2) & \Im(-\boldsymbol{\Theta}_1 + \boldsymbol{\Theta}_2) \\ \Im(\boldsymbol{\Theta}_1 + \boldsymbol{\Theta}_2) & \Re(\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2) \end{pmatrix}}_{\overline{\boldsymbol{\Theta}}} \underbrace{\begin{pmatrix} \Re(\check{\boldsymbol{h}}) \\ \Im(\check{\boldsymbol{h}}) \end{pmatrix}}_{\overline{\boldsymbol{h}}} = \boldsymbol{0}. \tag{35}$$

From (35), the vector $\overline{\boldsymbol{h}}$ belongs to the right null space of $\overline{\boldsymbol{\Theta}}$. In practice, $\overline{\boldsymbol{h}}$ is a linear combination of the $4N_tN_r$ right singular vectors of the matrix $\overline{\boldsymbol{\Theta}}$, denoted by $\boldsymbol{\beta}_i$, which are equal to the eigenvector of the Gramian $\overline{\boldsymbol{\Theta}}\overline{\boldsymbol{\Theta}}^H$

corresponding to the zero eigenvalue. Therefore, an estimate of $\overline{h}$ is given by

$$\widehat{\overline{h}} = \overline{\Phi}c, \tag{36}$$

where $\overline{\Phi} = [\beta_1, \beta_2, \ldots, \beta_{4N_tN_r}]$, and the $4N_tN_r \times 1$ vector c represents the ambiguity term to be estimated. The complex channel vector can also be obtained as

$$\widehat{\overline{h}} = \Phi c, \tag{37}$$

where Φ is obtained by combining the lines of $\overline{\Phi}$ in the following way:

$$\overline{\Phi} = \begin{pmatrix} \overline{\Phi}_{real} \\ \overline{\Phi}_{imag} \end{pmatrix} \rightarrow \Phi = \overline{\Phi}_{real} + j\overline{\Phi}_{imag}, \tag{38}$$

and j is the complex number satisfying $j^2 = -1$.

We mention that the matrices U_2 and U_4 do not depend on the received signal and can be computed offline prior to the transmission. It is also seen that the overestimated channel order L does not affect the estimation process. This is a common property with other subspace-based estimators [17].

4 Resolving the ambiguity term

As mentioned above, the subspace that contains the channels is obtained and the ambiguity term needs to be estimated to extract the exact coefficients. Different approaches can be applied to solve the ambiguity term c. To do so, we highlight the contribution of c on the received vector y. First, we separate the matrix Φ in two $N_tN_r(L+1) \times 4N_tN_r$ matrices Φ_i and Φ_s which contribute in the SI and intended channels, respectively (i.e., $\check{h}^{(i)} = \Phi_i c$ and $\check{h}^{(s)} = \Phi_s c$). By rearranging the elements of Φ_i as

$$\Phi_i = \begin{pmatrix} \Phi_{i,1}(0) \\ \Phi_{i,2}(0) \\ \vdots \\ \Phi_{i,N_t}(0) \\ \vdots \\ \Phi_{i,1}(L) \\ \Phi_{i,2}(L) \\ \vdots \\ \Phi_{i,N_t}(L) \end{pmatrix} \rightarrow \check{\Phi}_i = \begin{pmatrix} \Phi_{i,1}(0) & \ldots & \Phi_{i,N_t}(0) \\ \Phi_{i,1}(1) & \ldots & \Phi_{i,N_t}(1) \\ \vdots & & \vdots \\ \Phi_{i,1}(L) & \ldots & \Phi_{i,N_t}(L) \end{pmatrix}, \tag{39}$$

where each $\Phi_{i,q}(l)$ is a $N_r \times 4N_tN_r$ matrix, $\check{H}^{(i)} = [H^{(i)^T}(0), H^{(i)^T}(1), \ldots, H^{(i)^T}(L)]^T$ can be written as

$$\check{H}^{(i)} = \check{\Phi}_i(I_{N_t} \otimes c), \tag{40}$$

and $\check{H}^{(s)} = [H^{(s)^T}(0), H^{(s)^T}(1), \ldots, H^{(s)^T}(L)]^T$ can be also written as $\check{H}^{(s)} = \check{\Phi}_s(I_{N_t} \otimes c)$, where $\check{\Phi}_s$ is defined

in the same way as $\check{\Phi}_i$. $\check{H}^{(i)}$ and $\check{\Phi}_i$ are used to build the matrices $H^{(i)}$ and Ψ_i, respectively, having the same block structure as H in (13).

Next, we define the diagonal matrices K and A_p whose diagonal elements are $k = [k_{2,1}, \ldots, k_{2,N_t}]^T$ and $\alpha_p = [\alpha_{2p+1,1}, \ldots, \alpha_{2p+1,N_t}]^T$, respectively, and we denote by $x_{ip,p}(n) = [x_{1,ip,p}(n), \ldots, x_{N_t,ip,p}(n)]^T$, and $x_{ip,p} = [x_{ip,p}^T(0), \ldots, x_{ip,p}^T(N-1)]^T$. Using the previous notations and by developing $x = x_i + (I_N \otimes K)x_i^* + \sum_{p=1}^{P}(I_N \otimes A_p)x_{ip,p}$ in terms of the transmitter impairments, one can express the received signal in (14) as

$$y = \underbrace{\Psi_i(I_{NN_t} \otimes c)}_{H^{(i)}} x + \underbrace{\Psi_s(I_{NN_t} \otimes c)}_{H^{(s)}} s + w,$$

$$= \Psi_i(I_{NN_t} \otimes c)\left(x_i + (I_N \otimes K)x_i^* + \sum_{p=1}^{P}(I_N \otimes A_p)x_{ip,p}\right)$$

$$+ \Psi_s(I_{NN_t} \otimes c)s + w, \tag{41}$$

where Ψ_s and $H^{(s)}$ are defined in the same way as Ψ_i and $H^{(i)}$, respectively, and $s = [s^T(0), \ldots, s^T(N-1)]^T$. After some manipulations, one can easily verify that $(I_{NN_t} \otimes c)x_i = (x_i \otimes I_{4N_tN_r})c$ and $(I_{NN_t} \otimes c)s = (s \otimes I_{4N_tN_r})c$. Then, the received vector in (41) is rewritten as

$$y = \Psi_i\left(\left(x_i + (I_N \otimes K)x_i^* + \sum_{p=1}^{P}(I_N \otimes A_p)x_{ip,p}\right) \otimes I_{4N_tN_r}\right)c$$

$$+ \Psi_s\left(s \otimes I_{4N_tN_r}\right)c + w. \tag{42}$$

In (42), the received vector y is expressed as a linear function of the unknown vector c. This formulation makes the estimation of c more tractable. While the transmitted SI is known, the distorted parts $(I_N \otimes A_p)x_{ip,p}$ and $(I_N \otimes K)x_i^*$ of the SI from the cascade of the IQ mixer and PA need to be estimated. We begin by writing the following cost function $f(c, s, K, A_p) = ||y - \Psi_i((x_i + (I_N \otimes K)x_i^* + \sum_{p=1}^{P}(I_N \otimes A_p)x_{ip,p}) \otimes I_{4N_tN_r})c - \Psi_s(s \otimes I_{4N_tN_r})c||^2$ depending on c, K, A_p (for $p = 1, \ldots, P$) and s. Given an initial estimate $\widehat{c}$ of c, the minimization of $f(\widehat{c}, s, K, A_p)$ with respect to s, K and A_p can be recast as a least square (LS) problem. Then, using the solutions $\widehat{s}, \widehat{K}$ and $\widehat{A}_p$, we minimize $f(c, \widehat{s}, \widehat{K}, \widehat{A}_p)$ with respect to c. We iterate this procedure until the estimated parameters converge. An initial estimate of c is obtained using the LS criteria as

$$\widehat{c}_0 = (\Psi_i(x_i \otimes I_{4N_tN_r}))^{\#}y, \tag{43}$$

where the operator $(\cdot)^{\#}$ returns the pseudo-inverse of a given matrix. At the kth iteration, the estimate $\widehat{c}_{k-1}$

obtained at the previous iteration is used to find s, K and A_p (or equivalently k and α_p) as follows:

$$
\begin{pmatrix} \widehat{s}_k \\ \widehat{k}_k \\ \widehat{\alpha}_{1,k} \\ \vdots \\ \widehat{\alpha}_{P,k} \end{pmatrix} = \left[\boldsymbol{\Psi}_s \widehat{\boldsymbol{C}}_{k-1}, \ \boldsymbol{\Psi}_i \left(\mathrm{diag}\left(x_i^* \right) \boldsymbol{B} \right) \right.
$$

$$
\otimes \widehat{c}_{k-1}, \ \boldsymbol{\Psi}_i \left(\mathrm{diag}\left(x_{ip,1} \right) \boldsymbol{B} \right) \otimes \widehat{c}_{k-1}, \dots ,
$$

$$
\left. \boldsymbol{\Psi}_i \left(\mathrm{diag}\left(x_{ip,P} \right) \boldsymbol{B} \right) \otimes \widehat{c}_{k-1} \right]^{\#} \left(\boldsymbol{y} - \boldsymbol{\Psi}_i \widehat{\boldsymbol{C}}_{k-1} \boldsymbol{x}_i \right),
$$

$$(44)$$

where, for clarity, we introduce $\boldsymbol{B} = \boldsymbol{1}_N \otimes \boldsymbol{I}_{N_t}$ and $\widehat{\boldsymbol{C}}_{k-1} = \boldsymbol{I}_{NN_t} \otimes \widehat{c}_{k-1}$ and we use the equality $\left(((\boldsymbol{I}_N \otimes \boldsymbol{K}) \boldsymbol{x}_i^*) \otimes \boldsymbol{I}_{4N_tN_r} \right) \boldsymbol{c} = \left((\mathrm{diag}(\boldsymbol{x}_i^*) \boldsymbol{B}) \otimes \boldsymbol{c} \right) \boldsymbol{k}$. Then, $\widehat{s}_k$ is transformed in the frequency domain and each element of the frequency domain vector is projected to its closest discrete constellation point. The obtained vector is converted back to the time domain to obtain a better estimate $\widetilde{s}_k$ of s.

Then, an update of c at iteration k is obtained as:

$$
\widehat{c}_k = \left(\boldsymbol{\Psi}_i \left(\left(\boldsymbol{x}_i + (\boldsymbol{I}_N \otimes \widehat{\boldsymbol{K}}_k) + \sum_{p=1}^{P} \left(\boldsymbol{I}_N \otimes \widehat{\boldsymbol{A}}_{p,k} \right) \boldsymbol{x}_{ip,p} \right) \otimes \boldsymbol{I}_{4N_tNr} \right) \right.
$$

$$
\left. + \boldsymbol{\Psi}_s \left(\widetilde{s}_k \otimes \boldsymbol{I}_{4N_tN_r} \right) \right)^{\#} \boldsymbol{y}.
$$

$$(45)$$

If a set of P_{pilot}, pilot symbols are available at subcarriers indexed by $\mathcal{P} = \{ p_1, \dots, p_{P_{\mathrm{pilot}}} \}$, the intended transmit signal at antenna q can be represented as the sum of two signals:

$$
s_q^p(n) = \sum_{i=1}^{P_{\mathrm{pilot}}} S_q(p_i) e^{j2\pi p_i n/N},
$$

$$
s_q^d(n) = \sum_{k \notin \mathcal{P}} S_q(k) e^{j2\pi kn/N},
$$

$$(46)$$

where the first sequence $s_q^p(n)$ contains the pilot symbols and the second sequence $s_q^d(n)$ contains the unknown data symbols transmitted by other intended transmitter. Then, the received vector in (42) is rearranged as follows:

$$
\boldsymbol{y} = \boldsymbol{\Psi}_i \left(\left(\boldsymbol{x}_i + (\boldsymbol{I}_N \otimes \boldsymbol{K}) \boldsymbol{x}_i^* + \sum_{p=1}^{P} \left(\boldsymbol{I}_N \otimes \boldsymbol{A}_p \right) \boldsymbol{x}_{ip,p} \right) \otimes \boldsymbol{I}_{4N_tN_r} \right)
$$

$$
\boldsymbol{c} + \boldsymbol{\Psi}_s \left(\left(\boldsymbol{s}^p + \boldsymbol{s}^d \right) \otimes \boldsymbol{I}_{4N_tN_r} \right) \boldsymbol{c} + \boldsymbol{w}.
$$

$$(47)$$

where s^p and s^d are constructed in the same way as s and contain the pilot symbols and unknown symbols,

respectively. The initial estimate of c is modified to incorporate the pilot symbols as

$$
\widehat{c}_0 = \left(\boldsymbol{\Psi}_i (\boldsymbol{x}_i \otimes \boldsymbol{I}_{4N_tNr}) + \boldsymbol{\Psi}_s (\boldsymbol{s}^p \otimes \boldsymbol{I}_{4N_tN_r}) \right)^{\#} \boldsymbol{y}, \quad (48)
$$

and the estimates of s^d, K and A_p at iteration k are given by

$$
\begin{pmatrix} \widehat{s}_k^d \\ \widehat{k}_k \\ \widehat{\alpha}_{1,k} \\ \vdots \\ \widehat{\alpha}_{P,k} \end{pmatrix} = \left[\boldsymbol{\Psi}_s \widehat{\boldsymbol{C}}_{k-1}, \ \boldsymbol{\Psi}_i \left(\mathrm{diag}(x_i^*) \boldsymbol{B} \right) \right.
$$

$$
\otimes \widehat{c}_{k-1}, \ \boldsymbol{\Psi}_i \left(\mathrm{diag}(x_{ip,1}) \boldsymbol{B} \right) \otimes \widehat{c}_{k-1}, \dots ,
$$

$$
\boldsymbol{\Psi}_i \left(\mathrm{diag}(x_{ip,P}) \boldsymbol{B} \right) \otimes \widehat{c}_{k-1} \bigg]^{\#}
$$

$$
\times \left(\boldsymbol{y} - \boldsymbol{\Psi}_i \widehat{\boldsymbol{C}}_{k-1} \boldsymbol{x}_i - \Psi_s \widehat{\boldsymbol{C}}_{k-1} \boldsymbol{s}^p \right).
$$

$$(49)$$

As before, $\widehat{s}_k^d$ is converted to the frequency domain, demodulated then transformed to the time domain to obtain $\widetilde{s}_k^d$. The updated estimate of c at iteration k is obtained as:

$$
\widehat{c}_k = \left(\boldsymbol{\Psi}_i \left(\left(\boldsymbol{x}_i + (\boldsymbol{I}_N \otimes \widehat{\boldsymbol{K}}_k) \boldsymbol{x}_i^* + \sum_{p=1}^{P} \left(\boldsymbol{I}_N \otimes \widehat{\boldsymbol{A}}_{3,p} \right) \boldsymbol{x}_{ip,p} \right) \otimes \boldsymbol{I}_{4N_tN_r} \right) \right.
$$

$$
\left. + \boldsymbol{\Psi}_s \left(\left(\boldsymbol{s}^p + \widetilde{s}_k^d \right) \otimes \boldsymbol{I}_{4N_tN_r} \right) \right)^{\#} \boldsymbol{y}.
$$

$$(50)$$

In the following, we summarize the different steps of the proposed algorithm:

1. Compute the augmented covariance matrix $\boldsymbol{R}_{\widetilde{y}}$ by time averaging of T received samples as:

$$
\widehat{\boldsymbol{R}}_{\widetilde{y}} = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_t^* \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_t^* \end{pmatrix}^H
$$

2. Perform eigendecomposition of $\boldsymbol{R}_{\widetilde{y}}$ and take the p eigenvectors $\boldsymbol{v}_i$ corresponding to the smallest eigenvalue of $\boldsymbol{R}_{\widetilde{y}}$.
3. Construct the matrix $\overline{\boldsymbol{\Theta}}$ from $\boldsymbol{v}_i$ and compute the $4N_tN_r$ singular vectors of $\overline{\boldsymbol{\Theta}}$ corresponding to the zero singular value to form $\overline{\boldsymbol{\Phi}}$.
4. Build the matrices $\check{\boldsymbol{\Phi}}_i$ and $\check{\boldsymbol{\Phi}}_s$ as given in (39).
5. Estimate the ambiguity vector c by iterating between (44) and (45) if no pilot symbols are available or between (49) and (50) if a set of pilot symbols are available from the intended transceiver.

5 Simulation results

In this section, we provide some simulation results on the performance of the proposed estimation algorithm for a 2×2 MIMO full-duplex system. The transmitted bits are mapped to 4-QAM symbols, then passed through an OFDM modulator of length $N = 64$. The wireless channel is represented as a Rayleigh multipath fading channel with five equal-variance resolvable paths. Since the exact number of paths is supposed to be unknown, the algorithm is parametrized as if there are eight paths. In the following, the SNR is defined as the average intended-signal-to-thermal noise power ratio and the estimation mean square error (MSE) of H is $\text{MSE} = E\left(||H - \widehat{H}||^2\right)$. To model the RF impairments, a complete transmission chain is simulated. The PA coefficients are derived from the intercept points by taking the IIP3 = 20 dBm. For the IQ mixer, the ratio between the direct signal and the image is set to 28 dB which is specified in 3GPP LTE specifications [23]. The ADC is modelled as a 14-bit quantizer to incorporate the quantization noise. Therefore, no simplifications are made regarding the different impairments. Antenna separation can attenuate the SI by 40 dB while the RF cancellation stage reduces the direct path by 30 dB [1] leaving the weaker reflections and transceiver impairments to be reduced by the proposed digital algorithm.

The proposed algorithm is compared to different channel estimators: the least square (LS) and the maximum likelihood (ML) algorithms. For the LS estimator, the channel coefficients are obtained using the known *self* signal and the pilot symbols in the intended signal. It simply considers the unknown symbols as additive noise. The ML estimate is obtained by maximizing the following cost function:

$$L\left(H^{(i)}, H^{(s)}\right) = \log\left(\det(R)\right) - \left(y - H^{(i)}x - H^{(s)}s^p\right)^H$$
$$\times R^{-1}\left(y - H^{(i)}x - H^{(s)}s^p\right),$$

where $R = \alpha^2 H^{(s)H}H^{(s)} + \sigma^2 I_{N_r M}$. An iterative procedure to find the ML estimate was proposed in [35]. The covariance matrix is obtained by averaging 60 OFDM blocks. Figures 2 and 3 plot the MSE vs. SNR curves for the SI and intended channel estimations, respectively. In both figures, one pilot symbol, from the intended transceiver, is used to solve the ambiguity matrix. For comparison purpose, a perfect estimate of the ambiguity term c is obtained as $c_{perfect} = \arg\min_c ||\check{h} - \Phi c||_2^2$ and the corresponding curves are labelled by clairvoyant subspace. It is seen that, when one pilot symbol is used in the ML and LS estimators, the proposed subspace algorithm offers notably lower MSE over a large SNR range. We also represent the performance of the ML and LS estimators when 20% of the transmit symbols are known (pilot symbols equally spaced within one OFDM symbol) while keeping
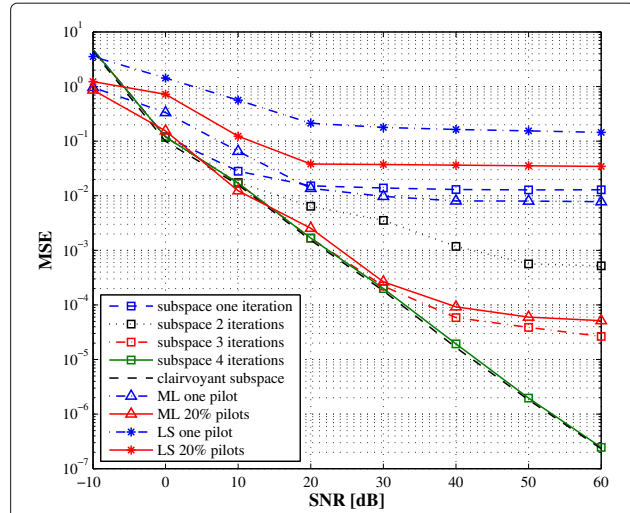
Fig. 2 SI channel estimation MSE vs. SNR with 60 received OFDM symbols

one pilot symbol for the subspace method[4]. In this case, the three algorithms give comparable performance at low SNR region with the expanse of lower bandwidth efficiency. As the SNR increases, the performance of the LS and ML estimators saturate due to the reduced number of pilot symbols and the presence of the unknown transmit signal from the intended transceiver which acts as an additive noise. While the subspace algorithm exploits the information bearing in the unknown data to find the signal subspace. The ambiguity term is first solved using the known transmit symbols, then the iterative decoding ambiguity estimation is applied to improve the estimation performance. From Figs. 2 and 3, three to four iterations
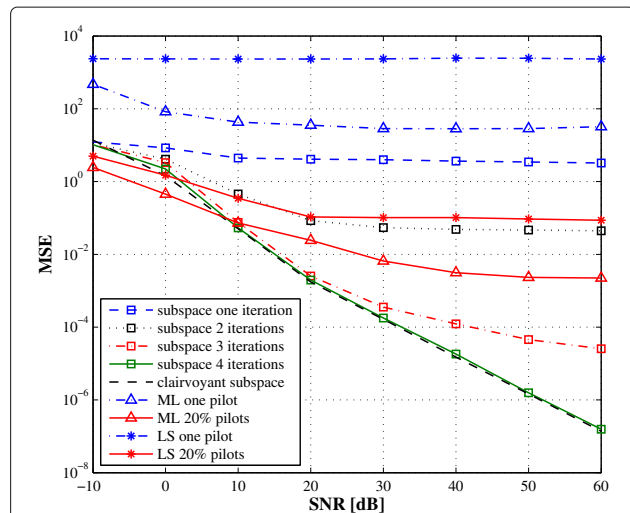
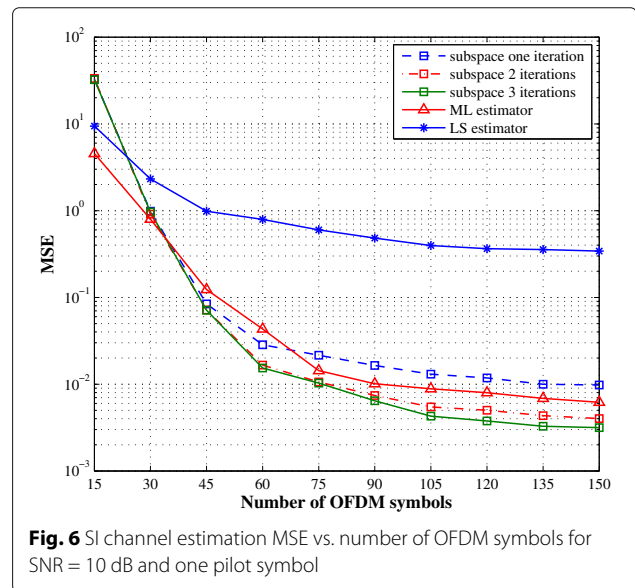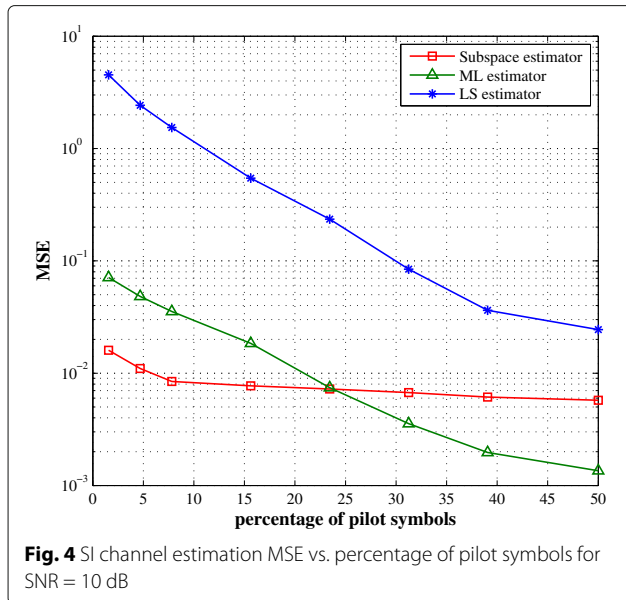**Fig. 3** Intended channel estimation MSE vs. SNR with 60 received OFDM symbols

are sufficient to converge and the performance is close to the performance when the ambiguity term **c** is perfectly obtained. Note that the ML solution is also obtained in an iterative way and for a fair comparison; we simulate the performance of the ML estimator after four iterations. As it can be expected, the estimate of the SI channel is more accurate than the estimate of the intended channel. This can be explained by the fact that the self-signal is known while one pilot symbol is known in the intended signal.
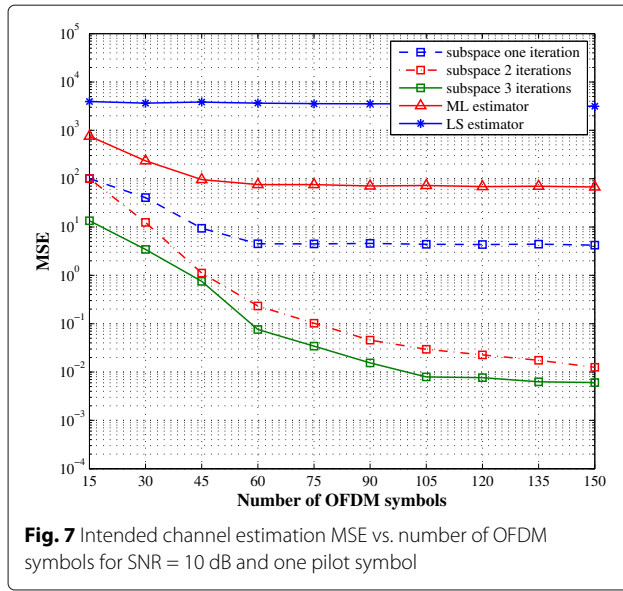
The number of pilot symbols is a critical issue in channel estimation since a large pilot sequence provides better estimation performance but reduces the bandwidth efficiency of the system. In Figs. 4 and 5, we compare the impact of the number of pilot symbols on the performance of the three estimators. We periodically place the pilot symbols within an OFDM symbol. Optimal pilot placement requires to verify all P_{pilot} combinations from N subcarriers and hence, leads to an NP-hard problem beyond the scope of this paper, and is left for future work. It can be seen from these figures that the subspace method is not greatly affected by the number of pilot symbols since the subspaces are obtained using the second-order statistics of the received signal and not the transmit signal itself. Clearly, the proposed algorithm outperforms the ML and LS estimators at a reduced number of pilots while this tendency is inverted when the number of pilots increases. However, a system with a large amount of pilot symbols is not of practical interest.

In Figs. 6 and 7, we evaluate the impact of the number of observed OFDM symbols on the estimation performance. For the three algorithms, we consider the transmission scheme where the number of pilot symbols is set to one and the SNR is 10 dB. As the subspace algorithm is based

**Fig. 5** Intended channel estimation MSE vs. percentage of pilot symbols for SNR = 10 dB
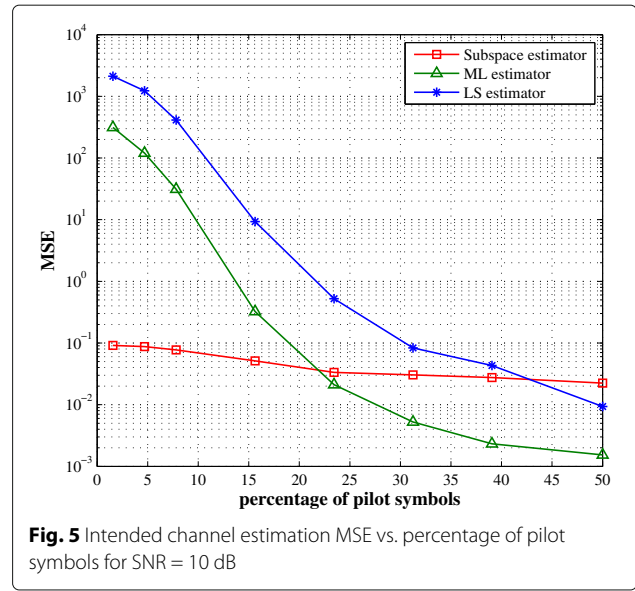
on estimates of the second-order statistic of the received signal, its performance varies with the number of OFDM symbols. All three algorithms are able to estimate the SI channel with an error floor for the LS. The ML and subspace algorithms offer the similar performance. On the other hand, the LS estimator fails to recover the intended channel, for any number of OFDM symbols. This can be explained by the fact that the number of unknowns (intended channel coefficients) is larger than the number of pilot symbols. Hence, it is not possible to use this method when the number of pilot symbols is small. The ML estimator presents also poor estimation performance for the intended channel, while the subspace method is

Fig. 4 SI channel estimation MSE vs. percentage of pilot symbols for SNR = 10 dB

Fig. 6 SI channel estimation MSE vs. number of OFDM symbols for SNR = 10 dB and one pilot symbol

Fig. 7 Intended channel estimation MSE vs. number of OFDM symbols for SNR = 10 dB and one pilot symbol

able to return a good channel estimate, with a better bandwidth efficiency compared to the other estimators, as soon as there are enough OFDM symbols to compute the covariance matrix.

Our primary motivation of this work is to develop an accurate channel estimator to cancel the SI signal. The performance of the SI-canceller are represented by its achieved output signal-to-residual-SI-and-noise power ratio (SINR) after SI cancellation vs. the input SNR. Ideally, if SI could be completely cancelled then the residual SI after cancellation is 0, and consequently, the output SINR equals the input SNR as shown by the dashed line "perfect cancellation" in Fig. 8. In other words, the "perfect

Fig. 8 Output SINR vs. input SNR after SI cancellation
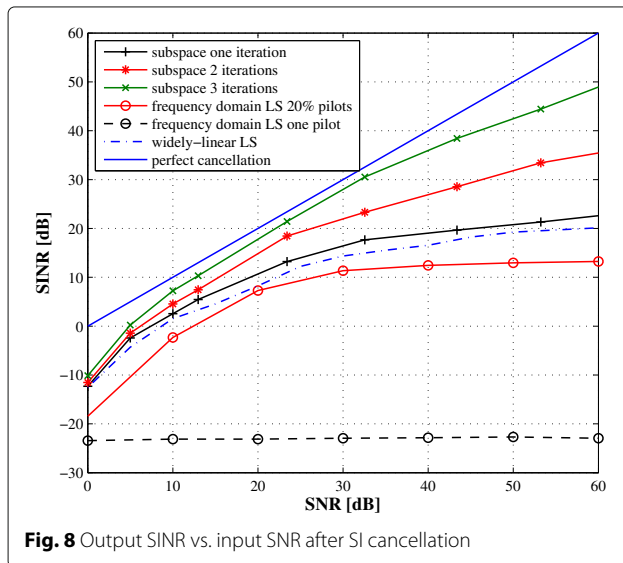
cancellation" is considered as the ideal upper-bound for the SINR. As shown in Fig. 8, with three iterations, the proposed subspace-based SI-canceller can offer an output SINR very close to the upper-bound over a large SNR range. At low SNR, the large estimation error results in a larger residual SI after cancellation, which ultimately affects the output SINR.

We also investigate in Fig. 8 a frequency domain method to estimate the different parameter using the pilot symbols on some subcarriers. We resort to the LS estimator to find the channel responses at the pilot subcarriers. Since the remaining subcarriers contain unknown symbols from the intended transceiver, the complete channel responses are obtained by linear interpolation of the estimated coefficients. Thus, the frequency domain approach uses only the portion of the signal containing pilots while the proposed approach exploits the whole received signal through the second-order statistics. Clearly, the performance of the frequency domain approach highly depends on the number of pilots (as shown in Fig. 8) since the interpolation cannot model the variance of the channel in the frequency domain. We also compare the proposed method with the widely linear LS estimator in [26]. Note that the algorithm in [26] ignores the PA nonlinearities and does not incorporate the intended signal in the estimation process. Some time frames are dedicated to transmit orthogonal pilot symbols for estimation purpose, where the transceiver receives only its own signal. Therefore, the widely linear LS estimator incurs an overhead and requires synchronization between the two transceivers. Besides, it shows a noise floor at high SNR because the PA nonlinearity is not considered during the estimation process. On the other hand, by exploiting the whole received signal through its second-order statistics, the proposed method offers good performance even with one pilot and still outperforms the frequency domain approach (even with much larger number of pilots). Figure 9 plots the bit error rate (BER) vs. SNR curves of the two approaches. For comparison, we include the case of perfect channel estimate. To improve the BER, the SINR should be kept as high as possible at the demodulator. To conclude, while the frequency domain approach is more intuitive, it needs a large number of pilots and is outperformed by the proposed method.

We evaluate the performance of the system in the presence of phase noise by simulation. Figures 10 and 11 plot respectively the SINR and the BER vs. the phase noise 3 dB bandwidth f_{3dB} for SNR = 20 dB and common oscillator at the transmitter and the receiver. The residual SI depends obviously on the quality of the oscillator represented by its f_{3dB}. Higher f_{3dB} results in a fast varying process. Clearly, the proposed method still offers good cancellation performance, which is degraded as f_{3dB} increases.
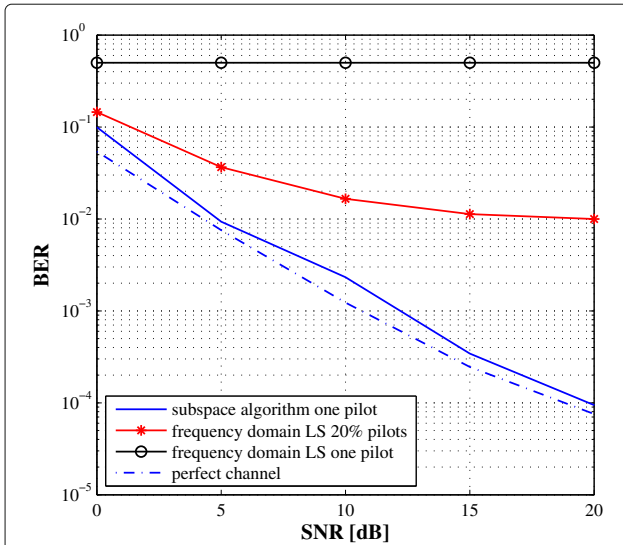
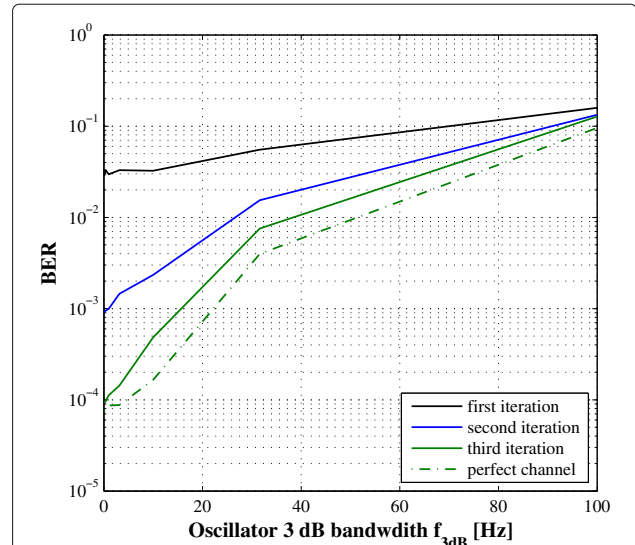**Fig. 9** BER vs. SNR comparison of the proposed and the frequency domain LS techniques

Fig. 11 BER vs. phase noise f_{3dB}

The PA nonlinearity effects on the performance of the proposed algorithm are also investigated through simulations. Figure 12 plots the resulting SINR after cancellation vs. the value of the PA third-order intercept point (IIP3) for SNR= 20 dB. For *perfect* cancellation, the resulting SINR after cancellation would be the SNR= 20 dB. A lower IIP3 indicates higher PA distortions (or poorer PA) and hence reduces the resulting SINR after cancellation. Figure 12 shows that as the IIP3 value increases, the cancellation performance is improved. However, for a sufficiently high IIP3 (e.g., 18 dBm or higher), the PA distortions are no longer dominant and the resulting SINR

after cancellation is unchanged. This can be explained by the fact that, when developing the algorithm, the third-order component of the signal $x_{q,ip3}(n) = x_q^{IQ}(n)|x_q^{IQ}(n)|^2$ is approximated by $x_q(n)|x_q(n)|^2$ to simplify the algorithm. This approximation only affects the algorithm performance when the nonlinear coefficients are sufficiently high.

6 Conclusions

In this paper, a subspace-based estimation has been proposed to jointly estimate the SI channel, the intended channel and the transmitter impairments for MIMO full-duplex systems. By exploiting the covariance and
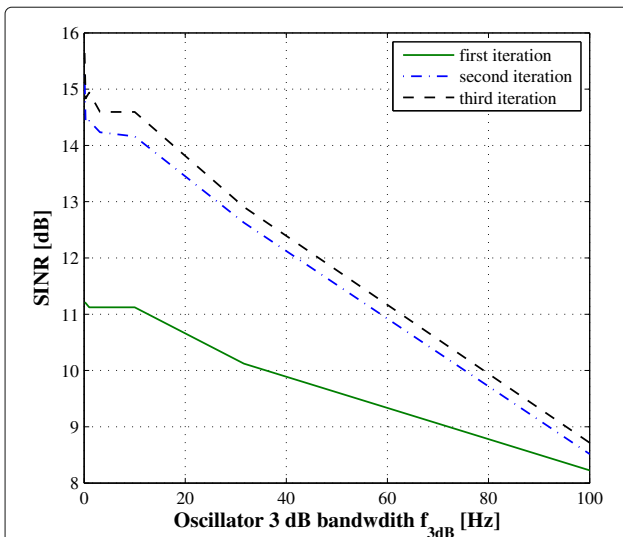

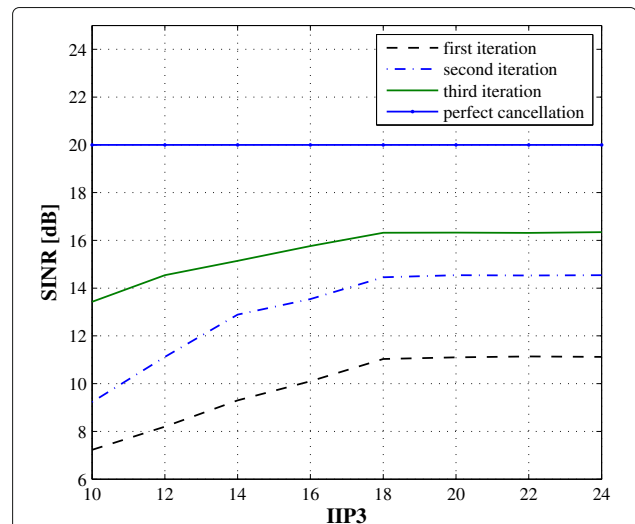**Fig. 10** SINR after SI cancellation vs. f_{3dB}

Fig. 12 SINR after SI cancellation vs. PA IIP3

pseudo-covariance matrix of the received signal, an effective way has been formulated to apply the subspace method for symmetric MIMO systems. The complete characterization of the second-order statistic of the received signal avoids the need of oversampling, required in traditional subspace methods. The subspace that contains the channels is blindly estimated and a short pilot sequence is needed to extract the channel coefficients from this subspace. The proposed method dramatically reduces the number of pilot symbols needed to identify the channel coefficients. Simulation results show that one pilot symbol is enough to obtain an accurate estimate while other methods are not able to recover the channel.

Endnotes

[1] The length of the cyclic prefix N_{cp} should be larger than the delay spread of the channel to eliminate the inter-symbol interference and inter-carrier interference. Therefore, if we know the length of the channel, we can set the cyclic prefix to be sufficiently large to satisfy $N_{cp} > L$. Since this information is in general not available, N_{cp} is chosen to guarantee $N_{cp} > L$. For example, if the distance between the two transceivers is 1 km, a cyclic prefix of 4 microsec is sufficient.

[2] Physically, the additive noise arises from the thermal agitation of the charge carriers in an electronic device and is independent from the input. It can also contain interference from other systems whose signals are independent from the transmit signal of the considered system.

[3] The previous condition is verified for independent channels between different antennas.

[4] The pilot symbols are equally spaced within one OFDM symbol.

Appendix 1: Eigenvalues of $R_{\widetilde{u}}$

Following the discussion in Section 3, we mention that M is of rank N, then it has N strictly positive eigenvalues, $\tau_1, \tau_2, \ldots, \tau_N$, and eigenvalue 0 of multiplicity N. And since the covariance matrix $R_{\widetilde{u}}$ is given by $\alpha^2 M \otimes I_{2N_t}$, it follows that $R_{\widetilde{u}}$ has also N eigenvalues $\tau_1, \tau_2, \ldots, \tau_N$ each of multiplicity $2N_t$ and eigenvalue 0 of multiplicity $2NN_t$. To find the non-zero eigenvalues, we solve the characteristic polynomial of M given by

$$\det\left(M - \tau I_{2N}\right) = 0. \tag{51}$$

First, if $\tau = 1$ is an eigenvalue of M, then it exists a vector $a \neq 0$ such that $Ma - a = 0$. It follows that $a(1) = a(2) = \cdots = a(2N) = 0$, which is in contradiction with $a \neq 0$. Therefore, 1 is not an eigenvalue of M.

By writing M as a block matrix:

$$M = \begin{pmatrix} I_N & M_{1,2} \\ M_{1,2} & I_N \end{pmatrix}, \tag{52}$$

the characteristic polynomial of M, for $\tau \neq 1$, is written as

$$\begin{aligned}\det(M - \tau I_{2N}) &= \det\left((1-\tau)I_N\right) \\ &\quad \times \det\left((1-\tau)I_N - M_{1,2}(1-\tau)^{-1}I_N M_{1,2}\right) \\ &= (1-\tau)^N\left(1 - \tau - (1-\tau)^{-1}\right)^N,\end{aligned} \tag{53}$$

where we used the fact that $M_{1,2}M_{1,2} = I_N$. Then, the solutions to $\det\left(M - \tau I_{2N}\right) = 0$ are 0 and 2. Therefore, all non-zero eigenvalues of M are equal to 2 and thus all the non-zero eigenvalues of $R_{\widetilde{u}}$ are equal to $2\alpha^2$.

Appendix 2: Precoding for complex modulation

To make it simple, we consider the matrices P and Q having the following block structure:

$$P = \begin{pmatrix} aI_{N/2} & 0I_{N/2} \\ 0I_{N/2} & bI_{N/2} \end{pmatrix},$$

$$Q = \begin{pmatrix} 0I_{N/2} & cI_{N/2} \\ dI_{N/2} & 0I_{N/2} \end{pmatrix}, \tag{54}$$

for given real numbers a, b, c and d. Similarly to the real modulation, we have $R_{\widetilde{u}} = M \otimes I_{2N_t}$ where M for complex modulation is given by

$$\begin{aligned}M &= \begin{pmatrix} PP^T + QQ^T & PQ^T + QP^T \\ PQ^T + QP^T & PP^T + QQ^T \end{pmatrix} \\ &= \begin{pmatrix} (a^2 + c^2) & 0 & 0 & (ad + bc) \\ 0 & (b^2 + d^2) & (ad + bc) & 0 \\ 0 & (ad + bc) & (a^2 + c^2) & 0 \\ (ad + bc) & 0 & 0 & (b^2 + d^2) \end{pmatrix} \otimes I_{N/2},\end{aligned}$$

for $a^2 + c^2 = b^2 + d^2$. Thus, for a, b, c and d satisfying $a^2 + c^2 = ad + bc$ and $b^2 + d^2 = ad + bc$, each line of M is repeated two times and $R_{\widetilde{u}}$ has rank $2NN_t$. As an example, we can take $a = 0.757$, $b = 0.5032$, $c = 0.4935$ and $d = 0.7506$.

References
1. JI Choi, M Jain, K Srinivasan, P Levis, S Katti, in *Proc. ACM MobiCom*. Achieving single channel, full duplex wireless communication (ACM, Chicago, 2010), pp. 1–12

2. M Duarte, A Sabharwal, V Aggarwal, R Jana, KK Ramakrishnan, CW Rice, NK Shankaranarayanan, Design and characterization of a full-duplex multiantenna system for WiFi networks. IEEE Trans. Veh. Technol. **63**(3), 1160–1177 (2014)

3. A Masmoudi, T Le-Ngoc, in *Proc. IEEE Global Telecommun. Conf.* Self-interference cancellation limits in full-duplex communication systems (IEEE, Washington DC, 2016)

4. MA Khojastepour, S Rangarajan, in *Proc. ASILOMAR Signals, Syst., Comput.* Wideband digital cancellation for full-duplex communications (IEEE, Pacific Frove, 2012), pp. 1300–1304

5. 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)

6. M Duarte, C Dick, A Sabharwal, Experiment-driven characterization of full-duplex wireless systems. IEEE Trans. Wireless Comm. **11**(12), 4296–4307 (2012)

7. J Ma, GY Li, J Zhang, T Kuze, H Iura, in *Proc. IEEE Global Telecommun. Conf.* A new coupling channel estimator for cross-talk cancellation at wireless relay stations, (Honolulu, 2009)

8. JR Krier, IF Akyildiz, in *Proc. IEEE Pers. Indoor and Mobile Radio Commun.* Active self-interference cancellation of passband signals using gradient descent (IEEE, London, 2013)

9. S Li, RD Murch, in *Proc. IEEE Global Telecommun. Conf.* Full-duplex wireless communication using transmitter output based echo cancellation, (2011), pp. 1–5

10. D Bliss, P Parker, A Margetts, in *Prog. IEEE Statistical Signal Processing.* Simultaneous transmission and reception for improved wireless network performance, (2007), pp. 478–482

11. BP Day, AR Margetts, DW Bliss, P Schniter, Full-duplex bidirectional MIMO: achievable rates under limited dynamic range. IEEE Trans. Signal Process. **60**(7), 3702–3713 (2012)

12. AC Cirik, J Zhang, M Haardt, Y Hua, in *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC).* Sum-rate maximization for bi-directional full-duplex MIMO systems under multiple linear constraints, (2014), pp. 389–393

13. Y Hua, P Liang, Y Ma, AC Cirik, Q Gao, A method for broadband full-duplex MIMO radio. IEEE Signal Process. Lett. **19**(12), 793–796 (2012)

14. A Masmoudi, T Le-Ngoc, in *Proc. IEEE Veh. Technol. Conf.* Self-interference mitigation using active signal injection full-duplex MIMO-OFDM systems (IEEE, Montreal, 2016)

15. J-J Van de Beek, O Edfors, M Sandell, SK Wilson, P Ola Borjesson, in *Proc. IEEE Veh. Technol. Conf.* On channel estimation in OFDM systems, (1995), pp. 815–819

16. H Minn, N Al-Dhahir, Optimal training signals for MIMO OFDM channel estimation. IEEE Trans. Wireless Comm. **5**(5), 1158–1168 (2006)

17. F Gao, Y Zeng, A Nallanathan, T-S Ng, Robust subspace blind channel estimation for cyclic prefixed MIMO ODFM systems: algorithm, identifiability and performance analysis. IEEE J. Select. Areas Comm. **26**(2), 378–388 (2008)

18. C-C Tu, B Champagne, Subspace-based blind channel estimation for MIMO-OFDM systems with reduced time averaging. IEEE Trans. Veh. Technol. **59**(3), 1539–1544 (2010)

19. E Moulines, P Duhamel, J-F Cardoso, S Mayrargue, Subspace methods for the blind identification of multichannel FIR filters. IEEE Trans. Signal Process. **43**(2), 516–525 (1995)

20. Y Zeng, T-S Ng, A semi-blind channel estimation method for multiuser multiantenna OFDM systems. IEEE Trans. Signal Process. **52**(5), 1419–1429 (2004)

21. E de Carvalho, DT Slock, Blind and semi-blind FIR multichannel estimation: (global) identifiability conditions. IEEE Trans. Signal Process. **52**(4), 1053–1064 (2004)

22. A Masmoudi, T Le-Ngoc, A maximum-likelihood channel estimator for self-interference cancellation in full-duplex systems. IEEE Trans. Veh. Technol. **65**(7), 5122–5132 (2016)

23. LTE; evolved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception (3GPP TS 36.101 version 11.2.0 release 11). ETSI, Sophia Antipolis Cedex, France (2012)

24. DW Bliss, TM Hancock, P Schniter, in *Proc. ASILOMAR Signals, Syst., Comput.* Hardware phenomenological effects on cochannel full-duplex MIMO relay performance (IEEE, Pacific Frove, 2012)

25. E Ahmed, A Eltawil, A Sabharwal, in *Proc. ASILOMAR Signals, Syst., Comput.* Self-interference cancellation with nonlinear distortion suppression for full-duplex systems (IEEE, Pacific Frove, 2013)

26. D Korpi, L Anttila, V Syrjala, M Valkama, Widely linear digital self-interference cancellation in direct-conversion full-duplex transceiver. IEEE J. Selected Areas Commun. **32**(9), 1674–1687 (2014)

27. A Masmoudi, T Le-Ngoc, in *Proc. IEEE Wireless Commun. and Netw. Conf.* Self-interference cancellation for full-duplex MIMO transceivers (IEEE, New Orleans, 2015)

28. WH Gerstacker, R Schober, A Lampe, Receivers with widely linear processing for frequency-selective channels. IEEE Trans. Commun. **51**(9), 1512–1523 (2003)

29. R Schober, WH Gerstacker, L-J Lampe, Data-aided and blind stochastic gradient algorithms for widely linear MMSE MAI suppression for DS-CDMA. IEEE Trans. Signal Process. **52**(3), 746–756 (2004)

30. M Kristensson, B Ottersten, D Slock, in *Proc. ASILOMAR Signals, Syst., Comput.* Blind subspace identification of a BPSK communication channel (IEEE, Pacific Frove, 1996)

31. A Masmoudi, T Le-Ngoc, in *Proc. IEEE Int. Conf. Commun.* A digital subspace-based self-interference cancellation in full-duplex MIMO transceivers (IEEE, London, 2015), pp. 4954–4959

32. A Masmoudi, T Le-Ngoc, in *Proc. IEEE Int. Conf. Commun.* Residual self-interference after cancellation in full-duplex systems (IEEE, Sydney, 2014)

33. JG McMichael, KE Kolodziej, in *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton).* Optimal tuning of analog self-interference cancellers for full-duplex wireless communication, (2012), pp. 246–251

34. FD Neeser, JL Massey, Proper complex random processes with applications to information theory. IEEE Trans. Inf. Theory. **39**(4), 1293–1302 (1993)

35. A Masmoudi, T Le-Ngoc, in *Proc. IEEE Veh. Technol. Conf.* A maximum-likelihood channel estimator in MIMO full-duplex systems (IEEE, Vancouver, 2014)