

An Augmented Nonlinear LMS for Digital Self-Interference Cancellation in Full-Duplex Direct-Conversion Transceivers

Zhe Li, *Student Member, IEEE*, Yili Xia, *Member, IEEE*, Wenjiang Pei, Kai Wang, and Danilo P. Mandic, *Fellow, IEEE*

Abstract—In future full-duplex communications, the cancellation of self-interference (SI) arising from hardware non-idealities will play an important role in the design of mobile-scale devices. To this end, we introduce an optimal digital SI cancellation solution for shared-antenna-based direct-conversion transceivers. To establish that the underlying widely linear signal model is not adequate for strong transmit signals, the impact of various circuit imperfections, including power amplifier (PA) distortion, frequency-dependent I/Q imbalance, quantization noise and thermal noise, on the performance of the conventional augmented least mean square (LMS) based SI canceller, is analyzed. In order to achieve a sufficient signal-to-interference-plus-noise ratio (SINR) when the nonlinear SI components are not negligible, we propose an augmented nonlinear LMS based SI canceller for a joint cancellation of both the linear and nonlinear SI components by virtue of a widely nonlinear model fit. A rigorous mean and mean square performance evaluation is conducted to justify the performance advantages of the proposed scheme over the conventional augmented LMS solution. Simulations on orthogonal frequency division multiplexing (OFDM)-based wireless local area network (WLAN) standard compliant waveforms support the analysis.

Index Terms—Full-duplex communication, I/Q imbalance, self-interference, augmented LMS, augmented nonlinear LMS, mean and mean square convergence analysis

I. INTRODUCTION

THE full-duplex (FD) technology aims at doubling the radio link data rate through simultaneous and bidirectional communication at the same center frequency, and is widely considered as a driving-force behind more spectrally efficient wireless networks and a potential candidate to fulfill the ambition of 5G to reach a 1000-fold gain in capacity [1], [2]. One of the major challenges in FD communications is the so-called self-interference (SI) problem, that is, a strong transmit signal coupled into the receiver (Rx) path. Since the transmitter (Tx)

and Rx chains are closely linked together in each transceiver node of FD communication systems, the SI power leaked into and reflected from the Tx chain could be even 50 dB to 110 dB higher than the Rx sensitivity level in either wireless local area network (WLAN) or cellular scenarios [3]–[5]. The design of FD transceivers has long been considered impossible for practical realizations and implementations, and it is only recently that their feasibility was experimentally demonstrated using the wireless open-access research platform (WARP) with WiFi waveforms [6]–[11]. Based on this promising result, it was recently suggested that a preferable FD network should consist of backhaul nodes operating in the FD mode and access nodes remaining in the legacy half-duplex (HD) mode [12]. However, recent studies have showed that operating access nodes in the FD mode significantly leverages the gain in degrees of freedom in either ergodic or fast-fading channel [13], and an imperative is to design a hardware structure suitable for mass-production. Owing to the physical constraints, such as small-size, low-cost and low-energy-consumption, direct-conversion transceivers are widely applied in HD wireless systems, and are also suitable for far-end device implementation in the context of FD communication systems.

In order to provide efficient SI cancellation, there exist numerous types of hardware solutions. According to the antenna placement strategies, these can be classified into separate-antennas-based [14], [15] and shared-antenna-based schemes [9], [10]. When each transceiver node is equipped with more than two separate antennas, SI attenuation can be achieved by improving electromagnetic insulation between the antennas. Owing to the inherent closed-loop within FD systems, the knowledge of the SI channel matrix can be obtained by either placing extra transmit antennas or allocating specific spatial resources [14], [15]. On the other hand, the shared-antenna-based design aims to separate the transmit and receive signals by sharing a common antenna [9], [10], the key component of which is a three port routing device, known as a circulator, used to isolate the incoming and outgoing signals. Requiring only off-the-shelf radio-frequency (RF) components, the shared-antenna structure stands out as a cost-effective and energy-saving choice for the design of mobile-scale FD transceivers. Demonstrations on the WARP have shown that in this way even 110 dB and 103 dB SI cancellation can be achieved in SISO [9] and MIMO systems respectively [10].

In a shared-antenna structure, it is necessary to consider

This work was partially supported by the National Natural Science Foundation of China under Grant 61401094 and Grant 61771124, the Natural Science Foundation of Jiangsu Province under Grant BK20140645, the Fundamental Research Funds for the Central Universities under Grant 2242016K41050, and the China Scholarship Council. (*Corresponding authors: Yili Xia; Wenjiang Pei.*)

Z. Li is with the School of Information Science and Engineering, Southeast University, 2 Sipailou, Nanjing 210096, P. R. China, and also with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: lizhe_nanjing@seu.edu.cn)

Y. Xia, W. Pei, and K. Wang are with the School of Information Science and Engineering, Southeast University, 2 Sipailou, Nanjing 210096, P. R. China. (e-mail: yili_xia@seu.edu.cn; wjpei@seu.edu.cn; kaiwang@seu.edu.cn)

D. P. Mandic is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: d.mandic@imperial.ac.uk).

further non-trivial analog and digital SI cancellation, due to the leakage of the circulator, single-path reflection from the antenna, and multi-path interference from the surrounding environment [16]. The purpose of analog SI cancellation is to prevent the saturation of the SI power level within the Rx low-noise amplifier (LNA), and meanwhile, to ensure that the difference between the power of residual SI and the received signal of interest does not exceed the dynamic range of an analog-to-digital converter (ADC) [17]. Subsequently, further digital cancellation is performed to deal with the residual SI components, as well as other RF circuit non-idealities, mainly including nonlinear distortion, I/Q imbalance and phase noise. The nonlinearity is largely caused by the power amplifier (PA), while I/Q imbalance and phase noise are mainly induced by the imperfect local oscillator (LO). The impact of PA nonlinear distortion on FD transceivers has been investigated in [18], [19], while the effect of phase noise was analyzed in [20], [21]. Since the I/Q imbalance is essentially reflected in the mismatch between in-phase and quadrature components of the complex-valued I/Q signal, it is also reflected in an image interference associated with the original signal [22]–[25].

The impact of the image interference caused by Tx I/Q imbalance on the SI cancellation has been studied in [26], indicating that it heavily limits the receiver path signal-to-interference-plus-noise ratio (SINR). However, due to size constraints of FD transceivers, the Rx and Tx are required to share a common imperfect LO, therefore, a more accurate analysis of image interference on SI cancellation should be performed by a joint consideration of both Tx and Rx I/Q imbalance. Motivated by this finding, a widely linear processing framework was developed in [27] to jointly suppress both the original transmit SI signal and its complex conjugate, i.e., the image interference, which arises due to the frequency-dependent I/Q imbalance in both the Tx and Rx. The cancellation parameters were subsequently estimated in the widely linear least squares sense. Although the model analysis in [27] has illustrated that under certain circumstances, e.g., for large transmit powers, the PA within the FD transceiver is likely to be operating close to or within its saturation region, consequently introducing third order nonlinear distortion, for mathematical simplicity, the higher order SI components have not been considered by the block-based SI canceller. A more general arbitrary nonlinear order PA and a simplified frequency-independent I/Q modulator were considered in [28], and the corresponding nonlinear SI cancellation was addressed by a model-fit widely nonlinear least squares approach.

By exploiting the advantages of adaptive estimation algorithms over block-based least squares ones, such as their lower computational complexities and faster adaptation for potential time-varying channels, the augmented (widely linear) least mean square (LMS) adaptive filtering algorithm [29]–[31] has been employed in a DSP-assisted analog SI cancellation process, and its theoretical performance in the presence of Tx and Rx IQ imbalance has been evaluated in [32]. For simplicity, the I/Q imbalance within transmitters and receivers were also considered to be frequency-independent in [32]. This is, however, not the case in wideband scenarios, since their frequency selectivity has been extensively reported and

justified in [23], [24]. Furthermore, although it has been illustrated by simulations that due to the undermodeling problem, the augmented LMS yields suboptimal SI cancellation results in the presence of PA nonlinearity, while a theoretical understanding of this suboptimality and ways of its mitigation are still lacking.

Therefore, in this paper, we first conduct a comprehensive mean and mean square performance analysis on the augmented LMS based SI canceller for wideband FD transceivers in the presence of both the PA nonlinear distortion and frequency-dependent image interference, to theoretically quantify its bias and variance increase, in the steady-state stage. The proposed analysis consequently facilitates a physical verification on how the augmented LMS SI canceller fails to achieve the required signal-to-noise ratio (SNR) when the transmit power is high enough. Next, in order to achieve a sufficient amount of SINR when the nonlinear SI components are not negligible, an augmented nonlinear LMS, whose underlying estimation framework generally takes into account both the nonlinear SI component and its associated image interference, is proposed for unbiased nonlinear SI cancellation, and a theoretical performance evaluation is conducted to demonstrate its performance advantages over the conventional augmented LMS. It is important to note that such further theoretical performance analysis is not an incremental step from the conventional augmented LMS, mainly due to the non-Gaussian nature of the higher order SI components in wideband scenarios. From the statistical perspective, this analysis also provides physical insights to the evaluation of the theoretical performance bounds on those block-based SI cancellers proposed in [27], [28]. Moreover, to facilitate the use of the proposed augmented nonlinear LMS based SI canceller in practical applications, a data pre-whitening scheme is employed to speed up its convergence. Simulations on orthogonal frequency division multiplexing (OFDM)-based WLAN standard compliant waveforms support the analysis.

Notations: Lowercase letters are used to denote scalars, a , boldface letters for column vectors, \mathbf{a} , and boldface uppercase letters for matrices, \mathbf{A} . The symbols $\mathbf{0}_N$ and $\mathbf{1}_N$ denote respectively an $N \times 1$ zero and unity vector. An $N \times N$ identity matrix is denoted by \mathbf{I}_N . The superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ denote respectively the complex conjugation, transpose, Hermitian transpose and matrix inversion operation. The operator $\text{Tr}\{\cdot\}$ represents the trace of a matrix, while the operators \otimes , $\|\cdot\|$ respectively denote the Kronecker product and Euclidean norm. The statistical expectation operator is denoted by $E\{\cdot\}$, matrix determinant by $\det\{\cdot\}$, while the operators $\Re\{\cdot\}$ and $\Im\{\cdot\}$ extract respectively the real and imaginary part of a complex variable and $j = \sqrt{-1}$. Matrix vectorization is designated by $\text{vec}\{\cdot\}$, which returns a column vector transformed by stacking the successive columns of matrix, and its inverse operation, i.e., restoring the matrix from the its vectorized form, is denoted by $\text{vec}^{-1}\{\cdot\}$. The extraction of matrix diagonal elements into a vector is denoted by $\text{diag}\{\cdot\}$. The operator $\lambda_{\max}\{\cdot\}$ returns the largest positive eigenvalue of a matrix.

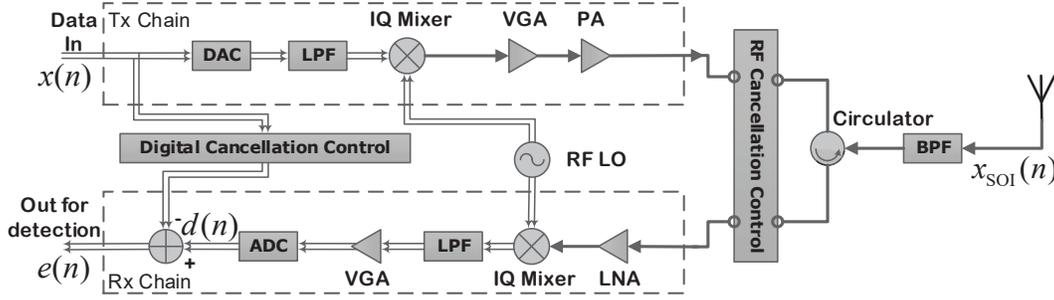


Fig. 1. The architecture of a shared-antenna FD transceiver.

II. FULL-DUPLEX TRANSCEIVER AND ITS WIDELY LINEAR BASEBAND EQUIVALENT MODEL

The structure of a typical shared-antenna FD direct-conversion transceiver is given in Fig. 1, and this structure is widely adopted in modern wireless transceivers, due to its simplicity [33]. In such an architecture, the leakage of the circulator, the single-path reflection from the antenna, and the multi-path interference from surrounding environment introduce plenty of residual self-interference (SI), which is firstly mitigated by the RF cancellation module and then suppressed by the digital baseband SI canceller [19], [27]. By considering the fact that low-cost RF components are preferable for the built-up of mobile transceivers, a precise baseband-equivalent system model that incorporates their prominent hardware non-idealities, such as Tx and Rx I/Q imbalance, PA distortion, ADC quantization noise and thermal noise, is a prerequisite for the digital SI cancellation control. This in fact yields a widely linear relation, which at a time instant n during the digital cancellation process, and between the observed signal $d(n)$ at the input of the canceller and its corresponding SI waveform $x(n)$, has the form [27]

$$d(n) = \mathbf{x}^T(n)\mathbf{h}^o + \mathbf{x}^H(n)\mathbf{g}^o + u(n) \quad (1)$$

where $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T$ is an SI vector of length M and is perfectly known by the receiver. The end-to-end channel impulse responses $\mathbf{h}^o = [h_1^o, h_2^o, \dots, h_M^o]^T$ and $\mathbf{g}^o = [g_1^o, g_2^o, \dots, g_M^o]^T$ are determined by frequency-dependent I/Q imbalance in both the transmitter and receiver, PA memory, and residual of analog cancellation. The composite noise term $u(n)$ represents the sum of interference components, including PA nonlinearity, thermal noise and quantization noise from an ADC, and is given by [27]

$$u(n) = \mathbf{x}_{\text{IMD}}^T(n)\mathbf{h}_{\text{IMD}}^o + \mathbf{x}_{\text{IMD}}^H(n)\mathbf{g}_{\text{IMD}}^o + v(n) + q(n) \quad (2)$$

where $\mathbf{x}_{\text{IMD}}(n) = [x_{\text{IMD}}(n), x_{\text{IMD}}(n-1), \dots, x_{\text{IMD}}(n-N+1)]^T$ and $x_{\text{IMD}}(n)$ represents third-order intermodulation (IMD) SI introduced by PA distortion, given by [27]

$$x_{\text{IMD}}(n) = k_{\text{TIQ}}^{3/2} |x(n)|^2 x(n) \quad (3)$$

The filter coefficients $\mathbf{h}_{\text{IMD}}^o = [h_{\text{IMD},1}^o, h_{\text{IMD},2}^o, \dots, h_{\text{IMD},N}^o]^T$ and $\mathbf{g}_{\text{IMD}}^o = [g_{\text{IMD},1}^o, g_{\text{IMD},2}^o, \dots, g_{\text{IMD},N}^o]^T$, where $N < M$, respectively represent the end-to-end channel impulse responses of the IMD SI component $\mathbf{x}_{\text{IMD}}(n)$ and its complex

conjugate $\mathbf{x}_{\text{IMD}}^*(n)$. In (2), the interference term $v(n)$ represents the thermal noise during the digital SI cancellation process, which is assumed to be a zero-mean complex-valued additive white Gaussian noise (AWGN), and its variance σ_v^2 is determined by [19], [27]

$$\sigma_v^2 = \frac{k_{\text{BB}} k_{\text{LNA}} k_{\text{RIQ}} p_{\text{sen}}}{\text{SNR}_{\text{req}}} \quad (4)$$

The quantization noise $q(n)$ in (2) is assumed to be another zero-mean AWGN process, whose variance σ_q^2 is computed as [34]

$$\sigma_q^2 = \frac{P_{\text{ADC}}}{10^{6.02\beta + 4.76 - \text{PAPR}/10}} \quad (5)$$

Physical meanings of other unmentioned parameters used from (3) to (5) are provided in Table I.

From (1), it is clear that in a feasible FD transceiver, the baseband signal before digital SI cancellation $d(n)$ is composed of various interference components, including SI $\mathbf{x}(n)$, IMD SI $\mathbf{x}_{\text{IMD}}(n)$, thermal noise $v(n)$ and their image counterparts, as well as quantization noise $q(n)$. In order to ascertain which components can be counted as primary interference under certain circumstances, simulations were carried out to illustrate their power relations within FD transceivers. According to the suggestions in [27], two types of practical FD transceivers were considered, the system parameters of which are listed in Table II. The main difference lies in their different analog SI cancellation capabilities. The RF separation and attenuation capabilities of the Type 1 transceiver are 40 dB and 30 dB respectively, while in a Type 2 transceiver, both were 10 dB lower. Consequently, the Type 2 transceiver exhibited an inferior analog SI cancellation capability compared to its Type 1 counterpart, weakening both the received signal of interest $x_{\text{SoI}}(n)$ and the thermal noise $v(n)$.

As shown in Fig. 2(a), in a Type 1 FD transceiver, the SI component $\mathbf{x}(n)$ and its image counterpart $\mathbf{x}^*(n)$ are both the dominant interference to the signal of interest in the entire transmit power range, and when the transmit power becomes higher, the IMD SI component $\mathbf{x}_{\text{IMD}}(n)$ linearly increased to become another major interference [27]. On the other hand, in Fig. 2(b), when the transmit power of a Type 2 FD transceiver went above 20 dBm, the thermal noise $v(n)$ became weaker than the quantization noise $q(n)$ and the image IMD SI $\mathbf{x}_{\text{IMD}}^*(n)$. This is because either a stronger nonlinear SI $\mathbf{x}_{\text{IMD}}(n)$ or a less efficient analog cancellation results in a lower receiver variable gain amplifier (VGA) gain k_{BB} .

TABLE I
USED NOTATIONS AND SYMBOLS.

Symbol	Denotation
k_{VGA}	Tx variable gain amplifier (VGA) gain
k_{BB}	Rx VGA gain
k_{LNA}	LNA gain
α_0	PA gain
α_1	Gain of PA nonlinearity
k_{TIQ}	Tx IQ mixer gain
k_{RIQ}	Rx IQ mixer gain
$f_{PA}(n)$	Memory polynomials of PA
$f_{RFE}(n)$	Estimation error of analog cancellation
p_{sen}	Receiver sensitivity
SNR_{req}	SNR requirement
p_{ADC}	Dynamic range of ADC
PAPR	Peak-to-average-power ratio
β	ADC bits

TABLE II
SYSTEM PARAMETERS OF TYPICAL FD TRANSCEIVERS.

Parameter	Value
Receiver sensitivity	-89 dBm
SNR requirement	15 dB
Thermal noise floor	-104 dBm
RF separation	40 dB (Type 1) 30 dB (Type 2)
RF attenuation	30 dB (Type 1) 20 dB (Type 2)
IRR	25 dB
Tx mixer gain	6 dB
Rx mixer gain	6 dB
PA gain	27 dB
PA IIP3	20 dBm
LNA gain	25 dB
Transmit power	-5 ~ 25 dBm
ADC dynamic range	7 dB
ADC bits	12
Peak-to-average-power ratio	10 dB

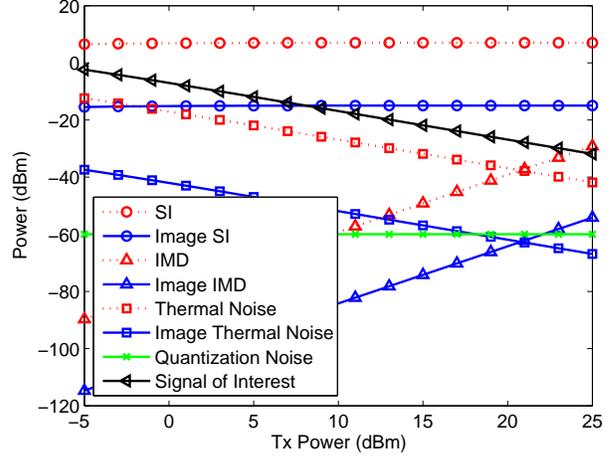
III. CONVENTIONAL AUGMENTED LMS BASED SI CANCELLER AND ITS PERFORMANCE ANALYSIS

In this section, we provide a comprehensive mean and mean square convergence analysis of the conventional augmented LMS, employed in [32] as a DSP-assisted analog SI cancellation process, in the presence of both frequency-dependent Tx and Rx I/Q imbalance and PA nonlinear distortion. For rigor, the proposed analysis covers both the cases of low and high transmit powers, as discussed in Section II.

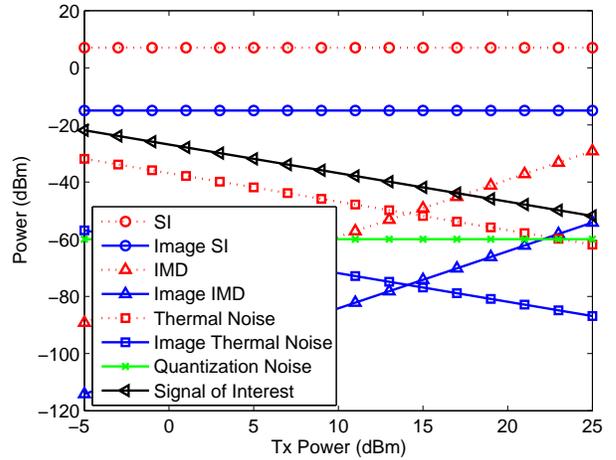
For the compactness of analysis, we first represent the widely linear model in (1) in an augmented form, given by

$$d(n) = \mathbf{x}^{aT}(n)\mathbf{w}^{ao} + u(n) \quad (6)$$

where $\mathbf{x}^a(n) = [\mathbf{x}^T(n), \mathbf{x}^H(n)]^T$ is the $2M \times 1$ augmented SI vector. In real-world wireless communications, the SI waveform $x(n)$ is always oversampled and bandlimited, and consequently non-white. However, for the simplicity of the analysis, we here assume $x(n)$ is critically sampled, so that, it can be modeled as a zero-mean proper white Gaussian random variable with variance σ_x^2 . Note that the Gaussianity and properness assumptions on $x(n)$ are valid for wideband OFDM waveforms. Indeed, the work in [35] has verified that a bandlimited uncoded OFDM symbol converges to a proper Gaussian random process as the number of subcarriers increases. The vector $\mathbf{w}^{ao} = [\mathbf{h}^{oT}, \mathbf{g}^{oT}]^T$ contains the augmented end-to-end system impulse responses, which models the transmit and receive frequency-dependent I/Q imbalance, PA distortion,



(a) Type 1 FD transceiver



(b) Type 2 FD transceiver

Fig. 2. Power comparison among different signal components in representative FD transceivers, before digital SI cancellation and against different levels of transmit powers. (a) Type 1 FD transceiver, and (b) Type 2 FD transceiver.

and the residual of analog SI cancellation.

The augmented LMS estimates the set of system parameters \mathbf{w}^{ao} by minimizing the MSE cost function $J^a(n)$, defined as [29]–[32]

$$J^a(n) = E[|e^a(n)|^2] = E[e^a(n)e^{a*}(n)] \quad (7)$$

where $e^a(n)$ is the instantaneous output error or residual SI, given by

$$e^a(n) = d(n) - \mathbf{x}^{aT}(n)\mathbf{w}^a(n) \quad (8)$$

and the augmented weight vector, that is, $\mathbf{w}^a(n)$, is updated as

$$\mathbf{w}^a(n+1) = \mathbf{w}^a(n) + \mu e^a(n)\mathbf{x}^{a*}(n) \quad (9)$$

where μ is the step size.

A. Mean Convergence Analysis

Upon introducing the $2M \times 1$ weight error vector

$$\tilde{\mathbf{w}}^a(n) = \mathbf{w}^a(n) - \mathbf{w}^{ao} \quad (10)$$

the output residual SI $e^a(n)$ in (8) now becomes

$$e^a(n) = u(n) - \mathbf{x}^{aT}(n)\tilde{\mathbf{w}}^a(n) \quad (11)$$

From (9), the recursion for the update of the weight error vector $\tilde{\mathbf{w}}^a(n)$ can be derived as

$$\tilde{\mathbf{w}}^a(n+1) = [\mathbf{I}_{2M} - \mu \mathbf{x}^{a*}(n) \mathbf{x}^{aT}(n)] \tilde{\mathbf{w}}^a(n) + \mu u(n) \mathbf{x}^{a*}(n) \quad (12)$$

By applying the statistical expectation operator $E[\cdot]$ to both sides of (12) and upon employing the standard independence assumptions [36], [37], that is, the composite noise $u(n)$ is statistically independent of any other variable within the augmented LMS, and $\tilde{\mathbf{w}}^a(n)$ is statistically independent of the augmented SI input vector $\mathbf{x}^a(n)$, we have

$$E[\tilde{\mathbf{w}}^a(n+1)] = (\mathbf{I}_{2M} - \mu \mathbf{R}^{a*}) E[\tilde{\mathbf{w}}^a(n)] + \mu E[u(n) \mathbf{x}^{a*}(n)] \quad (13)$$

where \mathbf{R}^a is the covariance matrix of the augmented SI vector $\mathbf{x}^a(n)$, defined as

$$\mathbf{R}^a = E[\mathbf{x}^a(n) \mathbf{x}^{aH}(n)] = \sigma_x^2 \mathbf{I}_{2M} \quad (14)$$

Therefore, the convergence of the augmented LMS in the mean weight error sense is guaranteed if the step-size μ satisfies [37]

$$0 < \mu < \frac{2}{\lambda_{\max}[\mathbf{R}^a]} = \frac{2}{\sigma_x^2} \quad (15)$$

From Fig. 2(a) and Fig. 2(b) in Section II, we observe that the relative power relationships among different SI components vary as the transmit power changes. Therefore, in order to accurately describe the statistical mean behavior of the augmented LMS based SI canceller, we consider the following two case studies.

1) *Low transmit power.* In this situation, both the nonlinear distortion component $\mathbf{x}_{\text{IMD}}(n)$ and the quantization noise $q(n)$ at the receiver are negligible, since their powers are much weaker than that of the thermal noise $v(n)$ in (2). Therefore, from (6), we have $u(n) \simeq v(n)$, and hence, at the steady-state, as, $n \rightarrow \infty$, from (13) we have

$$E[\tilde{\mathbf{w}}^a(\infty)] = \mathbf{0} \quad (16)$$

2) *High transmit power.* When the transmit power is high, the PA nonlinear distortion component $\mathbf{x}_{\text{IMD}}(n)$ becomes the third strongest interference among all the imperfections considered. After employing the independence assumptions and (2), the second term on the right hand side (RHS) of (13) can be derived as

$$\begin{aligned} E[u(n) \mathbf{x}^{a*}(n)] &= E[(\mathbf{x}_{\text{IMD}}^T(n) \mathbf{h}_{\text{IMD}}^0 + \mathbf{x}_{\text{IMD}}^H(n) \mathbf{g}_{\text{IMD}}^0) \mathbf{x}^{a*}(n)] \\ &= k_{\text{TIQ}}^{3/2} \left\{ \sum_{i=1}^N E[h_{\text{IMD},i} |x(n-i+1)|^2 x(n-i+1) \mathbf{x}^{a*}(n)] \right. \\ &\quad \left. + \sum_{i=1}^N E[g_{\text{IMD},i} |x(n-i+1)|^2 x^*(n-i+1) \mathbf{x}^{a*}(n)] \right\} \\ &= k_{\text{TIQ}}^{3/2} E[|x(n)|^4] [\mathbf{h}_{\text{IMD}}^{0T}, \mathbf{0}_{M-N}^T, \mathbf{g}_{\text{IMD}}^{0T}, \mathbf{0}_{M-N}^T]^T \\ &= 2k_{\text{TIQ}}^{3/2} \sigma_x^4 [\mathbf{h}_{\text{IMD}}^{0T}, \mathbf{0}_{M-N}^T, \mathbf{g}_{\text{IMD}}^{0T}, \mathbf{0}_{M-N}^T]^T \end{aligned} \quad (17)$$

The last step is performed with the help of the Gaussian fourth order moment factorizing theorem [38]–[40], and since $x(n)$ is proper (second order circular), we have $E[|x(n)|^4] = 2\sigma_x^4$.

From (13), the steady-state value of the weight error vector, that is, $E[\tilde{\mathbf{w}}^a(\infty)]$, can be evaluated as

$$\begin{aligned} E[\tilde{\mathbf{w}}^a(\infty)] &= (\mathbf{R}^{a*})^{-1} E[u(n) \mathbf{x}^{a*}(n)] \\ &= 2k_{\text{TIQ}}^{3/2} \sigma_x^2 [\mathbf{h}_{\text{IMD}}^{0T}, \mathbf{0}_{M-N}^T, \mathbf{g}_{\text{IMD}}^{0T}, \mathbf{0}_{M-N}^T]^T \end{aligned} \quad (18)$$

Remark 1: The upper bound on the step size μ for the mean convergence of augmented LMS for a low transmit power FD transceiver is identical to that for a high transmit power one. At the steady state, when the transmit power is low, the augmented LMS converges in the mean to the optimal weight coefficients associated with $\mathbf{x}^a(n)$, that is, \mathbf{w}^{a0} in (6), in an unbiased manner. However, as indicated by (18), when the transmit power is high, this yields a bias in the estimation of $2N$ out of $2M$ entries of the weight error vector $E[\tilde{\mathbf{w}}^a(\infty)]$, quantified by $2k_{\text{TIQ}}^{3/2} \sigma_x^2 [\mathbf{h}_{\text{IMD}}^{0T}, \mathbf{g}_{\text{IMD}}^{0T}]^T$. The level of this bias depends upon the level of undermodeling, that is, the transmitter mixer gain k_{TIQ} , the transmit SI signal power σ_x^2 , and the channel impulse responses associated with the IMD SI components, that is, $\mathbf{h}_{\text{IMD}}^0$ and $\mathbf{g}_{\text{IMD}}^0$.

B. Mean Square Convergence Analysis

From (7) and (11), and again by employing the standard independence assumptions stated in Section III-A, the MSE of augmented LMS based SI canceller, that is, $J^a(n)$, can be further evaluated as

$$\begin{aligned} J^a(n) &= E[\tilde{\mathbf{w}}^{aH}(n) \mathbf{x}^{a*}(n) \mathbf{x}^{aT}(n) \tilde{\mathbf{w}}^a(n)] + E[|u(n)|^2] \\ &\quad - E[u^*(n) \mathbf{x}^{aT}(n) \tilde{\mathbf{w}}^a(n)] - E[u(n) \tilde{\mathbf{w}}^{aH}(n) \mathbf{x}^{a*}(n)] \\ &= \text{Tr}[\mathbf{R}^{a*} \mathbf{K}^a(n)] + E[|u(n)|^2] - E[u^*(n) \mathbf{x}^{aT}(n)] E[\tilde{\mathbf{w}}^a(n)] \\ &\quad - E[u(n) \mathbf{x}^{aH}(n)] E[\tilde{\mathbf{w}}^{a*}(n)] \end{aligned} \quad (19)$$

where $\mathbf{K}^a(n) = E[\tilde{\mathbf{w}}^a(n) \tilde{\mathbf{w}}^{aH}(n)]$ is the covariance matrix of the augmented weight error vector $\tilde{\mathbf{w}}^a(n)$ [38]–[40]. It can be observed from (19) that the mean square convergence analysis of augmented LMS now rests upon both the first and second order statistical properties of $\tilde{\mathbf{w}}^a(n)$. To this end, we first apply the Hermitian operator $(\cdot)^H$ to both sides of (12), to yield

$$\tilde{\mathbf{w}}^{aH}(n+1) = \tilde{\mathbf{w}}^{aH}(n) [\mathbf{I}_{2M} - \mu \mathbf{x}^{a*}(n) \mathbf{x}^{aT}(n)] + \mu u^*(n) \mathbf{x}^{aT}(n) \quad (20)$$

Upon post-multiplying both sides of (12) by $\tilde{\mathbf{w}}^{aH}(n+1)$ in (20) and taking the statistical expectation $E[\cdot]$, the evolution of the weight error covariance matrix $\mathbf{K}^a(n)$ now becomes

$$\begin{aligned} \mathbf{K}^a(n+1) &= (1 - 2\mu\sigma_x^2) \mathbf{K}^a(n) + 2\mu \Re[\mathbf{Q}_3(n)] \\ &\quad + \mu^2 [\mathbf{Q}_1(n) + \mathbf{Q}_2(n) - \mathbf{Q}_4(n) - \mathbf{Q}_5(n)] \end{aligned} \quad (21)$$

where

$$\begin{aligned} \mathbf{Q}_1(n) &= E[|u(n)|^2 \mathbf{x}^{a*}(n) \mathbf{x}^{aT}(n)] \\ \mathbf{Q}_2(n) &= E[\mathbf{x}^{a*}(n) \mathbf{x}^{aT}(n) \tilde{\mathbf{w}}^a(n) \tilde{\mathbf{w}}^{aH}(n) \mathbf{x}^{a*}(n) \mathbf{x}^{aT}(n)] \\ \mathbf{Q}_3(n) &= E[u(n) \mathbf{x}^{a*}(n)] E[\tilde{\mathbf{w}}^{aH}(n)] \\ \mathbf{Q}_4(n) &= E[u(n) \mathbf{x}^{a*}(n) \mathbf{x}^{aH}(n) \mathbf{x}^{a*}(n)] E[\tilde{\mathbf{w}}^{aT}(n)] \\ \mathbf{Q}_5(n) &= E[u^*(n) \mathbf{x}^{a*}(n) \mathbf{x}^{aT}(n) \mathbf{x}^a(n)] E[\tilde{\mathbf{w}}^a(n)] \end{aligned} \quad (22)$$

It can be observed that the term $\mathbf{Q}_2(n)$ is independent of the IMD component $\mathbf{x}_{\text{IMD}}(n)$, and hence

$$\mathbf{Q}_2(n) = \sigma_x^4 \mathbf{K}^a(n) + \mathbf{P}^{a*} \mathbf{K}^{aT}(n) \mathbf{P}^a + 2M \sigma_x^4 \mathbf{I}_{2M} \text{Tr}[\mathbf{K}^a(n)] \quad (23)$$

where \mathbf{P}^a is the pseudocovariance matrix of the augmented SI vector $\mathbf{x}^a(n)$ [39]–[41], given by

$$\mathbf{P}^a = E[\mathbf{x}^a(n)\mathbf{x}^{aT}(n)] = \begin{bmatrix} \mathbf{0} & \sigma_x^2 \mathbf{I}_M \\ \sigma_x^2 \mathbf{I}_M & \mathbf{0} \end{bmatrix}$$

The term $\text{Tr}[\mathbf{K}^a(n)]$ on the RHS of (23) can be decomposed as $\mathbf{1}_{2M}^T \boldsymbol{\kappa}^a(n)$, where $\boldsymbol{\kappa}^a(n)$ is a $2M \times 1$ vector, whose entries are the diagonal elements of $\mathbf{K}^a(n)$, defined as

$$\boldsymbol{\kappa}^a(n) = [E[|\tilde{w}_1^a(n)|^2], E[|\tilde{w}_2^a(n)|^2], \dots, E[|\tilde{w}_{2M}^a(n)|^2]]^T \quad (24)$$

Then, based on (21), the evolution of $\boldsymbol{\kappa}^a(n)$ becomes

$$\begin{aligned} \boldsymbol{\kappa}^a(n+1) &= \underbrace{\{\mathbf{I}_{2M} - 2\mu\sigma_x^2 \mathbf{I}_{2M} + 2\mu^2\sigma_x^4 \mathbf{I}_{2M} + \mu^2\sigma_x^4 \mathbf{1}_{2M} \mathbf{1}_{2M}^T\}}_{\mathbf{F}^a} \boldsymbol{\kappa}^a(n) \\ &\quad + 2\mu \text{diag}\{\Re[\mathbf{Q}_3(n)]\} + \mu^2 \text{diag}\{\mathbf{Q}_1(n) - \mathbf{Q}_4(n) - \mathbf{Q}_5(n)\} \end{aligned} \quad (25)$$

The convergence of the recursion for the vector $\boldsymbol{\kappa}^a(n)$ in (25) is subject to two conditions: 1) the terms $\mathbf{Q}_1(n)$, $\mathbf{Q}_3(n)$, $\mathbf{Q}_4(n)$ and $\mathbf{Q}_5(n)$ are bounded, which is guaranteed if $E[\tilde{\mathbf{w}}^a(n)]$ is bounded; 2) all the eigenvalues of the transition matrix \mathbf{F}^a are less than unity [42], [43]. From (15), condition 1) holds when $0 < \mu < 2/\sigma_x^2$. Furthermore, the eigenvalues of \mathbf{F}^a , denoted by λ_i^a , where $i = 1, 2, \dots, 2M$, can be obtained by solving $\det[\mathbf{F}^a - \lambda_i^a \mathbf{I}_{2M}] = 0$, while from (25), it is easy to find that \mathbf{F}^a is Toeplitz, for which the diagonal elements are $1 - 2\mu + 3\mu^2\sigma_x^4$, and off-diagonal ones are $\mu^2\sigma_x^4$. Hence, after some algebraic manipulations, we have

$$\begin{aligned} \lambda_1^a &= 1 - \mu\sigma_x^2 + (2M + 2)\mu^2\sigma_x^4 \\ \lambda_j^a &= 1 - \mu\sigma_x^2 + 2\mu^2\sigma_x^4, \quad j = 2, 3, \dots, 2M \end{aligned}$$

Note that since $M \geq 1$, we have $\lambda_1^a > \lambda_j^a$ for $j = 2, 3, \dots, 2M$, and hence, condition 2) is satisfied if $\lambda_1^a < 1$, to yield

$$0 < \mu < \frac{1}{(M + 1)\sigma_x^2} \quad (26)$$

Remark 2: The upper bound in (26) is tighter than that in Condition 1), and therefore, the mean square convergence of the augmented LMS based SI canceller in the presence of frequency-dependent IQ imbalance and PA distortion is guaranteed if the step-size μ satisfies (26).

C. Steady State Analysis

Suppose that step-size μ is chosen such that the mean square stability of augmented LMS is guaranteed. Consider $n \rightarrow \infty$, then based on (19) and (24), the steady-state MSE $J^a(\infty)$ can be expressed as

$$J^a(\infty) = \sigma_x^2 \mathbf{1}_{2M}^T \boldsymbol{\kappa}^a(\infty) + E[|u(\infty)|^2] - 2\text{Tr}\{\Re[\mathbf{Q}_3(\infty)]\} \quad (27)$$

where, based on (25), $\boldsymbol{\kappa}^a(\infty)$ can be derived as

$$\begin{aligned} \boldsymbol{\kappa}^a(\infty) &= 2\mu(\mathbf{I}_{2M} - \mathbf{F}^a)^{-1} [\text{diag}\{\Re[\mathbf{Q}_3(\infty)]\} \\ &\quad + \mu \text{diag}\{\mathbf{Q}_1(\infty) - \mathbf{Q}_4(\infty) - \mathbf{Q}_5(\infty)\}] \end{aligned} \quad (28)$$

Similar to the mean convergence analysis in Section III-A, in order to evaluate the terms in (27) and (28), which involves

the overall noise $u(\infty)$, we need to consider the levels of the transmit power.

1) *Low transmit power.* In this case, the power of the thermal noise $v(\infty)$ in $u(\infty)$ is much higher than the power of the nonlinear distortion component $\mathbf{x}_{\text{IMD}}(\infty)$ and the quantization noise $q(\infty)$, so that, $u(n) \simeq v(n)$, and $E[|u(\infty)|^2] \simeq \sigma_v^2$. Upon inserting (16) into (22), we obtain $\mathbf{Q}_1(\infty) = \sigma_v^2 \sigma_x^2 \mathbf{I}_{2M}$, and $\mathbf{Q}_3(\infty) = \mathbf{Q}_4(\infty) = \mathbf{Q}_5(\infty) = \mathbf{0}$, therefore, $\boldsymbol{\kappa}^a(\infty)$ in (28) can be simplified as

$$\boldsymbol{\kappa}^a(\infty) = \frac{\mu\sigma_v^2}{2(1 - \mu(M + 1)\sigma_x^2)} \mathbf{1}_{2M}$$

Upon substituting into (27), this yields

$$J_{\text{low}}^a(\infty) = \frac{(1 - \mu\sigma_x^2)\sigma_v^2}{1 - \mu(M + 1)\sigma_x^2} \quad (29)$$

The achievable SINR is defined as a relative power ratio between the received signal of interest $x_{\text{SOI}}(n)$ and the residual SI $e^a(n)$. According to the analysis in [27], the power of $x_{\text{SOI}}(n)$ can be evaluated as

$$p_{x_{\text{SOI}}} = p_{\text{sen}} k_{\text{LNA}} k_{\text{BB}} k_{\text{RIQ}} \quad (30)$$

where the Rx VGA gain k_{BB} ensures the received signal fit within the voltage range of the ADC, and can be calculated as [19]

$$k_{\text{BB}} = \frac{p_{\text{ADC}}}{k_{\text{LNA}} k_{\text{RIQ}}} \frac{1}{[\alpha_0^2 k_{\text{VGA}} k_{\text{TIQ}} \sigma_x^2 + \alpha_1^2 k_{\text{VGA}}^3 k_{\text{TIQ}}^3 \sigma_x^6] \|f_{\text{RFE}}(n)\|^2 + p_{\text{sen}}}$$

From (4), (29) and (30), the achievable SINR of augmented LMS at the steady state can be evaluated as

$$\text{SINR}_{\text{low}} = \frac{p_{x_{\text{SOI}}}}{J_{\text{low}}^a(\infty)} = \frac{1 - \mu(M + 1)\sigma_x^2}{1 - \mu\sigma_x^2} \text{SNR}_{\text{req}} \quad (31)$$

2) *High transmit power.* In this case, with typical parameters of FD transceivers given in Table II, the power of the SI component σ_x^2 is guaranteed to be less than unity. Therefore, in (28), the Euclidean norm of $\mu \text{diag}\{\mathbf{Q}_1(\infty) - \mathbf{Q}_4(\infty) - \mathbf{Q}_5(\infty)\}$ is much smaller than that of $\text{diag}\{\Re[\mathbf{Q}_3(\infty)]\}$, as the terms $\mathbf{Q}_1(\infty)$, $\mathbf{Q}_4(\infty)$ and $\mathbf{Q}_5(\infty)$ contain a larger amount of higher order SI components. This situation is more pronounced when a smaller step-size μ is chosen, and based on the analysis in Appendix A, we have

$$\begin{aligned} &\mu \text{diag}\{\mathbf{Q}_1(\infty) - \mathbf{Q}_4(\infty) - \mathbf{Q}_5(\infty)\} + \text{diag}\{\Re[\mathbf{Q}_3(\infty)]\} \\ &= \mu(E[|v(\infty)|^2] + E[|q(\infty)|^2]) \text{diag}\{\mathbf{R}^{a*}\} + \text{diag}\{\Re[\mathbf{Q}_3(\infty)]\} \\ &\simeq \mu(\sigma_x^2 + \sigma_q^2)\sigma_v^2 \mathbf{1}_{2M} + 8k_{\text{TIQ}}^3 \sigma_x^4 \mathbf{p}_{\text{IMD}}^0 \end{aligned} \quad (32)$$

where the definition of $\mathbf{p}_{\text{IMD}}^0$ is given in (59) in Appendix A. Upon replacing (32) into (28), we arrive at

$$\boldsymbol{\kappa}^a(\infty) = \frac{\mu(\sigma_v^2 + \sigma_q^2)\mathbf{1}_{2M} + 8k_{\text{TIQ}}^3 \sigma_x^4 \mathbf{p}_{\text{IMD}}^0}{2(1 - \mu\sigma_x^2 - \mu M\sigma_x^2)} \quad (33)$$

According to the analysis in Appendix A, we have

$$\text{Tr}\{\Re[\mathbf{Q}_3(\infty)]\} = 4k_{\text{TIQ}}^3 \sigma_x^6 [\|\mathbf{h}_{\text{IMD}}^0\|^2 + \|\mathbf{g}_{\text{IMD}}^0\|^2] \quad (34)$$

and from (2),

$$E[|u(\infty)|^2] = \sigma_v^2 + \sigma_q^2 + 6k_{\text{TIQ}}^3 \sigma_x^6 [\|\mathbf{h}_{\text{IMD}}^o\|^2 + \|\mathbf{g}_{\text{IMD}}^o\|^2] \quad (35)$$

Therefore, upon replacing (33)-(35) into (27) and after a few algebraic manipulations, we arrive at

$$J_{\text{high}}^a(\infty) = \sigma_v^2 + \sigma_q^2 - 2k_{\text{TIQ}}^3 \sigma_x^6 [\|\mathbf{h}_{\text{IMD}}^o\|^2 + \|\mathbf{g}_{\text{IMD}}^o\|^2] + \frac{\mu M(\sigma_v^2 + \sigma_q^2)\sigma_x^2 + 4k_{\text{TIQ}}^3 \sigma_x^6 [\|\mathbf{h}_{\text{IMD}}^o\|^2 + \|\mathbf{g}_{\text{IMD}}^o\|^2]}{1 - \mu\sigma_x^2 - \mu M\sigma_x^2} \quad (36)$$

to give the achievable SINR of augmented LMS based SI canceller in the case of high transmit power in (37) at the top of the next page.

Remark 3: From (31) and (37), we observe that, for both low and high transmit powers, the achievable SINRs of augmented LMS at the steady state are monotonically decreasing functions of the step-size μ , the end-to-end channel impulse response length M , and the SI power σ_x^2 . Particularly, in the case of high transmit powers, the achievable SINR of augmented LMS is impaired by the non-negligible IMD SI $\mathbf{x}_{\text{IMD}}(n)$ and its image $\mathbf{x}_{\text{IMD}}^*(n)$, as well as their associated end-to-end channel impulse responses, and its degradation in SINR becomes more severe with an increase in transmit power.

IV. PROPOSED AUGMENTED NONLINEAR LMS BASED SI CANCELLER AND ITS PERFORMANCE ANALYSIS

From the mean and mean square analysis in Section III, it is clear that, for high transmit powers, the third order IMD component $\mathbf{x}_{\text{IMD}}(n)$, produced by PA distortion, may exceed the thermal noise floor, or even become stronger than the signal of interest $x_{\text{SOI}}(n)$. This leads to bias as well as suboptimality in SINR performance of the conventional augmented LMS based SI canceller. To address these issues, it is desirable to design an adaptive SI canceller which suppresses the SI, image SI and IMD components simultaneously. Upon revisiting the vectorized signal model in (1) and (6), instead of treating the IMD components $\mathbf{x}_{\text{IMD}}(n)$ and $\mathbf{x}_{\text{IMD}}^*(n)$ as a part of the aggregated interference $u(n)$, we can concatenate them with the SI component $\mathbf{x}(n)$ and its image counterpart $\mathbf{x}^*(n)$ to form a $(2M + 2N) \times 1$ vector $\mathbf{x}^b(n)$, given by

$$\mathbf{x}^b(n) = [\mathbf{x}^T(n), \mathbf{x}_{\text{IMD}}^T(n), \mathbf{x}^H(n), \mathbf{x}_{\text{IMD}}^H(n)]^T \quad (38)$$

Note that, compared with the augmented SI vector $\mathbf{x}^a(n)$ in (6), vector $\mathbf{x}^b(n)$ is no longer simply widely linear in $\mathbf{x}(n)$. In fact, by defining

$$\mathbf{x}^c(n) = [\mathbf{x}^T(n), \mathbf{x}^{dT}(n)]^T$$

where $\mathbf{x}^d(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$ contains the first N elements in $\mathbf{x}(n)$, $\mathbf{x}^b(n)$ in (38) can be represented in a widely nonlinear form in $\mathbf{x}^c(n)$ (or equivalently, in $\mathbf{x}(n)$) as [44], [45]

$$\mathbf{x}^b(n) = \Psi^T([\mathbf{x}^c(n), \mathbf{x}^{cH}(n)])$$

where $\Psi(\cdot)$ is a vectorized nonlinear function, Based on (3), $\Psi(\mathbf{x}^c(n))$ can be expressed element by element as

$$\Psi(\mathbf{x}^c(n)) = [\psi_1(x(n)), \psi_2(x(n-1)), \dots, \psi_M(x(n-M+1)), \psi_{M+1}(x(n)), \psi_{M+2}(x(n-1)), \dots, \psi_{M+N}(x(n-N+1))]^T$$

where

$$\psi_i(x) = \begin{cases} x, & i = 1, 2, \dots, M \\ k_{\text{TIQ}}^{3/2} |x|^2 x, & i = M+1, M+2, \dots, M+N \end{cases}$$

Based on the above discussion, we refer to $\mathbf{x}^b(n)$ as the augmented nonlinear SI vector, and for SI cancellation of FD transceivers in the presence of PA nonlinearity, we should take into account $\mathbf{x}^b(n)$ to give a model-fit widely nonlinear relation between the observed signal $d(n)$ and its corresponding SI waveform $x(n)$ in the form

$$d(n) = \mathbf{x}^{bT}(n)\mathbf{w}^{bo} + v(n) + q(n) \quad (39)$$

where $\mathbf{w}^{bo} = [\mathbf{h}^{oT}, \mathbf{h}_{\text{IMD}}^{oT}, \mathbf{g}^{oT}, \mathbf{g}_{\text{IMD}}^{oT}]^T$ is the $(2M + 2N) \times 1$ sufficient-length end-to-end filter impulse response of a FD transceiver. Similar to the conventional augmented LMS, the proposed augmented nonlinear LMS based SI canceller aims to estimate \mathbf{w}^{bo} by minimizing a mean square error cost function $J^b(n)$, defined as

$$J^b(n) = E[|e^b(n)|^2] = E[e^b(n)e^{b*}(n)] \quad (40)$$

where the instantaneous residual SI $e^b(n)$ is given by

$$e^b(n) = d(n) - \mathbf{x}^{bT}(n)\mathbf{w}^b(n) \quad (41)$$

and governs the update of the weight vector $\mathbf{w}^b(n)$ as

$$\mathbf{w}^b(n+1) = \mathbf{w}^b(n) + \mu e^b(n)\mathbf{x}^{b*}(n) \quad (42)$$

A. Mean Convergence Analysis

Upon introducing the $(2M + 2N) \times 1$ weight error vector

$$\tilde{\mathbf{w}}^b(n) = \mathbf{w}^b(n) - \mathbf{w}^{bo} \quad (43)$$

the output residual SI $e^b(n)$ in (41) becomes

$$e^b(n) = v(n) + q(n) - \mathbf{x}^{bT}(n)\tilde{\mathbf{w}}^b(n) \quad (44)$$

and the recursion of $\tilde{\mathbf{w}}^b(n)$ now obeys

$$\tilde{\mathbf{w}}^b(n+1) = [\mathbf{I}_{2M+2N} - \mu\mathbf{x}^{b*}(n)\mathbf{x}^{bT}(n)]\tilde{\mathbf{w}}^b(n) + \mu[v(n) + q(n)]\mathbf{x}^{b*}(n) \quad (45)$$

Upon applying the expectation operator $E[\cdot]$ on both sides of (45) and using the standard independence assumptions introduced in Section III-A, we arrive at

$$E[\tilde{\mathbf{w}}^b(n+1)] = (\mathbf{I}_{2M+2N} - \mu\mathbf{R}^{b*})E[\tilde{\mathbf{w}}^b(n)] \quad (46)$$

where $\mathbf{R}^b = E[\mathbf{x}^b(n)\mathbf{x}^{bH}(n)]$ is the covariance matrix of the augmented nonlinear SI input vector $\mathbf{x}^b(n)$. Based on (46), the step-size μ which guarantees the convergence of the proposed augmented nonlinear LMS in the mean sense should satisfy

$$|1 - \mu\lambda_k^b| < 1, \quad k = 1, \dots, 2M + 2N$$

where λ_k^b are the eigenvalues of \mathbf{R}^b . Note that although \mathbf{R}^b is Hermitian, its positive-definiteness is not always guaranteed due to the non-Gaussianity of $\mathbf{x}^b(n)$. To investigate this issue, we shall further decompose \mathbf{R}^b as

$$\mathbf{R}^b = E[\mathbf{x}^b(n)\mathbf{x}^{bH}(n)] = \begin{bmatrix} \mathbf{R}_0^b & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_0^b \end{bmatrix}$$

$$\text{SINR}_{\text{high}} = \frac{p_{\text{SOT}}}{J_{\text{high}}^a(\infty)} = \frac{1 - \mu(M+1)\sigma_x^2}{\left(\frac{1}{\text{SNR}_{\text{req}}} + \frac{\sigma_q^2}{k_{\text{BB}}k_{\text{LNA}}k_{\text{TIQ}}p_{\text{sen}}} \right) + \frac{2[1 + \mu(M+1)\sigma_x^2]k_{\text{TIQ}}^2\sigma_x^6(\|\mathbf{h}_{\text{IMD}}^0\|^2 + \|\mathbf{g}_{\text{IMD}}^0\|^2)}{k_{\text{BB}}k_{\text{LNA}}p_{\text{sen}}} } \quad (37)$$

where

$$\mathbf{R}_0^b = E[\mathbf{x}^e(n)\mathbf{x}^{eH}(n)] = \begin{bmatrix} E[|x(n)|^2]\mathbf{I}_M & \mathbf{\Omega}^T \\ \mathbf{\Omega} & k_{\text{TIQ}}^3 E[|x(n)|^6]\mathbf{I}_N \end{bmatrix}$$

$$\mathbf{\Omega} = [\mathbf{R}^d \quad \mathbf{0}_{N \times (M-N)}]$$

$$\mathbf{R}^d = E[\mathbf{x}^d(n)\mathbf{x}_{\text{IMD}}^H(n)] = k_{\text{TIQ}}^{3/2} E[|x(n)|^4]\mathbf{I}_N$$

in which $\mathbf{x}^e(n) = [\mathbf{x}^T(n), \mathbf{x}_{\text{IMD}}^T(n)]^T$. Now, by solving $\det[\mathbf{R}^b - \lambda_k^b \mathbf{I}_{2M+2N}] = 0$ and after some algebraic manipulations, we arrive at

$$\lambda_1^b = E[|x(n)|^2] = \sigma_x^2$$

$$\lambda_{2,3}^b = \frac{1}{2} \left\{ E[|x(n)|^2] + k_{\text{TIQ}}^3 E[|x(n)|^6] \pm \sqrt{(E[|x(n)|^2] + k_{\text{TIQ}}^3 E[|x(n)|^6])^2 - k_{\text{TIQ}}^3 (E[|x(n)|^4])^2} \right\}$$

$$= \frac{\sigma_x^2 + 6k_{\text{TIQ}}^3 \sigma_x^6 \pm \sigma_x^2 \sqrt{1 - 2k_{\text{TIQ}}^3 \sigma_x^4 + 36k_{\text{TIQ}}^6 \sigma_x^8}}{2} \quad (47)$$

where the algebraic multiplicities of λ_1^b , λ_2^b and λ_3^b are respectively $2M - 2N$, $2N$ and $2N$.

Remark 4: The covariance matrix \mathbf{R}^b is positive-definite since all the eigenvalues are positive. It is also worth noting that, although $x(n)$ is considered to be i.i.d. Gaussian, the positive definiteness of \mathbf{R}^b is still valid as long as $x(n)$ comes from any i.i.d. second order circular constellation, e.g., QPSK or M-QAM. Note that in (47), the largest eigenvalue of \mathbf{R}^b is λ_2^b , and therefore, the mean convergence bound on the step-size μ is given by

$$0 < \mu < \frac{2}{\lambda_{\max}[\mathbf{R}^b]}$$

$$= \frac{4}{\sigma_x^2 + 6k_{\text{TIQ}}^3 \sigma_x^6 + \sigma_x^2 \sqrt{1 - 2k_{\text{TIQ}}^3 \sigma_x^4 + 36k_{\text{TIQ}}^6 \sigma_x^8}} \quad (48)$$

which enables the proposed augmented nonlinear LMS based SI canceller to asymptotically achieve unbiased estimation of the optimal weight vector \mathbf{w}^{b0} , indicated by $E[\tilde{\mathbf{w}}^b(\infty)] = \mathbf{0}$ from (46), and independent on whether the transmit power of an FD transceiver is low or high.

B. Mean Square Convergence Analysis

Upon taking (44) into (40), and again employing the standard independence assumptions, the MSE of the proposed augmented nonlinear LMS based SI canceller can be further evaluated as

$$J^b(n) = \text{Tr}[\mathbf{R}^b \mathbf{K}^b(n)] + \sigma_v^2 + \sigma_q^2 \quad (49)$$

where $\mathbf{K}^b(n) = E[\tilde{\mathbf{w}}^b(n)\tilde{\mathbf{w}}^{bH}(n)]$ is the covariance matrix of the weight error vector $\tilde{\mathbf{w}}^b(n)$. To analyze its evolution, we

first multiply both sides of (45) by $\tilde{\mathbf{w}}^{bH}(n)$, and apply the statistical expectation operator $E[\cdot]$, to give

$$\mathbf{K}^b(n+1) = \mathbf{K}^b(n) + \mu^2(\sigma_v^2 + \sigma_q^2)\mathbf{R}^b$$

$$- \mu E[\mathbf{x}^{b*}(n)\mathbf{x}^{bT}(n)\tilde{\mathbf{w}}^b(n)\tilde{\mathbf{w}}^{bH}(n)]$$

$$- \mu E[\tilde{\mathbf{w}}^b(n)\tilde{\mathbf{w}}^{bH}(n)\mathbf{x}^{b*}(n)\mathbf{x}^{bT}(n)]$$

$$+ \mu^2 E[\mathbf{x}^{b*}(n)\mathbf{x}^{bT}(n)\tilde{\mathbf{w}}^b(n)\tilde{\mathbf{w}}^{bH}(n)\mathbf{x}^{b*}(n)\mathbf{x}^{bT}(n)] \quad (50)$$

Since the augmented nonlinear SI vector $\mathbf{x}^b(n)$ in (38) is non-Gaussian, the Gaussian fourth order moment factorizing theorem used in Section III is no longer applicable to evaluate the last term on the RHS of (50), and we therefore resort to matrix vectorization instead of matrix diagonalization to evaluate the mean square convergence behavior of the proposed augmented nonlinear LMS [46], [47]. By using the following matrix vectorization lemma for arbitrary matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$

$$\text{vec}\{\mathbf{ABC}\} = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}\{\mathbf{B}\}$$

it is straightforward to verify that the recursion for $\mathbf{K}^b(n)$ in (50) can be transformed into a linear vector relation as

$$\text{vec}\{\mathbf{K}^b(n+1)\} = \underbrace{(\mathbf{I}_{2M+2N} - \mu\mathbf{S} + \mu^2\mathbf{T})}_{\mathbf{F}^b} \text{vec}\{\mathbf{K}^b(n)\}$$

$$+ \mu^2(\sigma_v^2 + \sigma_q^2)\text{vec}\{\mathbf{R}^b\} \quad (51)$$

where

$$\mathbf{S} = \mathbf{I}_{2M+2N} \otimes \mathbf{R}^b + \mathbf{R}^b \otimes \mathbf{I}_{2M+2N}$$

$$\mathbf{T} = E[(\mathbf{x}^b(n)\mathbf{x}^{bH}(n)) \otimes (\mathbf{x}^{b*}(n)\mathbf{x}^{bT}(n))]$$

The condition on the step-size μ to guarantee the convergence of $\mathbf{K}^b(n)$ now relies on $|\lambda[\mathbf{F}^b]| < 1$. It has been proven in [47] that for positive definite \mathbf{R}^b and \mathbf{S} and nonnegative definite \mathbf{T} , the stability of the recursion in (51) is guaranteed when

$$0 < \mu < \min \left\{ \frac{1}{\lambda_{\max}[\mathbf{S}^{-1}\mathbf{T}]}, \frac{1}{\lambda_{\max}[\mathbf{\Gamma}]} \right\} \quad (52)$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} \frac{\mathbf{S}}{2} & -\frac{\mathbf{T}}{2} \\ \mathbf{I}_{(2M+2N)^2} & \mathbf{0} \end{bmatrix}$$

C. Steady State Analysis

Suppose the step-size μ is chosen to ensure the mean square stability of the proposed augmented nonlinear LMS. Then, from (49), its steady-state MSE $J^b(\infty)$ can be expressed as

$$J^b(\infty) = \text{Tr}[\mathbf{R}^b \mathbf{K}^b(\infty)] + \sigma_v^2 + \sigma_q^2 \quad (53)$$

where $\mathbf{K}^b(\infty)$ can be evaluated from (51) as

$$\mathbf{K}^b(\infty) = \text{vec}^{-1} \{ \mu^2(\sigma_v^2 + \sigma_q^2)(\mathbf{I}_{2M+2N} - \mathbf{F}^b)^{-1} \text{vec}\{\mathbf{R}^b\} \}$$

$$= \text{vec}^{-1} \{ \mu^2(\sigma_v^2 + \sigma_q^2)(\mu\mathbf{S} - \mu^2\mathbf{T})^{-1} \text{vec}\{\mathbf{R}^b\} \} \quad (54)$$

Upon substituting (54) into (53), we have

$$J^b(\infty) = (\sigma_v^2 + \sigma_q^2) \cdot \left(1 + \mu^2 \text{Tr}[\mathbf{R}^b \text{vec}^{-1}\{(\mu \mathbf{S} - \mu^2 \mathbf{T})^{-1} \text{vec}\{\mathbf{R}^b\}\}] \right)$$

Due to the existence of the fourth order moment matrix \mathbf{T} , a detailed evaluation of the steady state MSE $J^b(\infty)$ is much more difficult than that of the standard augmented LMS. However, as shown in (51), since matrix \mathbf{T} is multiplied by μ^2 within the matrix \mathbf{F}^b , then for a sufficient small step-size μ , we can neglect the term $\mu^2 \mathbf{T}$ in \mathbf{F}^b [46], [47]. In this way, the standard eigenvalue decomposition (EVD) of \mathbf{R}^b gives $\mathbf{R}^b = \mathbf{U} \mathbf{\Lambda}^b \mathbf{U}^H$, where \mathbf{U} is a unitary matrix and $\mathbf{\Lambda}^b = \text{diag}\{\lambda_1^b, \lambda_2^b, \dots, \lambda_{2M+2N}^b\}$ is a diagonal matrix comprising of the eigenvalues of \mathbf{R}^b , from (50), to give

$$\tilde{\mathbf{K}}^b(n+1) = \tilde{\mathbf{K}}^b(n) + \mu^2(\sigma_v^2 + \sigma_q^2) \mathbf{\Lambda}^b - \mu \mathbf{\Lambda}^b \tilde{\mathbf{K}}^b(n) - \mu \tilde{\mathbf{K}}^b(n) \mathbf{\Lambda}^b$$

where $\tilde{\mathbf{K}}^b(n) = \mathbf{U}^H \mathbf{K}^b(n) \mathbf{U}$, and its steady-state value $\tilde{\mathbf{K}}^b(\infty)$ can be obtained as

$$\tilde{\mathbf{K}}^b(\infty) = \frac{\mu(\sigma_v^2 + \sigma_q^2)}{2} \mathbf{I}_{2M+2N}$$

After substituting into (53), we obtain an approximated steady-state MSE, denoted by $J_{\text{ap}}^b(\infty)$, in the form

$$J_{\text{ap}}^b(\infty) = \text{Tr}[\mathbf{U} \mathbf{\Lambda}^b \tilde{\mathbf{K}}^b(\infty) \mathbf{U}^H] + \sigma_v^2 + \sigma_q^2 \\ = (\sigma_v^2 + \sigma_q^2) [\mu(M\sigma_x^2 + 6Nk_{\text{TIQ}}^3\sigma_x^6) + 1] \quad (55)$$

Now from (4), (30) and (55), the achievable SINR of the proposed augmented nonlinear LMS can be evaluated as in (56) at the top of the next page.

Remark 5: The achievable SINR of the proposed augmented nonlinear LMS based SI canceller in (56) is a monotonically decreasing function of the step-size μ , the length of SI channel impulse response M and that of IMD SI channel impulse response N , the transmitter mixer gain k_{TIQ} , and the SI power σ_x^2 . Moreover, in the situations of high transmit powers, owing to the model fitting advantage, the optimality of the proposed augmented nonlinear LMS for nonlinear SI cancellation can be also observed, since its achievable SINR is no longer impaired by the IMD channel impulse responses, that is, $\|\mathbf{h}_{\text{IMD}}^o\|^2$ and $\|\mathbf{g}_{\text{IMD}}^o\|^2$, which, however, remain the by-products of the undermodeling problem encountered by the conventional augmented LMS based SI canceller for FD transceivers in the joint presence of PA nonlinearity and frequency-dependent IQ imbalance.

V. COMPUTER SIMULATIONS

In order to validate performance advantages of the proposed augmented nonlinear LMS based digital SI canceller over the conventional augmented LMS one for full-duplex direct-conversion transceivers in the presence of PA nonlinear distortion and frequency-dependent IQ imbalance, simulations were conducted in the MATLAB programming environment. The simulated waveforms of the transmit SI $x(n)$ and the received signal of interest $x_{\text{SOI}}(n)$ were both considered to be

generated from OFDM transmission systems compliant with the wireless LAN (WLAN) 802.11 standards. The numbers of subcarriers and null subcarriers of the WLAN-OFDM transmission system were respectively $K = 64$ and $K_{\text{null}} = 14$. The length of cyclic prefix was $K_{\text{cp}} = 16$, the oversampling factor was $K_{\text{os}} = 4$, and the waveform bandwidth was $B_c = 20$ MHz, to give an OFDM symbol duration $T_{\text{sym}} = (K + K_{\text{cp}})/B_c = 4$ μs . The 16-QAM constellation scheme was used in each subcarrier. The residual analog cancellation error was subject to a 3-tap static Rayleigh distribution, whose detailed power-delay-profile is provided in [8]. The frequency-dependent transmitter and receiver I/Q imbalance impulse responses were both modeled as 2-tap static FIR filters [24]. In this way, the length of the end-to-end channel impulse responses for the SI vector $\mathbf{x}(n)$ and its image component $\mathbf{x}^*(n)$, that is, \mathbf{h}^o and \mathbf{g}^o , was fixed to $M = 5$, while that for the IMD components $\mathbf{x}_{\text{IMD}}(n)$ and $\mathbf{x}_{\text{IMD}}^*(n)$, that is, $\mathbf{h}_{\text{IMD}}^o$ and $\mathbf{g}_{\text{IMD}}^o$, was fixed to $N = 4$. All the simulation results were obtained by averaging 200 independent trials.

As stated in *Remark 1*, for high Tx powers, the conventional augmented LMS based SI canceller yields an unavoidable steady-state bias on the estimation of $2N$ out of $2M$ entries of the augmented end-to-end system impulse responses $\mathbf{w}^{\text{ao}} = [\mathbf{h}^{oT}, \mathbf{g}^{oT}]^T$, due to the arbitrary negligence of the IMD SI components, and this bias is quantified by (18). This analysis is supported by Fig. 3(a), in which the evolution of two representative weight error coefficients, that is, $\tilde{h}_1(n)$ and $\tilde{h}_2(n)$, are provided. A Type 2 FD transceiver with a transmit power at 25 dBm was considered, and two step-sizes $\mu = 0.05\mu_{\text{max}_1}$ and $\mu = 0.1\mu_{\text{max}_1}$ were used in augmented LMS, where μ_{max_1} is the upper bound of the step-size μ , which guarantees both the mean and mean square stability of augmented LMS, evaluated by using (26). Observe that the two weight error coefficients, $\tilde{h}_1(n)$ and $\tilde{h}_2(n)$ converged to their theoretical steady-state biased values, rather than 0. However, as discussed in Section IV, due to the appropriate model fitting, the proposed augmented nonlinear LMS based SI canceller was able to remove this bias, as illustrated in Fig. 3(b). By comparing the simulation results in Fig. 3(a) and Fig. 3(b), there was a cost in convergence paid by the proposed augmented nonlinear LMS in order to achieve unbiased nonlinear SI cancellation. This is because the higher order IMD components $\mathbf{x}_{\text{IMD}}(n)$ and $\mathbf{x}_{\text{IMD}}^*(n)$ were considered within its underlying estimation framework, and consequently, a higher eigenvalue spread occurred in its input covariance matrix \mathbf{R}^b , as compared with its linear counterpart.

We next validated the proposed mean square analysis of both augmented LMS and augmented nonlinear LMS based SI cancellers. Fig. 4(a) illustrates the theoretical and simulated achievable SINRs in the steady-state stage, with μ_{max_1} and $\mu = 0.1\mu_{\text{max}_1}$ and against different levels of transmit powers for a Type 2 FD transceiver. Observe that the empirical results were closely matched with the analytical ones, evaluated by using (37) and (56) respectively for augmented LMS and augmented nonlinear LMS. These also conform with the analysis in *Remark 3* and *Remark 5* which states that a smaller step-size μ enables a better steady-state SINR performance for both SI cancellers, but at the cost of slower convergence. Fig.

$$\text{SINR}_{\text{Proposed}} = \frac{p_{\text{XSOI}}}{J_{\text{ap}}^b(\infty)} = \frac{1}{\left(\frac{1}{\text{SNR}_{\text{req}}} + \frac{\sigma_q^2}{k_{\text{BB}} k_{\text{LNA}} k_{\text{TIQ}} p_{\text{sen}}} \right) [1 + \mu(M\sigma_x^2 + 6Nk_{\text{TIQ}}^3\sigma_x^6)]} \quad (56)$$

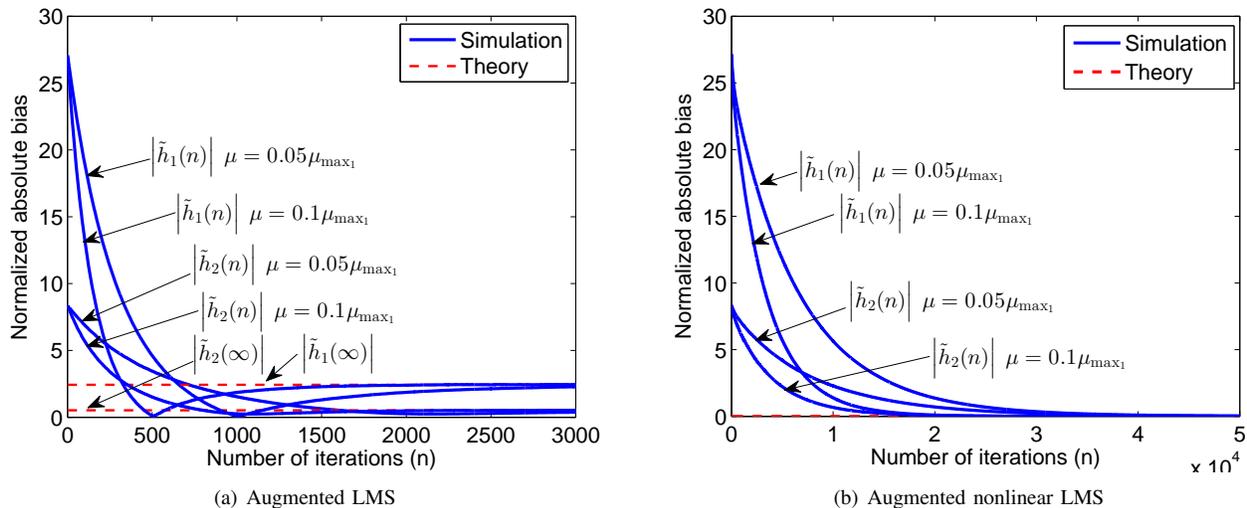


Fig. 3. Normalized absolute values of two representative error coefficients $\tilde{h}_1(n)$ and $\tilde{h}_2(n)$ for a Type 2 FD transceiver by using (a) augmented LMS and (b) augmented nonlinear LMS, where $\mu \in \{0.05\mu_{\text{max}1}, 0.1\mu_{\text{max}1}\}$.

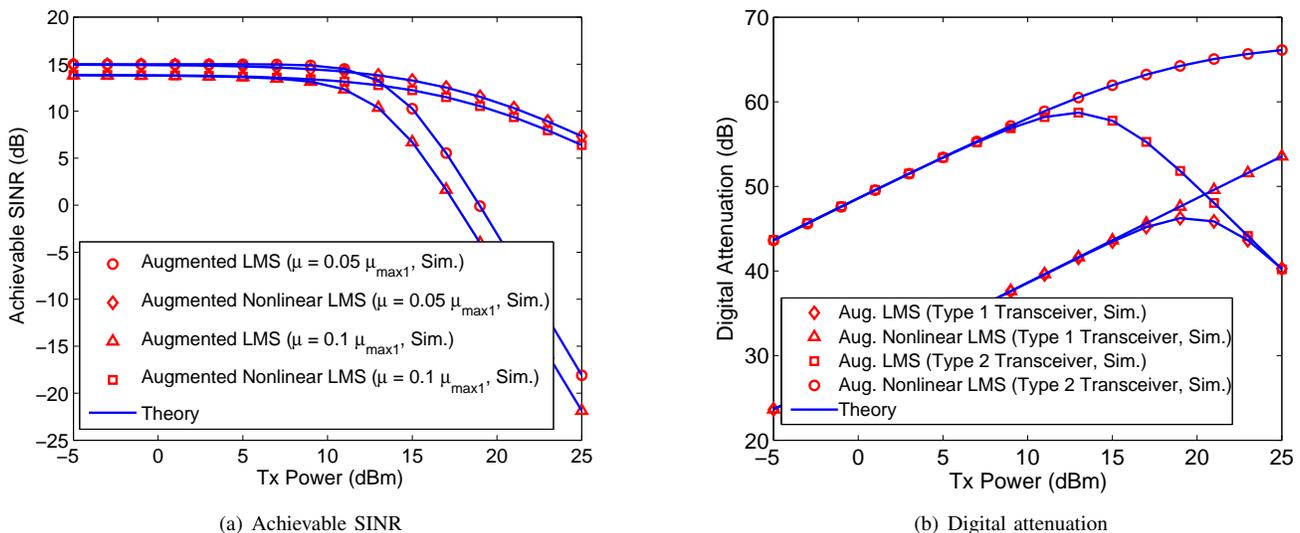


Fig. 4. Comparison of the theoretical and simulated steady-state mean square performances of both the conventional augmented LMS and the proposed augmented nonlinear LMS, measured in terms of (a) Achievable SINR, and (b) Digital attenuation.

4(a) also justifies the motivations to propose the augmented nonlinear LMS based SI canceller in the sense that its SINR performance is much better than that of augmented LMS in the high transmit power range where the IMD SI component $\mathbf{x}_{\text{IMD}}(n)$ and its image counterpart become dominant, since both components have been generically considered as a part of the augmented nonlinear input vector within the proposed augmented nonlinear LMS. As expected, when the transmit power was low enough, both SI cancellers provided a nearly

identical SINR performance, since the higher order components became negligible. This also resulted in the theoretical SINR performance of augmented LMS in (37), derived for a high transmit power, asymptotically converge to that in (31). The above discussion is also applicable for Fig. 4(b), where we compared the digital attenuation capability of the two considered SI cancellers after convergence with a step-size $\mu = 0.05\mu_{\text{max}1}$, for both Type 1 and Type 2 FD transceivers and against different levels of transmit powers. The digital

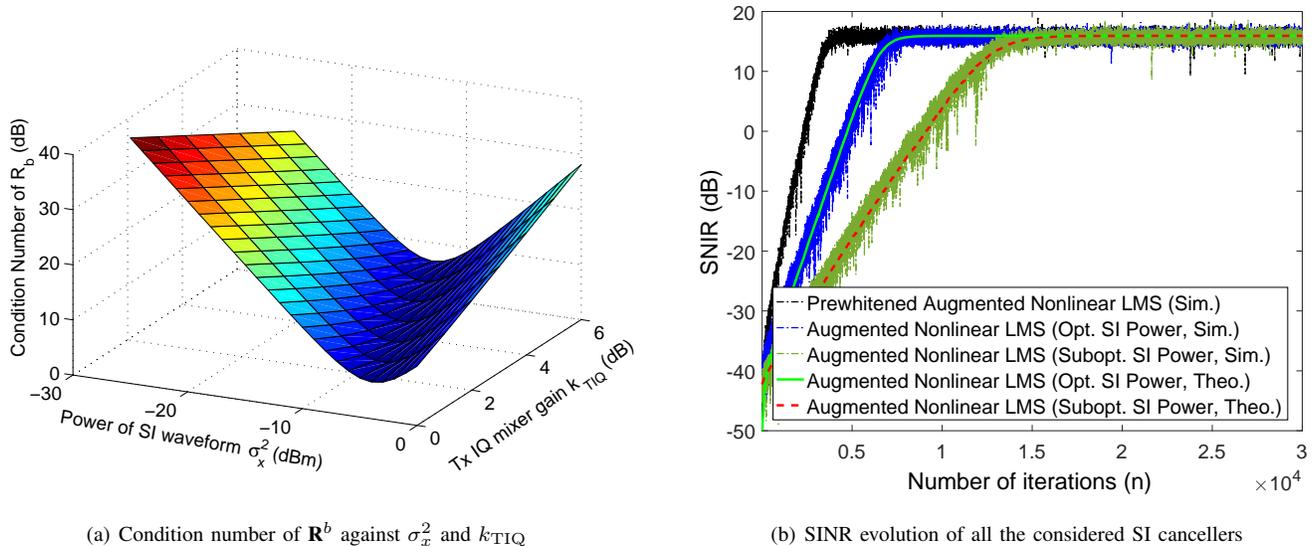


Fig. 5. Convergence speed analysis of the proposed augmented nonlinear LMS. (a) The variations of the condition number of \mathbf{R}^b against the SI power σ_x^2 and the Tx IQ mixer gain k_{TIQ} . (b) Performance comparison between the theoretical and simulated SINR of augmented nonlinear LMS and the simulated SINR of its prewhitened counterpart for a Type 2 FD transceiver, with $\mu = 0.005\mu_{\text{max}_2}$.

attenuation performance is a measure of the amount of SI before and after applying an SI canceller, defined as the power ratio between the desired signal $d(n)$ and the corresponding estimation error of an SI canceller [27]. Observe the excellent agreement between the simulated results and their theoretical evaluations, as well as the performance advantages of the proposed augmented nonlinear LMS over the conventional augmented LMS in the high transmit power range.

A. Further Convergence Speed Improvement of the Proposed Augmented Nonlinear LMS based SI Canceller

When higher order IMD components $\mathbf{x}_{\text{IMD}}(n)$ and $\mathbf{x}_{\text{IMD}}^*(n)$ are strong in a FD transceiver, a potential drawback encountered by adaptive SI cancellers lies in the slow convergence, incurred by the high eigenvalue spread within the input covariance matrix \mathbf{R}^b [18]. In fact, based on (47), the condition number of \mathbf{R}^b , the ratio between its maximum and minimum eigenvalues, is denoted by $\mathcal{C}_{\mathbf{R}^b}$, and can be calculated as

$$\mathcal{C}_{\mathbf{R}^b} = \frac{\lambda_2^b}{\lambda_3^b} = \frac{1 + 6k_{\text{TIQ}}^3\sigma_x^4 + \sqrt{1 - 2k_{\text{TIQ}}^3\sigma_x^4 + 36k_{\text{TIQ}}^6\sigma_x^8}}{1 + 6k_{\text{TIQ}}^3\sigma_x^4 - \sqrt{1 - 2k_{\text{TIQ}}^3\sigma_x^4 + 36k_{\text{TIQ}}^6\sigma_x^8}} \quad (57)$$

This reveals that the value of $\mathcal{C}_{\mathbf{R}^b}$ depends on both the Tx mixer gain, k_{TIQ} , and the SI power, σ_x^2 . Therefore, for the FD transceiver with a given k_{TIQ} , one possible solution to increase the convergence speed of the augmented nonlinear LMS is to synthetically scale σ_x^2 , so that, the minimum of $\mathcal{C}_{\mathbf{R}^b}$ can be achieved. In fact, as first proven in Appendix B and further illustrated in Fig. 5(a), although the global minimum of $\mathcal{C}_{\mathbf{R}^b}$ does exist, it is greater than unity. Owing to the closed-loop structure of the FD transceiver, the knowledge of the augmented nonlinear SI vector $\mathbf{x}^b(n)$ is ideally inherent in the receiver end. Therefore, a more efficient solution can be

established by using standard pre-whitening to decompose \mathbf{R}^b as $\mathbf{R}^b = \mathbf{U}\mathbf{\Lambda}^b\mathbf{U}^H$ and consequently to produce a whitened input vector, $\tilde{\mathbf{x}}^b(n) = \mathbf{\Phi}\mathbf{x}^b(n)$, where $\mathbf{\Phi} = (\mathbf{\Lambda}^b)^{-\frac{1}{2}}\mathbf{U}^H$. In this way, the speed of gradient descent of augmented nonlinear LMS is fixed and normalized, i.e., with a unity condition number, thus facilitating practical applications. This is supported by Fig. 5(b), which shows both the theoretical and simulated convergence behavior, measured in terms of SINR, of the proposed augmented nonlinear LMS based SI canceller and its data-whitening assisted version for a Type 2 FD transceiver with a Tx power at 15 dBm and a Tx IQ mixer gain $k_{\text{TIQ}} = 6$ dB. The step-size was $\mu = 0.005\mu_{\text{max}_2}$, where μ_{max_2} is its upper bound on μ which guarantees both the mean and mean square stability of augmented nonlinear LMS, evaluated by using (48) and (52), respectively. In this case, the optimal SI power σ_x^2 was roughly -13 dBm for the minimum eigenvalue spread. The simulation results for a suboptimal SI power, $\sigma_x^2 = -10$ dBm, are also provided. We observe that the proposed theoretical evaluation accurately described the empirical SINR evolution of augmented nonlinear LMS in both the transient and steady-state stages. We also observe that, with the optimal SI power, the augmented nonlinear LMS exhibited a faster convergence, about 7000 iterations, as compared with the considered suboptimal case. However, when the prewhitening scheme was employed, it merely required about 3000 iterations to arrive at the steady state.

VI. CONCLUSION

The impact of typical front-end non-idealities on a future 5G systems, such as nonlinear PA distortion, IQ imbalance, thermal noise and quantization noise, has been rigorously analyzed over both the mean and mean square performances of the conventional augmented LMS based digital SI canceller for fully-duplex direct-conversion transceivers. We have here quantified

the estimation bias and the second order performance sub-optimality exhibited by the augmented LMS in case of high transmit powers. To rectify these drawbacks, an augmented nonlinear LMS based SI canceller has been proposed, which naturally accounts for those higher order components by virtue of a widely nonlinear model fit. Its performance advantages over augmented LMS have been verified both theoretically and through numerical validation. To further increase convergence speed of the proposed scheme, the standard data pre-whitening scheme has also been employed in this context. Illustrative simulations on two representative types of FD transceivers for OFDM-based WLAN standard compliant waveforms support the analysis.

APPENDIX A

EVALUATION OF THE TERM $\text{diag}\{\mathbf{Q}_3(\infty)\}$ IN THE CASE OF HIGH TRANSMIT POWER

Based on (2) and (22), the diagonal elements of $\mathbf{Q}_3(n)$ can be derived as

$$\begin{aligned} & \text{diag}\{E[\mathbf{Q}_3(n)]\} \\ &= \text{diag}\{E[\mathbf{x}_{\text{IMD}}^T(n)\mathbf{h}_{\text{IMD}}^o(n)\mathbf{x}^{a*}(n)]E[\tilde{\mathbf{w}}^{aH}(n)]\} \\ &+ \text{diag}\{E[\mathbf{x}_{\text{IMD}}^H(n)\mathbf{g}_{\text{IMD}}^o(n)\mathbf{x}^{a*}(n)]E[\tilde{\mathbf{w}}^{aH}(n)]\} \\ &+ \text{diag}\{E[v(n)\mathbf{x}^{a*}(n)]E[\tilde{\mathbf{w}}^{aH}(n)]\} \end{aligned}$$

while the standard independence assumptions yield

$$\text{diag}\{E[v(n)\mathbf{x}^{a*}(n)]E[\tilde{\mathbf{w}}^{aH}(n)]\} = \mathbf{0} \quad (58)$$

Now, based on (17) and (58), we have

$$\begin{aligned} \text{diag}\{E[\mathbf{Q}_3(n)]\} &= [h_{\text{IMD},1}k_{\text{TIQ}}^{3/2}E[|x(n)|^4]E[\tilde{w}_1^*(n)], \\ &h_{\text{IMD},2}k_{\text{TIQ}}^{3/2}E[|x(n-1)|^4]E[\tilde{w}_2^*(n)], \dots, \\ &h_{\text{IMD},N}k_{\text{TIQ}}^{3/2}E[|x(n-N+1)|^4]E[\tilde{w}_N^*(n)], \mathbf{0}_{M-N}^T, \\ &g_{\text{IMD},1}k_{\text{TIQ}}^{3/2}E[|x(n)|^4]E[\tilde{w}_1^*(n)], \dots, \\ &g_{\text{IMD},N}k_{\text{TIQ}}^{3/2}E[|x(n)|^4]E[\tilde{w}_N^*(n)], \mathbf{0}_{M-N}^T] \end{aligned}$$

Note that $E[|x(n)|^4] = 2\sigma_x^4$, and hence, the steady-state evaluation of $\text{diag}\{E[\mathbf{Q}_3(n)]\}$, that is, $\text{diag}\{E[\mathbf{Q}_3(\infty)]\}$, is now subject to $E[\tilde{\mathbf{w}}^a(\infty)]$, and based on (18), this gives

$$\text{diag}\{\mathbf{Q}_3(\infty)\} = 4k_{\text{TIQ}}^3\sigma_x^6\mathbf{p}_{\text{IMD}}^o$$

where $\mathbf{p}_{\text{IMD}}^o$ is defined as

$$\begin{aligned} \mathbf{p}_{\text{IMD}}^o &= [|h_{\text{IMD},1}|^2, |h_{\text{IMD},2}|^2, \dots, |h_{\text{IMD},N}|^2, \mathbf{0}_{M-N}^T, \\ &|g_{\text{IMD},1}|^2, |g_{\text{IMD},2}|^2, \dots, |g_{\text{IMD},N}|^2, \mathbf{0}_{M-N}^T]^T \quad (59) \end{aligned}$$

Also note that $\Re[\mathbf{Q}_3(\infty)] = \mathbf{Q}_3(\infty)$, since it is real-valued.

APPENDIX B

MINIMIZATION OF $\mathcal{C}_{\mathbf{R}^b}$

By defining a positive variable $\epsilon = k_{\text{TIQ}}^3\sigma_x^4$, the condition number $\mathcal{C}_{\mathbf{R}^b}$ in (57) becomes

$$\mathcal{C}_{\mathbf{R}^b} = \frac{1 + 6\epsilon + \sqrt{1 - 2\epsilon + 36\epsilon^2}}{1 + 6\epsilon - \sqrt{1 - 2\epsilon + 36\epsilon^2}}$$

After some mathematical manipulations, its first derivative with respect to ϵ can be expressed as

$$\frac{\partial \mathcal{C}_{\mathbf{R}^b}}{\partial \epsilon} = \frac{14(6\epsilon - 1)}{(1 + 6\epsilon - \sqrt{1 - 2\epsilon + 36\epsilon^2})^2 \sqrt{1 - 2\epsilon + 36\epsilon^2}}$$

It can be easily verified that this function has a zero only at $\epsilon = \frac{1}{6}$, and its denominator is always positive for $\epsilon > 0$. Therefore, for $\epsilon \in (0, \frac{1}{6})$, we have $\partial \mathcal{C}_{\mathbf{R}^b} / \partial \epsilon < 0$, while for $\epsilon \in (\frac{1}{6}, \infty)$, we have $\partial \mathcal{C}_{\mathbf{R}^b} / \partial \epsilon > 0$, indicating that the global minimum of $\mathcal{C}_{\mathbf{R}^b}$ exists at $\epsilon = \frac{1}{6}$. After some algebraic manipulations, the minimum condition number is found to be $\frac{17+4\sqrt{15}}{7} \approx 4.64$.

REFERENCES

- [1] S. Hong, J. Brand, J. Choi, M. Jain, J. Mehlman, S. Katti, and P. Levis, "Applications of self-interference cancellation in 5G and beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb. 2014.
- [2] S. Goyal, P. Liu, S. S. Panwar, R. A. DiFazio, R. Yang, and E. Bala, "Full duplex cellular systems: Will doubling interference prevent doubling capacity?" *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 121–127, May 2015.
- [3] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Commun. Surv. Tutorials*, vol. 17, no. 4, pp. 2017–2046, Fourth quarter 2015.
- [4] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.
- [5] K. E. Kolodziej, J. G. McMichael, and B. T. Perry, "Multitap RF canceller for in-band full-duplex wireless communications," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 6, pp. 4321–4334, Jun. 2016.
- [6] J. Choi, M. Jain, and K. Srinivasan, "Achieving single channel, full duplex wireless communication," in *Proc. ACM MOBICOM*, Sep. 2010, pp. 1–12.
- [7] M. Jain, J. I. Choi, T. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. ACM MOBICOM*, Sep. 2011, pp. 301–312.
- [8] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.
- [9] D. Bharadia, E. McMillin, and S. Katti, "Full duplex radios," *ACM SIGCOMM Cmp. Commun. Rev.*, vol. 43, no. 4, pp. 375–386, Aug. 2013.
- [10] D. Bharadia, S. Katti, and I. Nsdi, "Full duplex MIMO radios," in *Proc. 11th USENIX Symp. NSDI*, Apr. 2014, pp. 359–372.
- [11] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 2, pp. 680–694, Feb. 2014.
- [12] E. Everett, M. Duarte, C. Dick, and A. Sabharwal, "Empowering full-duplex wireless communication by exploiting directional diversity," in *Proc. 45th Asilomar Conf. Signals, Syst., Comput. (ASILOMAR)*, Nov. 2011, pp. 2002–2006.
- [13] A. Sahai, S. Diggavi, and A. Sabharwal, "On degrees-of-freedom of full-duplex uplink/downlink channel," in *Proc. IEEE ITW*, Sep. 2013, pp. 1–5.
- [14] T. Snow, C. Fulton, and W. J. Chappell, "Transmit-receive duplexing using digital beamforming system to cancel self-interference," *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 12, pp. 3494–3503, Dec. 2011.
- [15] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Dec. 2011.
- [16] D. Korpi, J. Tamminen, M. Turunen, T. Huusari, Y. S. Choi, L. Anttila, S. Talwar, and M. Valkama, "Full-duplex mobile device: Pushing the limits," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 80–87, Sep. 2016.
- [17] Y. S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.
- [18] L. Anttila, D. Korpi, V. Syrjälä, and M. Valkama, "Cancellation of power amplifier induced nonlinear self-interference in full-duplex transceivers," *Proc. 47th Asilomar Conf. Signals, Syst., Comput. (ASILOMAR)*, pp. 1193–1198, 2013.

- [19] D. Korpi, T. Riihonen, V. Syrjälä, L. Anttila, M. Valkama, and R. Wichman, "Full-duplex transceiver system calculations: Analysis of ADC and linearity challenges," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 7, pp. 3821–3836, Jul. 2014.
- [20] V. Syrjälä, M. Valkama, L. Anttila, T. Riihonen, and D. Korpi, "Analysis of oscillator phase-noise effects on self-interference cancellation in full-duplex OFDM radio transceivers," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 6, pp. 2977–2990, Jun. 2014.
- [21] X. Quan, Y. Liu, S. Shao, C. Huang, and Y. Tang, "Impacts of phase noise on digital self-interference cancellation in full-duplex communications," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1881–1893, Apr. 2017.
- [22] A. Tarighat, R. Bagheri, and A. H. Sayed, "Compensation schemes and performance analysis of IQ imbalances in OFDM receivers," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3257–3268, Aug. 2005.
- [23] L. Anttila, M. Valkama, and M. Renfors, "Frequency-selective I/Q mismatch calibration of wideband direct-conversion transmitters," *IEEE Trans. Circuits Syst. II*, vol. 55, no. 4, pp. 359–363, Apr. 2008.
- [24] —, "Circularity-based I/Q imbalance compensation in wideband direct-conversion receivers," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2099–2113, Jul. 2008.
- [25] Z. Li, Y. Xia, W. Pei, K. Wang, Y. Huang, and D. P. Mandic, "Noncircular measurement and mitigation of I/Q imbalance for OFDM-based WLAN transmitters," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 3, pp. 383–393, Jan. 2017.
- [26] S. Li and R. D. Murch, "An investigation into baseband techniques for communication systems," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 9, pp. 4794–4806, Sep. 2014.
- [27] D. Korpi, L. Anttila, V. Syrjala, and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion full-duplex transceiver," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1674–1687, Sep. 2014.
- [28] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wirel. Commun. Netw.*, vol. 2017, no. 1, p. 24, 2017.
- [29] S. Javidi, M. Pedzisz, S. L. Goh, and D. P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," in *Proc. 1st IARP Workshop Cogn. Inform. Process.*, Jun. 2008, pp. 54–57.
- [30] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. John Wiley & Sons, 2009.
- [31] Y. Xia, S. C. Douglas, and D. P. Mandic, "Adaptive frequency estimation in smart grid applications: Exploiting noncircularity and widely linear adaptive estimators," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 44–54, Oct. 2012.
- [32] M. Sakai, H. Lin, and K. Yamashita, "Self-interference cancellation in full-duplex wireless with IQ imbalance," *Phys. Commun.*, vol. 18, pp. 2–14, Dec. 2016.
- [33] P. I. Mak, S. P. U, and R. P. Martins, "Transceiver architecture selection: Review, state-of-the-art survey and case study," *IEEE Circuits Syst. Mag.*, vol. 7, no. 2, pp. 6–25, Sep. 2007.
- [34] Q. Gu, *RF System Design of Transceivers for Wireless Communications*. New York, NJ, USA: Springer-Verlag, 2006.
- [35] S. Wei, D. L. Goeckel, and P. A. Kelly, "Convergence of the complex envelope of bandlimited OFDM signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4893–4904, Oct. 2010.
- [36] B. Fisher and N. J. Bershad, "The complex LMS adaptive algorithm—Transient weight mean and covariance with application to the ALE," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 1, pp. 34–44, Feb. 1983.
- [37] S. C. Douglas and D. P. Mandic, "Performance analysis of the conventional complex LMS and augmented complex LMS algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2010, pp. 3794–3797.
- [38] D. P. Mandic, S. Kanna, and S. C. Douglas, "Mean square analysis of the CLMS and ACLMS for non-circular signals: The approximate uncorrelating transform approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, no. 2, Mar. 2015, pp. 3531–3535.
- [39] Y. Xia and D. Mandic, "Complementary mean square analysis of augmented CLMS for second order noncircular Gaussian signals," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1413–1417, Sep. 2017.
- [40] —, "A full mean square analysis of CLMS for second-order noncircular inputs," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5578–5590, Nov. 2017.
- [41] Y. Xia and D. P. Mandic, "Augmented performance bounds on strictly linear and widely linear estimators with complex data," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 507–514, Jan. 2018.
- [42] K. Mayyas, "Performance analysis of the deficient length LMS adaptive algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2727–2740, Aug. 2005.
- [43] Y. Xia, S. C. Douglas, and D. P. Mandic, "Performance analysis of the deficient length augmented CLMS algorithm for second order noncircular complex signals," *Signal Process.*, vol. 144, pp. 214–225, 2018.
- [44] Y. Xia, B. Jelfs, M. M. Van Hulle, J. C. Principe, and D. P. Mandic, "An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 74–83, Nov. 2011.
- [45] D. Xu, H. Zhang, and D. P. Mandic, "Convergence analysis of an augmented algorithm for fully complex-valued neural networks," *Neural Netw.*, vol. 69, pp. 44–50, Sep. 2015.
- [46] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 639–652, Mar. 2003.
- [47] A. H. Sayed, *Adaptive Filters*. John Wiley & Sons, 2011.



Zhe Li (S'16) received the B.S. degree from the Nanjing University of Posts and Telecommunication, Nanjing, China, in 2011, and the M.S. degree from Southeast University, Nanjing, in 2014, where he is currently pursuing the Ph.D. degree with the School of Information and Engineering. From 2016 to 2017, He was a visiting Ph.D. student with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K. His research interests include complex-valued statistical signal processing and its applications in communications.



Yili Xia (M'11) received the B.Eng. degree in Information Engineering from Southeast University, Nanjing, China, in 2006, and the M.Sc. (Distinction) degree in Communications & Signal Processing from Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., in 2007. He obtained his Ph.D. degree in adaptive signal processing from Imperial College London in 2011.

Since 2013 he has been an Associate Professor with the School of Information and Engineering, Southeast University, Nanjing, China, where he is currently the Deputy Head of the Department of Information and Signal Processing Engineering. His research interests include complex and hyper-complex statistical analysis, linear and nonlinear adaptive filters, as well as their applications on communications and power systems.



Wenjiang Pei received the M.S. and Ph.D. degrees in instrumentation and measurement from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1995 and 1997, respectively.

He is currently a Professor with the School of Information and Engineering, Southeast University, Nanjing. His research interests signal processing and hardware instrumentation for communications.



Kai Wang received the Ph.D. degree in signal processing from the School of Information and Engineering, Southeast University, Nanjing, China, in 2009. He is currently an Associate Professor in signal processing with Southeast University.

His current research interests include parameter estimation and signal processing for communications.



Danilo P. Mandic (M'99-SM'03-F'12) received the Ph.D. degree in nonlinear adaptive signal processing from Imperial College London, London, U.K., in 1999, where he is currently a Professor in Signal Processing.

He has been working in the area of nonlinear adaptive signal processing, multivariate data analysis and nonlinear dynamics. He has been a Guest Professor at Katholieke Universiteit Leuven, Leuven, Belgium, the Tokyo University of Agriculture and Technology, Tokyo, Japan, and Westminster University, London, and a Frontier Researcher at RIKEN, Japan. His publication record includes two research monographs entitled *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability* (1st ed., Aug. 2001) and *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models* (1st ed., Wiley, Apr. 2009), an edited book titled *Signal Processing Techniques for Knowledge Extraction and Information Fusion* (Springer, 2008) and more than 200 publications on signal and image processing.

Prof. Mandic has been a member of the IEEE Technical Committee on Signal Processing Theory and Methods, and has been an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS, and the International Journal of Mathematical Modelling and Algorithms. He is currently an Associate Editor for the IEEE SIGNAL PROCESSING MAGAZINE and IEEE TRANSACTIONS ON INFORMATION AND SIGNAL PROCESSING OVER NETWORKS. He has produced award winning papers and products resulting from his collaboration with Industry.