SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

August 2017, Vol. 60 082303:1–082303:13
doi: 10.1007/s11432-016-0386-x

Compressed sensing-based time-domain channel estimator for full-duplex OFDM systems with IQ-imbalances

Hai YU[1], Feng SHU[1,2,3,4*], You YOU[1], Jin WANG[1], Tingting LIU[1], Xiaohu YOU[3], Jinhui LU[1], Jianxin WANG[1] & Xiaohua ZHU[1]

[1]*School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing* 210094*, China;*
[2]*National Key Laboratory of Electromagnetic Environment, China Research Institute of Radiowave Propagation, Qingdao* 266107*, China;*
[3]*National Mobile Communications Research Laboratory, Southeast University, Nanjing* 210096*, China;*
[4]*College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou* 350002*, China*

Received September 28, 2016; accepted November 9, 2016; published online March 6, 2017

Abstract In full-duplex orthogonal frequency-division multiplexing (OFDM) systems with in-phase and quadrature (IQ) imbalances, a time-domain least squares (TD-LS) channel estimator is proposed for estimating both the source-to-destination (intended) and the destination-to-destination (self-interference) channels. To further improve the performance, an adaptive orthogonal matching pursuit (OMP) is proposed and its sparsity is estimated by a threshold method. Finally, the full-duplex interference is removed using serial interference cancellation and sphere decoding is used to complete the maximum likelihood detection. Simulation results demonstrate that in terms of both the mean square error (MSE) and the bit error rate (BER), the proposed adaptive OMP performs better than the TD-LS by exploiting the sparse property of the channel. Additionally, compared to the gradient projection, the proposed adaptive OMP is better in the low and medium signal-to-noise ratio (SNR) regions and marginally worse in the high SNR region.

Keywords full-duplex, OFDM, IQ imbalance, channel estimation, time-domain, least squares, adaptive OMP, gradient projection

Citation Yu H, Shu F, You Y, et al. Compressed sensing-based time-domain channel estimator for full-duplex OFDM systems with IQ-imbalances. Sci China Inf Sci, 2017, 60(8): 082303, doi: 10.1007/s11432-016-0386-x

1 Introduction

Conventional wireless communication systems operate in a half-duplex (HD) mode with an uplink and downlink, using orthogonal channels such as different frequency bands or time slots. Of late, full-duplex has become immensely popular owing to its high spectrum efficiency that is double that of the HD mode [1–7]. In the full-duplex mode, non-orthogonal/overlapped simultaneous transmission and reception are implemented in the same time and frequency bands. One of the challenging issues in full-duplex

* Corresponding author (email: shufeng@njust.edu.cn)

systems is the estimation of both the self-interference channel and the intended channel with a high-precision for reducing the full-duplex self-interference as much as possible [8,9]. Based on the expectation maximization approach, a blind channel estimator for simultaneously estimating the self-interference and the intended channel parameters in full-duplex wireless systems without requiring a training sequence is proposed [10]. Refs. [11, 12] have proposed a maximum-likelihood (ML) algorithm that performs a joint estimation of both the self-interference and the intended channels, and an iterative procedure was constructed for improving the estimation performance.

In-phase and quadrature (IQ) imbalance in a direct-conversion transmitter and receiver structure has a serious impact on the OFDM system performance [13]. In half-duplex OFDM systems, channel estimation and IQ compensation have been investigated completely and extensively. A simple least squares (LS) channel estimation is proposed for OFDM systems with transmit and receive IQ-imbalances in [14–16]. Refs. [17, 18] have made a joint estimate of the transmit IQ-imbalance, channel impulse response (CIR), and the receive IQ-imbalance from the frequency-domain. In a sparse wireless channel, a sparse time-domain channel estimator using parallel iterative shrinkage was designed to fully exploit the sparse property of the wireless channel and a bit error rate (BER) performance close to that of the ideal channel was achieved [19].

In conventional half-duplex OFDM systems without IQ-imbalances, compressed sensing-based channel estimation algorithms mainly fall into two categories. The first includes typical algorithms based on convex optimization such as the basis pursuit (BP) [20, 21], gradient projection (GP) [22], iterative shrinkage, etc. [19]. The second involves greedy methods with a lower-complexity such as the matching pursuit (MP) [23–25], orthogonal matching pursuit (OMP) [26], etc. Compared to the MP, the OMP converges more rapidly. However, the main disadvantage of the OMP is that it needs to know the channel sparsity in advance. In this paper, we solve this problem by estimating the channel sparsity before employing the OMP.

To the best of our knowledge, channel estimation and pilot optimization for full-duplex OFDM systems with IQ-imbalances continue to be open problems. In such systems, high-precision channel estimations are critical for reducing the effect of both the self-interference and the IQ-imbalance distortion. In this paper, we investigate the channel estimation in full-duplex OFDM systems with IQ-imbalances at both the transmitter and receiver. Our main contributions are as follows:

- We first derive the mathematical model for the time-domain channel estimation. This model provides the basic building blocks for all the channel estimators in this paper. Then, we propose a time-domain least square(TD-LS) channel estimator for full-duplex OFDM systems with IQ-imbalances that aims to estimate the mixture channel parameters combining the effects of the IQ-imbalances and the CIR. Both the intended and self-interference channels are estimated, simultaneously and jointly, by the proposed TD-LS. The latter is used to remove the full-duplex interference and the former is for detection after interference cancellation.
- To exploit the sparse property of the channel for improving the channel estimation performance, an adaptive OMP algorithm is proposed. In this method, the sparsity for the OMP is computed by a threshold method and the threshold is determined by estimating the noise variance. Thus, the adaptive OMP can be applied to practical OFDM systems.

This paper is organized as follows: The full-duplex system model with IQ-imbalances is described in Section 2. Section 3 presents the mathematical model for time-domain channel estimation. In Section 4, two channel estimators, the TD-LS and the adaptive OMP are developed. The simulation results and discussions are presented in Section 5. Finally, Section 6 concludes this paper.

Notations. Throughout this paper, matrices and vectors are denoted by bold upper case and bold lower case letters, respectively. The signs $(\bullet)^{\mathrm{H}}$, $(\bullet)^*$, $(\bullet)^{\mathrm{T}}$, $(\bullet)^{-1}$, $\mathrm{tr}(\bullet)$, $\|\bullet\|_F$, and $\det(\bullet)$ denote the matrix conjugate transpose, conjugate, transpose, inverse, trace, norm-2, and the determinant, respectively. The notation, $\mathcal{E}\{\bullet\}$, refers to the expectation operation. The symbols, $\boldsymbol{I}_N$ and $\boldsymbol{0}_{N\times M}$, denote the all-zero $N \times N$ identity matrix and the $N \times M$ matrix, respectively. $\otimes$ denotes the Kronecker product of two matrices. Diag $\{\boldsymbol{a}\}$ denotes the operation of placing all the elements of vector $\boldsymbol{a}$ over the diagonal of the

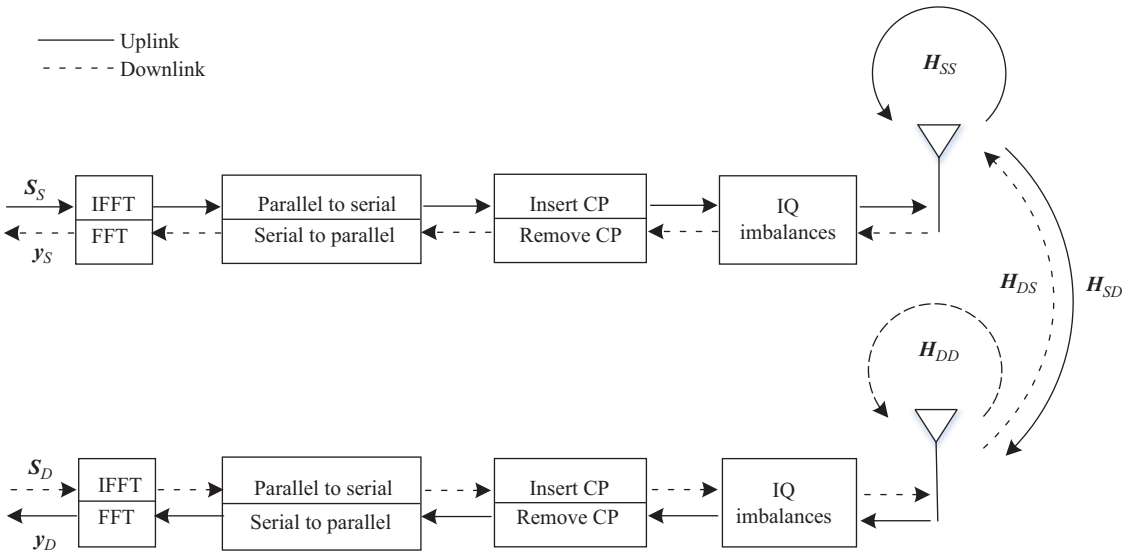

Figure 1 Full-duplex system model with IQ-imbalance.

diagonal matrix.

2 System model

We consider a full-duplex point-to-point OFDM system with IQ imbalance, as shown in Figure 1. The received signal at the destination node consists of the intended signal from the source node, the self-interference signal from itself, and the additive white Gaussian noise. Without loss of generality, we consider the source-to-destination link.

In Figure 1, the transmitted mth OFDM symbols from source-to-destination and destination-to-destination are denoted as

$$\boldsymbol{s}_S(m,:) = [s_S(m,1)\ s_S(m,2)\ \cdots\ s_S(m,N)]^{\mathrm{T}}, \tag{1}$$

$$\boldsymbol{s}_D(m,:) = [s_D(m,1)\ s_D(m,2)\ \cdots\ s_D(m,N)]^{\mathrm{T}}, \tag{2}$$

where N is the number of total subcarriers. To simplify the discussion, block fading is assumed in the following, i.e., the channel is assumed to be constant within a frame but varying from one frame to another. Here, we assume that the RF self-interference cancellation has been done. We will perform the baseband self-interference cancellation after the channel estimation. We obtain the final received mth OFDM symbol with a full-duplex interference using the following operations: inverse fast Fourier transform (IFFT), parallel-to-serial conversion, inserting the cyclic prefix (CP), IQ-imbalance distortion at the transmitter, multi-path fading channel with a full-duplex interference, IQ-imbalance distortion at the receiver, removing the CP, serial-to-parallel conversion, and FFT.

$$\begin{aligned}\boldsymbol{y}_D(m,:) = &\left(\mu_{r,D}\mu_{t,S}\mathrm{Diag}\{\boldsymbol{H}_{SD}\} + \nu_{r,D}\nu^*_{t,S}\mathrm{Diag}\{\boldsymbol{H}^{\#}_{SD}\}\right)\boldsymbol{s}_S(m,:)\\ &+\left(\mu_{r,D}\nu_{t,S}\mathrm{Diag}\{\boldsymbol{H}_{SD}\} + \nu_{r,D}\mu^*_{t,S}\mathrm{Diag}\{\boldsymbol{H}^{\#}_{SD}\}\right)\boldsymbol{s}^{\#}_S(m,:)\\ &+\left(\mu_{r,D}\mu_{t,D}\mathrm{Diag}\{\boldsymbol{H}_{DD}\} + \nu_{r,D}\nu^*_{t,D}\mathrm{Diag}\{\boldsymbol{H}^{\#}_{DD}\}\right)\boldsymbol{s}_D(m,:)\\ &+\left(\mu_{r,D}\nu_{t,D}\mathrm{Diag}\{\boldsymbol{H}_{DD}\} + \nu_{r,D}\mu^*_{t,D}\mathrm{Diag}\{\boldsymbol{H}^{\#}_{DD}\}\right)\boldsymbol{s}^{\#}_D(m,:) + \boldsymbol{\omega}(m,:),\end{aligned} \tag{3}$$

where the superscript # is defined as

$$\boldsymbol{X}^{\#} = [X^*(1)\ X^*(N)\ \cdots\ X^*(N/2+2)\ X^*(N/2+1)\ X^*(N/2)\ \cdots\ X^*(2)]^{\mathrm{T}} \tag{4}$$

with

$$\boldsymbol{X} = [X(1) \;\; X(2) \;\; \cdots \;\; X(N/2) \;\; X(N/2+1) \;\; X(N/2+2) \;\; \cdots \;\; X(N)]^{\mathrm{T}}, \tag{5}$$

$$\mu_{t,S} = \cos(\theta_{t,S}/2) + j\alpha_{t,S}\sin(\theta_{t,S}/2), \quad \nu_{t,S} = \alpha_{t,S}\cos(\theta_{t,S}/2) - j\sin(\theta_{t,S}/2), \tag{6}$$

$$\mu_{t,D} = \cos(\theta_{t,D}/2) + j\alpha_{t,D}\sin(\theta_{t,D}/2), \quad \nu_{t,D} = \alpha_{t,D}\cos(\theta_{t,D}/2) - j\sin(\theta_{t,D}/2), \tag{7}$$

$$\mu_{r,D} = \cos(\theta_{r,D}/2) + j\alpha_{r,D}\sin(\theta_{r,D}/2), \quad \nu_{r,D} = \alpha_{r,D}\cos(\theta_{r,D}/2) - j\sin(\theta_{r,D}/2), \tag{8}$$

where $\theta_{t,S}$, $\alpha_{t,S}$, $\theta_{t,D}$, and $\alpha_{t,D}$ are the transmit phase and amplitude imbalances between the I and Q branches for transmitter and receiver; $\theta_{r,D}$ and $\alpha_{r,D}$ are the receive phase and amplitude imbalance parameters between the I and Q branches in the receiver.

$$\boldsymbol{H}_{SD} = \boldsymbol{F}_{N\times N}\begin{pmatrix} \boldsymbol{h}_{SD} \\ \boldsymbol{0}_{(N-L)\times 1} \end{pmatrix}, \tag{9}$$

$$\boldsymbol{H}_{DD} = \boldsymbol{F}_{N\times N}\begin{pmatrix} \boldsymbol{h}_{DD} \\ \boldsymbol{0}_{(N-L)\times 1} \end{pmatrix}, \tag{10}$$

where $\boldsymbol{F}_{N\times N}$ is the normalized discrete Fourier transform matrix, $\boldsymbol{h}_{SD} = [h_{SD}(1)\, h_{SD}(2) \;\cdots\; h_{SD}(L)]^{\mathrm{T}}$ is the CIR of the source-to-destination, $\boldsymbol{h}_{DD} = [h_{DD}(1)\, h_{DD}(2) \;\cdots\; h_{DD}(L)]^{\mathrm{T}}$ is the CIR of the destination-to-destination with L being the length of the CP, and $\boldsymbol{w}(m,:)$ indicates the distorted noise over the mth OFDM symbol:

$$\boldsymbol{w}(m,:) = \mu_r \boldsymbol{v}_D(m,:) + \nu_r \boldsymbol{v}_D^{\#}(m,:), \tag{11}$$

where $\boldsymbol{v}_D(m,:)$ is the additive white Gaussian noise vector. For the convenience of deriving the following, we define the matrix $\boldsymbol{F}_{P\times Q}$ with the entry (p,q) being

$$\boldsymbol{F}_{P\times Q}(p,q) = \frac{1}{\sqrt{P}}\exp\left(\frac{-j2\pi(p-1)(q-1)}{P}\right), \quad j=\sqrt{-1}, p\in\{1,2,\ldots,P\}, q\in\{1,2,\ldots,Q\}. \tag{12}$$

3 Problem formulation

For the purpose of channel estimation, the system model in (3) is further organized as

$$\begin{aligned}\boldsymbol{y}_D(m,:) = {} & \mathrm{Diag}\{\boldsymbol{s}_S(m,:)\}(\mu_{r,D}\mu_{t,S}\boldsymbol{H}_{SD} + \nu_{r,D}\nu_{t,S}^{*}\boldsymbol{H}_{SD}^{\#}) \\ & + \mathrm{Diag}\{\boldsymbol{s}_S^{\#}(m,:)\}(\mu_{r,D}\nu_{t,S}\boldsymbol{H}_{SD} + \nu_{r,D}\mu_{t,S}^{*}\boldsymbol{H}_{SD}^{\#}) \\ & + \mathrm{Diag}\{\boldsymbol{s}_D(m,:)\}(\mu_{r,D}\mu_{t,D}\boldsymbol{H}_{DD} + \nu_{r,D}\nu_{t,D}^{*}\boldsymbol{H}_{DD}^{\#}) \\ & + \mathrm{Diag}\{\boldsymbol{s}_D^{\#}(m,:)\}(\mu_{r,D}\nu_{t,D}\boldsymbol{H}_{DD} + \nu_{r,D}\mu_{t,D}^{*}\boldsymbol{H}_{DD}^{\#}) + \boldsymbol{\omega}(m,:).\end{aligned} \tag{13}$$

We define

$$\boldsymbol{H}_{SD}^{a} = \mu_{r,D}\mu_{t,S}\boldsymbol{H}_{SD} + \nu_{r,D}\nu_{t,S}^{*}\boldsymbol{H}_{SD}^{\#}, \tag{14}$$

$$\boldsymbol{H}_{SD}^{b} = \mu_{r,D}\nu_{t,S}\boldsymbol{H}_{SD} + \nu_{r,D}\mu_{t,S}^{*}\boldsymbol{H}_{SD}^{\#}, \tag{15}$$

$$\boldsymbol{H}_{DD}^{a} = \mu_{r,D}\mu_{t,D}\boldsymbol{H}_{DD} + \nu_{r,D}\nu_{t,D}^{*}\boldsymbol{H}_{DD}^{\#}, \tag{16}$$

$$\boldsymbol{H}_{DD}^{b} = \mu_{r,D}\nu_{t,D}\boldsymbol{H}_{DD} + \nu_{r,D}\mu_{t,D}^{*}\boldsymbol{H}_{DD}^{\#} \tag{17}$$

with

$$\boldsymbol{H}_{SD}^{a}=\boldsymbol{F}_{N\times N}\left(\begin{array}{c}\mu_{r,D}\mu_{t,S}\boldsymbol{h}_{SD}+\nu_{r,D}\nu_{t,S}^{*}\boldsymbol{h}_{SD}^{*}\\ \boldsymbol{0}_{(N-L)\times 1}\end{array}\right)=\boldsymbol{F}_{N\times L}\boldsymbol{h}_{SD}^{a}, \tag{18}$$

$$\boldsymbol{H}_{SD}^{b}=\boldsymbol{F}_{N\times N}\left(\begin{array}{c}\mu_{r,D}\nu_{t,S}\boldsymbol{h}_{SD}+\nu_{r,D}\mu_{t,S}^{*}\boldsymbol{h}_{SD}^{*}\\ \boldsymbol{0}_{(N-L)\times 1}\end{array}\right)=\boldsymbol{F}_{N\times L}\boldsymbol{h}_{SD}^{b}, \tag{19}$$

$$\boldsymbol{H}_{DD}^{a}=\boldsymbol{F}_{N\times N}\left(\begin{array}{c}\mu_{r,D}\mu_{t,D}\boldsymbol{h}_{DD}+\nu_{r,D}\nu_{t,D}^{*}\boldsymbol{h}_{DD}^{*}\\ \boldsymbol{0}_{(N-L)\times 1}\end{array}\right)=\boldsymbol{F}_{N\times L}\boldsymbol{h}_{DD}^{a}, \tag{20}$$

$$\boldsymbol{H}_{DD}^{b}=\boldsymbol{F}_{N\times N}\left(\begin{array}{c}\mu_{r,D}\nu_{t,D}\boldsymbol{h}_{DD}+\nu_{r,D}\mu_{t,D}^{*}\boldsymbol{h}_{DD}^{*}\\ \boldsymbol{0}_{(N-L)\times 1}\end{array}\right)=\boldsymbol{F}_{N\times L}\boldsymbol{h}_{DD}^{b}. \tag{21}$$

Substituting (18)–(21) in (13), we obtain the following form:

$$\boldsymbol{y}_{D}(m,:)= \tag{22}$$
$$\underbrace{\left[\underbrace{\mathrm{Diag}\{\boldsymbol{s}_{S}(m,:)\}\ \mathrm{Diag}\{\boldsymbol{s}_{S}^{\#}(m,:)\}\ \mathrm{Diag}\{\boldsymbol{s}_{D}(m,:)\}\ \mathrm{Diag}\{\boldsymbol{s}_{D}^{\#}(m,:)\}}_{\boldsymbol{s}_{P_m}}\right](\boldsymbol{I}_{4}\otimes\boldsymbol{F}_{N\times L})}_{\boldsymbol{A}_m}\underbrace{\left(\begin{array}{c}\boldsymbol{h}_{SD}^{a}\\ \boldsymbol{h}_{SD}^{b}\\ \boldsymbol{h}_{DD}^{a}\\ \boldsymbol{h}_{DD}^{b}\end{array}\right)}_{\boldsymbol{x}}+\boldsymbol{w}(m,:).$$

If the first M OFDM symbols per frame are used as a training sequence to estimate the CIR, the above model can be readily extended as

$$\underbrace{\left(\begin{array}{c}\boldsymbol{y}_{D}(1,:)\\ \vdots\\ \boldsymbol{y}_{D}(M,:)\end{array}\right)}_{\boldsymbol{y}_{D}}=\underbrace{\left(\begin{array}{c}\boldsymbol{A}_{1}\\ \vdots\\ \boldsymbol{A}_{M}\end{array}\right)}_{\boldsymbol{A}}\boldsymbol{x}+\underbrace{\left(\begin{array}{c}\boldsymbol{w}(1,:)\\ \vdots\\ \boldsymbol{w}(M,:)\end{array}\right)}_{\boldsymbol{w}}. \tag{23}$$

When $MN\geqslant 4L$ and $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ is nonsingular, the above equation has a unique solution.

4 Proposed channel estimation methods

In this section, we present three ways to solve the equation shown in (23). First, a time-domain least-squares channel estimator is developed by utilizing the time-domain property of the channel. Subsequently, two compressed sensing channel estimators are designed to further improve the performance of the channel estimation by exploiting the sparse property of the channel. Finally, sphere decoding is applied to implement maximum likelihood detection after the serial interference cancellation.

4.1 Proposed time-domain least-squares estimator

With respect to (23), if $MN>4L$, the LS channel estimator is given by

$$\hat{\boldsymbol{x}}_{TD-LS}=\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{y}_{D} \tag{24}$$

that forms a channel estimation error, as follows:

$$\Delta\boldsymbol{x}=\boldsymbol{x}-\hat{\boldsymbol{x}}_{TD-LS}=\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{w}, \tag{25}$$

with a covariance matrix:

$$\boldsymbol{R}_{\boldsymbol{x}} = \mathcal{E}\{\Delta\boldsymbol{x}^{\mathrm{H}}\Delta\boldsymbol{x}\} = (\mu_{r,D}^2 + \nu_{r,D}^2)(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}. \tag{26}$$

The problem of minimizing the MSE of the TD-LS estimator in (24) is formalized as

$$\begin{aligned} \min_{(\boldsymbol{s}_S(m,:),\boldsymbol{s}_D(m,:))} \quad & \mathrm{tr}\{\boldsymbol{R}_{\boldsymbol{x}}\} \\ \text{s.t.} \quad & \boldsymbol{s}_S(m,:), \boldsymbol{s}_D(m,:) \in \mathbb{C}_{MQAM}^N,\ m \in \{1,2,\ldots,M\}, \\ & \sum_{m=1}^{M} \boldsymbol{s}_S^{\mathrm{H}}(m,:)\boldsymbol{s}_S(m,:) \leqslant MNP_S,\ \sum_{m=1}^{M} \boldsymbol{s}_D^{\mathrm{H}}(m,:)\boldsymbol{s}_D(m,:) \leqslant MNP_D, \end{aligned} \tag{27}$$

where $\mathbb{C}_{\mathrm{MQAM}}^N$ is the N-D discrete space with each dimensional element selected from the points of the signal constellation of the MQAM; P_S and P_D are the average transmit powers of the source and destination, respectively, per time-frequency grid,

$$\mathrm{tr}\{\boldsymbol{R}_{\boldsymbol{x}}\} = \left(\mu_{r,D}^2 + \nu_{r,D}^2\right)\mathrm{tr}\{(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\} \tag{28}$$

with

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \left(\boldsymbol{I}_4 \otimes \boldsymbol{F}_{N\times L}^{\mathrm{H}}\right)\left(\sum_{m=1}^{M} \boldsymbol{s}_{P_m}^{\mathrm{H}}\boldsymbol{s}_{P_m}\right)\left(\boldsymbol{I}_4 \otimes \boldsymbol{F}_{N\times L}\right). \tag{29}$$

The optimization problem in (27) is cast as an integer optimization. Owing to its high dimension, it is an NP-hard problem. Using simulation, we find that the BER performance of the proposed TD-LS gradually improves, as the conditional number of $\sum_{m=1}^{M} \boldsymbol{s}_{P_m}^{\mathrm{H}}\boldsymbol{s}_{P_m}$ approaches unity. In the simulation, we select a suitable pilot pattern with a conditional number, $\sum_{m=1}^{M} \boldsymbol{s}_{P_m}^{\mathrm{H}}\boldsymbol{s}_{P_m}$, close to unity using a random search method.

4.2 Proposed adaptive OMP estimator

In this paper, the wireless channel is supposed to be sparse; thus the vectors, $\boldsymbol{h}_{SD}$ and $\boldsymbol{h}_{DD}$, in (9) and (10) are sparse. Hence, the large vector $\boldsymbol{x}$ in (23) that is to be estimated is also sparse. For the measurement matrices $\boldsymbol{A}$ we establish the restricted isometry property (RIP) to guarantee the accuracy of the reconstruction of the compressed sensing with an overwhelming probability. We select $\boldsymbol{x}$ as the basic training sequence with zero-mean, constant-amplitude, random-phase variables; the various nodes take different cyclically shifting versions of the basic sequence. For example,

$$\boldsymbol{s}_S(m,:) = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{N-t_1+1} \\ \boldsymbol{I}_{t_1-1} & \boldsymbol{0} \end{bmatrix} \boldsymbol{x}, \tag{30}$$

$$\boldsymbol{s}_D(m,:) = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I}_{N-t_2+1} \\ \boldsymbol{I}_{t_2-1} & \boldsymbol{0} \end{bmatrix} \boldsymbol{x}, \tag{31}$$

where $t_1, t_2 \in \{1,2,\ldots,N\}$ are the different shift indices. It is clear that $\boldsymbol{s}_S^{\#}(m,:)$ and $\boldsymbol{s}_D^{\#}(m,:)$ are the reverse cyclically shifting versions of $\boldsymbol{x}^*$.

Theorem 1. Let $\{x_i\}_{i=1}^N$ be a basic sequence of independent and identically distributed (i.i.d) zero-mean, constant-amplitude, random-phase variables with a variance, $1/N$. For any $\delta_s \in (0,1)$, the measurement matrix $\boldsymbol{A}$ satisfies the RIP with a probability, $1 - \exp\left(-(\frac{c_1\sqrt{N}}{S} - \frac{1}{N\sqrt{N}})^2\right)$, where c_1 is a constant depending only on δ_s.

Proof. See Appendix A.

In what follows, the OMP is adopted for exploiting the sparse property. However, the OMP needs to know the number of nonzero elements (called the sparsity of $\boldsymbol{x}$) in advance. In general, it is not easy

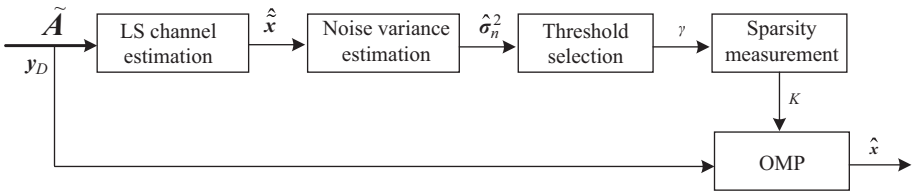

Figure 2 Block diagram of the proposed adaptive OMP algorithm.

to know the number of nonzero elements or the sparsity in an estimated vector. Considering that the power of the CIR concentrates on the inside of the cyclic prefix, the outside of the CP can be used to estimate the variance of the channel noise. Scaling the estimated noise variance, a suitable threshold for measuring the sparsity of $\boldsymbol{x}$ can be approximately decided. Based on the estimated sparsity, we design the adaptive OMP shown in Figure 2.

Below, we present the calculation of the sparsity of $\boldsymbol{x}$ in (23). Similarly, Eqs. (18)–(21) are rewritten as

$$\boldsymbol{H}_{SD}^{a} = \boldsymbol{F}_{N\times\frac{N}{4}} \begin{pmatrix} \boldsymbol{h}_{SD}^{a} \\ \boldsymbol{0}_{(\frac{N}{4}-L)\times 1} \end{pmatrix}, \tag{32}$$

$$\boldsymbol{H}_{SD}^{b} = \boldsymbol{F}_{N\times\frac{N}{4}} \begin{pmatrix} \boldsymbol{h}_{SD}^{b} \\ \boldsymbol{0}_{(\frac{N}{4}-L)\times 1} \end{pmatrix}, \tag{33}$$

$$\boldsymbol{H}_{DD}^{a} = \boldsymbol{F}_{N\times\frac{N}{4}} \begin{pmatrix} \boldsymbol{h}_{DD}^{a} \\ \boldsymbol{0}_{(\frac{N}{4}-L)\times 1} \end{pmatrix}, \tag{34}$$

$$\boldsymbol{H}_{DD}^{b} = \boldsymbol{F}_{N\times\frac{N}{4}} \begin{pmatrix} \boldsymbol{h}_{DD}^{b} \\ \boldsymbol{0}_{(\frac{N}{4}-L)\times 1} \end{pmatrix}. \tag{35}$$

Using the above four expressions, the system model in (22) is represented as

$$\boldsymbol{y}_D(m,:) = \underbrace{\boldsymbol{s}_{P_m}\left(\boldsymbol{I}_4 \otimes \boldsymbol{F}_{N\times N/4}\right)}_{\tilde{\boldsymbol{A}}_m} \underbrace{\begin{pmatrix} \boldsymbol{h}_{SD}^{a} \\ \boldsymbol{0}_{(\frac{N}{4}-L)\times 1} \\ \boldsymbol{h}_{SD}^{b} \\ \boldsymbol{0}_{(\frac{N}{4}-L)\times 1} \\ \boldsymbol{h}_{DD}^{a} \\ \boldsymbol{0}_{(\frac{N}{4}-L)\times 1} \\ \boldsymbol{h}_{DD}^{b} \\ \boldsymbol{0}_{(\frac{N}{4}-L)\times 1} \end{pmatrix}}_{\tilde{\boldsymbol{x}}} + \boldsymbol{w}(m,:). \tag{36}$$

Similar to (23), we have

$$\underbrace{\begin{pmatrix} \boldsymbol{y}_D(1,:) \\ \vdots \\ \boldsymbol{y}_D(M,:) \end{pmatrix}}_{\boldsymbol{y}_D} = \underbrace{\begin{pmatrix} \tilde{\boldsymbol{A}}_1 \\ \vdots \\ \tilde{\boldsymbol{A}}_M \end{pmatrix}}_{\tilde{\boldsymbol{A}}} \tilde{\boldsymbol{x}} + \underbrace{\begin{pmatrix} \boldsymbol{w}(1,:) \\ \vdots \\ \boldsymbol{w}(M,:) \end{pmatrix}}_{\boldsymbol{w}}, \tag{37}$$

which is simplified as

$$\boldsymbol{y}_D = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} + \boldsymbol{w}. \tag{38}$$

If and only if $(\tilde{\boldsymbol{A}}^{\mathrm{H}}\tilde{\boldsymbol{A}})$ is nonsingular, the unique LS estimate of $\tilde{\boldsymbol{x}}$ in (38) is given by

$$\hat{\tilde{\boldsymbol{x}}} = \left(\tilde{\boldsymbol{A}}^{\mathrm{H}}\tilde{\boldsymbol{A}}\right)^{-1}\tilde{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{y}_D. \tag{39}$$

On the basis of (36), (37), and (39), we design the noise variance estimator as follows:

$$\hat{\sigma}_w^2 = \frac{1}{4(N-L)}\sum_j\sum_i |\hat{\tilde{\boldsymbol{x}}}(i+jN/4)|^2, \quad i \in (L+1, N/4)\,,\ j \in \{0,1,2,3\}\,, \tag{40}$$

which gives the estimated sparsity K of $\boldsymbol{x}$ in (23):

$$\hat{K} = \sum_{i=1}^{N} I\left(\hat{\tilde{x}}(n)\right), \tag{41}$$

where the function $I(\cdot)$ is the indicator function defined as

$$I\left(\hat{x}'(n)\right) = \begin{cases} 1, & |\hat{\tilde{x}}(n)| \geqslant g; \\ 0, & |\hat{\tilde{x}}(n)| < g \end{cases} \tag{42}$$

with $g^2 = \gamma\hat{\sigma_w^2}$, where the optimal value of g is attained by simulation in the next section. Finally, the detailed procedure for the proposed adaptive OMP is summarized below.

Algorithm 1 Proposed adaptive OMP algorithm

Input: $\boldsymbol{A}$, $\boldsymbol{z}$
Output: $\hat{\tilde{\boldsymbol{x}}}$
Initialization: $K = 0$
Steps:
1: Compute $\hat{\tilde{\boldsymbol{x}}}$ using (39) and estimate $\hat{\sigma}_w^2$ using (40).
2: Set the nice threshold γ by simulation.
3: Estimate the K-sparsity for the OMP using (41).
4: Call the OMP method in [26] to obtain the sparse version of $\boldsymbol{x}$ in (23).

4.3 Gradient projection method

The solving of $\boldsymbol{x}$ in (23) can be transformed into the following optimization problem:

$$\min_{\boldsymbol{x}} \ \| \boldsymbol{x} \|_0, \quad \text{s.t.} \ \ \boldsymbol{y}_D = \boldsymbol{A}\boldsymbol{x} \tag{43}$$

that is an NP-hard problem. Replacing the l_0-norm by the l_1-norm in the objective function leads to the following optimization:

$$\min_{\boldsymbol{x}} \ \frac{1}{2}\| \boldsymbol{y}_D - \boldsymbol{A}\boldsymbol{x} \|_2^2 + \beta \| \boldsymbol{x} \|_1 \tag{44}$$

that can be further cast in a real-valued form:

$$\min_{\overline{\boldsymbol{x}}} \ \frac{1}{2}\| \overline{\boldsymbol{z}} - \overline{\boldsymbol{A}}\overline{\boldsymbol{x}} \|_2^2 + \beta \| \overline{\boldsymbol{x}} \|_1, \tag{45}$$

where

$$\overline{\boldsymbol{z}} = \begin{bmatrix} \mathrm{Re}(\boldsymbol{y}_D) \\ \mathrm{Im}(\boldsymbol{y}_D) \end{bmatrix}, \quad \overline{\boldsymbol{x}} = \begin{bmatrix} \mathrm{Re}(\boldsymbol{x}) \\ \mathrm{Im}(\boldsymbol{x}) \end{bmatrix}, \quad \overline{\boldsymbol{A}} = \begin{bmatrix} \mathrm{Re}(\boldsymbol{A}) & -\mathrm{Im}(\boldsymbol{A}) \\ \mathrm{Im}(\boldsymbol{A}) & \mathrm{Re}(\boldsymbol{A}) \end{bmatrix}. \tag{46}$$

Let us define $\overline{\boldsymbol{x}} = \boldsymbol{u} - \boldsymbol{v}$, where $\boldsymbol{u} \succeq \mathbf{0}_{8L\times 1}$ and $\boldsymbol{u} \succeq \mathbf{0}_{8L\times 1}$ with $\succeq$ are generalized inequalities. The optimization problem (45) is converted into a quadratic program:

$$\begin{aligned} \min_{\boldsymbol{u},\ \boldsymbol{v}} \quad & \frac{1}{2} \parallel \overline{\boldsymbol{z}} - \overline{\boldsymbol{A}}(\boldsymbol{u} - \boldsymbol{v}) \parallel_2^2 + \beta \mathbf{1}_{8L}^{\mathrm{T}} \boldsymbol{u} + \beta \boldsymbol{I}_{8L}^{\mathrm{T}} \boldsymbol{v}, \\ \text{s.t.} \quad & \boldsymbol{u} \succeq \mathbf{0}_{8L\times 1}, \boldsymbol{v} \succeq \mathbf{0}_{8L\times 1}, \end{aligned} \tag{47}$$

where $\mathbf{1}_{8L}$ is an $8L-$D column vector containing all ones,

$$\begin{aligned} \overline{\overline{\boldsymbol{x}}} &= \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix}, \quad \boldsymbol{b} = \overline{\boldsymbol{A}}^{\mathrm{T}} \overline{\boldsymbol{z}}, \\ \boldsymbol{c} &= \beta \mathbf{1}_{16L} + \begin{bmatrix} -\boldsymbol{b} \\ \boldsymbol{b} \end{bmatrix}, \end{aligned} \tag{48}$$

and

$$\boldsymbol{B} = \begin{bmatrix} \overline{\boldsymbol{A}}^{\mathrm{T}} \overline{\boldsymbol{A}} & -\overline{\boldsymbol{A}}^{\mathrm{T}} \overline{\boldsymbol{A}} \\ -\overline{\boldsymbol{A}}^{\mathrm{T}} \overline{\boldsymbol{A}} & \overline{\boldsymbol{A}}^{\mathrm{T}} \overline{\boldsymbol{A}} \end{bmatrix}. \tag{49}$$

The above optimization finally becomes

$$\begin{aligned} \min_{\overline{\overline{\boldsymbol{x}}}} \quad & F(\overline{\overline{\boldsymbol{x}}}) = \frac{1}{2} \overline{\overline{\boldsymbol{x}}}^{\mathrm{T}} \boldsymbol{B} \overline{\overline{\boldsymbol{x}}} + \boldsymbol{c}^{\mathrm{T}} \overline{\overline{\boldsymbol{x}}}, \\ \text{s.t.} \quad & \overline{\overline{\boldsymbol{x}}} \succeq \mathbf{0}_{16L\times 1}, \end{aligned} \tag{50}$$

where $\overline{\overline{\boldsymbol{x}}} \succeq \mathbf{0}_{16L\times 1}$ implies that $\overline{\overline{\boldsymbol{x}}}$ is a positive semi-definite matrix. The above optimization problem can be solved by the GP method in [22].

4.4 Interference cancellation and detection

As we have estimated the CIRs of the source-to-destination and the destination-to-destination channels in the previous subsections, it is easy to obtain the corresponding frequency-domain transfer functions $\boldsymbol{H}_{SD}^{a}$, $\boldsymbol{H}_{SD}^{b}$, $\boldsymbol{H}_{DD}^{a}$, and $\boldsymbol{H}_{DD}^{b}$ using (14)–(17). Considering that the destination receiver has a knowledge of its own transmit data completely, after removing the self-interference, the received mth OFDM symbol is represented as

$$\tilde{\boldsymbol{y}}_D(m,:) = \boldsymbol{y}_D(m,:) - \mathrm{Diag}\{\hat{\boldsymbol{H}}_{DD}^{a}\} \boldsymbol{s}_D(m,:) - \mathrm{Diag}\{\hat{\boldsymbol{H}}_{DD}^{b}\} \boldsymbol{s}_D^{\#}(m,:), \tag{51}$$

where $m > M$. Based on the above expression, the ML detector corresponding to the nth pair of subchannels is represented as

$$\begin{aligned} &\begin{pmatrix} \hat{\boldsymbol{s}}_S(m,n) \\ \hat{\boldsymbol{s}}_S^*(m,N+2-n) \end{pmatrix} \\ &= \underset{\left[\boldsymbol{s}_S(m,\ n),\ \boldsymbol{s}_S^*(m,\ N+2-n)\right]^{\mathrm{T}}}{\arg\min} \left\{ \begin{pmatrix} \triangle \boldsymbol{y}_D(m,n) \\ \triangle \boldsymbol{y}_D^*(m,N+2-n) \end{pmatrix}^{\mathrm{H}} \begin{pmatrix} \triangle \boldsymbol{y}_D(m,n) \\ \triangle \boldsymbol{y}_D^*(m,N+2-n) \end{pmatrix} \right\}, \end{aligned} \tag{52}$$

where

$$\begin{aligned} \begin{pmatrix} \triangle \boldsymbol{y}_D(m,n) \\ \triangle \boldsymbol{y}_D^*(m,N+2-n) \end{pmatrix} &= \begin{pmatrix} \tilde{\boldsymbol{y}}_D(m,n) \\ \tilde{\boldsymbol{y}}_D^*(m,N+2-n) \end{pmatrix} \\ &- \begin{bmatrix} \hat{\boldsymbol{H}}_{SD}^{a}(n) & \hat{\boldsymbol{H}}_{SD}^{b}(n) \\ \left(\hat{\boldsymbol{H}}_{SD}^{b}\right)^{*}(N+2-n) & \left(\hat{\boldsymbol{H}}_{SD}^{a}\right)^{*}(N+2-n) \end{bmatrix} \begin{pmatrix} \boldsymbol{s}_S(m,n) \\ \boldsymbol{s}_S^*(m,N+2-n) \end{pmatrix}. \end{aligned} \tag{53}$$

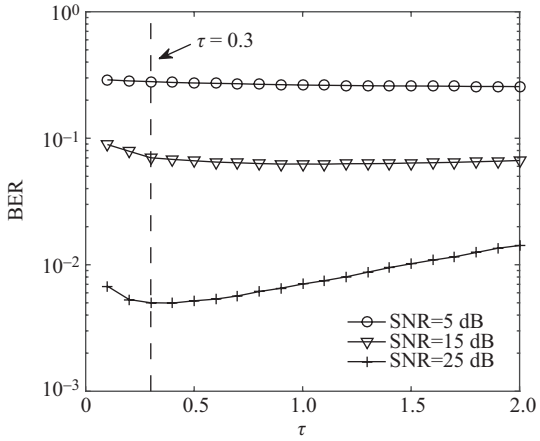

Figure 3 Curves of the BER versus the β of the GP in [22] with three different SNRs.

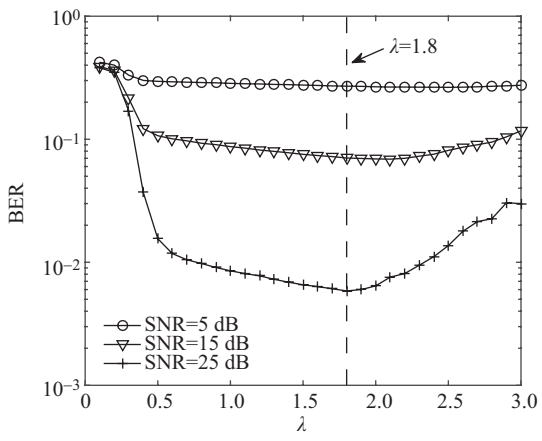

Figure 4 Curves of the BER vesus the γ of the proposed adaptive OMP with three different SNRs.

It is obvious that the ML detection in (52) requires J^2 complex multiplications (CMs) per pair of sub-channels, where J is the size of the signal constellation. As J increases, the computational complexity of the ML detector becomes a power increasing function of J. To reduce the complexity of the ML detection, the sphere decoding method (SD) in [27] is adopted to provide a low-complexity implementation of the ML detection, while approximately achieving the same performance as that of the ML.

5 Simulation and discussion

In our simulation, the system parameters are set as follows: $N = 256$, $L = 48$, signal bandwidth BW = 5 MHz, digital modulation is 64 QAM, and carrier frequency $f_c = 2$ GHz. A typical EVA channel model with the maximum path delay being 2.51 μs in the LTE standard is employed in the following simulation. $\alpha_{t,S}$, $\alpha_{t,D}$, and $\alpha_{r,D}$ mentioned below are set to 1 dB, whereas $\theta_{t,S}$, $\theta_{t,D}$, and $\theta_{r,D}$ are set to 2°.

Figure 3 shows the curves of the BER versus the β of the GP algorithm for three different values of the SNR, 5 dB, 15 dB, and 25 dB. It is evident that each SNR has an optimal β that achieves a minimum BER. From Figure 3, the value of β is selected as 0.3; the GP achieves almost an optimal BER performance for all the three SNRs, 5 dB, 15 dB, and 25 dB. Thus, in the following simulation, the value of β in the GP is set to 0.3.

Figure 4 illustrates the curves of the BER versus the γ of the proposed adaptive OMP algorithm for three different SNRs, 5 dB, 15 dB, and 25 dB. From Figure 4, it can be seen that the proposed adaptive OMP achieves an almost optimal BER performance at $\gamma = 1.8$ for the three different SNRs. Thus, the value of γ in the proposed adaptive OMP is set to 1.8, in the following simulation.

Figure 5 makes a performance comparison amongst the TD-LS, adaptive OMP, and GP. The last two methods demonstrate a better BER performance than the TD-LS. The proposed adaptive OMP performs marginally better than the GP in the low and medium SNR regions, and slightly worse than the GP in the high SNR region.

Figure 6 demonstrates the curves of the MSE versus the SNR for the GP, TD-LS, and adaptive the OMP, where MSE indicates the mean square error [28]. The same performance trend as shown in Figure 5 is obtained. The MSE curves of the GP and adaptive OMP intersect at the same SNR point as the BER curves.

6 Conclusion

We have investigated the channel estimation in full-duplex OFDM systems with IQ-imbalances. Two

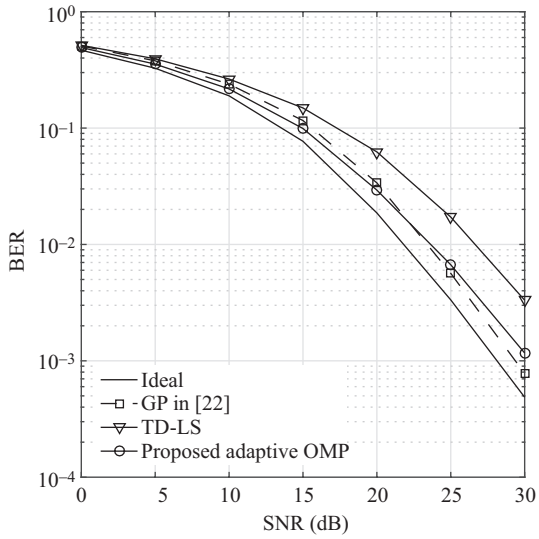

Figure 5 Curves of the BER versus the SNRs of different channel estimators.

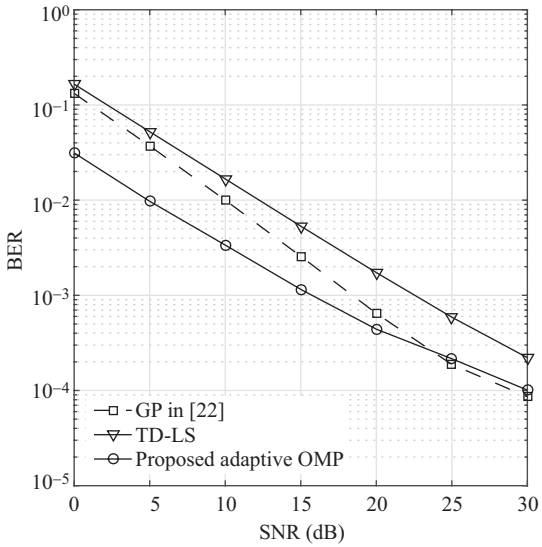

Figure 6 Curves of the MSE versus the SNRs of different channel estimators.

high-performance channel estimators, the TD-LS and the adaptive OMP, have been proposed to estimate both the self-interference (full-duplex) and the intended channels. The estimated self-interference channel is used to cancel the full-duplex self-interference, and the estimated intended channel works along with the sphere decoding to detect useful data symbols at the receiver. In accordance with the simulation results, the proposed adaptive OMP achieves a BER performance close to the ideal case and performs better than the TD-LS and GP in the low and medium SNR regions.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 61271230, 61472190, 6147210, 61501238), Open Research Fund of National Key Laboratory of Electromagnetic Environment, China Research Institute of Radiowave Propagation (Grant No. 201500013), Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (Grant No. 2013D02), and Research Fund for Doctoral Program of Higher Education of China (Gant No. 20113219120019)

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Riihonen T, Werner S, Wichman R, et al. On the feasibility of full-duplex relaying in the presence of loop interference. In: Proceedings of IEEE 10th Workshop on Signal Processing Advances in Wireless Communications, Perugia, 2009. 275–279
2 Ju H, Oh E, Hong D. Catching resource-devouring worms in next-generation wireless relay systems: two-way relay and full-duplex relay. IEEE Commun Mag, 2009, 47: 58–65
3 Bliss D W, Parker P A, Margetts A R. Simultaneous transmission and reception for improvedwireless network performance. In: Proceedings of IEEE/SP 14th Workshop on Statistical Signal Processing, Madison, 2007. 478–482
4 Kim T M, Yang H J, Paulraj A J. Distributed sum-rate optimization for full-duplex MIMO System under limited dynamic range. IEEE Signal Proc Let, 2013, 20: 555–558
5 Riihonen T, Werner S, Wichman R. Hybrid full-duplex/half-duplex relaying with transmit power adaptation. IEEE Trans Wirel Commun, 2011, 10: 3074–3085
6 Ahmed M, Le-Ngoc T. Channel estimation and self-interference cancellation in full-duplex Communication Systems. IEEE Trans Veh Technol, 2016, 66: 321–334
7 Zhang G P, Yang K, Liu P, et al. Using full duplex relaying in device-to-device (D2D) based wireless multicast services: a two-user case. Sci China Inf Sci, 2015, 58: 082301
8 Day B P, Margetts A R, Bliss D W, et al. Full-duplex bidirectional MIMO: achievable rates under limited dynamic range. IEEE Trans Signal Proces, 2011, 60: 3702–3713
9 Sabharwal A, Schniter P, Guo D N, et al. In-band full-duplex wireless: challenges and opportunities. IEEE J Sel Area Comm, 2013, 32: 1637–1652

10 Koohian A, Mehrpouyan H, Ahmadian M, et al. Bandwidth efficient channel estimation for full duplex communication systems. In: Proceedings of IEEE International Conference on Communications, London, 2015. 4710–4714
11 Masmoudi A, Le-Ngoc T. A maximum-likelihood channel estimator in MIMO full-duplex systems. In: Proceedings of IEEE Vehicular Technology Conference, Vancouver, 2014. 2863–2865
12 Masmoudi A, Le-Ngoc T. A maximum-likelihood channel estimator for self-interference cancellation in full-duplex systems. IEEE Trans Veh Technol, 2015, 65: 5122–5132
13 Li Y B. In-phase and Quadrature Imbalance—Modeling, Estimation, and Compensation. New York: Springer-Verlag, 2014
14 Tarighat A, Sayed A H. Joint compensation of transmitter and receiver impairments in OFDM systems. IEEE Trans Wirel Commun, 2007, 6: 240–247
15 Zou Q Y, Tarighat A, Sayed A H. Joint compensation of IQ imbalance and phase noise in OFDM wireless systems. IEEE Trans Commun, 2009, 57: 404-414
16 Tarighat A, Bagheri R, Sayed A H. Compensation schemes and performance analysis of IQ imbalances in OFDM receivers. IEEE Trans Signal Proces, 2005, 53: 3257–3268
17 Lopez-Estraviz E, De Rore S, Horlin F, et al. Pilot design for joint channel and frequency-dependent transmit/receive IQ imbalance estimation and compensation in OFDM-based transceivers. In: Proceedings of IEEE International Conference on Communications, Glasgow, 2007. 4861–4866
18 Minn H, Munoz D. Pilot Designs for channel estimation of MIMO OFDM systems with frequency-dependent I/Q imbalances. IEEE Trans Commun, 2010, 58: 2252–2264
19 Shu F, Zhao J H, You X H, et al. An efficient sparse channel estimator combining time-domain LS and iterative shrinkage for OFDM systems with IQ-imbalances. Sci China Inf Sci, 2012, 55: 2604–2610
20 Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. Siam Review, 2001, 43: 129–159
21 Zhang Y. Theory of compressive sensing via ℓ1-minimization: a non-RIP analysis and extensions. J Oper Res Soc China, 2013, 1:79–105
22 Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE J Sel Top Signal, 2007, 1: 586–597
23 Cotter S F, Rao B D. Sparse channel estimation via matching pursuit with application to equalization. IEEE Trans Commun, 2002, 50: 374–377
24 Wang D M, Han B, Zhao J H, et al. Channel estimation algorithms for broadband MIMO-OFDM sparse channel. In: Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communication, Beijing, 2003. 1929–1933
25 Wu C J, Lin D W. A group matching pursuit algorithm for sparse channel estimation for OFDM transmission. In: Proceedings of IEEE International Conference on Acoust, Speech and Signal Process, Toulouse, 2006. 429–432
26 Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: Proceedings of 23th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, 1993. 40–44
27 Hassibi B, Vikalo H. On the sphere-decoding algorithm I. Expected complexity. IEEE Trans Signal Proces, 2005, 53: 2806–2818
28 Shu F, Lee J, Wu L N, et al. Time-frequency channel estimation for digital amplitude modulation broadcasting systems based on OFDM. IEEE Proc Commun, 2003, 150: 259–264

Appendix A Proof of Theorem 1

Proof. According to the definition in Candès' literature[1)], for each integer $S = 1, 2, \ldots$, let us define the isometry constant δ_S of a given matrix $\boldsymbol{A}$ as the smallest number such that

$$(1 - \delta_S)\|\boldsymbol{x}\|_2^2 \leqslant \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leqslant (1 + \delta_S)\|\boldsymbol{x}\|_2^2 \tag{A1}$$

holds for all the S-sparse vectors. Then, matrix $\boldsymbol{A}$ is said to have an S-RIP with $\delta_S \in (0, 1)$.

Therefore, $\boldsymbol{A}$ satisfies the RIP, if the singular values of all submatrices $\boldsymbol{A}_T$ formed by selecting no more than S columns indexed by $T \subset \{1, \ldots, 4L\}$ from the set of all the columns of $\boldsymbol{A}$, are within $(\sqrt{1 - \delta_S}, \sqrt{1 + \delta_S})$. Notice that the singular values of $\boldsymbol{A}_T$ are the square root of the eigenvalues of the Gram matrix, $\boldsymbol{G}_T = \boldsymbol{A}_T^{\mathrm{H}}\boldsymbol{A}_T$.

According to the Gĕrsgorin's Disc Theorem in Varga's literature[2)], the eigenvalues of $\boldsymbol{G}_T$ lie in the union of the S discs centered at $\boldsymbol{G}_T(i, i)$ with a radius, $\sum_{j\neq i, j=1}^{S} |\boldsymbol{G}_T(i, j)|$. As all the diagonal elements of $\boldsymbol{G}_T$ are equal to $1/N$, we can obtain $\delta_d = 1 - 1/N$. For a certain value δ_o which satisfies $\delta_d + \delta_o = \delta_S \in (0, 1)$, if every off-diagonal element satisfies $|\boldsymbol{G}_T(i, j)| < \delta_o/S$, then, the radius of each disc will be $\sum_{j\neq i, j=1}^{S} |\boldsymbol{G}_T(i, j)| < \delta_o$. Thus, all the eigenvalues of $\boldsymbol{G}_T$ are in the interval $(1 - \delta_d - \delta_o, 1 - \delta_d + \delta_o) \subset (1 - \delta_S, 1 + \delta_S)$.

For a given subset T the off-diagonal element $\boldsymbol{G}_T(i, j)$ is the inner product between the T_i-th and T_j-th columns of matrix $\boldsymbol{A}$, where $T_i = k_i L + p_i$, $k_i \in [0, 3]$, and $p_i \in [1, L]$; $T_j = k_j L + p_i$, $k_j \in [0, 3]$ and $p_j \in [1, L]$. Because the source and destination nodes take different cyclic shifted versions of the basic sequence, $\boldsymbol{x}_{k_j}$ and $\boldsymbol{x}_{k_i}$ should be the cyclic shift version of $\boldsymbol{x}$ or the reverse cycle shifted version of $\boldsymbol{x}^*$. Without loss of generality, assuming $T_i < T_j$, the off-diagonal element

1) Candès E J. The restricted isometry property and its implications for compressed sensing. Compt Rend Mathem, 2008, 346: 589–592
2) Varga R S. Geršgorin and His Circles. Heidelberg: Springer-Verlag, 2011

becomes

$$\boldsymbol{G}_T(i,j)=\frac{1}{N}\Big[x^*_{k_j,1}x_{k_i,1}\boldsymbol{F}_{N\times L}(1,p_j-p_i)+x^*_{k_j,2}x_{k_i,2}\boldsymbol{F}_{N\times L}(2,p_j-p_i)\\+\cdots+x^*_{k_j,N}x_{k_i,N}\boldsymbol{F}_{N\times L}(N,p_j-p_i)\Big]. \tag{A2}$$

We consider two scenarios below:
Case 1. $k_i=k_j, p_i\neq p_j$.

$$\boldsymbol{G}_T(i,j)=\frac{1}{N^2}\sum_{n=1}^{N}\boldsymbol{F}_{N\times L}(n,p_j-p_i)=0. \tag{A3}$$

Case 2. $k_i\neq k_j$.

It is obvious that the entries of (A2) are not independent because x_n may exist in two different terms. For extreme cases such as $\boldsymbol{x}_{k_j}=\boldsymbol{x}^{\#}_{k_i}$, we have

$$\boldsymbol{G}_T(i,j)=\frac{1}{N}\Big[x^*_{k_i,1}x_{k_i,1}\boldsymbol{F}_{N\times L}(1,p_j-p_i)+x^*_{k_i,N}x_{k_i,2}\boldsymbol{F}_{N\times L}(2,p_j-p_i)\\+x^*_{k_i,N-1}x_{k_i,3}\boldsymbol{F}_{N\times L}(2,p_j-p_i)\\+\cdots+x^*_{k_i,N/2+1}x_{k_i,N/2+1}\boldsymbol{F}_{N\times L}(2,p_j-p_i)\\+\cdots+x^*_{k_i,2}x_{k_i,N}\boldsymbol{F}_{N\times L}(N,p_j-p_i)\Big]. \tag{A4}$$

Observing the above right-hand expression, we find the first and middle terms are constant, and the second and last terms both depend on $x_{k_i,2}$ and $x_{k_i,N}$; however, the second and third terms are independent. Therefore, we split the summation of the two i.i.d random variables and each split sum is formed by grouping the alternating terms. Rewriting the summation as

$$\boldsymbol{G}_T(i,j)=\sum_{i}^{N}z_n \tag{A5}$$

that can be decomposed as

$$\boldsymbol{G}_T(i,j)=z_1+z_{N/2+1}+\underbrace{\sum_{n=2}^{N/2}z_n}_{\boldsymbol{G}^1_T(i,j)}+\underbrace{\sum_{n=N/2+2}^{N}z_n}_{\boldsymbol{G}^2_T(i,j)}. \tag{A6}$$

Applying Lemma 4 in Haupt's literature[3)], we have

$$\begin{aligned}P\left(|\boldsymbol{G}_T(i,j)|\geqslant\frac{\delta_o}{S}\right)&\leqslant P\left(|\boldsymbol{G}^1_T(i,j)+\boldsymbol{G}^2_T(i,j)|\geqslant\frac{\delta_o}{S}-\frac{2}{N^2}\right)\\&\leqslant P\left(|\boldsymbol{G}^1_T(i,j)|\geqslant\frac{\delta_o}{2S}-\frac{1}{N^2}or|\boldsymbol{G}^2_T(i,j)|\geqslant\frac{\delta_o}{2S}-\frac{1}{N^2}\right)\\&\leqslant 2\max\left\{P\left(|\boldsymbol{G}^1_T(i,j)|\geqslant\frac{\delta_o}{2S}-\frac{1}{N^2}\right),P\left(|\boldsymbol{G}^2_T(i,j)|\geqslant\frac{\delta_o}{2S}-\frac{1}{N^2}\right)\right\}\\&\leqslant 4\exp\left(-\frac{(\frac{N\delta_o}{2S}-\frac{1}{N})^2}{N-2}\right)<4\exp\left(-(\frac{\sqrt{N}\delta_o}{2S}-\frac{1}{N\sqrt{N}})^2\right).\end{aligned} \tag{A7}$$

In general, considering all the off-diagonal elements, the number of nonzero entries is $12L^2$; applying the above inequality yields:

$$P\left(\bigcup_{i\neq j}|\boldsymbol{G}_T(i,j)|\geqslant\frac{\delta_o}{S}\right)<48L^2\exp\left(-\left(\frac{\sqrt{N}\delta_o}{2S}-\frac{1}{N\sqrt{N}}\right)^2\right), \tag{A8}$$

which can be viewed as the required condition on the off-diagonal elements of the Gram matrix. Assuming $\delta_o=\delta_S/4$, we have

$$P\left(\boldsymbol{A}\text{ does not satisfy RIP}\right)<48L^2\exp\left(-\left(\frac{\sqrt{N}\delta_S}{8S}-\frac{1}{N\sqrt{N}}\right)^2\right). \tag{A9}$$

Defining $c_1<\delta_S/8$ and $c_2=48L^2$, the matrix $\boldsymbol{A}$ satisfies the RIP with a probability exceeding

$$1-\exp\left(-\left(\frac{c_1\sqrt{N}}{S}-\frac{1}{N\sqrt{N}}\right)^2\right) \tag{A10}$$

for a sufficiently large N. This completes the proof of Theorem 1.

3) Haupt J, Bajwa W U, Raz G, et al. Toeplitz compressed sensing matrices with applications to sparse channel estimation. IEEE Trans Inform Theory, 2010, 56: 5862–5875