Joint Least Squares Estimation of Frequency, DC Offset, I-Q Imbalance, and Channel in MIMO Receivers

Chen-Jiu Hsu, Racy Cheng, and Wern-Ho Sheen, *Member, IEEE*

***Abstract*—Multiple-input–multiple-output (MIMO) receivers with direct-conversion radio-frequency (RF) architecture are investigated in this paper. Direct conversion is a low-cost RF design that requires fewer external components in chip implementation. Nevertheless, it introduces extra RF impairments such as I-Q imbalance and dc offset in addition to frequency offset that is commonly encountered in all RF architectures. This paper proposes to do the joint least squares (LS) estimation of frequency, dc offset, I-Q imbalance, and a channel in MIMO receivers to improve performance; frequency-dependent and frequency-independent I-Q imbalances are included. Previously, RF impairments were separately estimated in MIMO receivers, which leads to inferior performance. In particular, a receiver architecture that facilitates the joint estimation of the frequency, the dc offset, the I-Q imbalance, and the channel is proposed. The LS criterion is then applied to obtain the joint estimators, with a special training-sequence design to reduce complexity. Simplified estimators on the frequency and dc offset are also proposed with almost no loss in performance. Finally, the estimators are shown through analysis to be unbiased and approach the Cramér–Rao lower bound (CRLB) for signal-to-noise ratios (SNRs) of interest. The analysis is verified by computer simulations.**

***Index Terms*—Direct-conversion RF architecture, joint least squares (LS) estimations, multiple-input–multiple-output (MIMO) receiver.**

I. Introduction

USING multiple transmit and receive antennas in a wireless communication system, a.k.a. the multiple-input–multiple-output (MIMO) system, is capable of providing diversity gain, array gain (power gain), and/or degree-of-freedom gain over the single-input–single-output (SISO) systems [1]–[3]. Space-time coding, beam forming, and spatial multiplexing are modes of operations for exploiting the diversity gain, the array gain, and the degree-of-freedom gain, respectively. MIMO has been one of the key technologies for enabling a high-data-rate, high-spectral-efficiency transmission in wireless communications.

Manuscript received December 27, 2007; revised April 23, 2008. First published September 19, 2008; current version published May 11, 2009. The review of this paper was coordinated by Dr. T. Taniguchi.

C.-J. Hsu and W.-H. Sheen are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: linch.ee89@nctu.edu.tw; whsheen@cm.nctu.edu.tw).

R. Chen is with the Department of Electronic Engineering, Minghsin University of Science and Technology, Hsinchu 304, Taiwan (e-mail: racy@must.edu.tw).

Digital Object Identifier 10.1109/TVT.2008.2005989

Using direct-conversion radio-frequency (RF) architecture, on the other hand, is a low-cost design that requires fewer external components in chip implementation [4], [5]. Direct-conversion RF architecture, nevertheless, introduces extra RF impairments such as I-Q imbalance and dc offset, in addition to frequency offset that is commonly encountered in all types of RF architectures [4], [5]. Generally, at the receiver, RF parameters (impairments) need to be estimated and compensated from the received signal before it can be forwarded to the next step of signal processing for symbol detection and decoding.

Receiver designs with direct-conversion RF have been an important research topic (first for SISO systems [6]–[15] and then extended to MIMO systems [16]–[19]). In [6], the frequency-independent I-Q imbalance from a transmitter and a receiver was investigated along with the dc offset. Performance degradation was evaluated, and adaptive compensation schemes were proposed by using gradient-type adaptive algorithms. However, no frequency offset was considered in the work. In [7]–[9], statistical signal processing-based (blind) techniques were proposed to adaptively cancel the self-image interference due to frequency-dependent [8], [9] and frequency-independent I-Q imbalances [7], respectively, without considering the frequency offset and the dc offset. No training data are needed in these schemes at the expense of a slow convergence rate. In [10]–[12], frequency-dependent and frequency-independent I-Q imbalances were investigated for orthogonal frequency-division multiplexing (OFDM) systems at the absence of the frequency offset and the dc offset. In [10], the I-Q imbalance was compensated by using a simple two-tap adaptive frequency-domain equalizer, whereas in [11] and [12], pre-fast Fourier transform (pre-FFT) and post-FFT schemes were proposed, with the former for frequency-independent and the latter for both frequency-dependent and frequency-independent I-Q imbalances at the receiver; in addition, a predistortion scheme at the transmitter was proposed in [12] for compensating the transmitter I-Q imbalance. In [13], the I-Q imbalance was jointly investigated with the frequency offset, where a finite impulse response (FIR) filter was used for compensating the frequency-dependent I-Q imbalance and an asymmetric phase compensator for the frequency-independent one. Gil *et al.* [14] developed a joint estimation of the frequency offset, the channel, the (frequency-independent) I-Q imbalance, and the dc offset in the time domain. Under the assumption of white Gaussian noise at the output of the sampler, the maximum-likelihood (ML) criterion was used to derive the solution. Finally, in

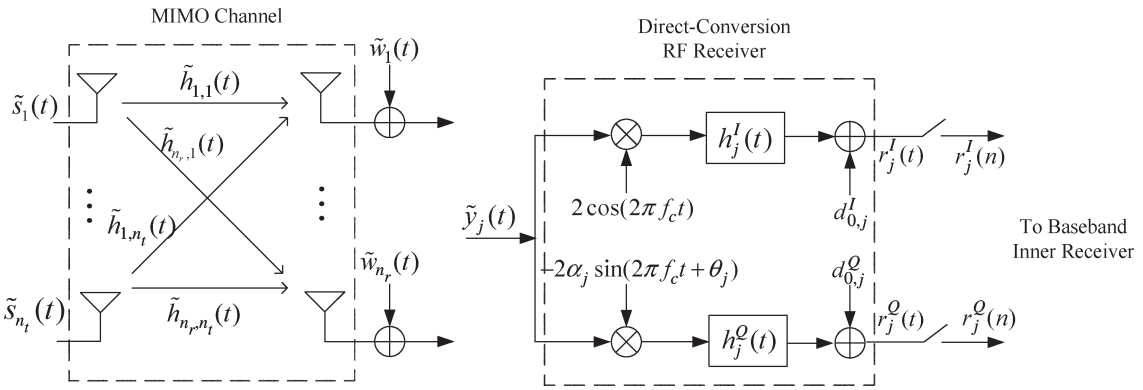

Fig. 1. MIMO signal model with direct-conversion RF architecture.

[15], a joint adaptive compensation for transmitter and receiver I-Q imbalances was investigated for OFDM systems under the effect of the frequency offset.

Recently, receiver designs with direct-conversion RF have drawn greater attention in MIMO systems [16]–[19]. In [16]–[19], the I-Q imbalance was jointly investigated with symbol detection for MIMO-OFDM systems, with no consideration of the frequency offset and the dc offset. Using an extended channel model that incorporates the I-Q imbalance, symbol detection was performed on the extended channel without an explicit estimation/compensation of the I-Q imbalance. Nonadaptive [e.g., least squares (LS)] or adaptive (e.g., recursive LS and least mean squares) filtering can be used for estimating the extended channel prior to symbol detection [16]–[19]. Likewise, MIMO detectors such as maximum likelihood (ML), zero forcing, and minimum mean square error (MMSE) [3] can be employed for symbol detection. In [17]–[19], transmitter and receiver I-Q imbalances were considered, whereas in [16], only the receiver I-Q imbalance was of interest. Note that using the extended channel increases system dimension and, hence, detection complexity as compared to the receiver in which RF impairments are estimated and compensated before the symbol detection.

So far, most MIMO receiver designs with direct-conversion RF treat the I-Q imbalance as the only impairment, assuming that all others have been estimated separately and compensated beforehand. Unfortunately, under the same optimization criterion, a separate estimation of RF parameters leads to inferior performance and may not be feasible if the estimation of one parameter is seriously degraded at the presence of others. In this paper, we propose to do the joint estimation of the frequency, the dc offset, the I-Q imbalance, and the channel in MIMO receivers to improve its performance. First, a receiver architecture that facilitates the joint estimation of the frequency, the dc offset, the I-Q imbalance, and the channel is proposed; frequency-dependent and frequency-independent I-Q imbalances are included. Second, the LS criterion is applied to obtain the joint estimators, with a special training-sequence design to reduce complexity. Simplified estimators on the frequency and dc offset are also proposed with almost no loss in performance. Lastly, the LS estimators are shown through analysis to be unbiased and approach the Cramér–Rao lower bound (CRLB) for signal-to-noise ratios (SNRs) of interest.

The rest of this paper is organized as follows. The MIMO signal model with the effects of the frequency offset, the dc offset, the I-Q imbalance, and the multipath fading channel is described in Section II. Section III proposes a receiver architecture that facilitates the joint estimation of RF parameters. The detail of the joint LS estimation is derived in Section IV, with its performance analyzed in Section V. Numerical results are given in Section VI, and conclusions are given in Section VII.

II. System Model

Fig. 1 depicts the MIMO signal model with direct-conversion RF architecture. n_t and n_r are the number of transmit and receive antennas, respectively. Consider a block transmission with a prefix to avoid interblock interference. The bandpass-transmitted signal from the ith transmit antenna is $\tilde{s}_i(t) = \mathrm{Re}\{s_i(t)e^{\boldsymbol{j}2\pi f_c t}\}$, with the baseband signal

$$s_i(t) = \sum_k \sum_{n=-N_g}^{N-1} s_{i,k}(n) g_T\left(t - (k(N+N_g)+n)T_s\right)$$

where f_c is the carrier frequency, $\boldsymbol{j} = \sqrt{-1}$, $s_{i,k}(n)$ is the transmitted kth block symbol with transmit power $\sigma_s^2 = (1/(N+N_g))\sum_{i=1}^{n_t}\sum_{n=-N_g}^{N-1}|s_{i,k}(n)|^2$, N_g is the length of the prefix, N is the length of useful data, $g_T(t)$ is the transmit filter with unit power, and T_s is the symbol time. For a data-aided estimation as considered here, $s_{i,k}(n)$ is perfectly known to the receiver. A total of $P \geq 1$ training blocks will be assumed in this paper, starting from the zeroth block $(k = 0)$.

Let $\tilde{h}_{j,i}(t) = \mathrm{Re}\{h_{j,i}(t)e^{\boldsymbol{j}2\pi f_c t}\}$ denote the channel response from the ith transmit to the jth receive antenna, and let $h_{j,i}(t)$ be its baseband equivalent. The bandpass signal that is received from the jth antenna is given by

$$\tilde{y}_j(t) = \mathrm{Re}\left\{y_j(t)e^{\boldsymbol{j}2\pi(f_c+\Delta f)t}\right\} + \tilde{w}_j(t)$$

where $y_j(t) = \sum_{i=1}^{n_t} s_i(t) \otimes h_{j,i}(t)$, $\tilde{w}_j(t) = \mathrm{Re}\{w_{0,j}(t)e^{\boldsymbol{j}2\pi f_c t}\}$ is the bandpass additive white Gaussian noise [$w_{0,j}(t)$ is its baseband equivalent], Δf is the frequency offset that is the same for all receive branches, and $\otimes$ denotes the operation of convolution.

Taking into account the frequency-independent I-Q imbalance at the RF front end, the complex sinusoidal signal coming into the mixer of the jth branch is given by

$$\begin{aligned} c_j(t) &= 2\cos(2\pi f_c t) - \boldsymbol{j} 2\alpha_j \sin(2\pi f_c t + \theta_j) \\ &= \gamma_j e^{-\boldsymbol{j}2\pi f_c t} + \varphi_j e^{\boldsymbol{j}2\pi f_c t} \end{aligned}$$

where α_j and θ_j are the gain and the phase imbalance, respectively, $\gamma_j = (1 + \alpha_j e^{-\boldsymbol{j}\theta_j})$, and $\varphi_j = (1 - \alpha_j e^{\boldsymbol{j}\theta_j})$. Note that α_j and θ_j are different from one branch to another. After down conversion, $h_j^I(t)$ and $h_j^Q(t)$ are the baseband filters that are used to remove out-of-band noise and high-frequency components. If $h_j^I(t) \neq h_j^Q(t)$, we say that there exists frequency-dependent I-Q imbalance. The frequency-dependent I-Q imbalance is mostly encountered in a wide-band RF receiver because it is generally difficult to maintain baseband filters to have the same response over a wide frequency range [8], [13].

After a simple manipulation, baseband signal $r_j(t)$ is given by

$$\begin{aligned} r_j(t) &\doteq r_j^I(t) + \boldsymbol{j} r_j^Q(t) \\ &= h_{+,j}(t) \otimes \left[y_j(t) e^{\boldsymbol{j}2\pi\Delta f t} + w_{0,j}(t) \right] \\ &\quad + h_{-,j}(t) \otimes \left[y_j(t) e^{\boldsymbol{j}2\pi\Delta f t} + w_{0,j}(t) \right]^* + d_{0,j} \end{aligned} \quad (1)$$

where $h_{+,j}(t) = 1/2 \cdot [h_j^I(t) + h_j^Q(t)\alpha_j e^{-\boldsymbol{j}\theta_j}]$, $h_{-,j}(t) = 1/2 \cdot [h_j^I(t) - h_j^Q(t)\alpha_j e^{\boldsymbol{j}\theta_j}]$, $d_{0,j} = d_{0,j}^I + \boldsymbol{j} d_{0,j}^Q$ is the dc offset, and $[x]^*$ denotes the complex conjugate of x. After sampling, the end-to-end equivalent discrete system can be modeled as (up to a constant for the case of no aliasing)

$$\begin{aligned} r_j(n) &= h_{+,j}(n) \otimes \left[y_j(n) e^{\boldsymbol{j}2\pi\nu n} + w_{0,j}(n) \right] \\ &\quad + h_{-,j}(n) \otimes \left[y_j(n) e^{\boldsymbol{j}2\pi\nu n} + w_{0,j}(n) \right]^* + d_{0,j} \end{aligned} \quad (2)$$

where $h_{+,j}(n) = 1/2 \cdot [h_j^I(n) + h_j^Q(n)\alpha_j e^{-\boldsymbol{j}\theta_j}]$, $h_{-,j}(n) = 1/2 \cdot [h_j^I(n) - h_j^Q(n)\alpha_j e^{\boldsymbol{j}\theta_j}]$, $\nu = \Delta f T_S$ is the normalized frequency offset, $y_j(n) = \sum_{i=1}^{n_t} s_i(n) \otimes h_{j,i}(n)$, and $w_{0,j}(n)$ is the zero-mean additive white Gaussian noise with $\sigma_w^2 \doteq E[|w_{0,j}(n)|^2]$, $j = 1, \ldots, n_r$. For use in Section V, $r_j(n)$ is rewritten as

$$r_j(n) = r_{j,d}(n) + r_{j,r}(n) \quad (3)$$

where $r_{j,d}(n) = h_{+,j}(n) \otimes y_j(n) e^{\boldsymbol{j}2\pi\nu n} + h_{-,j}(n) \otimes y_j^*(n) e^{-\boldsymbol{j}2\pi\nu n} + d_{0,j}$ is the deterministic part, and $r_{j,r}(n) = h_{+,j}(n) \otimes w_{0,j}(n) + h_{-,j}(n) \otimes w_{0,j}^*(n)$ is the random part due to noise. $\{r_{j,r}(n)\}$ is generally the colored Gaussian noise.

Equation (2) says that the effect of the I-Q imbalance introduces self-image interference in the received signal. With $h_j^I(n) = h_j^Q(n) = h_j(n)$, (2) degenerates to the case of no frequency-dependent I-Q imbalance, with $h_{+,j}(n) = 1/2 \cdot \gamma_j h_j(n)$ and $h_{-,j}(n) = 1/2 \cdot \varphi_j h_j(n)$. In addition, if $\alpha_j = 1$, $\theta_j = 0$, and $h_j^I(n) = h_j^Q(n) = h_j(n)$, we have the no I-Q imbalance case, i.e.,

$$r_j(n) = h_j(n) \otimes \left[y_j(n) e^{\boldsymbol{j}2\pi\nu n} + w_{0,j}(n) \right] + d_{0,j}.$$

III. Receiver Architecture

Motivated by (2), one way to process the received signal is to cancel out first the self-image interference due to the I-Q imbalance. By introducing filter $\rho_j(n)$, we have

$$\begin{aligned} & r_j(n) - \rho_j(n) \otimes r_j^*(n) \\ &= \left(h_{+,j}(n) - \rho_j(n) \otimes h_{-,j}^*(n) \right) \otimes \left(y_j(n) e^{\boldsymbol{j}2\pi\nu n} + w_{0,j}(n) \right) \\ &\quad + \underbrace{\left(h_{-,j}(n) - \rho_j(n) \otimes h_{+,j}^*(n) \right)}_{=0} \\ &\quad \otimes \left(y_j(n) e^{\boldsymbol{j}2\pi\nu n} + w_{0,j}(n) \right)^* + \left(d_{0,j} - \rho_j(n) \otimes d_{0,j}^* \right). \end{aligned} \quad (4)$$

To completely cancel out the self-image interference, $\rho_j(n) = (h_{+,j}^*(n))^{-1} \otimes h_{-,j}(n)$, where $(h_{+,j}^*(n))^{-1}$ is the inverse filter of $h_{+,j}^*(n)$. [For the case of no frequency-dependent I-Q imbalance, $\rho_j(n) = \varphi_j / \gamma_j^*$ as is given in [14].] Thus, (4) becomes

$$\begin{aligned} & r_j(n) - \rho_j(n) \otimes r_j^*(n) \\ &= \underbrace{\left(h_{+,j}(n) - \rho_j(n) \otimes h_{-,j}^*(n) \right)}_{=g_j(n)} \\ &\quad \otimes \left(y_j(n) e^{\boldsymbol{j}2\pi\nu n} + w_{0,j}(n) \right) + \left(d_{0,j} - \rho_j(n) \otimes d_{0,j}^* \right) \\ &= e^{\boldsymbol{j}2\pi\nu n} \left(\sum_{i=1}^{n_t} s_i(n) \otimes g_{j,i}(n) \right) + d_j + w_j(n) \end{aligned} \quad (5)$$

where $g_j(n) \doteq h_{+,j}(n) - \rho_j(n) \otimes h_{-,j}^*(n)$, $g_{j,i}(n) \doteq h_{j,i}(n) \otimes (g_j(n) e^{-\boldsymbol{j}2\pi\nu n})$, $d_j = d_{0,j} - \rho_j(n) \otimes d_{0,j}^*$, and $w_j(n) = g_j(n) \otimes w_{0,j}(n)$. $g_{j,i}(n)$ is the overall impulse response from the ith transmit to the jth receive antenna after canceling out the self-image interference, d_j is the equivalent dc offset, and $w_j(n)$ is the additive Gaussian noise, but generally not white. In (5), $\rho_j(n)$, $g_{j,i}(n)$, ν, and d_j are the deterministic unknown parameters to be estimated. In what follows, $\rho_j(n)$ and $g_{j,i}(n)$ will be approximated as FIR filters with large enough taps, although they are generally infinite impulse response ones, as can be seen in (4) and (5).

Following (5), we propose the receiver architecture as in Fig. 2; after canceling the self-image interference, the dc offset is compensated next, and then, the compensation for the frequency offset follows. Parameters $\rho_j(n)$, $g_{j,i}(n)$, ν, and d_j will be jointly estimated in the LS sense, which will be discussed in Section IV, and the MIMO detection is done with the MMSE detector based on estimated channel responses $\{\hat{g}_{j,i}(n)\}$ along with the compensated received signals from all branches. Some other types of MIMO detectors can be used as well [20].

IV. Joint Least Squares Estimation

Here, the set of parameters $\rho_j(n)$, $g_{j,i}(n)$, ν, and d_j are jointly estimated in the sense of LS. To do that, $\rho_j(n)$

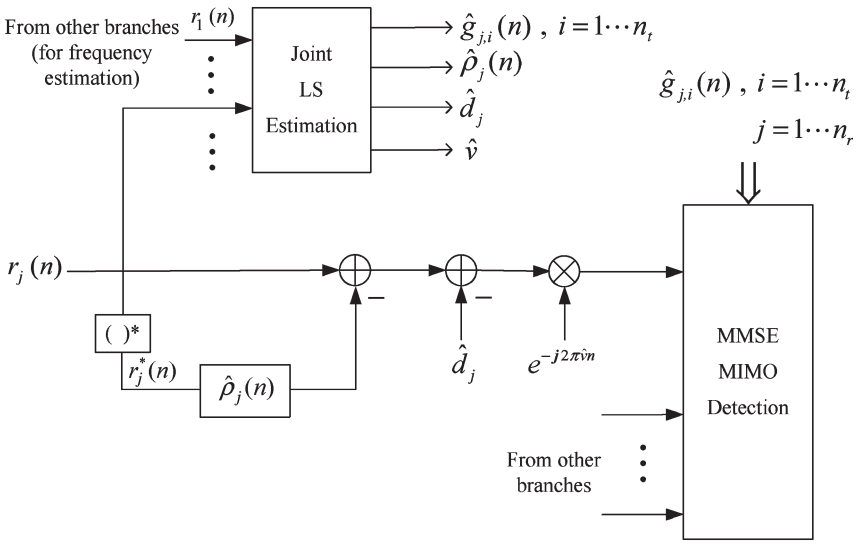

Fig. 2. MIMO receiver with joint LS estimation/compensation of RF parameters.

and $g_{j,i}(n)$ are approximated by FIR filters $\boldsymbol{\rho}_j = [\rho_j(0), \rho_j(1), \ldots, \rho_j(L_{\boldsymbol{\rho}_j}-1)]^T$ and $\mathbf{g}_{j,i} = [g_{j,i}(0), g_{j,i}(1), \ldots, g_{j,i}(L_{\mathbf{g}_{j,i}}-1)]^T$, with length $L_{\boldsymbol{\rho}_j}$ and $L_{\mathbf{g}_{j,i}}$, respectively.[1] In the following, $L_{\boldsymbol{\rho}_j} = L_{\boldsymbol{\rho}}$, $\forall j$, and $L_{\mathbf{g}_{j,i}} = L_{\mathbf{g}}$, $\forall i, j$. In addition, $L_{\boldsymbol{\rho}}$ and $L_{\mathbf{g}}$ are assumed to be large enough (to contain 99% of the energy) here and in the next sections, and, thus, the approximation error is negligible. The impact of $L_{\boldsymbol{\rho}}$ and $L_{\mathbf{g}}$ will be investigated by computer simulations in Section VI.

A. LS Estimators

Consider the case of $L_{\mathbf{g}} \le N_g$; thus, there will be no interblock interference. Let $\mathbf{r}_j(k) = [r_{j,k}(0), r_{j,k}(1), \ldots, r_{j,k}(N-1)]^T$ be the useful part of the kth received block, where $r_{j,k}(n) \doteq r_j(k(N_g+N)+n)$; let $\mathbf{R}_j(k)$ be the $N \times L_{\boldsymbol{\rho}}$ received signal matrix with the (m,l)th entry $[\mathbf{R}_j(k)]_{m,l} = r_{j,k}(m-l)$ for $0 \le m \le N-1$ and $0 \le l \le L_{\boldsymbol{\rho}}-1$; and let $\mathbf{S}_i(k)$ be the $N \times L_{\mathbf{g}}$ signal matrix with $[\mathbf{S}_i(k)]_{m,l} = s_{i,k}(m-l)$ for $0 \le m \le N-1$ and $0 \le l \le L_{\mathbf{g}}-1$. From (5), the useful part of the jth received signal can be written as the following vector form:

$$\mathbf{r}_j(k) - \mathbf{R}_j^*(k)\boldsymbol{\rho}_j = \Gamma_k(\nu)\left(\sum_{i=1}^{n_t}\mathbf{S}_i(k)\mathbf{g}_{j,i}\right) + d_j\mathbf{1}_N + \mathbf{w}_j(k) \tag{6}$$

for $k = 0, \ldots, P-1$, where $\Gamma_k(\nu) = e^{\boldsymbol{j}2\pi k(N_g+N)\nu}$. $\mathrm{diag}\{1, e^{\boldsymbol{j}2\pi\nu}, \ldots, e^{\boldsymbol{j}2\pi\nu(N-1)}\}$ is a diagonal matrix, $\mathbf{1}_N$ is the vector of ones with dimension N, and $\mathbf{w}_j(k) = [w_{j,k}(0), w_{j,k}(1), \ldots, w_{j,k}(N-1)]^T$, with $w_{j,k}(n) = w_j(k(N_g+N)+n)$.

Furthermore, let $\mathbf{r}_j = [\mathbf{r}_j^T(0), \mathbf{r}_j^T(1), \ldots, \mathbf{r}_j^T(P-1)]^T$, $\mathbf{R}_j = [\mathbf{R}_j^T(0), \mathbf{R}_j^T(1), \ldots, \mathbf{R}_j^T(P-1)]^T$, $\mathbf{S}(k) = [\mathbf{S}_1(k), \mathbf{S}_2(k), \ldots, \mathbf{S}_{n_t}(k)]$, $\mathbf{g}_j = [\mathbf{g}_{j,1}^T, \mathbf{g}_{j,2}^T, \ldots, \mathbf{g}_{j,n_t}^T]^T$, and $\mathbf{w}_j = [\mathbf{w}_j^T(0), \mathbf{w}_j^T(1), \ldots, \mathbf{w}_j^T(P-1)]^T$. The total useful received signal for training is

$$\mathbf{r}_j - \mathbf{R}_j^*\boldsymbol{\rho}_j = \Gamma(\nu)\mathbf{S}\mathbf{g}_j + d_j\mathbf{1} + \mathbf{w}_j \tag{7}$$

where $\mathbf{S} = [\mathbf{S}^T(0), \mathbf{S}^T(1), \ldots, \mathbf{S}^T(P-1)]^T$, and $\Gamma(\nu) = \mathrm{diag}\{\Gamma_0(\nu), \Gamma_1(\nu), \ldots, \Gamma_{P-1}(\nu)\}$. From (7), the joint LS estimates of all parameters are obtained by minimizing the cost function[2]

$$\Lambda(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j, \tilde{d}_j, \tilde{\mathbf{g}}_j, j = 1, \ldots, n_r) = \sum_{j=1}^{n_r}\Lambda_j(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j, \tilde{d}_j, \tilde{\mathbf{g}}_j) \tag{8}$$

with

$$\Lambda_j(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j, \tilde{d}_j, \tilde{\mathbf{g}}_j) = \left\|\mathbf{r}_j - \mathbf{R}_j^*\tilde{\boldsymbol{\rho}}_j - \tilde{d}_j\mathbf{1} - \Gamma(\tilde{\nu})\mathbf{S}\tilde{\mathbf{g}}_j\right\|^2. \tag{9}$$

Recall that each branch of the receiver has its own I-Q imbalance, dc offset, and channel response; however, the frequency offset is the same for all branches. Therefore, given fixed trial frequency offset $\tilde{\nu}$, $\boldsymbol{\rho}_j$, d_j, and $\mathbf{g}_j$ can be estimated by simply minimizing cost function $\Lambda_j(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j, \tilde{d}_j, \tilde{\mathbf{g}}_j)$. In other words, $\boldsymbol{\rho}_j$, d_j, and $\mathbf{g}_j$ can be independently estimated from one branch to another. On the other hand, (8) can be used to jointly estimate the frequency offset to increase the performance by exploiting the diversity and the power gain that are inherent in MIMO systems.

The LS solution of (8) or (9) can be successively obtained as follows [14]. First, under fixed trial $(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j, \tilde{d}_j)$, the LS estimate of channel $\hat{\mathbf{g}}_j$ is given by [21]

$$\hat{\mathbf{g}}_j(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j, \tilde{d}_j) = (\Gamma(\tilde{\nu})\mathbf{S})^{\dagger}\left(\mathbf{r}_j - \mathbf{R}_j^*\tilde{\boldsymbol{\rho}}_j - \tilde{d}_j\mathbf{1}\right) \tag{10}$$

[1]Throughout this paper, bold uppercase letters denote matrices, and bold lowercase letters denote vectors. $(\cdot)^T$ and $(\cdot)^H$ represent the operations of conjugate transpose and transpose of a matrix or a vector, respectively. In addition, $\mathbf{I}$ and $\mathbf{1}$ denote $\mathbf{I}_{N\cdot P}$ and $\mathbf{1}_{N\cdot P}$, respectively.

[2]The estimators that jointly maximize (8) are the ML estimators, provided that $\mathbf{w}_j$ is white Gaussian noise, which is a case that is not true here. The joint ML estimation of ν, $\boldsymbol{\rho}_j$, d_j, and $\mathbf{g}_j$ in this case is very complicated if not impossible.

where $(\mathbf{X})^{\dagger}$ denotes the pseudoinverse of $\mathbf{X}$. Using singular value decomposition

$$\mathbf{X} = \mathbf{U}\begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^H$$

we have

$$(\mathbf{X})^{\dagger} = \mathbf{V}\begin{pmatrix} \mathbf{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^H$$

where $\mathbf{\Sigma} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$, r and $\{\lambda_k\}$ are the rank and singular values of $\mathbf{X}$, respectively, and $\mathbf{U}$ and $\mathbf{V}$ are some unitary matrices. If $\mathbf{X}$ has a full column rank, then $(\mathbf{X})^{\dagger} = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H$.

Second, substituting $\hat{\mathbf{g}}_j$ in (9), we have

$$\Lambda_j(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j, \tilde{d}_j) = \left\| (\mathbf{I} - \mathbf{C}(\tilde{\nu}))\left(\mathbf{r}_j - \mathbf{R}_j^*\tilde{\boldsymbol{\rho}}_j - \tilde{d}_j\mathbf{1}\right)\right\|^2 \quad (11)$$

where $\mathbf{C}(\tilde{\nu}) = \Gamma(\tilde{\nu})\mathbf{B}\Gamma^H(\tilde{\nu})$, with $\mathbf{B} = \mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H$. Note that $\mathbf{C}(\tilde{\nu})$ and $\mathbf{B}$ are projection matrices. By minimizing (11) with respect to $\tilde{d}_j$, given fixed $(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j)$, the LS estimate for d_j is given by

$$\hat{d}_j = \left((\mathbf{I} - \mathbf{C}(\tilde{\nu}))\,\mathbf{1}\right)^{\dagger}(\mathbf{I} - \mathbf{C}(\tilde{\nu}))\left(\mathbf{r}_j - \mathbf{R}_j^*\tilde{\boldsymbol{\rho}}_j\right). \quad (12)$$

In addition, it is easy to show that

$$\mathbf{f}^H(\tilde{\nu}) \doteq \left((\mathbf{I} - \mathbf{C}(\tilde{\nu}))\,\mathbf{1}\right)^{\dagger} = \begin{cases} \frac{\mathbf{1}^H(\mathbf{I}-\mathbf{C}(\tilde{\nu}))}{\|(\mathbf{I}-\mathbf{C}(\tilde{\nu}))\mathbf{1}\|^2}, & \text{if } \|(\mathbf{I} - \mathbf{C}(\tilde{\nu}))\,\mathbf{1}\|^2 \neq 0 \\ \mathbf{0}^T, & \text{otherwise} \end{cases} \quad (13)$$

$$\hat{d}_j = \mathbf{f}^H(\tilde{\nu})\left(\mathbf{r}_j - \mathbf{R}_j^*\tilde{\boldsymbol{\rho}}_j\right). \quad (14)$$

Third, substituting $\hat{d}_j$ in (11), one obtains

$$\Lambda_j(\tilde{\nu}, \tilde{\boldsymbol{\rho}}_j) = \left\|(\mathbf{I} - \mathbf{C}(\tilde{\nu}))\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\tilde{\nu})\right)(\mathbf{r}_j - \mathbf{R}^*\tilde{\boldsymbol{\rho}}_j)\right\|^2 \quad (15)$$

and the LS estimate $\hat{\boldsymbol{\rho}}_j$ is given by

$$\hat{\boldsymbol{\rho}}_j = \mathbf{P}_j(\tilde{\nu})(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\tilde{\nu}))\,\mathbf{r}_j \quad (16)$$

where $\mathbf{P}_j(\tilde{\nu}) = [(\mathbf{I} - \mathbf{C}(\tilde{\nu}))(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\tilde{\nu}))\mathbf{R}_j^*]^{\dagger}$. Finally, substituting $\hat{\boldsymbol{\rho}}_j$ in (15) and using (8), we have

$$\Lambda(\tilde{\nu}) = \sum_{j=1}^{n_r}\Lambda_j(\tilde{\nu}) \quad (17)$$

with $\Lambda_j(\tilde{\nu}) = \|(\mathbf{I} - \mathbf{C}(\tilde{\nu}))(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\tilde{\nu}))(\mathbf{r}_j - \mathbf{R}_j^*\hat{\boldsymbol{\rho}}_j(\tilde{\nu}))\|^2$, and the LS estimate $\hat{\nu}$ is given by

$$\hat{\nu} = \arg\min_{\tilde{\nu}}\{\Lambda(\tilde{\nu})\}. \quad (18)$$

Generally, no closed form is available for $\hat{\nu}$; an exhaustive search has to be performed for the solution.

From Appendix A, it is shown that

$$\Lambda_j(\tilde{\nu}) = \mathbf{r}_j^H\mathbf{Q}(\tilde{\nu})\mathbf{r}_j - \left(\mathbf{R}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{r}_j\right)^H\left[\mathbf{R}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{R}_j^*\right]^{-1}\mathbf{R}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{r}_j \quad (19)$$

where $\mathbf{Q}(\tilde{\nu}) = (\mathbf{I} - \mathbf{1}\mathbf{f}^H(\tilde{\nu}))^H(\mathbf{I} - \mathbf{C}(\tilde{\nu}))(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\tilde{\nu}))$. Coincidentally, in (19), the relevant elements to be calculated are in the form of $\mathbf{v}^H\mathbf{Q}(\tilde{\nu})\mathbf{u}$, where $\mathbf{v} = [\,v_0 \quad v_1 \quad \cdots \quad v_{N\cdot P-1}\,]^T$ and $\mathbf{u} = [\,u_0 \quad u_1 \quad \cdots \quad u_{N\cdot P-1}\,]^T$ are $N \cdot P \times 1$ vectors. In addition, from Appendix A

$$\mathbf{v}^H\mathbf{Q}(\tilde{\nu})\mathbf{u} = \mathbf{v}^H\mathbf{u} - \mathbf{v}^H\mathbf{C}(\tilde{\nu})\mathbf{u} - \frac{\left(\mathbf{v}^H\mathbf{1} - \mathbf{v}^H\mathbf{C}(\tilde{\nu})\mathbf{1}\right)\left(\mathbf{1}^H\mathbf{u} - \mathbf{1}^H\mathbf{C}(\tilde{\nu})\mathbf{u}\right)}{\mathbf{1}^H\left(\mathbf{I} - \mathbf{C}(\tilde{\nu})\right)\mathbf{1}}. \quad (20)$$

Define $\mathbf{v}^T = [\mathbf{v}_1^T, \mathbf{v}_2^T, \ldots, \mathbf{v}_P^T]$, $\mathbf{u}^T = [\mathbf{u}_1^T, \mathbf{u}_2^T, \ldots, \mathbf{u}_P^T]$, and

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1P} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2P} \\ \vdots & \vdots & & \vdots \\ \mathbf{B}_{P1} & \mathbf{B}_{P2} & \cdots & \mathbf{B}_{PP} \end{bmatrix}$$

where $\mathbf{v}_r$ and $\mathbf{u}_s$ are $N \times 1$ vectors, and $\mathbf{B}_{rs}$ is an $N \times N$ matrix, with $1 \le r$ and $s \le P$. It can be shown in the following that

$$\mathbf{v}^H\mathbf{C}(\tilde{\nu})\mathbf{u} = \sum_{r=1}^{P}\sum_{s=1}^{P} e^{-\boldsymbol{\jmath}2\pi(s-r)(N+N_g)\tilde{\nu}} \times\left[-\eta_{r,s}(0) + \sum_{m=0}^{N-1}\left(\eta_{r,s}(m)e^{-\boldsymbol{\jmath}2\pi m\tilde{\nu}} + \varsigma_{r,s}(m)e^{\boldsymbol{\jmath}2\pi m\tilde{\nu}}\right)\right] \quad (21)$$

where

$$\eta_{r,s}(m) = \sum_{k=m}^{N-1}[\mathbf{B}_{rs}]_{k-m,k}v_{r,k-m}^*u_{s,k}$$

$$\varsigma_{r,s}(m) = \sum_{k=m}^{N-1}[\mathbf{B}_{rs}]_{k,k-m}\nu_{r,k}^*u_{s,k-m}.$$

$x_{r,k} = [\mathbf{x}_r]_k$ and $\mathbf{x} = \mathbf{v}$, $\mathbf{u}$. Thus, the terms $\mathbf{v}^H\mathbf{C}(\tilde{\nu})\mathbf{u}$, $\mathbf{1}^H\mathbf{C}(\tilde{\nu})\mathbf{u}$, and $\mathbf{v}^H\mathbf{C}(\tilde{\nu})\mathbf{1}$ in (20) are in the same form of (21), which can be efficiently calculated by using FFT in searching for $\hat{\nu}$.

Given $\hat{\nu}$ in (18), the other estimates are obtained as follows:

$$\hat{\boldsymbol{\rho}}_j = \mathbf{P}_j(\hat{\nu})\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\hat{\nu})\right)\mathbf{r}_j \quad (22)$$

$$\hat{d}_j = \mathbf{f}^H(\hat{\nu})\left(\mathbf{r}_j - \mathbf{R}_j^*\hat{\boldsymbol{\rho}}_j\right) \quad (23)$$

$$\hat{\mathbf{g}}_j = (\Gamma(\hat{\nu})\mathbf{S})^{\dagger}\left(\mathbf{r}_j - \mathbf{R}_j^*\hat{\boldsymbol{\rho}}_j - \hat{d}_j\mathbf{1}\right). \quad (24)$$

B. Low-Complexity Implementation

From (14), (16), and (21), it is observed that the calculation of projection matrix $\mathbf{B} = \mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H$ in $\mathbf{C}(\tilde{\nu}) = \Gamma(\tilde{\nu})\mathbf{B}\Gamma^H(\tilde{\nu})$ plays a key role in determining the complexity of the estimators of $\hat{d}_j$, $\hat{\boldsymbol{\rho}}_j$, and $\hat{\nu}$. Motivated by the design for SISO systems in [23], we design training format $\mathbf{S}$ as in (25), shown at the bottom of the next page, where $\mathbf{A}$ is an $n_tL_{\mathbf{g}} \times n_tL_{\mathbf{g}}$ full-rank matrix, $N = K \cdot n_tL_{\mathbf{g}}$ with $K \ge 1$, and $\{\phi_k\}_{k=0}^{P-1}$ are the parameters for optimizing the estimation performance. Example $\{\phi_k\}$ for $K = P = 2$ will be given in Section VI; nevertheless, the issue of their optimum design will

not be pursued any further in this paper. This way, projection matrix $\mathbf{B}$ becomes

$$\mathbf{B}=\frac{1}{P\cdot K}\begin{bmatrix}\mathbf{F} & e^{\boldsymbol{j}(\phi_0-\phi_1)}\mathbf{F} & \cdots & e^{\boldsymbol{j}(\phi_0-\phi_{P-1})}\mathbf{F}\\ e^{\boldsymbol{j}(\phi_1-\phi_0)}\mathbf{F} & \mathbf{F} & \cdots & e^{\boldsymbol{j}(\phi_1-\phi_{P-1})}\mathbf{F}\\ \vdots & \vdots & & \vdots\\ e^{\boldsymbol{j}(\phi_{P-1}-\phi_0)}\mathbf{F} & e^{\boldsymbol{j}(\phi_{P-1}-\phi_1)}\mathbf{F} & \cdots & \mathbf{F}\end{bmatrix} \tag{26}$$

with

$$\mathbf{F}=\begin{bmatrix}\mathbf{I}_{n_tL_{\mathbf{g}}} & \mathbf{I}_{n_tL_{\mathbf{g}}} & \cdots & \mathbf{I}_{n_tL_{\mathbf{g}}}\\ \mathbf{I}_{n_tL_{\mathbf{g}}} & \mathbf{I}_{n_tL_{\mathbf{g}}} & \cdots & \mathbf{I}_{n_tL_{\mathbf{g}}}\\ \vdots & \vdots & & \vdots\\ \mathbf{I}_{n_tL_{\mathbf{g}}} & \mathbf{I}_{n_tL_{\mathbf{g}}} & \cdots & \mathbf{I}_{n_tL_{\mathbf{g}}}\end{bmatrix}_{N\times N}$$

which contains the K^2 matrices of $\mathbf{I}_{n_tL_{\mathbf{g}}}$. Recall that $\mathbf{A}$ is a full-rank square matrix, and, therefore, its projection matrix $\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$ is equal to $\mathbf{I}_{n_tL_{\mathbf{g}}}$. Since $\mathbf{B}$ is a sparse matrix now, $\mathbf{C}(\tilde{\nu})=\Gamma(\tilde{\nu})\mathbf{B}\Gamma^H(\tilde{\nu})$ can be calculated much easily. In particular, using (26), $\mathbf{v}^H\mathbf{C}(\tilde{\nu})\mathbf{u}$ in (21) becomes

$$\begin{aligned}\mathbf{v}^H\mathbf{C}(\tilde{\nu})\mathbf{u}=&\sum_{r=1}^{P}\sum_{s=1}^{P}e^{-\boldsymbol{j}2\pi(s-r)(N+N_g)\tilde{\nu}}\\ &\times\left[-\eta_{r,s}(0)+\sum_{m=0}^{K-1}\left(\eta_{r,s}(m)e^{-\boldsymbol{j}2\pi mn_tL_{\mathbf{g}}\tilde{\nu}}+\varsigma_{r,s}(m)e^{\boldsymbol{j}2\pi mn_tL_{\mathbf{g}}\tilde{\nu}}\right)\right]\end{aligned} \tag{27}$$

where

$$\begin{aligned}\eta_{r,s}(m)=&\sum_{k=mn_tL_{\mathbf{g}}}^{N-1}[\mathbf{B}_{rs}]_{k-mn_tL_{\mathbf{g}},k}v^*_{r,k-mn_tL_{\mathbf{g}}}u_{s,k}\\ =&\sum_{k=mn_tL_{\mathbf{g}}}^{N-1}\frac{e^{\boldsymbol{j}(\phi_{r-1}-\phi_{s-1})}}{P\cdot K}[\mathbf{F}]_{k-mn_tL_{\mathbf{g}},k}v^*_{r,k-mn_tL_{\mathbf{g}}}u_{s,k}\\ =&\frac{e^{\boldsymbol{j}(\phi_{r-1}-\phi_{s-1})}}{P\cdot K}\sum_{k=mn_tL_{\mathbf{g}}}^{N-1}v^*_{r,k-mn_tL_{\mathbf{g}}}u_{s,k}\\ \varsigma_{r,s}(m)=&\frac{e^{\boldsymbol{j}(\phi_{r-1}-\phi_{s-1})}}{P\cdot K}\sum_{k=mn_tL_{\mathbf{g}}}^{N-1}v^*_{r,k}u_{s,k-mn_tL_{\mathbf{g}}}.\end{aligned}$$

Hence, the complexity of the carrier-frequency offset estimation is reduced by about $n_tL_{\mathbf{g}}$ times. However, the frequency range that can be estimated is also reduced by the same factor with this design. For SISO systems ($n_t=n_r=1$) with $P=1$, (25) degenerates to the one in [23]. Furthermore, for the undesirable case of $\phi_0=\phi_1=\cdots=\phi_{P-1}$, $\|(\mathbf{I}-\mathbf{C}(\tilde{\nu}))\mathbf{1}\|^2|_{\tilde{\nu}=0}=0$ in (13). In other words, with this design of the training sequence, it is not able to estimate the dc offset when the frequency offset is zero because $d_j\mathbf{1}$ is now located in the space that is spanned by the column vectors of $\mathbf{S}$, and its effect is included in the estimate $\hat{\mathbf{g}}_j$. During the real data transmission, however, the receiver needs the estimate $\hat{d}_j$ for the dc offset compensation.

In addition, from the simulation results in Section VI, the frequency-dependent I-Q imbalance has little effect on the estimation of ν and d_j, and can be neglected with almost no loss in performance. With this observation, we propose the simplified estimators for the frequency and dc offset by replacing $\mathbf{R}_j$ as $\mathbf{r}_j$ in (14) and (19) as follows, i.e., the case of $L_{\boldsymbol{\rho}}=1$

$$\hat{\nu}_s=\arg\min_{\tilde{\nu}}\left\{\Lambda_s(\tilde{\nu})=\sum_{j=1}^{n_r}\Lambda_{j,s}(\tilde{\nu})\right\} \tag{28}$$

with

$$\Lambda_{j,s}(\tilde{\nu})=\mathbf{r}_j^H\mathbf{Q}(\tilde{\nu})\mathbf{r}_j-\frac{\left|\mathbf{r}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{r}_j\right|^2}{\mathbf{r}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{r}_j^*} \tag{29}$$

where no matrix inversion is needed, and

$$\hat{d}_{j,s}(\hat{\nu}_s)=\mathbf{f}^H(\hat{\nu}_s)\left(\mathbf{r}_j-\mathbf{r}_j^*\hat{\rho}_j(\hat{\nu}_s)\right) \tag{30}$$

with $\hat{\rho}_j(\hat{\nu}_s)=\mathbf{r}_j^T\mathbf{Q}(\hat{\nu}_s)\mathbf{r}_j/\mathbf{r}_j^T\mathbf{Q}(\hat{\nu}_s)\mathbf{r}_j^*$. It will be shown in Section VI that there is no degradation on the BER performance when using the simplified estimators in (28) and (30).

V. Performance Analysis

In this section, the mean and the variance of estimators $\hat{\nu}$, $\hat{\boldsymbol{\rho}}_j$, $\hat{d}_j$, and $\hat{\mathbf{g}}_j$ are analyzed under the conditions of SNR $\gg 1$ and $N\gg 1$, where SNR $\doteq\sigma_s^2/\sigma_w^2$. We start with $\hat{\nu}$. From [22] and [23], using the fact that $\hat{\nu}$ is close to true carrier frequency offset ν for SNR $\gg 1$ and $N\gg 1$, one has

$$E\{\hat{\nu}\}\approx\nu-\frac{E\{\dot{\Lambda}(\nu)\}}{E\{\ddot{\Lambda}(\nu)\}} \tag{31}$$

$$E\{\hat{\nu}-\nu)^2\}\approx\frac{E\left\{\left[\dot{\Lambda}(\nu)\right]^2\right\}}{\left[E\left\{\ddot{\Lambda}(\nu)\right\}\right]^2} \tag{32}$$

where $\dot{\Lambda}(\nu)=\partial\Lambda(\nu)/\partial\nu$, and $\ddot{\Lambda}(\nu)=\partial^2\Lambda(\nu)/\partial^2\nu$. Unfortunately, it is quite cumbersome to evaluate (31) and (32) with

$$\mathbf{S}=\left[\underbrace{\underbrace{\left[e^{\boldsymbol{j}\phi_0}\mathbf{A}\right]^H\cdots\left[e^{\boldsymbol{j}\phi_0}\mathbf{A}\right]^H}_{K}\underbrace{\left[e^{\boldsymbol{j}\phi_1}\mathbf{A}\right]^H\cdots\left[e^{\boldsymbol{j}\phi_1}\mathbf{A}\right]^H}_{K}\cdots\underbrace{\left[e^{\boldsymbol{j}\phi_{P-1}}\mathbf{A}\right]^H\cdots\left[e^{\boldsymbol{j}\phi_{P-1}}\mathbf{A}\right]^H}_{K}}_{K\cdot P}\right]^H \tag{25}$$

$\Lambda(\nu)$ given in (17), where ν is jointly estimated with $\hat{\boldsymbol{\rho}}_j$, $\hat{d}_j$, and $\hat{\mathbf{g}}_j$. To simplify the analysis, the effects of the I-Q imbalance and the dc offset on the frequency estimation will be neglected here, i.e., $\boldsymbol{\rho}_j = \mathbf{0}$ and $d_j = 0$. By setting $\mathbf{P}_j(\tilde{\nu}) = \mathbf{0}$ and $\mathbf{f}(\tilde{\nu}) = \mathbf{0}$, (17) becomes

$$\begin{aligned}\Lambda(\tilde{\nu}) &= \sum_{j=1}^{n_r} \left\|\left(\mathbf{I} - \mathbf{C}(\tilde{\nu})\right)\mathbf{r}_j\right\|^2 \\ &= \sum_{j=1}^{n_r} \left(\|\mathbf{r}_j\|^2 - \mathbf{r}_j^H \mathbf{C}(\tilde{\nu})\mathbf{r}_j\right) \\ &= \sum_{j=1}^{n_r} \left(\|\mathbf{r}_j\|^2 - \mathbf{r}_j^H \Gamma(\tilde{\nu})\mathbf{B}\Gamma^H(\tilde{\nu})\mathbf{r}_j\right). \end{aligned} \tag{33}$$

It will be shown in Section VI by computer simulations that the simplified analysis is very accurate for the ranges of $\boldsymbol{\rho}_j$ and d_j of practical interest. From Appendix B, it is shown that $E\{(\hat{\nu} - \nu)\} \approx 0$ (an unbiased estimator), and

$$E\left\{(\hat{\nu} - \nu)^2\right\} = \sigma_w^2 \frac{1}{2\sum\limits_{j=1}^{n_r} \mathbf{z}_j^H (\mathbf{I} - \mathbf{B})\mathbf{z}_j} \tag{34}$$

where $\mathbf{z}_j = \boldsymbol{\jmath} 2\pi \boldsymbol{\Phi}\mathbf{S}\mathbf{g}_j$, $\boldsymbol{\Phi} = \mathrm{diag}\{\kappa(1), \kappa(2), \ldots, \kappa(P)\}$, and $\kappa(p) = [(N_g+N)p+1, (N_g+N)p+2, \ldots, (N_g+N)p+N]$.

Next, we analyze estimators $\hat{\boldsymbol{\rho}}_j$, $\hat{d}_j$, and $\hat{\mathbf{g}}_j$ with no influence of the frequency offset; that is, $\hat{\nu} = \nu$ is assumed in the analysis. It will be shown in Section VI by computer simulations that the analysis predicts the MSE performance very well even when $\hat{\nu} \neq \nu$. For estimator $\boldsymbol{\rho}_j$, from (7) and (16), it is easy to show that

$$\hat{\boldsymbol{\rho}}_j(\nu) = \mathbf{P}_j(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)\left(\mathbf{R}_j^* \boldsymbol{\rho}_j + \Gamma(\nu)\mathbf{S}\mathbf{g}_j + d_j\mathbf{1} + \mathbf{w}_j\right). \tag{35}$$

In addition, by using (45) and the identities

$$\begin{aligned}&\mathbf{P}_j(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)\mathbf{R}_j^* \boldsymbol{\rho}_j = \boldsymbol{\rho}_j \\ &\mathbf{P}_j(\nu)\underbrace{\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)\mathbf{1}}_{=\mathbf{0}} d_j = \mathbf{0} \\ &\mathbf{P}_j(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)\Gamma(\nu)\mathbf{S}\mathbf{g}_j \\ &\quad = \left[\mathbf{R}_j^T \mathbf{Q}(\tilde{\nu})\mathbf{R}_j^*\right]^{-1} \mathbf{R}_j^T \underbrace{\mathbf{Q}(\tilde{\nu})\Gamma(\nu)\mathbf{S}}_{=\mathbf{0}}\mathbf{g}_j = \mathbf{0}\end{aligned}$$

one has

$$\hat{\boldsymbol{\rho}}_j = \boldsymbol{\rho}_j + \mathbf{P}_j(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)\mathbf{w}_j. \tag{36}$$

Following the notation in (3), $\mathbf{R}_j$ can be rewritten as $\mathbf{R}_j = \mathbf{R}_{j,d} + \mathbf{R}_{j,r}$, where $\mathbf{R}_{j,d}$ and $\mathbf{R}_{j,r}$ are the deterministic and random parts of $\mathbf{R}_j$, respectively. Taking the expectation on both sides of (36), we have

$$E\{\hat{\boldsymbol{\rho}}_j|\nu\} \approx \boldsymbol{\rho}_j + \mathbf{P}_{j,d}(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)E\{\mathbf{w}_j\} = \boldsymbol{\rho}_j \tag{37}$$

where $\mathbf{P}_{j,d}(\nu)$ is the deterministic part of $\mathbf{P}_j(\nu)$. The approximation is justifiable with SNR $\gg 1$. Furthermore, the MSE is derived as

$$\begin{aligned}&E\left\{|\hat{\boldsymbol{\rho}}_j - \boldsymbol{\rho}_j|^2|\nu\right\} \\ &\quad = \mathrm{tr}\left\{E\left\{(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j)(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j)^H|\nu\right\}\right\} \\ &\quad = \mathrm{tr}\Big\{E\Big\{\mathbf{P}_j(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right) \\ &\qquad\qquad \times \mathbf{w}_j\mathbf{w}_j^H\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)^H \mathbf{P}_j^H(\nu)\Big\}\Big\} \\ &\quad \approx \mathrm{tr}\left\{\mathbf{P}_{j,d}(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)\mathbf{K}_{\mathbf{w}_j}\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)^H \mathbf{P}_{j,d}^H(\nu)\right\}\end{aligned} \tag{38}$$

where $\mathrm{tr}\{\mathbf{X}\}$ denotes the trace of matrix $\mathbf{X}$, and $\mathbf{K}_{\mathbf{w}_j} \doteq E\{\mathbf{w}_j\mathbf{w}_j^H\}$ is the correlation matrix of $\mathbf{w}_j$. The explicit expression of $\mathbf{K}_{\mathbf{w}_j}$ can be found in Appendix C.

For the estimator $\hat{d}_j$, from (14)

$$\begin{aligned}\hat{d}_j(\hat{\boldsymbol{\rho}}_j, \nu) &= \mathbf{f}^H(\nu)\left(\mathbf{r}_j - \mathbf{R}_j^* \hat{\boldsymbol{\rho}}_j\right) \\ &= \underbrace{\mathbf{f}^H(\nu)\mathbf{1}}_{=1} d_j + \mathbf{f}^H(\nu)\mathbf{R}_j^*(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j) \\ &\quad + \underbrace{\mathbf{f}^H(\nu)\Gamma(\nu)\mathbf{S}}_{=\mathbf{0}}\mathbf{g}_j + \mathbf{f}^H(\nu)\mathbf{w}_j \\ &= d_j + \mathbf{f}^H(\nu)\mathbf{R}_j^*(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j) + \mathbf{f}^H(\nu)\mathbf{w}_j.\end{aligned} \tag{39}$$

Similar to (37), we have $E\{\hat{d}_j - d_j|\nu\} \approx 0$, and under SNR $\gg 1$, the MSE is given by

$$\begin{aligned}&E\left\{|\hat{d}_j - d_j|^2|\nu\right\} \\ &\quad \approx E\Big\{\left(\mathbf{f}^H(\nu)\mathbf{R}_{j,d}^*(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j) + \mathbf{f}^H(\nu)\mathbf{w}_j\right) \\ &\qquad \times \left(\mathbf{f}^H(\nu)\mathbf{R}_{j,d}^*(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j) + \mathbf{f}^H(\nu)\mathbf{w}_j\right)^H\Big\} \\ &\quad = \mathbf{f}^H(\nu)\mathbf{R}_{j,d}^* E\left\{(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j)(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j)^H|\nu\right\}\mathbf{R}_{j,d}^T\mathbf{f}(\nu) \\ &\qquad + \mathbf{f}^H(\nu)\mathbf{K}_{\mathbf{w}_j}\mathbf{f}(\nu) \\ &\qquad + 2\mathrm{Re}\left\{\mathbf{f}^H(\nu)\mathbf{R}_{j,d}^* E\left\{(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j)\mathbf{w}_j^H|\nu\right\}\mathbf{f}(\nu)\right\}\end{aligned} \tag{40}$$

where we have used the approximation

$$\begin{aligned}E\left\{(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j)\mathbf{w}_j^H|\nu\right\} &= E\left\{\mathbf{P}_j(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)\mathbf{w}_j\mathbf{w}_j^H\right\} \\ &\approx \mathbf{P}_{j,d}(\nu)\left(\mathbf{I} - \mathbf{1}\mathbf{f}^H(\nu)\right)\mathbf{K}_{\mathbf{w}_j}.\end{aligned}$$

Finally, for the estimator $\hat{\mathbf{g}}_j$

$$\begin{aligned}\hat{\mathbf{g}}_j(\hat{d}_j, \hat{\boldsymbol{\rho}}_j, \nu) &= (\Gamma(\nu)\mathbf{S})^\dagger\left(\mathbf{r}_j - \mathbf{R}_j^*\hat{\boldsymbol{\rho}}_j - \hat{d}_j\mathbf{1}\right) \\ &= (\Gamma(\nu)\mathbf{S})^\dagger\Big(\mathbf{R}_j^*\boldsymbol{\rho}_j + \Gamma(\nu)\mathbf{S}\mathbf{g}_j \\ &\qquad + d_j\mathbf{1} + \mathbf{w}_j - \mathbf{R}_j^*\hat{\boldsymbol{\rho}}_j - \hat{d}_j\mathbf{1}\Big) \\ &= \mathbf{g}_j + \mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{R}_j^*(\boldsymbol{\rho}_j - \hat{\boldsymbol{\rho}}_j) \\ &\quad + \mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{1}(d_j - \hat{d}_j) + \mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{w}_j.\end{aligned} \tag{41}$$

Similarly, $E\{\hat{\mathbf{g}}_j - \mathbf{g}_j|\nu\} \approx \mathbf{0}$, and the MSE is given by

$$\begin{aligned}
&E\left\{\|\hat{\mathbf{g}}_j - \mathbf{g}_j\|^2|\nu\right\} \\
&\quad\approx \mathrm{tr}\left\{\mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{K}_{\mathbf{w}_j}\Gamma(\nu)(\mathbf{S}^\dagger)^H\right\} \\
&\qquad+\mathrm{tr}\Big\{\mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{R}^*_{j,d}E\Big\{(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)^H|\nu\Big\} \\
&\qquad\quad\times\mathbf{R}^T_{j,d}\Gamma(\nu)(\mathbf{S}^\dagger)^H\Big\} \\
&\qquad+\mathrm{tr}\left\{\mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{1}E\left\{|\hat{d}_j-d_j|^2|\nu\right\}\mathbf{1}^H\Gamma(\nu)(\mathbf{S}^\dagger)^H\right\} \\
&\qquad+2\mathrm{Re}\Big(\mathrm{tr}\Big\{\mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{R}^*_{j,d}E\left\{(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)(d_j-\hat{d}_j)^H|\nu\right\} \\
&\qquad\quad\times\mathbf{1}^H\Gamma(\nu)(\mathbf{S}^\dagger)^H\Big\}\Big) \\
&\qquad+2\mathrm{Re}\Big(\mathrm{tr}\Big\{\mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{R}^*_{j,d}E\left\{(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)\mathbf{w}_j^H|\nu\right\} \\
&\qquad\quad\times\Gamma(\nu)(\mathbf{S}^\dagger)^H\Big\}\Big) \\
&\qquad+2\mathrm{Re}\Big(\mathrm{tr}\Big\{\mathbf{S}^\dagger\Gamma^H(\nu)\mathbf{1}E\left\{(d_j-\hat{d}_j)\mathbf{w}_j^H|\nu\right\} \\
&\qquad\quad\times\Gamma(\nu)(\mathbf{S}^\dagger)^H\Big\}\Big). \qquad (42)
\end{aligned}$$

In deriving (42), we have used the following approximations:

$$\begin{aligned}
&E\left\{(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)(d_j-\hat{d}_j)^H|\nu\right\} \\
&\quad\approx E\left\{(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)\left(\mathbf{f}^H(\nu)\mathbf{R}^*_{j,d}(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)+\mathbf{f}^H(\nu)\mathbf{w}_j\right)^H\right\} \\
&\quad= E\left\{(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)^H|\nu\right\}\mathbf{R}^T_{j,d}\mathbf{f}(\nu) \\
&\qquad+E\left\{(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)\mathbf{w}_j^H|\nu\right\}\mathbf{f}(\nu) \\
&E\left\{(d_j-\hat{d}_j)\mathbf{w}_j^H|\nu\right\} \\
&\quad\approx \mathbf{f}^H(\nu)\mathbf{R}^*_{j,d}E\left\{(\boldsymbol{\rho}_j-\hat{\boldsymbol{\rho}}_j)\mathbf{w}_j^H|\nu\right\}+\mathbf{f}^H(\nu)\mathbf{K}_{\mathbf{w}_j}.
\end{aligned}$$

VI. Numerical Results

The performance of the proposed estimators is evaluated for an uncoded MIMO OFDM system. Table I gives the system parameters. The transmission is done on a packet-by-packet basis, with the training portion consisting of two OFDM symbols at the beginning of each packet. A wide-sense stationary uncorrelated scattering discrete channel is considered, with impulse response $h_{j,i}(\tau)=\sum_{l=0}^{L}h_{j,i}(l)\delta(\tau-lT_s)$, where $L+1$ is the length of the channel, and $\{h_{j,i}(l)\}$ are tap gains that are mutually independent complex Gaussian random variables with zero mean and variance σ_l^2. Exponential multipath intensity profile is employed with $\sigma_l^2=\sigma_0^2\cdot\exp(-lT_s/T_{\mathrm{RMS}})$, where T_{RMS} is the root-mean-square delay spread and, to maintain the unit power gain, $\sigma_0^2=1-\exp(-T_s/T_{\mathrm{RMS}})$. The channel remains unchanged during a packet. The parameters are set as $T_{\mathrm{RMS}}=50$ ns, $L=10$, and $L_{\mathbf{g}}=16$. In Figs. 3–7, the training sequence is the one that is given in [25] for the case of $n_t=2$. In Fig. 8, square matrix $\mathbf{A}$ of the low-complexity training sequence is designed as $5230F641_H$ given in [26] for the first transmit antenna and its circular shift by $L_{\mathbf{g}}$ for the second transmit antenna. In addition, $K=P=2$, $\phi_0=0$, and $\phi_1=\pi/2$.

TABLE I
System Parameters

Channel Bandwidth	20MHz
FFT length (N), cyclic prefix length (N_g)	$N=64,\ N_g=16$
OFDM-Symbol Time (T_{OFDM}), Symbol Time (T_s)	$T_{OFDM}=4\mu s,\ T_S=50ns$
Subcarrier Spacing ($1/NT_s$)	0.3125MHz
Number of Transmit and Receive Antenna (n_t, n_r)	$n_t=2,\ n_r=2,3$
Frequency-independent I-Q imbalance (α_j,θ_j)	$(\alpha_1=1.08,\theta_1=5^o)$ $(\alpha_2=1.09,\theta_2=6^o)$ $(\alpha_3=1.1,\theta_3=7^o)$
$\{h_j^I(n), h_j^Q(n)\}$: 2rd order Butterworth filter with cut-off frequency (f_j^I, f_j^Q) MHz	$(f_1^I=8, f_1^Q=8.3)$ $(f_2^I=7.9, f_2^I=8.2)$ $(f_3^I=8.1, f_3^I=8.4)$
Frequency-offset $\triangle f$	Fixed $0.25/NT_S$ for static channel, uniform over $(-0.5/NT_S,\ 0.5/NT_S)$ for fading channel.
DC-offset ($d_{0,j}$), with signal power normalized to 1	$d_{0,1}=0.2\times(1+\boldsymbol{j})/\sqrt{2}$ $d_{0,2}=0.15\times(1+\boldsymbol{j})/\sqrt{2}$ $d_{0,3}=0.1\times(1+\boldsymbol{j})/\sqrt{2}$

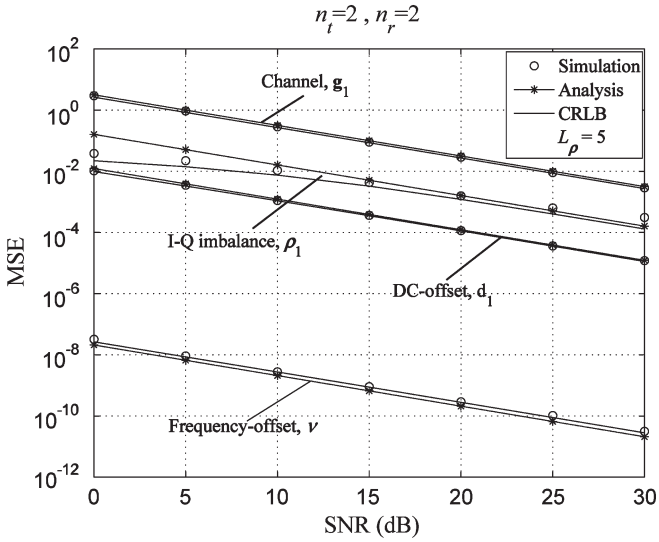

Fig. 3. MSE performance in a static channel.

In Fig. 3, simulations are given to verify the MSE analysis with the system $n_t=n_r=2$ under the static channel $h_{j,i}(l)=\sqrt{(1-\exp(-1))(1-\exp(-l))},\ \forall j,\ i,\ 0\le l\le 10$. Only the results of the first receive antenna are shown; similarity is observed for other receive antennas. As is shown, the analysis predicts the MSE performance very well for all the estimators in the SNRs of interest. (Note that the simulations for estimators $\hat{\boldsymbol{\rho}}_j$, $\hat{d}_j$, and $\hat{\mathbf{g}}_j$ have used the real estimated frequency $\hat{\nu}$, which may not be equal to the true frequency ν.) In addition, the variance of the estimators approaches the respective CRLB. One observation that is worthy of noting is that the analysis of the frequency estimation is done under the perfect condition of no

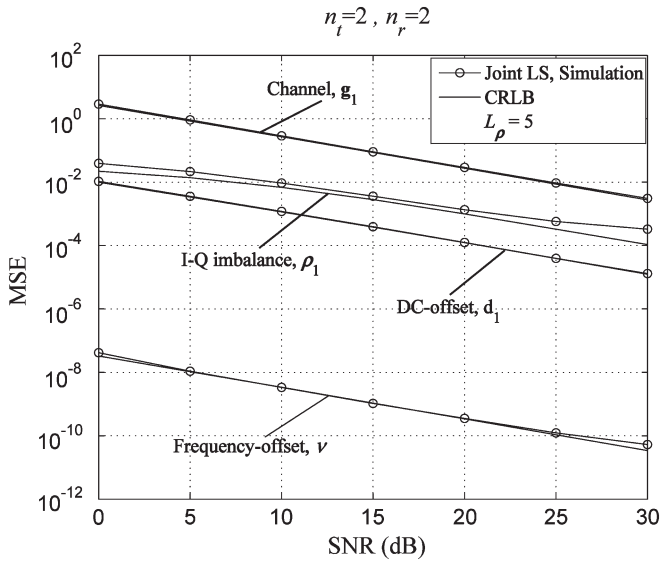

Fig. 4. MSE performance of the joint estimators in Rayleigh fading channels.

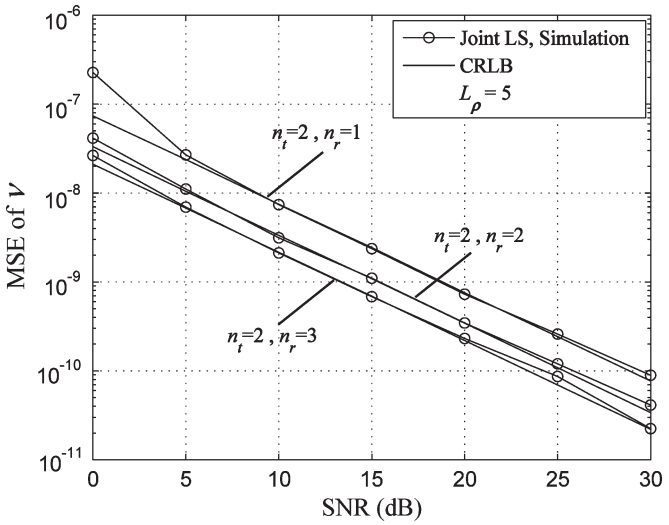

Fig. 5. MSE performance of the frequency estimator in Rayleigh fading channels.

dc offset and I-Q imbalance. This explains why the analytical MSE of the frequency estimator is slightly smaller than its corresponding CRLB in Fig. 3. In Fig. 4, the MSE performance is evaluated for the Rayleigh fading channel. In this case, the frequency offset is set to uniformly vary between -0.5 and 0.5 of subcarrier spacing. Again, the estimators performed very closely to the CRLBs. In Fig. 4, the variance of $\hat{\rho}$ tends to have a floor at the high SNR region. This may be attributed to having a modeling error by using $L_\rho = 5$ in this case. The error, however, does not cause too much loss in the BER performance, as shown in Fig. 7.

The performance of the frequency-offset estimation is shown in Fig. 5 with a different number of received antennas. It is clearly shown that more than one receive antenna branch can be used in the estimation to improve the performance by exploiting the power and the diversity gain that are offered by multiple receive antennas.

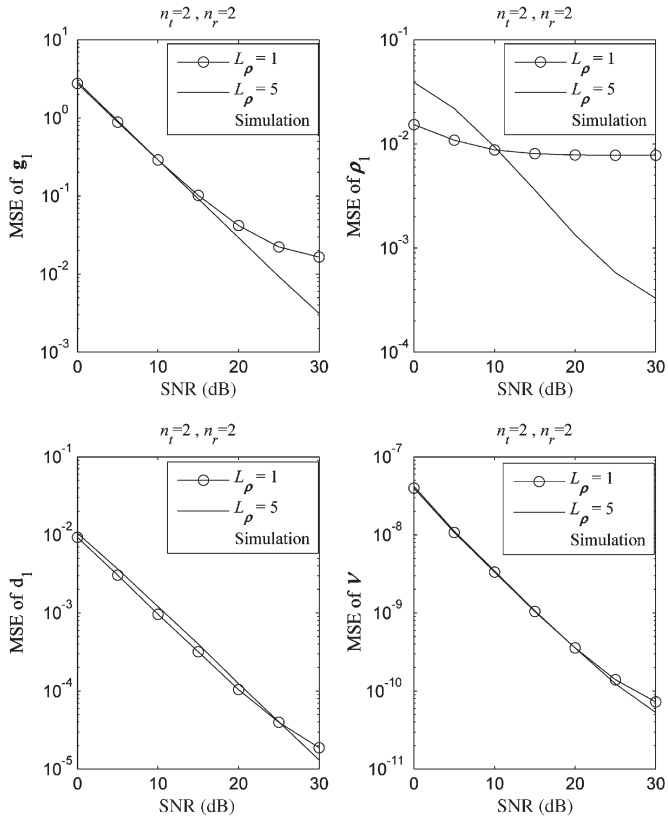

Fig. 6. Effects of L_ρ on the MSE in Rayleigh fading channels.

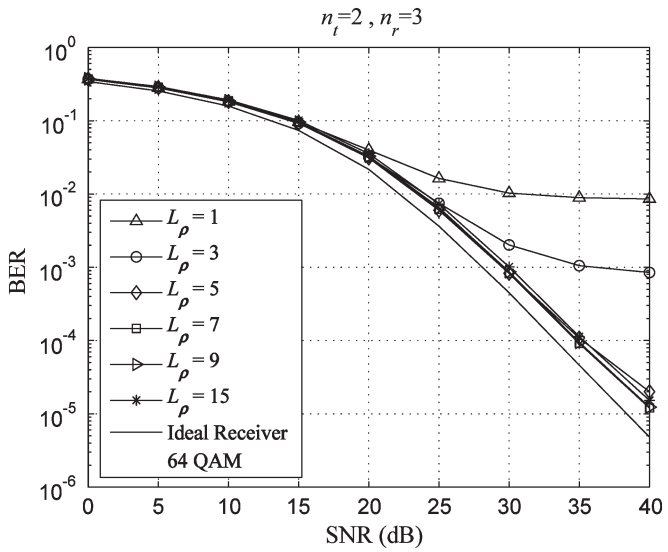

Fig. 7. Effects of L_ρ on the BER performance in Rayleigh fading channels.

Fig. 6 investigates the effect of L_ρ on the MSE performance of the joint estimators by computer simulations. As can be seen, L_ρ has a significant impact on the channel and I-Q imbalance estimation, particularly at the high SNR region, and that causes an error floor in the BER performance, as can be seen in Fig. 7. Nevertheless, it affects the estimation of the dc offset and the frequency offset in a very insignificant way; this motivates us to use the simplified estimators proposed in (28) and (30).

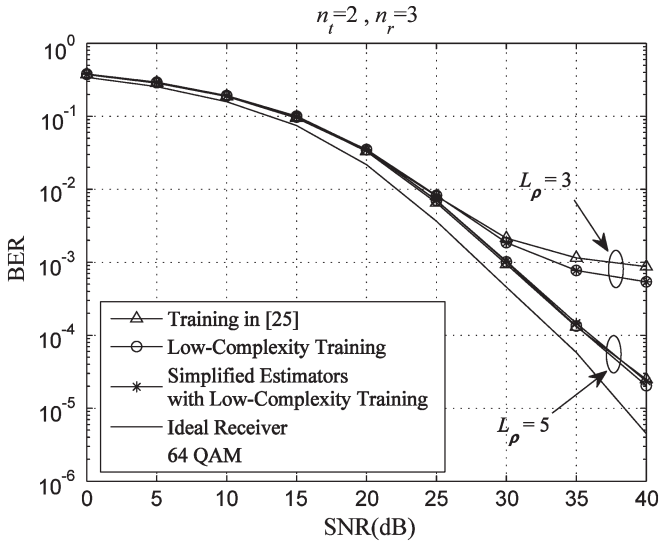

Fig. 8. BER performance with low-complexity and/or simplified estimators in Rayleigh fading channels.

Fig. 7 shows the impact of $L_{\boldsymbol{\rho}}$ on the BER performance in the Rayleigh fading channel. The modulation scheme is the 64 quadratic-amplitude modulation. The receiver is the one given in Fig. 2; after compensating the I-Q imbalance and the dc and frequency offset, MMSE MIMO detection is performed based on the channel estimation coming out of the joint LS estimators. Clearly, the modeling error due to the use of a small $L_{\boldsymbol{\rho}}$ incurs an error floor in the BER performance, as predicted in Fig. 6, where a small $L_{\boldsymbol{\rho}}$ results in a large MSE for estimating the channel and the I-Q imbalance. On the other hand, an $L_{\boldsymbol{\rho}}$ that is too large, e.g., $L_{\boldsymbol{\rho}} = 15$, slightly degrades the BER performance, as can be seen in the figure, due to the extra noise that is induced by using a large filter length. In the figure, an ideal receiver is the one with the perfect RF compensation. Finally, in Fig. 8, we show the BER performance by using low-complexity training sequence and/or simplified frequency and dc offset estimators. The low-complexity training works very well, and almost no performance loss is observed with the low-complexity implementations. In fact, the low-complexity training performs a little better than the training that is used in [25] for $L_{\boldsymbol{\rho}} = 3$.

VII. Conclusion

The theory of the joint LS estimation of the frequency, the I-Q imbalance, the dc offset, and the channel has been developed for MIMO receivers with direct-conversion RF architecture. Frequency-independent and frequency-dependent I-Q imbalances have been included. Previously, RF parameters were estimated separately, and that leads to inferior performance. The estimators have been shown through analysis to be unbiased and approach the CRLB for the SNRs of interest. Special attention has been paid to the implementation complexity issue; several measures have been proposed, including a special training-sequence design and low-complexity estimators for the frequency and dc offset. Simulation results have shown that the performance degradation is negligible when using the low-complexity designs.

Appendix A
Derivation of (19)

From (16) and a simple manipulation

$$\hat{\boldsymbol{\rho}}_j(\tilde{\nu}) = \left[\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\left(\mathbf{I}-\mathbf{1}\mathbf{f}^H(\tilde{\nu})\right)\mathbf{R}_j^*\right]^{\dagger}\left(\mathbf{I}-\mathbf{1}\mathbf{f}^H(\tilde{\nu})\right)\mathbf{r}_j$$
$$= \left[\mathbf{R}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{R}_j^*\right]^{-1}\mathbf{R}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{r}_j \tag{43}$$

where

$$\mathbf{Q}(\tilde{\nu}) = \mathbf{Q}^H(\tilde{\nu}) = \left(\mathbf{I}-\mathbf{1}\mathbf{f}^H(\tilde{\nu})\right)^H\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\left(\mathbf{I}-\mathbf{1}\mathbf{f}^H(\tilde{\nu})\right).$$

Using (43) in (15), we have

$$\Lambda_j(\tilde{\nu}) = \left\|\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\left(\mathbf{I}-\mathbf{1}\mathbf{f}^H(\tilde{\nu})\right)\left(\mathbf{r}_j-\mathbf{R}_j^*\hat{\boldsymbol{\rho}}_j(\tilde{\nu})\right)\right\|^2$$
$$= \left(\mathbf{r}_j^H-\hat{\boldsymbol{\rho}}_j^H(\tilde{\nu})\mathbf{R}_j^T\right)\mathbf{Q}(\tilde{\nu})\left(\mathbf{r}_j-\mathbf{R}_j^*\hat{\boldsymbol{\rho}}_j(\tilde{\nu})\right)$$
$$= \mathbf{r}_j^H\mathbf{Q}(\tilde{\nu})\mathbf{r}_j-\left(\mathbf{R}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{r}_j\right)^H$$
$$\times\left[\mathbf{R}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{R}_j^*\right]^{-1}\mathbf{R}_j^T\mathbf{Q}(\tilde{\nu})\mathbf{r}_j. \tag{44}$$

Furthermore, from (13), for $\mathbf{f}(\tilde{\nu}) \neq \mathbf{0}$

$$\mathbf{Q}(\tilde{\nu}) = \left(\mathbf{I}-\mathbf{f}(\tilde{\nu})\mathbf{1}^H\right)\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\left(\mathbf{I}-\mathbf{1}\mathbf{f}^H(\tilde{\nu})\right)$$
$$= \left(\mathbf{I}-\frac{\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\mathbf{1}\mathbf{1}^H}{\left\|\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\mathbf{1}\right\|^2}\right)\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)$$
$$\times\left(\mathbf{I}-\frac{\mathbf{1}\mathbf{1}^H\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)}{\left\|\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\mathbf{1}\right\|^2}\right)$$
$$= \left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)-\frac{\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\mathbf{1}\mathbf{1}^H\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)}{\left\|\left(\mathbf{I}-\mathbf{C}(\tilde{\nu})\right)\mathbf{1}\right\|^2}. \tag{45}$$

Appendix B
Derivation of $E\{(\hat{\nu}-\nu)\}$ and $E\{(\hat{\nu}-\nu)^2\}$

From (33), it can be shown that

$$\dot{\Lambda}(\nu) = \boldsymbol{\jmath}2\pi\cdot\sum_{j=1}^{n_r}\mathbf{r}_j^H\Gamma(\nu)\mathbf{D}\Gamma^H(\nu)\mathbf{r}_j \tag{46}$$

$$\ddot{\Lambda}(\nu) = 4\pi^2\cdot\sum_{j=1}^{n_r}\mathbf{r}_j^H\Gamma(\nu)\mathbf{E}\Gamma^H(\nu)\mathbf{r}_j \tag{47}$$

with $\mathbf{D} = \boldsymbol{\Phi}\mathbf{B}-\mathbf{B}\boldsymbol{\Phi}$ and $\mathbf{E} = 2\boldsymbol{\Phi}\mathbf{B}\boldsymbol{\Phi}-\mathbf{B}\boldsymbol{\Phi}^2-\boldsymbol{\Phi}^2\mathbf{B}$, where $\boldsymbol{\Phi} = \mathrm{diag}\{\kappa(1),\kappa(2),\ldots,\kappa(P)\}$, and $\kappa(p) = [(N_g+N)p+1,(N_g+N)p+2,\ldots,(N_g+N)p+N]$. Recall that $\mathbf{B} = \mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H$. Using $\mathbf{r}_j = \Gamma(\nu)\mathbf{S}\mathbf{g}_j+\mathbf{w}_j$ in (46) and (47), one obtains

$$\dot{\Lambda}(\nu) = \boldsymbol{\jmath}2\pi\cdot\sum_{j=1}^{n_r}\mathbf{g}_j^H\mathbf{S}^H\mathbf{D}\tilde{\mathbf{w}}_j+\tilde{\mathbf{w}}_j^H\mathbf{D}\mathbf{S}\mathbf{g}_j+\tilde{\mathbf{w}}_j^H\mathbf{D}\tilde{\mathbf{w}}_j \tag{48}$$

$$\ddot{\Lambda}(\nu) = 4\pi^2\cdot\sum_{j=1}^{n_r}\mathbf{g}_j^H\mathbf{S}^H\mathbf{E}\mathbf{S}\mathbf{g}_j+\mathbf{g}_j^H\mathbf{S}^H\mathbf{E}\tilde{\mathbf{w}}_j+\tilde{\mathbf{w}}_j^H\mathbf{E}\mathbf{S}\mathbf{g}_j$$
$$+\tilde{\mathbf{w}}_j^H\mathbf{E}\tilde{\mathbf{w}}_j \tag{49}$$

where $\tilde{\mathbf{w}}_j = \Gamma^H(\nu)\mathbf{w}_j$ with $E\{\tilde{\mathbf{w}}_j\tilde{\mathbf{w}}_j^H\} \approx \sigma_w^2\mathbf{I}$. In addition

$$\begin{aligned} E\left\{\dot{\Lambda}(\nu)\right\} &= \boldsymbol{\jmath}2\pi \cdot E\left\{\sum_{j=1}^{n_r} \tilde{\mathbf{w}}_j^H \mathbf{D}\tilde{\mathbf{w}}_j\right\} \\ &= \boldsymbol{\jmath}2\pi \cdot \sum_{j=1}^{n_r} \mathrm{tr}\left\{\mathbf{D}\cdot E\left\{\tilde{\mathbf{w}}_j\tilde{\mathbf{w}}_j^H\right\}\right\} \\ &\approx \boldsymbol{\jmath}2\pi\sigma_w^2 \cdot \sum_{j=1}^{n_r} \mathrm{tr}\{\mathbf{D}\} \\ &= 0 \end{aligned} \tag{50}$$

$$\begin{aligned} E\left\{\ddot{\Lambda}(\nu)\right\} &= 4\pi^2 \cdot \sum_{j=1}^{n_r} \mathbf{g}_j^H\mathbf{S}^H\mathbf{E}\mathbf{S}\mathbf{g}_j + E\left\{\tilde{\mathbf{w}}_j^H\mathbf{E}\tilde{\mathbf{w}}_j\right\} \\ &\approx 4\pi^2 \cdot \sum_{j=1}^{n_r} \mathbf{g}_j^H\mathbf{S}^H\mathbf{E}\mathbf{S}\mathbf{g}_j + \sigma_w^2\mathrm{tr}\{\mathbf{E}\} \\ &= 4\pi^2 \cdot \sum_{j=1}^{n_r} \mathbf{g}_j^H\mathbf{S}^H\mathbf{E}\mathbf{S}\mathbf{g}_j \\ &= 2\cdot \sum_{j=1}^{n_r} \mathbf{z}_j^H(\mathbf{B}-\mathbf{I})\mathbf{z}_j \end{aligned} \tag{51}$$

where $\mathbf{z}_j = \boldsymbol{\jmath}2\pi\boldsymbol{\Phi}\mathbf{S}\mathbf{g}_j$. On the other hand

$$\begin{aligned} &E\left\{\left(\dot{\Lambda}(\nu)\right)^2\right\} \\ &= -4\pi^2 E\left\{\left(\sum_{j=1}^{n_r} \mathbf{g}_j^H\mathbf{S}^H\mathbf{D}\tilde{\mathbf{w}}_j + \tilde{\mathbf{w}}_j^H\mathbf{D}\mathbf{S}\mathbf{g}_j + \tilde{\mathbf{w}}_j^H\mathbf{D}\tilde{\mathbf{w}}_j\right)^2\right\} \\ &\approx -4\pi^2 \cdot \sum_{j=1}^{n_r} 2\mathbf{g}_j^H\mathbf{S}^H\mathbf{D}E\left\{\tilde{\mathbf{w}}_j\tilde{\mathbf{w}}_j^H\right\}\mathbf{D}\mathbf{S}\mathbf{g}_j \\ &= 2\sigma_w^2 \cdot \sum_{j=1}^{n_r} \mathbf{z}_j^H(\mathbf{I}-\mathbf{B})\mathbf{z}_j^H. \end{aligned} \tag{52}$$

The approximation in (52) is justifiable for SNR $\gg 1$, and the last equality is obtained by using $\mathbf{S}^H(\boldsymbol{\Phi}\mathbf{B}-\mathbf{B}\boldsymbol{\Phi}) = \mathbf{S}^H(\boldsymbol{\Phi}\mathbf{B}-\boldsymbol{\Phi})$, $(\boldsymbol{\Phi}\mathbf{B}-\mathbf{B}\boldsymbol{\Phi})\mathbf{S} = (\boldsymbol{\Phi}-\mathbf{B}\boldsymbol{\Phi})\mathbf{S}$, and the fact that $(\mathbf{I}-\mathbf{B})$ is a projection matrix. Therefore, from (31), (32), and (50)–(52), we have $E\{(\hat{\nu}-\nu)\} \approx 0$ and

$$E\left\{(\hat{\nu}-\nu)^2\right\} \approx \sigma_w^2 \frac{1}{2\sum_{j=1}^{n_r} \mathbf{z}_j^H(\mathbf{I}-\mathbf{B})\mathbf{z}_j}. \tag{53}$$

APPENDIX C
DERIVATION OF THE CRLB

In Section V, it is shown that $\hat{\nu}$, $\hat{\boldsymbol{\rho}}_j$, $\hat{d}_j$, and $\hat{\mathbf{g}}_j$ are unbiased estimators provided that SNR $\gg 1$ and $N \gg 1$. Here, we derive the CRLB for those estimators under the same conditions. The derivation follows that given in [14] and [24]. Define $\boldsymbol{\omega}^T = (\nu, \boldsymbol{\lambda}_1^T, \ldots, \boldsymbol{\lambda}_{n_r^T})$, with $\boldsymbol{\lambda}_j^T = (\boldsymbol{\rho}_j^T, d_j, \mathbf{g}_j^T, \boldsymbol{\rho}_j^H, d_j^*, \mathbf{g}_j^H)$ with length $L_{\boldsymbol{\lambda}}$. From (7), since $\mathbf{w}_j$ for $j = 1, \ldots, n_r$ are independent of each other, the probability density function of observation $\mathbf{r}_j$, given $\boldsymbol{\omega}$, is shown as follows:

$$\begin{aligned} f(\boldsymbol{\omega}) &\doteq f(\mathbf{r}_1, \ldots, \mathbf{r}_{n_r}|\boldsymbol{\omega}) \\ &= \prod_{j=1}^{n_r} p(\mathbf{r}_j - \mathbf{R}_j^*\boldsymbol{\rho}_j|\mathbf{g}_j, d_j, \boldsymbol{\rho}_j, \nu) \\ &= \prod_{j=1}^{n_r} \frac{1}{\pi^N \det(\mathbf{K}_{\mathbf{w}_j})} \\ &\quad \times \exp\left\{-\left(\mathbf{r}_j - \mathbf{R}_j^*\boldsymbol{\rho}_j - d_j\mathbf{1} - \Gamma(\nu)\mathbf{S}\mathbf{g}_j\right)^H \right. \\ &\quad \left. \times\, \mathbf{K}_{\mathbf{w}_j}^{-1}\left(\mathbf{r}_j - \mathbf{R}_j^*\boldsymbol{\rho}_j - d_j\mathbf{1} - \Gamma(\nu)\mathbf{S}\mathbf{g}_j\right)\right\} \end{aligned} \tag{54}$$

where $\mathbf{K}_{\mathbf{w}_j} \doteq E\{\mathbf{w}_j\mathbf{w}_j^H\}$ is the correlation matrix of $\mathbf{w}_j$. Recall that $\mathbf{w}_j = [\,\mathbf{w}_j^T(0) \quad \mathbf{w}_j^T(1) \quad \cdots \quad \mathbf{w}_j^T(P-1)\,]^T$ with $w_j(n) = g_j(n) \otimes w_{0,j}(n)$. We have $\mathbf{w}_j(k) = \boldsymbol{\Phi}_{g_j}\mathbf{w}_{0,j}(k)$ and $\mathbf{w}_{0,j}(k) = [w_{0,j,k}(-L_{\mathbf{g}_j}+1), \ldots, w_{0,j,k}(0), w_{0,j,k}(1), \ldots, w_{0,j,k}(N-1)]^T$ with $w_{0,j,k}(n) \doteq w_{0,j}(k(N_g+N)+n)$ and $\boldsymbol{\Phi}_{g_j}$, shown at the bottom of the page, where $L_{\mathbf{g}_j}$ is the length of $g_j(n)$. Therefore, $\mathbf{w}_j = \tilde{\boldsymbol{\Phi}}_j\mathbf{w}_{0,j}$ with $\mathbf{w}_{0,j} = [\mathbf{w}_{0,j}^T(0) \quad \mathbf{w}_{0,j}^T(1) \quad \cdots \quad \mathbf{w}_{0,j}^T(P-1)]^T$ and $\tilde{\boldsymbol{\Phi}}_j = \mathrm{diag}\{\underbrace{\boldsymbol{\Phi}_{g_j}, \ldots, \boldsymbol{\Phi}_{g_j}}_{P}\}$. Finally, we can obtain $\mathbf{K}_{\mathbf{w}_j} = \sigma_w^2\tilde{\boldsymbol{\Phi}}_j\tilde{\boldsymbol{\Phi}}_j^H$.

Let $\mathbf{M} = E\{(\partial \ln f(\boldsymbol{\omega})/\partial\boldsymbol{\omega}^T)^H(\partial \ln f(\boldsymbol{\omega})/\partial\boldsymbol{\omega}^T)\}$ be the Fisher information matrix. From [24], the CRLB for each respective parameter is given by

$$\mathrm{var}(\hat{\nu}) \geq [\mathbf{M}^{-1}]_{11}$$

$$E\left\{\|\hat{\boldsymbol{\rho}}_j - \boldsymbol{\rho}_j\|^2\right\} \geq \sum_{k=1+(j-1)L_{\boldsymbol{\lambda}}+1}^{1+(j-1)L_{\boldsymbol{\lambda}}+L_{\boldsymbol{\rho}}} [\mathbf{M}^{-1}]_{kk}$$

$$\mathrm{var}(\hat{d}_j) \geq [\mathbf{M}^{-1}]_{kk|k=1+(j-1)L_{\boldsymbol{\lambda}}+L_{\boldsymbol{\rho}}+1}$$

$$E\left\{\|\hat{\mathbf{g}}_j - \mathbf{g}_j\|^2\right\} \geq \sum_{k=1+(j-1)L_{\boldsymbol{\lambda}}+L_{\boldsymbol{\rho}}+2}^{1+(j-1)L_{\boldsymbol{\lambda}}+L_{\boldsymbol{\rho}}+1+L_{\mathbf{g}}} [\mathbf{M}^{-1}]_{kk}.$$

$$\boldsymbol{\Phi}_{g_j} = \begin{bmatrix} g_j(L_{\mathbf{g}_j}-1) & \cdots & g_j(0) & 0 & \cdots & 0 \\ 0 & g_j(L_{\mathbf{g}_j}-1) & \cdots & g_j(0) & 0 & \cdots & 0 \\ & \ddots & & \ddots & & \ddots & \\ 0 & \cdots & 0 & g_j(L_{\mathbf{g}_j}-1) & \cdots & & g_j(0) \end{bmatrix}_{N\times(N+L_{\mathbf{g}_j}-1)}$$

By definition

$$\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^T} = \left(\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \nu}, \frac{\partial \ln f(\boldsymbol{\omega})}{\partial \boldsymbol{\lambda}_1^T}, \ldots, \frac{\partial \ln f(\boldsymbol{\omega})}{\partial \boldsymbol{\lambda}_{n_r}^T} \right) \quad (55)$$

$$\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \boldsymbol{\lambda}_j^T} = \left(\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \boldsymbol{\rho}_j^T}, \frac{\partial \ln f(\boldsymbol{\omega})}{\partial d_j}, \frac{\partial \ln f(\boldsymbol{\omega})}{\partial \mathbf{g}_j^T}, \frac{\partial \ln f(\boldsymbol{\omega})}{\partial \boldsymbol{\rho}_j^H}, \frac{\partial \ln f(\boldsymbol{\omega})}{\partial d_j^*}, \frac{\partial \ln f(\boldsymbol{\omega})}{\partial \mathbf{g}_j^H} \right). \quad (56)$$

After some derivation, the partial derivative of $\ln f(\boldsymbol{\omega})$ with respect to each parameter is shown to be

$$\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \nu} = \sum_{j=1}^{n_r} \left(\mathbf{w}_j^H \mathbf{K}_{\mathbf{w}_j}^{-1} \Gamma(\nu) \mathbf{z}_j + \mathbf{z}_j^H \Gamma^H(\nu) \mathbf{K}_{\mathbf{w}_j}^{-1} \mathbf{w}_j \right) \quad (57)$$

$$\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \boldsymbol{\rho}_j^T} = \left(\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \boldsymbol{\rho}_j^H} \right)^* = \mathbf{w}_j^H \mathbf{K}_{\mathbf{w}_j}^{-1} \mathbf{R}_j^* \quad (58)$$

$$\frac{\partial \ln f(\boldsymbol{\omega})}{\partial d_j} = \left(\frac{\partial \ln f(\boldsymbol{\omega})}{\partial d_j^*} \right)^* = \mathbf{w}_j^H \mathbf{K}_{\mathbf{w}_j}^{-1} \mathbf{1} \quad (59)$$

$$\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \mathbf{g}_j^T} = \left(\frac{\partial \ln f(\boldsymbol{\omega})}{\partial \mathbf{g}_j^H} \right)^* = \mathbf{w}_j^H \mathbf{K}_{\mathbf{w}_j}^{-1} \Gamma(\nu) \mathbf{S}. \quad (60)$$

From (55)–(60), matrix $\mathbf{M}$ can be evaluated.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wirel. Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[4] A. A. Abidi, "Direct-conversion radio transceivers for digital communications," *IEEE J. Solid-State Circuits*, vol. 30, no. 12, pp. 1399–1410, Dec. 1995.

[5] B. Razavi, *RF Microelectronics*. Englewood Cliffs, NJ: Prentice–Hall, 1998.

[6] J. K. Cavers and M. W. Liao, "Adaptive compensation for imbalance and offset losses in direct conversion transceivers," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 581–588, Nov. 1993.

[7] M. Valkama, M. Renfors, and V. Koivunen, "Advanced methods for I/Q imbalance compensation in communication receivers," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2335–2344, Oct. 2001.

[8] M. Valkama, M. Renfors, and V. Koivunen, "Compensation of frequency-selective I/Q imbalances in wideband receivers: Models and algorithms," in *Proc. IEEE 3rd Workshop SPAWC*, Taoyuan, Taiwan, Mar. 2001, pp. 42–45.

[9] K. P. Pun, J. E. Franca, C. Azeredo-Leme, C. F. Chan, and C. S. Choy, "Correction of frequency-dependent I/Q mismatches in quadrature receivers," *Electron. Lett.*, vol. 37, no. 23, pp. 1415–1417, Nov. 2001.

[10] A. Schuchert, R. Hasholzner, and P. Antoine, "A novel IQ imbalance compensation scheme for the reception of OFDM signals," *IEEE Trans. Consum. Electron.*, vol. 47, no. 3, pp. 313–318, Aug. 2001.

[11] A. Tarighat, R. Bagheri, and A. H. Sayed, "Compensation schemes and performance analysis of IQ imbalance in OFDM receivers," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3257–3268, Aug. 2005.

[12] A. Tarighat and A. H. Sayed, "Joint compensation of transmitter and receiver impairments in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 240–247, Jan. 2007.

[13] G. Xing, M. Shen, and H. Liu, "Frequency offset and I/Q imbalance compensation for direct conversion receivers," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 673–680, Mar. 2005.

[14] G. T. Gil, I. H. Sohn, J. K. Park, and Y. H. Lee, "Joint ML estimation of carrier frequency, channel, I/Q mismatch, and DC offset in communication receivers," *IEEE Trans. Veh. Technol.*, vol. 54, no. 1, pp. 338–349, Jan. 2005.

[15] D. Tandur and M. Moonen, "Joint adaptive compensation of transmitter and receiver IQ imbalance under carrier frequency offset in OFDM-based systems," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5246–5252, Nov. 2007.

[16] A. Tarighat and A. H. Sayed, "MIMO OFDM receivers for systems with IQ imbalances," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3583–3596, Sep. 2005.

[17] R. M. Rao and B. Daneshrad, "I/Q mismatch cancellation for MIMO OFDM systems," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Barcelona, Spain, Sep. 2004, pp. 2710–2714.

[18] H. Kamata, K. Sakaguchi, and K. Araki, "An effective IQ imbalance compensation scheme for MIMO-OFDM communication system," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Berlin, Germany, Sep. 2005, pp. 181–185.

[19] T. C. W. Schenk, P. F. M. Smulders, and E. R. Fledderus, "Estimation and compensation of frequency selective TX/RX IQ imbalance in MIMO OFDM systems," in *Proc. ICC*, Jun. 2006, pp. 251–256.

[20] T. Kailath, H. Vikalo, and B. Hassibi, "MIMO receiver algorithms," in *Space-Time Wireless Systems*, H. Bolcskei, D. Gesbert, C. Papadias, and A. van der Veen, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[21] J. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications and Control*. Englewood Cliffs, NJ: Prentice–Hall, 1995.

[22] M. H. Meyers and L. E. Franks, "Joint carrier phase and symbol timing recovery for PAM systems," *IEEE Trans. Commun.*, vol. COM-28, no. 8, pp. 1121–1129, Aug. 1980.

[23] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1580–1589, Sep. 2000.

[24] A. van den Bos, "A Cramer–Rao lower bound for complex parameters," *IEEE Trans. Signal Process.*, vol. 42, no. 10, p. 2859, Oct. 1994.

[25] TGn Sync Group, *IEEE P802.11 Wireless LAN - TGn Sync Proposal Technical Specification*, Jan. 2005. Proposal of IEEE 802.11n, IEEE Document 802.11-04/889r4.

[26] C. Tellambura, M. G. Parker, Y. J. Guo, S. J. Shepherd, and S. K. Barton, "Optimal sequences for channel estimation using discrete Fourier transform techniques," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 230–238, Feb. 1999.

Chen-Jiu Hsu was born in Hsinchu, Taiwan, in 1982. He received the B.S. degree in electronics engineering and the M.S. degree in communication engineering in 2004 and 2006, respectively, from the National Chiao Tung University, Hsinchu, where he is currently working toward the Ph.D. degree in communication engineering.

His current research interests include multiple-input–multiple-output transceiver design, particularly in radio-frequency front-end impairment-compensation algorithms.

Racy Cheng received the B.S. degree in mathematics from the National Central University, Taoyuan, Taiwan, in 1988, the M.S. degree in electrophysics from Polytechnic University, Brooklyn, NY, in 1992, and the Ph.D. degree in electrical engineering from the State University of New York at Stony Brook in 1997.

From 1997 to 2000, he was with the Computer and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, working on Digital Enhanced Cordless Telecommunications and Global System for Mobile Communications projects. From 2000 to 2005, he was with InProComm, Hsinchu, where he led a team for the 802.11g Baseband/Medium Access Control Application Specific Integrated Circuit design. He is currently with Minghsin University of Science and Technology, Hsinchu. In addition to teaching, he is also a Technical Consultant for research institutes and industrial companies. The projects he is working on range from 4G cellular systems to proprietary wireless communication devices. His research interest includes digital signal processing with applications to wireless communication and multimedia.

Wern-Ho Sheen (M'91) received the B.S. degree from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 1982, the M.S. degree from the National Chiao Tung University, Hsinchu, Taiwan, in 1984, and the Ph.D. degree from Georgia Institute of Technology, Atlanta, in 1991.

From 1993 to 2001, he was with the National Chung Cheng University, Chiayi, Taiwan, where he held the positions of Professor with the Department of Electrical Engineering and Managing Director with the Center for Telecommunication Research. Since 2001, he has been a Professor with the Department of Communication Engineering, National Chiao Tung University. He has been extensively consulting for the industry and research institutes in Taiwan. His research interests include the general areas of communication theory, cellular mobile and personal radio systems, adaptive signal processing for wireless communications, spread-spectrum communications, and very-large-scale integration design for wireless communications systems.