Channel Estimation and Equalization for Terahertz Receiver With RF Impairments

Ziyuan Sha and Zhaocheng Wang, *Fellow, IEEE*

***Abstract*— The radio frequency (RF) impairments of analog devices have been regarded as an important factor degrading the performance of Terahertz (THz) communications, where in-phase/quadrature (IQ) imbalance and phase noise (PN) are the two typical RF impairments at THz transceiver. In this paper, we investigate the channel estimation (CE) and equalization for THz receiver in the presence of wideband IQ imbalance and PN, where single-carrier frequency-domain equalization is used. Since PN is inserted between channel impulse response (CIR) and wideband IQ imbalance at the receiver, the CIR and IQ imbalance cannot be treated together as an effective channel. Therefore, a novel two-stage CE method is derived to separately estimate the CIR and wideband IQ imbalance parameters, and its corresponding channel equalization method is designed to compensate both wideband IQ imbalance and PN, and equalizes the CIR. Theoretical and simulation results validate the efficiency of our proposed CE, and the proposed equalization method is shown to outperform the state-of-the-art methods.**

***Index Terms*— Terahertz communication, single-carrier frequency-domain equalization, phase noise, IQ imbalance, channel estimation, equalization.**

I. Introduction

TERAHERTZ (THz) band provides enormous spectrum resources for ultra-high speed wireless communications, leading to a theoretical data rate in the order of 1Tbps [1]–[3]. So far, many encouraging experimental results on THz communications have been reported [4]–[12], where the achievable data rate have reached 100Gbps [7], [12]. Nevertheless, it is difficult to implement wireless transceiver in THz band. The first challenge is the severe signal attenuation caused by high spreading loss and molecular absorption. Given the low output power of THz transmitters, it demands extremely high antenna gain to balance the link budget. Therefore, most experiments equip highly directional antennas (e.g. horn antenna, parabolic antenna) at both Tx and Rx sides. Alternatively, ultra-massive antenna array [13] could be utilized to steer the beam direction with sufficiently high power gain.

The second challenge is the design of radio frequency (RF) devices for THz band [1]–[3]. Since THz band is a frequency gap between millimeter-wave and infrared band, the generation of THz signal is difficult for both electronic and photonic devices, leading to severe RF impairments. The reported RF impairments [9]–[12] mainly include the non-linearity of power amplifier (PA), the imbalance between in-phase/quadrature (IQ) branches and the phase noise (PN) of local oscillator (LO). The LO signal in THz transceivers is usually generated by feeding a low-frequency LO signal to a frequency multiplier [42], where the frequency multiplier with multiplication factor M increases the power of PN by a factor of M^2 [42], [43]. Hence, the PN in THz transceivers would be stronger than the millimeter-wave or lower frequency devices. Meanwhile, the large bandwidth in THz communications causes wideband IQ imbalance in THz circuits [11]. These effects impose distortions on the received signal and degrade the system performance, which could be the predominant factor limiting link performance instead of thermal noise [11]. To mitigate the RF impairments, digital signal processing methods are commonly employed, including the signal pre-distortion at Tx side and the signal compensation at Rx side. Especially, for the low-cost devices in THz communications [3] which suffer more severe RF impairments, the well-designed channel estimation and equalization techniques could provide an efficient way of Rx signal processing to mitigate the severe RF impairments.

Since the THz PA usually has a low output power and suffers the non-linearity effect, we prefer single-carrier (SC) transmission instead of its orthogonal frequency division multiplexing (OFDM) counterpart. It is well-known that the SC waveform has a lower peak-to-average power ratio (PAPR) than the OFDM waveform, which alleviates the non-linearity effect of PA and ensures a higher Tx power as well. Furthermore, the SC frequency-domain equalization (SC-FDE) architecture could be adopted. Although the time-domain Viterbi or Bahl-Cocke-Jelinek-Raviv algorithms may achieve maximum likelihood (ML) equalization, these two algorithms have excessively high complexity. In contrast, SC-FDE is an efficient equalization scheme with low complexity, which is preferable for THz communications due to the ultra-high data throughput. However, the design of low-complexity SC-FDE scheme under RF impairments would be a challenging problem.

A. Related Works

The signal processing methodologies towards RF impairments have been extensively investigated over the past two decades. For Tx side, the IQ imbalance and PA non-linearity usually can be cancelled via digital pre-distortion [14]–[17],

Manuscript received June 29, 2020; revised November 15, 2020 and February 17, 2021; accepted March 1, 2021. Date of publication April 8, 2021; date of current version May 18, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1801501 and in part by the National Natural Science Foundation of China under Grant 61871253. *(Corresponding author: Zhaocheng Wang.)*

The authors are with the Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: shazy17@mails.tsinghua.edu.cn; zcwang@tsinghua.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2021.3071824.

Digital Object Identifier 10.1109/JSAC.2021.3071824

TABLE I
COMPARISON BETWEEN DIFFERENT SIGNAL MODELS

Impairment Type	Baseband Signal Model	Parameter Definition
(i) No impairment	$y[n] = h[n] * x[n] + w[n]$	$y[n]$: Rx Signal, $x[n]$: Tx Signal, $h[n]$: CIR, $w[n]$: AWGN
(ii) NB IQ imbalance	$y[n] = \mu(h[n] * x[n]) + \nu(h[n] * x[n])^* + \mu w[n] + \nu w^*[n]$	μ, ν: NB IQ imbalance parameters
(iii) WB IQ imbalance	$y[n] = c_1[n] * (h[n] * x[n]) + c_2[n] * (h[n] * x[n])^* + c_1[n] * w[n] + c_2[n] * w^*[n]$	$c_1[n], c_2[n]$: WB IQ imbalance parameters
(iv) Tx PN + Rx PN	$y[n] = e^{-j\theta_R[n]}(h[n] * (x[n]e^{j\theta_T[n]})) + w[n]$	$\theta_R[n]$: Rx PN, $\theta_T[n]$: Tx PN
(v) Rx PN + NB IQ imbalance	$y[n] = \mu e^{-j\theta_R[n]}(h[n] * x[n]) + \nu(e^{-j\theta_R[n]}(h[n] * x[n]))^* + \mu w[n] + \nu w^*[n]$	-
(vi) Tx PN + Rx PN + WB IQ imbalance	$y[n] = c_1[n] * (e^{-j\theta_R[n]}(h[n] * (x[n]e^{j\theta_T[n]}))) + c_2[n] * (e^{-j\theta_R[n]}(h[n] * (x[n]e^{j\theta_T[n]})))^* + c_1[n] * w[n] + c_2[n] * w^*[n]$	-

where the Tx signal is distorted in digital domain to neutralize the analog RF impairments. Since our focus is the processing schemes at Rx side, the Tx processing would not be discussed in detail here.

For Rx side, we give a brief review of the baseband signal models under various types of RF impairments in Table I, where we use $*$ to represent the linear convolution operation and $(\cdot)^*$ to represent the conjugate operation.

As shown in model (ii), (iii) of Table I, the IQ imbalance models can be categorized as narrowband (NB), i.e. frequency-independent [18]–[20], and wideband (WB), i.e. frequency-dependent [21]–[25], where the WB model takes the mismatch of filters into account and results in a more accurate but complicated model. In the literature, for both NB and WB IQ imbalance, the channel impulse response (CIR) and IQ imbalance parameters are usually treated together as the effective channel state information (CSI). As a result, the channel estimation (CE) and equalization are performed according to the effective CSI. Especially, for SC-FDE systems, the IQ imbalance compensation methods are investigated in [23], [24].

The signal model of PN is given by model (iv). The compensation of PN is also a widely studied topic for OFDM and SC-FDE systems, e.g. [26]–[31], where a common idea is to use decision-directed method. Specifically, the detected data symbol and the estimated PN can be utilized to improve each other alternately. It should be noted that the effect of PN is a random phase rotation in time domain, which exhibits inter-carrier interference (ICI) in frequency domain [30], [31]. Therefore, the estimation of PN is convenient for SC-FDE [29] because the phase rotation can be easily estimated by the detected data symbol in time domain. Nevertheless, as shown in model (vi), the influence of PN becomes much more complicated when the WB IQ imbalance is present, so the above studies cannot be adopted to our research directly.

Furthermore, the joint compensation of IQ imbalance and PN is investigated in [32]–[35] based on model (v). It can be seen that only the NB IQ imbalance is considered, and the Tx PN is ignored as well. On the other hand, in THz communications, the IQ imbalance could be frequency-dependent [11], [12] due to the large bandwidth, so the WB IQ imbalance model should be more accurate. However, the signal model with both WB IQ imbalance and PN, corresponding to model (vi), is not well-investigated in the literature. Only [36] studied the compensation of IQ sample timing mismatch for OFDM systems, yet an efficient estimation and compensation scheme for both WB IQ imbalance and PN has not been proposed.

B. Our Contributions

Since the WB IQ imbalance and strong PN are the two dominant RF impairments in THz circuits, we study the CE and equalization problems with the coexistence of these two RF impairments. As demonstrated in Section I-A, the coexistence of WB IQ imbalance and PN leads to the previously unsolved signal model (vi) in the existing studies. Therefore, we propose the CE method to handle both WB IQ imbalance and PN for the first time. Different from the existing studies [21], [24], we cannot treat the CIR and WB IQ imbalance together as the effective channel in model (vi) due to the Rx PN. Hence, a novel two-stage CE algorithm is proposed to separately estimate the CIR coefficients and WB IQ imbalance parameters.

In Stage-1, we estimate the total CSI which is the product of the CIR and WB IQ imbalance parameters in frequency domain. Since it is feasible for THz communication to transmit multiple pilot blocks to improve the CE accuracy, we propose a CE method based on extended Kalman filter (EKF) which tracks the PN of each pilot block so that we could combine all pilot blocks. Moreover, a pilot sequence design method based on Chu sequence is proposed and an approximate Cramer-Rao lower bound (CRLB) is derived.

In Stage-2, we propose to extract the CIR coefficients and WB IQ imbalance parameters from the total CSI for the first time, which is essentially a blind deconvolution problem. We employ a straightforward gradient descent demixing method [37] to solve this problem, where the design parameters of the analog filters are required to initialize the start point for gradient descent.

Finally, based on the estimated CIR and IQ imbalance parameters, a low-complexity channel equalization scheme is proposed. Specifically, we compensate the IQ imbalance and equalize the CIR in the presence of PN. Then, an iterative algorithm is adopted to jointly compensate PN and detect data symbol. Our simulations verify that the proposed equalization scheme has better bit error rate (BER) performance than the

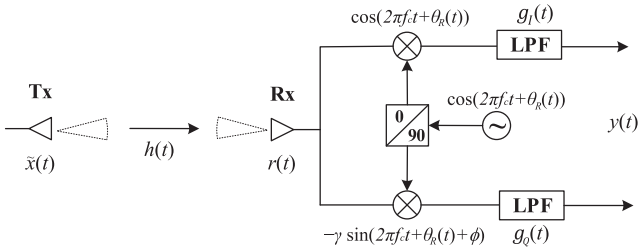

Fig. 1. System block diagram, where f_c denotes the carrier frequency.

existing equalization methods designed for the previous signal models (iii), (iv), (v).

The rest of this paper is organized as follow. Section II introduces our system model. Section III and IV propose the Stage-1 and the Stage-2 CE, respectively. Our equalization scheme is proposed in Section V. The simulation results are demonstrated in Section VI, and the conclusions are drawn in Section VII. We use normal-face letters to denote scalars, and use boldface lower-case and upper-case letters to denote column vectors and matrices, respectively. Matrix transpose, conjugate, conjugate transpose, inverse, trace, and pseudo-inverse are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{-1}$, tr$(\cdot)$ and $(\cdot)^\dagger$, respectively. $\angle(\cdot)$, $|\cdot|$, $||\cdot||$ represent the phase angle of a scalar, the magnitude of a scalar and the l_2-norm of a vector, $*$, $\circledast$, $\odot$ represent linear convolution, circular convolution and element-wise multiplication. diag$\{\boldsymbol{p}\}$ denotes the diagonal matrix whose diagonal elements are formed by vector $\boldsymbol{p}$, and $\boldsymbol{I}_n$ denotes a $n \times n$ identity matrix.

II. System Model

We consider a single-input single-output THz communication link with Tx PN, Rx PN and Rx IQ imbalance, while the Tx IQ imbalance and PA non-linearity are assumed to be compensated [17]. In this section, the time-domain signal model is introduced firstly, and then the frequency-domain model for SC-FDE is derived. Especially, the condition of converting linear convolution to circular convolution under PN is discussed, which is usually ignored in the existing researches.

A. Time-Domain Signal Model

The system block diagram is shown in Fig. 1, where highly directional antennas with sharp radiation patterns are equipped at both Tx and Rx sides. Meanwhile, our system model can also represent a single-RF-chain transmission scenario where ultra-large antenna arrays are used to generate directional beams.

We denote the continuous-time baseband Tx signal as $x(t)$ and the Tx PN as $\theta_T(t)$. The equivalent Tx signal $\tilde{x}(t)$ with PN [36] is given by

$$\tilde{x}(t) = x(t)e^{j\theta_T(t)}. \tag{1}$$

The equivalent received baseband signal $r(t)$ can be expressed as

$$r(t) = h(t) * \tilde{x}(t) + w(t), \tag{2}$$

where $h(t)$, $w(t)$ represent the continuous-time baseband CIR and the additive white Gaussian noise (AWGN), respectively.

Next, $r(t)$ passes through a direct quadrature down-converter with WB IQ imbalance and PN as shown in Fig. 1, where the Rx PN is denoted by $\theta_R(t)$. Because of the IQ imbalance, the amplitudes of the in-phase and quadrature LO are not equal, and their phase difference is not strictly 90° as well. Therefore, we define the amplitude imbalance parameter as γ and the phase imbalance parameter as ϕ, which results in a quadrature LO signal expressed as $-\gamma \sin(2\pi f_c t + \theta_R(t) + \phi)$. Moreover, a mismatch between the two low-pass filters (LPFs) of the IQ branches exists, and we denote the impulse responses of the in-phase LPF and the quadrature LPF as $g_I(t)$ and $g_Q(t)$, respectively.

Consequently, the actual baseband Rx signal $y(t)$ after down conversion is given by

$$y(t) = c_1(t) * (r(t)e^{-j\theta_R(t)}) + c_2(t) * (r(t)e^{-j\theta_R(t)})^*, \tag{3}$$

where we have

$$c_1(t) = \frac{1}{2}(g_I(t) + \gamma e^{-j\phi} g_Q(t)),$$
$$c_2(t = \frac{1}{2}(g_I(t) - \gamma e^{j\phi} g_Q(t)).$$

The derivation of (3) is given in Appendix A. It can be observed that $r(t)$ firstly multiplies the Rx PN $e^{-j\theta_R(t)}$ and then convolutes with the WB IQ imbalance parameters $c_1(t)$ or $c_2(t)$, which indicates that $h(t)$ and $c_1(t)$ ($c_2(t)$) are separated.

Assume that the Tx signal $x(t)$ satisfies Nyquist criterion. We combine (1)-(3), and convert the signal model into discrete-time form as

$$\begin{aligned} y[n] = {} & c_1[n] * (e^{-j\theta_R[n]}(h[n] * (x[n]e^{j\theta_T[n]}))) \\ & + c_2[n] * (e^{-j\theta_R[n]}(h[n] * (x[n]e^{j\theta_T[n]})))^* + \tilde{w}[n], \end{aligned} \tag{4}$$

where $y(t)$, $r(t)$, $\tilde{x}(t)$, $x(t)$, $h(t)$, $c_1(t)$, $c_2(t)$, $w(t)$ are converted into discrete-time forms, e.g. $y[n] = y(nT_s)$, and T_s denotes the symbol interval. The noise term is given by $\tilde{w}[n] = c_1[n] * w[n] + c_2[n] * w^*[n]$, where the AWGN $w[n] \sim \mathcal{CN}(0, \sigma^2)$. Note that the Rx PN is omitted in $\tilde{w}[n]$ because it does not change the statistical property of noise.

The random-walk PN model [26] is applied for the Tx and Rx PN, which is given by

$$\begin{aligned} \theta_T[n+1] &= \theta_T[n] + \Delta\theta_T[n+1], \\ \theta_R[n+1] &= \theta_R[n] + \Delta\theta_R[n+1], \end{aligned} \tag{5}$$

where $\Delta\theta_T[n+1] \sim N(0, \sigma^2_{\theta_T})$, $\Delta\theta_R[n+1] \sim N(0, \sigma^2_{\theta_R})$. The variance $\sigma^2_{\theta_T}$ and $\sigma^2_{\theta_R}$ are derived by $\sigma^2_{\theta_T} = 2\pi f_{3\mathrm{dB},T} T_s$, $\sigma^2_{\theta_R} = 2\pi f_{3\mathrm{dB},R} T_s$, where $f_{3\mathrm{dB},T}$, $f_{3\mathrm{dB},R}$ represent the one-side 3dB bandwidth of the Lorentzian spectrum of the Tx and Rx LO, respectively.

B. Channel Model and Noise Model

We denote the maximum length of CIR $h[n]$ as L_h, corresponding to the delay spread of channel. According to [38], there could be multiple transmission paths in the THz environment. However, as illustrated in Fig. 2, since the highly

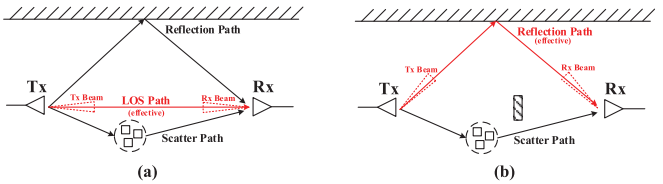

Fig. 2. Channel model, where only one transmission path (or cluster) is effective due to the extremely sharp beams at Tx and Rx sides. (a) The LOS path is the effective transmission path. (b) A reflection path is chosen as the effective transmission path, where the LOS path is blocked.

directional beams are used at both Tx and Rx sides and the channel is single-input single-output, there is commonly only one effective transmission path (or cluster) inside the beamwidth, which results in a small delay spread. In addition, as reported in [39], a double-reflected ray may extend the delay spread of the line-of-sight (LOS) path, e.g. the path reflected by the Rx antenna and reflected again by the Tx antenna. However, the double-reflected path has 20dB higher attenuation than the LOS path, which could be ignored. Consequently, a small delay spread can be assumed and results in a small maximum CIR length L_h. This assumption could facilitate the CE procedure and ensure a reasonable length of cyclic prefix (CP) as well. Besides, the channel is considered as time-invariant.

Due to the highly directional beams, the characteristics of the equivalent baseband THz channel become relatively simple since there is commonly only one effective path or cluster. Therefore, in our design of CE and equalization, the CIR is considered as a complex sequence with a short length, corresponding to a typical directional THz channel with small delay spread. Except the length of CIR, we do not introduce further constraints on the THz channel to guarantee the generality of our proposed CE and equalization methods. Meanwhile, the severe RF impairments of THz hardware components become the major challenge in CE and equalization.

On the other hand, apart from the thermal noise, the noise in THz band may contain molecular absorption noise [40] whose power spectral density (PSD) is determined by the spectrum of the medium absorption coefficient. However, like most experiments on THz communications [4]–[12], we assume that the communication band lies in a transmission window [2] below 1THz, which ensures that the spectrum of molecular absorption is almost flat. Hence, the molecular absorption noise and the thermal noise can be represented together by the AWGN term $w[n]$.

C. Frequency-Domain Signal Model

An important prerequisite for frequency-domain equalization is to convert linear convolution into circular convolution. To this end, CP can be added into a SC block. However, due to the presence of PN, the circular property may not retain. As illustrated in Fig. 3(a), since the CP and the tail of a SC block may multiply different Tx PN, the circular property does not hold when the Tx signal $\tilde{x}[n]$ convolutes with channel $h[n]$. Similarly, the convolution between the Rx signal $\tilde{r}[n]$ and the WB IQ imbalance parameters may be not able to convert into circular convolution either.

To circumvent this problem, two solutions are given here. Firstly, we can set a small length for each SC block. In this way, the PN at the CP and the PN at the tail of a SC block can be approximately considered as same. However, small block length leads to very high overhead of CP, so we only apply short SC block for pilot block transmission. Secondly, zero-padding (ZP) SC block can be utilized as shown in Fig. 3(b), where the tail of each SC block is padded with zeros. Since zero signal is immune to phase rotation, the circular property would retain. Therefore, we use ZP SC blocks for data transmission.

We denote the length of SC block and CP (or ZP) as N, N_{CP}, respectively, and the maximum length of WB IQ imbalance parameters $c_1[n]$, $c_2[n]$ as L_c, where $N_{CP} \geq L_c + L_h - 1$. The vector form of time-domain Tx signal is defined as $\boldsymbol{x} = [x[0], \ldots, x[N-1]]^T$, where $x[n]$ is time-domain pilot or data symbol. It should be noted that $x[n] = 0$ for $n = N - N_{CP}, \ldots, N-1$ when ZP SC block is applied.

Using the circular property of $x[n]$, the vector form of time-domain Rx signal $\boldsymbol{y} = [y[0], \ldots, y[N-1]]^T$ can be expressed as

$$\begin{aligned}\boldsymbol{y} = \boldsymbol{C}_1\big(\mathrm{diag}\{\boldsymbol{p}_R\}\boldsymbol{H}\mathrm{diag}\{\boldsymbol{p}_T\}\boldsymbol{x}\big) \\ + \boldsymbol{C}_2\big(\mathrm{diag}\{\boldsymbol{p}_R\}\boldsymbol{H}\mathrm{diag}\{\boldsymbol{p}_T\}\boldsymbol{x}\big)^* + \tilde{\boldsymbol{w}}, \quad (6)\end{aligned}$$

where $\boldsymbol{H}$, $\boldsymbol{C}_1$, $\boldsymbol{C}_2$ are $N \times N$ circulant matrices whose first columns are

$$\begin{aligned}&[h[0], \ldots, h[L_h-1], 0, \ldots, 0]^T, \\ &[c_1[0], \ldots, c_1[L_c-1], 0, \ldots, 0]^T, \\ &[c_2[0], \ldots, c_2[L_c-1], 0, \ldots, 0]^T,\end{aligned}$$

respectively, and $\boldsymbol{p}_T = [e^{j\theta_T[0]}, \ldots, e^{j\theta_T[N-1]}]^T$, $\boldsymbol{p}_R = [e^{-j\theta_R[0]}, \ldots, e^{-j\theta_R[N-1]}]^T$. The noise term $\tilde{\boldsymbol{w}} \approx \boldsymbol{C}_1\boldsymbol{w} + \boldsymbol{C}_2\boldsymbol{w}^*$, where $\boldsymbol{w} = [w[0], \ldots, w[N-1]]^T$ and the Rx PN is omitted in $\tilde{\boldsymbol{w}}$.

Denote the vector forms of the CIR and WB IQ imbalance parameters as $\boldsymbol{h} = [h[0], \ldots, h[L_h-1]]^T$, $\boldsymbol{c}_1 = [c_1[0], \ldots, c_1[L_c-1]]^T$, $\boldsymbol{c}_2 = [c_2[0], \ldots, c_2[L_c-1]]^T$. Then, according to the property of circulant matrix, $\boldsymbol{H}$, $\boldsymbol{C}_1$, $\boldsymbol{C}_2$ can be decomposed into

$$\begin{aligned}\boldsymbol{H} &= \frac{1}{N}\boldsymbol{F}^H\mathrm{diag}\{\boldsymbol{h}_F\}\boldsymbol{F}, \\ \boldsymbol{C}_1 &= \frac{1}{N}\boldsymbol{F}^H\mathrm{diag}\{\boldsymbol{c}_{F,1}\}\boldsymbol{F}, \\ \boldsymbol{C}_2 &= \frac{1}{N}\boldsymbol{F}^H\mathrm{diag}\{\boldsymbol{c}_{F,2}\}\boldsymbol{F}, \quad (7)\end{aligned}$$

where $\boldsymbol{F}$ is $N \times N$ DFT matrix with element $e^{-j2\pi n_1 n_2/N}$ in the n_1th row and the n_2th column, and $\boldsymbol{h}_F = \boldsymbol{F}_h\boldsymbol{h}$, $\boldsymbol{c}_{F,1} = \boldsymbol{F}_c\boldsymbol{c}_1$, $\boldsymbol{c}_{F,2} = \boldsymbol{F}_c\boldsymbol{c}_2$ represent the frequency-domain CIR and WB IQ parameters, $\boldsymbol{F}_h$, $\boldsymbol{F}_c$ represent the submatrix formed by the first L_h, L_c columns of $\boldsymbol{F}$, respectively.

By substituting (7) into (6), the frequency-domain signal model is given by

$$\begin{aligned}\boldsymbol{y}_F = \mathrm{diag}\{\boldsymbol{c}_{F,1}\}\boldsymbol{P}_R\mathrm{diag}\{\boldsymbol{h}_F\}\boldsymbol{P}_T\boldsymbol{x}_F \\ + \mathrm{diag}\{\boldsymbol{c}_{F,2}\}\tilde{\boldsymbol{P}}_R\mathrm{diag}\{\boldsymbol{h}_F^{\#}\}\tilde{\boldsymbol{P}}_T\boldsymbol{x}_F^{\#} + \tilde{\boldsymbol{w}}_F, \quad (8)\end{aligned}$$

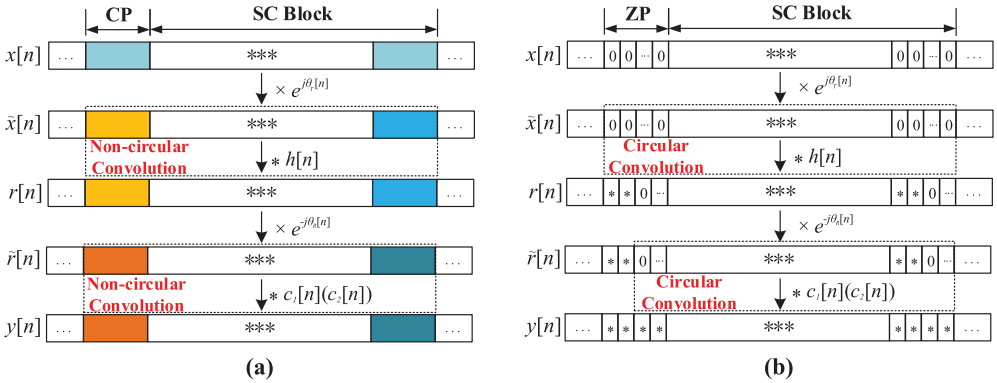

Fig. 3. An interpretation of signal model, where the noise term is omitted and $\tilde{r}[n] = r[n]e^{-j\theta_R[n]}$. (a) CP SC block: the circular property is spoiled by PN, (b) ZP SC block: the circular property retains due to the zero signal.

where $\boldsymbol{x}_F = \boldsymbol{F}\boldsymbol{x}$, $\boldsymbol{y}_F = \boldsymbol{F}\boldsymbol{y}$, $\tilde{\boldsymbol{w}}_F = \boldsymbol{F}\tilde{\boldsymbol{w}}$ are the frequency-domain Tx, Rx signal and noise, respectively, and $(\cdot)^{\#}$ represents the conjugate symmetry operation, e.g. $\boldsymbol{x}_F^{\#} = [x_F^*[0], x_F^*[N-1], \ldots, x_F^*[1]]^T$. Besides, we have $\boldsymbol{P}_T = \frac{1}{N}\boldsymbol{F}\text{diag}\{\boldsymbol{p}_T\}\boldsymbol{F}^H$, $\boldsymbol{P}_R = \frac{1}{N}\boldsymbol{F}\text{diag}\{\boldsymbol{p}_R\}\boldsymbol{F}^H$, $\tilde{\boldsymbol{P}}_T = \frac{1}{N}\boldsymbol{F}\text{diag}\{\boldsymbol{p}_T^*\}\boldsymbol{F}^H$, $\tilde{\boldsymbol{P}}_R = \frac{1}{N}\boldsymbol{F}\text{diag}\{\boldsymbol{p}_R^*\}\boldsymbol{F}^H$. It should be noted that $\boldsymbol{P}_T$, $\boldsymbol{P}_R$, $\tilde{\boldsymbol{P}}_T$, $\tilde{\boldsymbol{P}}_R$ are also circulant matrices whose first columns are $\frac{1}{N}\boldsymbol{F}\boldsymbol{p}_T$, $\frac{1}{N}\boldsymbol{F}\boldsymbol{p}_R$, $\frac{1}{N}\boldsymbol{F}\boldsymbol{p}_T^*$, $\frac{1}{N}\boldsymbol{F}\boldsymbol{p}_T^*$, respectively.

III. Stage-1 Channel Estimation

In the Stage-1 CE, the product of the frequency-domain CIR and WB IQ imbalance parameters is referred to as the frequency-domain total CSI, which is defined as $\boldsymbol{h}_{F,1} = \boldsymbol{h}_F \odot \boldsymbol{c}_{F,1}$, $\boldsymbol{h}_{F,2} = \boldsymbol{h}_F^{\#} \odot \boldsymbol{c}_{F,2}$ according to (8), and the corresponding time-domain total CSI is $\boldsymbol{h} \circledast \boldsymbol{c}_1$, $\boldsymbol{h}^* \circledast \boldsymbol{c}_2$, respectively. However, given that L_c is small and we assume L_h to be small as well, the circular convolutions $\boldsymbol{h} \circledast \boldsymbol{c}_1$, $\boldsymbol{h}^* \circledast \boldsymbol{c}_2$ are equivalent to linear convolutions $\boldsymbol{h} * \boldsymbol{c}_1$, $\boldsymbol{h}^* * \boldsymbol{c}_2$ since only the first $L_c + L_h - 1$ elements are non-zero in circular convolution output. Therefore, we define the time-domain total CSI as $\boldsymbol{h}_1 = \boldsymbol{h} * \boldsymbol{c}_1$, $\boldsymbol{h}_2 = \boldsymbol{h}^* * \boldsymbol{c}_2$, which are two length-$(L_h + L_c - 1)$ vectors, and we have $\boldsymbol{h}_{F,1} = \boldsymbol{F}_H\boldsymbol{h}_1$, $\boldsymbol{h}_{F,2} = \boldsymbol{F}_H\boldsymbol{h}_2$, where $\boldsymbol{F}_H$ is the sub-matrix formed by the first $L_h + L_c - 1$ columns of $\boldsymbol{F}$.

Since the bandwidth can be ultra-wide in THz communications, it is possible to transmit a long pilot sequence consisting of multiple pilot blocks to enhance CE accuracy. Therefore, we propose an EKF-based CE method to combine multiple pilot blocks. Besides, the pilot sequence design and an approximate CRLB for Stage-1 CE are also presented in this section.

A. EKF-Based Channel Estimation

We assume that M consecutive short pilot blocks with CP are transmitted for CE, the time-domain Tx, Rx signal of the mth pilot is denoted as $\boldsymbol{x}_m$, $\boldsymbol{y}_m$ respectively, and the corresponding frequency domain signal is $\boldsymbol{x}_{F,m}$, $\boldsymbol{y}_{F,m}$. As mentioned in Section II-C, since the pilot blocks are short (N is small), the Tx and Rx PN vary slowly within a pilot block, so they can be approximated as constant for each pilot block. Therefore, we use common phase error (CPE) $\theta_{T,m}$ and $\theta_{R,m}$ to represent the Tx and Rx PN for the mth pilot block, respectively. We consider $\boldsymbol{P}_T \approx e^{j\theta_{T,m}}\boldsymbol{I}_N$, $\boldsymbol{P}_R \approx e^{-j\theta_{R,m}}\boldsymbol{I}_N$, $\tilde{\boldsymbol{P}}_T \approx e^{-j\theta_{T,m}}\boldsymbol{I}_N$, $\tilde{\boldsymbol{P}}_R \approx e^{j\theta_{R,m}}\boldsymbol{I}_N$, and hence the frequency-domain signal model (8) becomes

$$\begin{aligned}\boldsymbol{y}_{F,m} &\approx e^{-j\theta_m}\text{diag}\{\boldsymbol{c}_{F,1}\}\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{x}_{F,m} \\ &\quad + e^{j\theta_m}\text{diag}\{\boldsymbol{c}_{F,2}\}\text{diag}\{\boldsymbol{h}_F^{\#}\}\boldsymbol{x}_{F,m}^{\#} + \tilde{\boldsymbol{w}}_{F,m} \\ &= e^{-j\theta_m}\text{diag}\{\boldsymbol{x}_{F,m}\}\boldsymbol{F}_H\boldsymbol{h}_1 \\ &\quad + e^{j\theta_m}\text{diag}\{\boldsymbol{x}_{F,m}^{\#}\}\boldsymbol{F}_H\boldsymbol{h}_2 + \tilde{\boldsymbol{w}}_{F,m}, \end{aligned} \tag{9}$$

where $m = 0, \ldots, M-1$, and $\theta_m = \theta_{R,m} - \theta_{T,m}$.

Note that there exists a phase ambiguity in the estimation of θ_m, $\boldsymbol{h}_1$, $\boldsymbol{h}_2$, i.e. for an arbitrary θ, the estimation values $\theta_m - \theta$, $\boldsymbol{h}_1 e^{-j\theta}$, $\boldsymbol{h}_2 e^{j\theta}$ are also consistent with (9). However, in the equalization procedure (see Section V), this phase ambiguity leads to a phase rotation of the signal, which can be treated as PN and compensated via the phase-tracking pilots. So, a phase ambiguity in the total CSI would be eliminated in the subsequent channel equalization. Therefore, in the Stage-1 CE, we set $\theta = \theta_0$ and estimate $\boldsymbol{h}_1' = \boldsymbol{h}_1 e^{-j\theta_0}$, $\boldsymbol{h}_2' = \boldsymbol{h}_2 e^{j\theta_0}$ instead of $\boldsymbol{h}_1$, $\boldsymbol{h}_2$.

Denote the frequency-domain Rx signal vector encompassing all M pilot blocks as $\bar{\boldsymbol{y}}_F = [\boldsymbol{y}_{F,0}^T, \ldots, \boldsymbol{y}_{F,M-1}^T]^T$, and from (9), we have

$$\bar{\boldsymbol{y}}_F \approx \bar{\boldsymbol{X}} \begin{bmatrix} \boldsymbol{F}_H & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_H \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_1' \\ \boldsymbol{h}_2' \end{bmatrix} + \bar{\boldsymbol{w}}_F, \tag{10}$$

where $\bar{\boldsymbol{w}}_F = [\tilde{\boldsymbol{w}}_{F,0}^T, \ldots, \tilde{\boldsymbol{w}}_{F,M-1}^T]^T$, and

$$\bar{\boldsymbol{X}} = \begin{bmatrix} e^{-j\theta_0'}\text{diag}\{\boldsymbol{x}_{F,0}\} & e^{j\theta_0'}\text{diag}\{\boldsymbol{x}_{F,0}^{\#}\} \\ \vdots & \vdots \\ e^{-j\theta_{M-1}'}\text{diag}\{\boldsymbol{x}_{F,M-1}\} & e^{j\theta_{M-1}'}\text{diag}\{\boldsymbol{x}_{F,M-1}^{\#}\} \end{bmatrix}, \tag{11}$$

and we define $\theta_m' = \theta_m - \theta_0$ for $m = 0, \ldots, M-1$ ($\theta_0' = 0$). It can be seen from (10) that $\boldsymbol{h}_1'$ and $\boldsymbol{h}_2'$ can be easily estimated via least-square (LS) estimation if we have the value of θ_m'.

To estimate θ_m', we use the first two pilot blocks to acquire an initial estimation of $\boldsymbol{h}_1'$ and $\boldsymbol{h}_2'$ by assuming $\theta_1' = \theta_0' = 0$,

the estimated $\hat{\boldsymbol{h}}_1^{'}$ and $\hat{\boldsymbol{h}}_2^{'}$ are given by

$$\begin{bmatrix}\hat{\boldsymbol{h}}_1^{'}\\ \hat{\boldsymbol{h}}_2^{'}\end{bmatrix} = \left(\begin{bmatrix}\text{diag}\{\boldsymbol{x}_{F,0}\} & \text{diag}\{\boldsymbol{x}_{F,0}^{\#}\}\\ \text{diag}\{\boldsymbol{x}_{F,1}\} & \text{diag}\{\boldsymbol{x}_{F,1}^{\#}\}\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_H & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{F}_H\end{bmatrix}\right)^{\dagger}\begin{bmatrix}\boldsymbol{y}_{F,0}\\ \boldsymbol{y}_{F,1}\end{bmatrix}. \tag{12}$$

Based on $\hat{\boldsymbol{h}}_1^{'}$ and $\hat{\boldsymbol{h}}_2^{'}$, the EKF is employed to estimate θ_m'. According to (5), the state equation of θ_m' can be expressed as

$$\theta_m' = \theta_{m-1}' + \Delta\theta', \quad m = 1, \ldots, M-1, \tag{13}$$

where $\Delta\theta' \sim \mathcal{N}(0, \sigma_\theta^2)$, $\sigma_\theta^2 = (N_{cp}+N)(\sigma_{\theta_T}^2 + \sigma_{\theta_R}^2)$. On the other hand, since the effect of phase noise on $\boldsymbol{y}_{F,m}$ is nonlinear, the first-order Taylor expansion is adopted to approximate the observation equation (9). Denote the estimated value of θ_m' as $\hat{\theta}_m'$, and the observation equation is approximated by

$$\boldsymbol{y}_{F,m} \approx \boldsymbol{\beta} + \boldsymbol{\alpha}(\theta_m' - \hat{\theta}_{m-1}') + \tilde{\boldsymbol{w}}_{F,m}, \tag{14}$$

where $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta} = e^{-j\hat{\theta}_{m-1}'}\text{diag}\{\boldsymbol{x}_{F,m}\}\boldsymbol{F}_H\boldsymbol{h}_1 + e^{j\hat{\theta}_{m-1}'}\text{diag}\{\boldsymbol{x}_{F,m}^{\#}\}\boldsymbol{F}_H\boldsymbol{h}_2,$$

and $\boldsymbol{\alpha} = \frac{\partial\boldsymbol{\beta}}{\partial\hat{\theta}_{m-1}'}$.

According to (13) and (14), by maximizing the a posterior probability $p(\theta_m'|\hat{\theta}_{m-1}', \boldsymbol{y}_{F,m}) \propto p(\hat{\theta}_{m-1}'|\theta_m')p(\boldsymbol{y}_{F,m}|\theta_m')$, the EKF for tracking θ_m' can be derived by

$$\begin{aligned}\hat{\theta}_m' &= \hat{\theta}_{m-1}' + \frac{(\sigma_{\theta,m-1}^2+\sigma_\theta^2)\sum_{k=0}^{N-1}2\text{Re}\{(y_{F,m}[k]-\beta[k])^*\alpha[k]\}}{\sigma_F^2 + 2(\sigma_{\theta,m-1}^2+\sigma_\theta^2)},\\ \sigma_{\theta,m}^2 &= \frac{(\sigma_{\theta,m-1}^2+\sigma_\theta^2)\sigma_F^2}{\sigma_F^2 + 2(\sigma_{\theta,m-1}^2+\sigma_\theta^2)},\end{aligned} \tag{15}$$

where $\sigma_{\theta,m}^2$ represents the variance of $\hat{\theta}_m'$, σ_F^2 represents the variance of each element in $\tilde{\boldsymbol{w}}_{F,m}$, and we set $\sigma_F^2 = N\sigma^2$. Besides, $\hat{\theta}_m'$ and $\sigma_{\theta,m}^2$ are initialized as $\hat{\theta}_0' = 0$, $\sigma_{\theta,0}^2 = 0$, and the $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ derived in (12) are used to compute $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

After acquiring the estimation $\hat{\theta}_m'$, we can derive an estimation $\hat{\boldsymbol{X}}$ of $\bar{\boldsymbol{X}}$ according to (11), and based on LS criterion, the estimation $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ are expressed as

$$\begin{bmatrix}\hat{\boldsymbol{h}}_1^{'}\\ \hat{\boldsymbol{h}}_2^{'}\end{bmatrix} = \left(\hat{\boldsymbol{X}}\begin{bmatrix}\boldsymbol{F}_H & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{F}_H\end{bmatrix}\right)^{\dagger}\bar{\boldsymbol{y}}_F. \tag{16}$$

Additionally, the estimation of $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ can be further improved by a phase calibration process using $\boldsymbol{y}_{F,0}$. Specifically, we can obtain a phase calibration θ by solving

$$\theta = \arg\min_\theta ||\boldsymbol{y}_{F,0} - e^{-j\theta}\hat{\boldsymbol{h}}_1^{'} - e^{j\theta}\hat{\boldsymbol{h}}_2^{'}||^2.$$

Then, $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ could be replaced by $e^{-j\theta}\hat{\boldsymbol{h}}_1^{'}$, $e^{j\theta}\hat{\boldsymbol{h}}_2^{'}$, respectively. However, since a phase ambiguity in $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ is insignificant for equalization, this calibration is unnecessary for practical systems.

B. Pilot Sequence Design

The design of M pilot sequences $\boldsymbol{x}_m$, $m = 0, \ldots, M-1$ is important for the CE procedure. It can be seen from (16) that a $\hat{\boldsymbol{X}}$ with large condition number would significantly amplify the noise in $\bar{\boldsymbol{y}}_F$ and results in performance loss. Ideally, $\hat{\boldsymbol{X}}$ is expected to be a semi-unitary matrix with condition number of 1. However, since the phase noise term θ_m' is uncertain, we only consider the original pilot matrix $\boldsymbol{X}$, corresponding to the $\bar{\boldsymbol{X}}$ where all θ_m' are set to be zero. The design objective is to find $\boldsymbol{X}$ satisfying $\boldsymbol{X}^H\boldsymbol{X} = C\boldsymbol{I}_{2N}$, where C is a constant.

We propose a method based on Chu sequence to construct $\boldsymbol{X}$ with various block length N and pilot block number M. The time-domain sequence $\boldsymbol{x}_m$ is designed as

$$x_m[n] = e^{-j\pi\frac{n(n+2qm)}{N}}, \quad m = 0, \ldots, M-1, \tag{17}$$

where q is an integer parameter to be determined. By restricting N to an even number, we have the properties $x_{F,m}[k] = x_m^*[k]x_{F,m}[0]$ and $x_{F,m}[0] = e^{j\pi\frac{(qm)^2}{N}}x_{F,0}[0]$ for the frequency-domain sequence $\boldsymbol{x}_{F,m}$, which are derived in Appendix B.

Since $\boldsymbol{X}$ are constructed by diagonal matrices, to ensure $\boldsymbol{X}^H\boldsymbol{X} = C\boldsymbol{I}_{2N}$, we only need to let the inner products G_k between the kth column and the $(N+k)$th column of $\boldsymbol{X}$ be zero for all $k = 1, \ldots, N$. Using the two properties above, G_k can be written as

$$\begin{aligned}G_k &= \sum_{m=0}^{M-1} x_{F,m}^*[k-1]x_{F,m}^*[N-k+1]\\ &= \sum_{m=0}^{M-1} x_m[k-1]x_m[N-k+1](x_{F,m}^*[0])^2\\ &= e^{-j\frac{2\pi(k-1)^2}{N}}\sum_{m=0}^{M-1}(x_{F,m}^*[0])^2\\ &= e^{-j\frac{2\pi(k-1)^2}{N}}(x_{F,0}^*[0])^2\sum_{m=0}^{M-1}e^{-j\pi\frac{2q^2m^2}{N}}.\end{aligned}$$

It can be seen that we can ensure $G_k = 0$ for all k by letting $\sum_{m=0}^{M-1} e^{-j\pi\frac{2q^2m^2}{N}} = 0$.

Note that if we let $2q^2/N$ be an odd number, then we have $\sum_{m=0}^{M-1} e^{-j\pi\frac{2q^2m^2}{N}} = 0$ for an arbitrary even number M. Based on this observation, many pairs of (N, q) can be designed, e.g.

$$(N, q) = (24, 6), (32, 4), (40, 10), (50, 5), (54, 9), \ldots.$$

Meanwhile, M is an even number, indicating that $2, 4, 6, 8, \ldots$ consecutive pilot blocks can be transmitted. Besides, the Chu sequence is constant-amplitude in time domain, which is also favorable for the Tx PA.

C. CRLB for Stage-1 Channel Estimation

The estimation error of (9) results from the noise term $\tilde{\boldsymbol{w}}_{F,m}$ and the approximation of constant PN over single pilot block. To derive the CRLB of $\boldsymbol{h}_1'$, $\boldsymbol{h}_2'$, θ_m', we use the technique similar to [26], [33], where the approximation error of PN is treated as AWGN.

Compare (8) with (9), we can rewrite (8) as (18),

$$\begin{aligned}\boldsymbol{y}_{F,m} &= e^{-j\theta_m}\text{diag}\{\boldsymbol{c}_{F,1}\}\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{x}_{F,m} \\ &+ e^{j\theta_{T,m}}\text{diag}\{\boldsymbol{c}_{F,1}\}\boldsymbol{E}_R\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{x}_F \\ &+ e^{-j\theta_{R,m}}\text{diag}\{\boldsymbol{c}_{F,1}\}\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{E}_T\boldsymbol{x}_F \\ &+ \text{diag}\{\boldsymbol{c}_{F,1}\}\boldsymbol{E}_R\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{E}_T\boldsymbol{x}_F \\ &+ e^{j\theta_m}\text{diag}\{\boldsymbol{c}_{F,2}\}\text{diag}\{\boldsymbol{h}_F^{\#}\}\boldsymbol{x}_{F,m}^{\#} \\ &+ e^{-j\theta_{T,m}}\text{diag}\{\boldsymbol{c}_{F,2}\}\tilde{\boldsymbol{E}}_R\text{diag}\{\boldsymbol{h}_F^{\#}\}\boldsymbol{x}_F^{\#} \\ &+ e^{j\theta_{R,m}}\text{diag}\{\boldsymbol{c}_{F,2}\}\text{diag}\{\boldsymbol{h}_F^{\#}\}\tilde{\boldsymbol{E}}_T\boldsymbol{x}_F^{\#} \\ &+ \text{diag}\{\boldsymbol{c}_{F,2}\}\tilde{\boldsymbol{E}}_R\text{diag}\{\boldsymbol{h}_F^{\#}\}\tilde{\boldsymbol{E}}_T\boldsymbol{x}_F^{\#} + \tilde{\boldsymbol{w}}_{F,m}, \quad (18)\end{aligned}$$

where $\boldsymbol{P}_T$, $\boldsymbol{P}_R$, $\tilde{\boldsymbol{P}}_T$, $\tilde{\boldsymbol{P}}_R$ are split into $\boldsymbol{P}_T = e^{j\theta_{T,m}}\boldsymbol{I}_N + \boldsymbol{E}_T$, $\boldsymbol{P}_R = e^{-j\theta_{R,m}}\boldsymbol{I}_N + \boldsymbol{E}_R$, $\tilde{\boldsymbol{P}}_T = e^{-j\theta_{T,m}}\boldsymbol{I}_N + \tilde{\boldsymbol{E}}_T$, $\tilde{\boldsymbol{P}}_R = e^{j\theta_{R,m}}\boldsymbol{I}_N + \tilde{\boldsymbol{E}}_R$. The CPE $\theta_{T,m}$, $\theta_{R,m}$ are given by $\theta_{T,m} = \angle\left(\frac{1}{N}\sum_{n=0}^{N-1} e^{j\theta_T[n]}\right)$ and $\theta_{R,m} = \angle\left(\frac{1}{N}\sum_{n=0}^{N-1} e^{-j\theta_R[n]}\right)$, respectively.

Recall that $\boldsymbol{P}_T$, $\boldsymbol{P}_R$, $\tilde{\boldsymbol{P}}_T$, $\tilde{\boldsymbol{P}}_R$ are circulant matrices. Because the length of pilot block N is short, it can be verified that the differences between the diagonal elements of $\boldsymbol{P}_T$, $\boldsymbol{P}_R$, $\tilde{\boldsymbol{P}}_T$, $\tilde{\boldsymbol{P}}_R$ and $e^{j\theta_{T,m}}$, $e^{-j\theta_{R,m}}$, $e^{-j\theta_{T,m}}$, $e^{j\theta_{R,m}}$ can be ignored. Hence, $\boldsymbol{E}_T$, $\boldsymbol{E}_R$, $\tilde{\boldsymbol{E}}_T$, $\tilde{\boldsymbol{E}}_R$ represent the matrices formed by the small non-diagonal elements of $\boldsymbol{P}_T$, $\boldsymbol{P}_R$, $\tilde{\boldsymbol{P}}_T$, $\tilde{\boldsymbol{P}}_R$. Therefore, the two terms $\text{diag}\{\boldsymbol{c}_{F,1}\}\boldsymbol{E}_R\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{E}_T\boldsymbol{x}_F$ and $\text{diag}\{\boldsymbol{c}_{F,2}\}\tilde{\boldsymbol{E}}_R\text{diag}\{\boldsymbol{h}_F^{\#}\}\tilde{\boldsymbol{E}}_T\boldsymbol{x}_F^{\#}$ in (18) are negligible compared to other terms.

Next, we treat the term $e^{j\theta_{T,m}}\text{diag}\{\boldsymbol{c}_{F,1}\}\boldsymbol{E}_R\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{x}_F$ as AWGN and consider its average power. We assume that the time-domain Tx power of pilot is σ_x^2, so the average power of each element in $\boldsymbol{x}_F$ is $\sigma_{F,x}^2 = N\sigma_x^2$. We denote the average power over all elements in $\boldsymbol{h}_F$, $\boldsymbol{c}_{F,1}$, $\boldsymbol{c}_{F,2}$ as P_h, P_{c1}, P_{c2}, respectively. Since $\boldsymbol{E}_R$ is circulant matrix, we denote the sum of the average power of the elements in its each row as P_E. The average power of each element in $e^{j\theta_{T,m}}\text{diag}\{\boldsymbol{c}_{F,1}\}\boldsymbol{E}_R\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{x}_F$ is approximately given by $P_{c1}P_EP_h\sigma_{F,x}^2$. Similarly, the average power of the terms $e^{-j\theta_{R,m}}\text{diag}\{\boldsymbol{c}_{F,1}\}\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{E}_T\boldsymbol{x}_F$, $e^{-j\theta_{T,m}}\text{diag}\{\boldsymbol{c}_{F,2}\}\tilde{\boldsymbol{E}}_R\text{diag}\{\boldsymbol{h}_F^{\#}\}\boldsymbol{x}_F^{\#}$, and $e^{j\theta_{R,m}}\text{diag}\{\boldsymbol{c}_{F,2}\}\text{diag}\{\boldsymbol{h}_F^{\#}\}\tilde{\boldsymbol{E}}_T\boldsymbol{x}_F^{\#}$ in (18) can also be derived. Together with the noise term $\tilde{\boldsymbol{w}}_{F,m}$, the total noise power in (18) can be approximated as

$$\sigma_{\text{Total}}^2 = 2(P_{c1}+P_{c2})P_EP_h\sigma_{F,x}^2 + (P_{c1}+P_{c2})\sigma_F^2. \quad (19)$$

The computation of P_E is given in Appendix C.

Then, we are ready to compute the CRLB for (10). The real-value vector formed by unknown parameters is defined as

$$\boldsymbol{\xi} = [\text{Re}\{\boldsymbol{h}_1'\}^T, \text{Re}\{\boldsymbol{h}_2'\}^T, \text{Im}\{\boldsymbol{h}_1'\}^T, \text{Im}\{\boldsymbol{h}_2'\}^T, \boldsymbol{\theta}'^T]^T,$$

where $\boldsymbol{\theta}' = [\theta_1', \ldots, \theta_{M-1}']^T$. According to [33], the Fisher information matrix $\boldsymbol{I}(\boldsymbol{\xi})$ is given by

$$\boldsymbol{I}(\boldsymbol{\xi}) = \frac{2}{\sigma_{\text{Total}}^2}\sum_{k=0}^{NM}\text{Re}\left\{\frac{\partial\mu[k]}{\partial\boldsymbol{\xi}}\left(\frac{\partial\mu[k]}{\partial\boldsymbol{\xi}}\right)^H\right\}, \quad (20)$$

where $\mu[k]$ denotes the kth element in $\boldsymbol{\mu}$, which is defined as

$$\boldsymbol{\mu} = \bar{\boldsymbol{X}}\begin{bmatrix}\boldsymbol{F}_H & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{F}_H\end{bmatrix}\begin{bmatrix}\boldsymbol{h}_1'\\ \boldsymbol{h}_2'\end{bmatrix}.$$

Since $\left(\text{E}[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^H] - \boldsymbol{I}^{-1}(\boldsymbol{\xi})\right)$ is positive semi-definite for unbiased estimation, the CRLB of the total MSE of $\boldsymbol{h}_1'$ and $\boldsymbol{h}_2'$ can be derived as

$$\text{E}[||\hat{\boldsymbol{h}}_1' - \boldsymbol{h}_1'||^2 + ||\hat{\boldsymbol{h}}_2' - \boldsymbol{h}_2'||^2] \geq \text{tr}\left([\boldsymbol{I}^{-1}(\boldsymbol{\xi})]_{1:4L_H,1:4L_H}\right), \quad (21)$$

where $[\boldsymbol{I}^{-1}(\boldsymbol{\xi})]_{1:4L_H,1:4L_H}$ represents the sub-matrix formed by the first $4L_H$ rows and the first $4L_H$ columns of $\boldsymbol{I}^{-1}(\boldsymbol{\xi})$, and $L_H = L_c + L_h - 1$.

IV. Stage-2 Channel Estimation

To facilitate the subsequent compensation of WB IQ imbalance in channel equalization, we need to further separate $\boldsymbol{h}$ and $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ from $\hat{\boldsymbol{h}}_1'$, $\hat{\boldsymbol{h}}_2'$ in Stage-2 CE. According to $\boldsymbol{h}_1' = (\boldsymbol{h}*\boldsymbol{c}_1)e^{-j\theta_0}$, $\boldsymbol{h}_2' = (\boldsymbol{h}^**\boldsymbol{c}_2)e^{j\theta_0}$, it can be seen that the estimation of $\boldsymbol{h}$, $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ is essentially a deconvolution problem. Since a deconvolution problem could have multiple possible solutions, it is difficult to acquire the accurate estimation of $\boldsymbol{h}$, $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, even if we have the perfect estimation of $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ without the phase ambiguity θ_0. To solve this problem, we propose a method based on a gradient descent deconvolution method [37]. The basic idea is to use the priori information of LPF parameters to acquire a coarse initial estimation of $\boldsymbol{h}$, $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ from $\hat{\boldsymbol{h}}_1'$, $\hat{\boldsymbol{h}}_2'$, and then gradient descent is employed to approach the true value from the initial estimation. In this way, the gradient descent could converge to an estimation value close to the true value.

A. Estimation Ambiguity

Before introducing the Stage-2 CE method, we should consider the ambiguity of the estimation of $\boldsymbol{h}$, $\boldsymbol{c}_1$, $\boldsymbol{c}_2$. There is another ambiguity of scaling factor besides the phase ambiguity θ_0. Specifically, the solution of our deconvolution problem can be written as

$$\boldsymbol{h}' = \alpha e^{-j\theta_0}\boldsymbol{h}, \quad \boldsymbol{c}_1' = \alpha^{-1}\boldsymbol{c}_1, \quad \boldsymbol{c}_2' = (\alpha^*)^{-1}\boldsymbol{c}_2,$$

where α is an arbitrary non-zero complex scaling factor. It can be seen that we have $\boldsymbol{h}_1' = \boldsymbol{h}' * \boldsymbol{c}_1'$, $\boldsymbol{h}_2' = \boldsymbol{h}'^* * \boldsymbol{c}_2'$ with the arbitrary scaling factor α. Therefore, instead of $\boldsymbol{h}$, $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, we consider $\boldsymbol{h}'$, $\boldsymbol{c}_1'$, $\boldsymbol{c}_2'$ as the CSI to be estimated, which are not determined values, but a set of values with an arbitrary scaling factor α and a phase ambiguity θ_0.

Furthermore, the phase ambiguity θ_0 can be relaxed to an arbitrary phase because it will be compensated as part of PN in the equalization procedure (see Section V-C). So, we can further rewrite $\boldsymbol{h}'$ as

$$\boldsymbol{h}' = \alpha e^{-j\theta}\boldsymbol{h},$$

where θ is allowed to be an arbitrary phase. Therefore, the $\boldsymbol{h}'$, $\boldsymbol{c}_1'$, $\boldsymbol{c}_2'$ with an arbitrary pair of (α, θ) can be considered as the true value of the Stage-2 CE.

B. Proposed Stage-2 CE Method

Firstly, we perform a coarse initial estimation of $\boldsymbol{h}'$, $\boldsymbol{c}_1'$, $\boldsymbol{c}_2'$, where we acquire an approximation of $\boldsymbol{c}_1'$ via the priori information of LPF parameters, and then derive the estimation

of $\boldsymbol{h}'$ and $\boldsymbol{c}_2'$ according to their frequency-domain relationships $\boldsymbol{h}_{F,1}' = (\boldsymbol{F}_h\boldsymbol{h}') \odot (\boldsymbol{F}_c\boldsymbol{c}_1')$ and $\boldsymbol{h}_{F,2}' = (\boldsymbol{F}_h\boldsymbol{h}')^{\#} \odot (\boldsymbol{F}_c\boldsymbol{c}_2')$, respectively.

From the definition of $c_1[n]$ in (3), we have

$$c_1[n] = \frac{1}{2}(g_I[n] + \gamma e^{-j\phi} g_Q[n]), \quad n = 0, \ldots, L_c - 1,$$

where $g_I[n]$, $g_Q[n]$ are the equivalent baseband impulse response of the LPFs. Although the accurate value of $g_I[n]$, $g_Q[n]$ may fluctuate, the approximate values of $g_I[n]$, $g_Q[n]$ are accessible based on the LPF design parameters. Hence, we assume that the approximation of $g_I[n]$ and $g_Q[n]$ can be obtained according to the priori information of LPF parameters, and we denote this approximation as $\hat{g}[n]$. Then, given that γ is close to 1 and ϕ is small, we set the initial estimation of $c_1[n]$ to be $\hat{c}_1[n] = \hat{g}[n]$, and its vector form is denoted as $\hat{\boldsymbol{c}}_1'$.

Then, we convert the time-domain total CSI into frequency domain as $\hat{\boldsymbol{h}}_{F,1}^{'} = \boldsymbol{F}_H\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_{F,2}^{'} = \boldsymbol{F}_H\hat{\boldsymbol{h}}_2^{'}$. According to the frequency-domain relationship $\boldsymbol{h}_{F,1}' = (\boldsymbol{F}_h\boldsymbol{h}') \odot (\boldsymbol{F}_c\boldsymbol{c}_1')$, the initial estimation of $\boldsymbol{h}_1'$ based on LS estimation is given by

$$\hat{\boldsymbol{h}}^{'} = \frac{1}{N}\boldsymbol{F}_h^H \left(\hat{\boldsymbol{h}}_{F,1}'./(\boldsymbol{F}_c\hat{\boldsymbol{c}}_1')\right), \tag{22}$$

where $./$ represents element-wise division. On the other hand, according to $\boldsymbol{h}_{F,2}' = (\boldsymbol{F}_h\boldsymbol{h}')^{\#} \odot (\boldsymbol{F}_c\boldsymbol{c}_2')$, the initial value of $\boldsymbol{c}_2'$ based on linear minimum mean square error (LMMSE) estimation is given by

$$\hat{\boldsymbol{c}}_2' = \frac{1}{N}\boldsymbol{F}_c^H \left(\hat{\boldsymbol{h}}_{F,2}' \odot \boldsymbol{g}_F\right), \tag{23}$$

where the kth element of $\boldsymbol{g}_F$ is

$$g_F[k] = \frac{\hat{h}_F'[N-k]}{|\hat{h}_F'[N-k]|^2 + \delta_h}, \quad k = 0, \ldots, N-1,$$

where $\hat{h}_F'[N-k]$ denotes the $(N-k)$th element of $\hat{\boldsymbol{h}}_F^{'} = \boldsymbol{F}_h\hat{\boldsymbol{h}}^{'}$. Since the first- and second-order statistic properties required by LMMSE estimation might be unavailable, we set $\delta_h = \frac{1}{10N}||\hat{\boldsymbol{h}}_F^{'}||^2$. The introduce of δ_h is to avoid noise amplification for small $|\hat{h}_F'[N-k]|$.

Next, we improve the initial estimation $\hat{\boldsymbol{h}}^{'}$, $\hat{\boldsymbol{c}}_1'$, $\hat{\boldsymbol{c}}_2'$ via gradient descent. The objective function $J(\hat{\boldsymbol{h}}^{'}, \hat{\boldsymbol{c}}_1', \hat{\boldsymbol{c}}_2')$ to be minimized is defined as

$$\begin{aligned} J(\hat{\boldsymbol{h}}^{'}, \hat{\boldsymbol{c}}_1', \hat{\boldsymbol{c}}_2') = &||\hat{\boldsymbol{h}}_{F,1}^{'} - (\boldsymbol{F}_h\hat{\boldsymbol{h}}^{'}) \odot (\boldsymbol{F}_c\hat{\boldsymbol{c}}_1')||^2 \\ &+ ||\hat{\boldsymbol{h}}_{F,2}^{'} - (\boldsymbol{F}_h\hat{\boldsymbol{h}}^{'})^{\#} \odot (\boldsymbol{F}_c\hat{\boldsymbol{c}}_2')||^2. \end{aligned} \tag{24}$$

The conjugate gradient of $J(\hat{\boldsymbol{h}}^{'}, \hat{\boldsymbol{c}}_1', \hat{\boldsymbol{c}}_2')$ with respect to $\hat{\boldsymbol{h}}^{'}$, $\hat{\boldsymbol{c}}_1'$, $\hat{\boldsymbol{c}}_2'$ can be derived from (24) as

$$\begin{aligned} \frac{\partial J}{\partial \hat{\boldsymbol{h}}^{'*}} = &-(\text{diag}\{\boldsymbol{F}_c\hat{\boldsymbol{c}}_1'\}\boldsymbol{F}_h)^H(\hat{\boldsymbol{h}}_{F,1}^{'} - (\boldsymbol{F}_h\hat{\boldsymbol{h}}^{'}) \odot (\boldsymbol{F}_c\hat{\boldsymbol{c}}_1')) \\ &- (\text{diag}\{\boldsymbol{F}_c\hat{\boldsymbol{c}}_2'\}\boldsymbol{F}_h)^T(\hat{\boldsymbol{h}}_{F,2}^{'} - (\boldsymbol{F}_h\hat{\boldsymbol{h}}^{'})^{\#} \odot (\boldsymbol{F}_c\hat{\boldsymbol{c}}_2'))^*, \end{aligned} \tag{25}$$

$$\frac{\partial J}{\partial \hat{\boldsymbol{c}}_1^{'*}} = -(\text{diag}\{\boldsymbol{F}_h\hat{\boldsymbol{h}}^{'}\}\boldsymbol{F}_c)^H(\hat{\boldsymbol{h}}_{F,1}^{'} - (\boldsymbol{F}_h\hat{\boldsymbol{h}}^{'}) \odot (\boldsymbol{F}_c\hat{\boldsymbol{c}}_1')), \tag{26}$$

$$\frac{\partial J}{\partial \hat{\boldsymbol{c}}_2^{'*}} = -(\text{diag}\{\boldsymbol{F}_h\hat{\boldsymbol{h}}^{'*}\}\boldsymbol{F}_c)^H(\hat{\boldsymbol{h}}_{F,2}^{'} - (\boldsymbol{F}_h\hat{\boldsymbol{h}}^{'})^{\#} \odot (\boldsymbol{F}_c\hat{\boldsymbol{c}}_2')). \tag{27}$$

Therefore, the descent direction is given by

$$\boldsymbol{t} = -\left[\left(\frac{\partial J}{\partial \hat{\boldsymbol{h}}^{'*}}\right)^T, \left(\frac{\partial J}{\partial \hat{\boldsymbol{c}}_1^{'*}}\right)^T, \left(\frac{\partial J}{\partial \hat{\boldsymbol{c}}_2^{'*}}\right)^T\right]^T.$$

Then, a common gradient descent algorithm can be applied to minimize the objective function. Specifically, for each iteration in the gradient descent, the initial point is denoted as $\boldsymbol{t}_0 = [\hat{\boldsymbol{h}}^{'T}, \hat{\boldsymbol{c}}_1^{'T}, \hat{\boldsymbol{c}}_2^{'T}]^T$. We compute the corresponding descent direction $\boldsymbol{t}$ at $\boldsymbol{t}_0$, and search a parameter $c > 0$ to minimize $J(\boldsymbol{t}_0 + c\boldsymbol{t})$ so that $(\boldsymbol{t}_0 + c\boldsymbol{t})$ is set to be the initial point for the next iteration. The gradient descent algorithm stops when the decrease of $J(\hat{\boldsymbol{h}}^{'}, \hat{\boldsymbol{c}}_1', \hat{\boldsymbol{c}}_2')$ is less than a threshold δ_t or $||\boldsymbol{t}||_2$ is smaller than a threshold ϵ_t.

As a summary of Section III and IV, the entire proposed two-stage CE method is described in Algorithm 1, which provides a novel CE procedure to derive $\hat{\boldsymbol{h}}^{'}, \hat{\boldsymbol{c}}_1', \hat{\boldsymbol{c}}_2'$ separately under the coexistence of WB IQ imbalance and strong PN in THz circuits.

Algorithm 1 Two-Stage CE Algorithm for Signal Model (8)

Require: $\boldsymbol{x}_{F,m}$, $\boldsymbol{y}_{F,m}$, $m = 0, \ldots, M-1$, $\hat{g}[n]$;
1: **Stage-1 CE**
2: Derive the initial estimation $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ of the total CSI according to (12);
3: Set $\hat{\theta}_0' = 0$, $\sigma_{\theta,0}^2 = 0$;
4: **for** $m = 1 : (M-1)$ **do**
5: Compute $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ according to (14), where $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ are used to replace $\boldsymbol{h}_1'$, $\boldsymbol{h}_2'$;
6: Employ EKF to track $\hat{\theta}_m'$, $\sigma_{\theta,m}^2$ for pilot block m according to (15);
7: **end for**
8: Construct $\hat{\boldsymbol{X}}$ according to (11), where θ_m' are replaced by $\hat{\theta}_m'$, $m = 0, \ldots, M-1$;
9: Derive the estimation $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ according to (16);
10: **Stage-2 CE**
11: Set the initial estimation of $\boldsymbol{c}_1'$ to $\hat{c}_1'[n] = \hat{g}[n]$;
12: Use $\hat{\boldsymbol{h}}_1^{'}$, $\hat{\boldsymbol{h}}_2^{'}$ to derive the initial estimation of $\boldsymbol{h}'$, $\boldsymbol{c}_2'$ according to (22), (23);
13: Set $\boldsymbol{t}_0 = [\hat{\boldsymbol{h}}^{'T}, \hat{\boldsymbol{c}}_1^{'T}, \hat{\boldsymbol{c}}_2^{'T}]^T$;
14: **repeat**
15: Compute the conjugate gradient of $J(\hat{\boldsymbol{h}}^{'}, \hat{\boldsymbol{c}}_1', \hat{\boldsymbol{c}}_2')$ at $\boldsymbol{t}_0$ according to (25)-(27), whose opposite direction gives the descent direction $\boldsymbol{t}$;
16: Search step length parameter c to minimize $J(\boldsymbol{t}_0 + c\boldsymbol{t})$, and update $\boldsymbol{t}_0 = \boldsymbol{t}_0 + c\boldsymbol{t}$;
17: **until** $||\boldsymbol{t}||_2 < 10^{-3}$
18: Derive the final estimation $[\hat{\boldsymbol{h}}^{'T}, \hat{\boldsymbol{c}}_1^{'T}, \hat{\boldsymbol{c}}_2^{'T}]^T = \boldsymbol{t}_0$;

V. Channel Equalization

In this section, we propose a low-complexity channel equalization method which considers the coexistence of WB IQ

imbalance and strong PN in THz communications. The perfect CSI $\boldsymbol{h}'$, $\boldsymbol{c}'_1$, $\boldsymbol{c}'_2$ is assumed to be used in the equalization. As discussed in Section IV-A, the scaling factor ambiguity and phase ambiguity in $\boldsymbol{h}'$, $\boldsymbol{c}'_1$, $\boldsymbol{c}'_2$ should be considered, so we have

$$\boldsymbol{h}' = \alpha e^{-j\theta}\boldsymbol{h}, \quad \boldsymbol{c}'_1 = \alpha^{-1}\boldsymbol{c}_1 \ \boldsymbol{c}'_2 = (\alpha^*)^{-1}\boldsymbol{c}_2, \tag{28}$$

where α and θ are two unknown parameters. Meanwhile, the ZP SC block is adopted for data transmission to ensure that linear convolution can be converted to circular convolution under PN as demonstrated in Section II-C.

A. IQ Imbalance Compensation

Firstly, the WB IQ imbalance is compensated. According to the frequency-domain signal model (8), the frequency-domain Rx signal $\boldsymbol{y}_{F,h}$ without IQ imbalance is defined as

$$\boldsymbol{y}_{F,h} = \boldsymbol{P}_R \text{diag}\{\boldsymbol{h}_F\}\boldsymbol{P}_T\boldsymbol{x}_F. \tag{29}$$

We need to recover $\boldsymbol{y}_{F,h}$ from $\boldsymbol{y}_F$ to compensate the WB IQ imbalance.

According to $\boldsymbol{P}_T = \frac{1}{N}\boldsymbol{F}\text{diag}\{\boldsymbol{p}_T\}\boldsymbol{F}^H$, for an arbitrary N-dimension vector $\boldsymbol{x}$, we have

$$\begin{aligned}(\boldsymbol{P}_T\boldsymbol{x})^{\#} &= \left(\boldsymbol{F}\left(\text{diag}\{\boldsymbol{p}_T\}\frac{1}{N}\boldsymbol{F}^H\boldsymbol{x}\right)\right)^{\#} \\ &= \tilde{\boldsymbol{P}}_T\boldsymbol{F}\left(\frac{1}{N}\boldsymbol{F}^H\boldsymbol{x}\right)^* = \tilde{\boldsymbol{P}}_T\boldsymbol{x}^{\#},\end{aligned}$$

where the property of DFT $\boldsymbol{F}\boldsymbol{x}^* = (\boldsymbol{F}\boldsymbol{x})^{\#}$ is used, and similarly, we have $(\boldsymbol{P}_R\boldsymbol{x})^{\#} = \tilde{\boldsymbol{P}}_R\boldsymbol{x}^{\#}$. Therefore, $\boldsymbol{y}^{\#}_{F,h}$ can be written as

$$\begin{aligned}\boldsymbol{y}^{\#}_{F,h} &= \tilde{\boldsymbol{P}}_R(\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{P}_T\boldsymbol{x}_F)^{\#} = \tilde{\boldsymbol{P}}_R\text{diag}\{\boldsymbol{h}^{\#}_F\}(\boldsymbol{P}_T\boldsymbol{x}_F)^{\#} \\ &= \tilde{\boldsymbol{P}}_R\text{diag}\{\boldsymbol{h}^{\#}_F\}\tilde{\boldsymbol{P}}_T\boldsymbol{x}^{\#}_F.\end{aligned} \tag{30}$$

Combining (8), (29) and (30), the frequency-domain signal model can be simplified as

$$\boldsymbol{y}_F = \text{diag}\{\boldsymbol{c}_{F,1}\}\boldsymbol{y}_{F,h} + \text{diag}\{\boldsymbol{c}_{F,2}\}\boldsymbol{y}^{\#}_{F,h} + \tilde{\boldsymbol{w}}_F.$$

However, we need to take the ambiguity of CSI into account. Defining the $\boldsymbol{y}_{F,h}$ with ambiguity as $\boldsymbol{y}'_{F,h} = \alpha\boldsymbol{y}_{F,h}$, we arrive at

$$\boldsymbol{y}_F = \text{diag}\{\boldsymbol{c}'_{F,1}\}\boldsymbol{y}'_{F,h} + \text{diag}\{\boldsymbol{c}'_{F,2}\}\boldsymbol{y}'^{\#}_{F,h} + \tilde{\boldsymbol{w}}_F, \tag{31}$$

where $\boldsymbol{c}'_{F,1} = \alpha^{-1}\boldsymbol{c}_{F,1} = \boldsymbol{F}_c\boldsymbol{c}'_1$, $\boldsymbol{c}'_{F,2} = (\alpha^*)^{-1}\boldsymbol{c}_{F,2} = \boldsymbol{F}_c\boldsymbol{c}'_2$.

Then, we use LMMSE criterion to recover $\boldsymbol{y}'_{F,h}$ from (31). It can be seen that the IQ imbalance introduces a conjugate symmetry interference term $\boldsymbol{y}'^{\#}_{F,h}$ in frequency domain, so we need to treat the two scenarios of $k = 0, \frac{N}{2}$ and $k \neq 0, \frac{N}{2}$ respectively.

For $k = 1, \ldots, \frac{N}{2} - 1$, we have

$$\begin{bmatrix} y_F[k] \\ y^*_F[N-k] \end{bmatrix} = \begin{bmatrix} c'_{F,1}[k] & c'_{F,2}[k] \\ c'^*_{F,2}[N-k] & c'^*_{F,1}[N-k] \end{bmatrix} \cdot \begin{bmatrix} y'_{F,h}[k] \\ y'^*_{F,h}[N-k] \end{bmatrix} + \begin{bmatrix} \tilde{w}_F[k] \\ \tilde{w}^*_F[N-k] \end{bmatrix}.$$

Based on the LMMSE criterion, $y'_{F,h}[k]$ and $y'^*_{F,h}[N-k]$ are recovered by

$$\begin{bmatrix} \hat{y}'_{F,h}[k] \\ \hat{y}'^*_{F,h}[N-k] \end{bmatrix} = \boldsymbol{C}^H_k\left(\boldsymbol{C}_k\boldsymbol{C}^H_k + \frac{\tilde{\sigma}^2_F}{\sigma^2_{F,y}}\boldsymbol{I}_2\right)^{-1} \times \begin{bmatrix} y_F[k] \\ y^*_F[N-k] \end{bmatrix}, \tag{32}$$

where we set $\tilde{\sigma}^2_F \approx N\sigma^2$ and $\sigma^2_{F,y} \approx \frac{1}{N}||\boldsymbol{y}_F||^2 - \tilde{\sigma}^2_F$, and $\boldsymbol{C}_k$ is defined as

$$\boldsymbol{C}_k = \begin{bmatrix} c'_{F,1}[k] & c'_{F,2}[k] \\ c'^*_{F,2}[N-k] & c'^*_{F,1}[N-k] \end{bmatrix}.$$

On the other hand, for $k = 0, \frac{N}{2}$, we have

$$y_F[k] = c'_{F,1}[k]y'_{F,h}[k] + c'_{F,2}[k]y'^*_{F,h}[k] + \tilde{w}_F[k],$$

which can be rewritten as

$$\begin{aligned}&(c'^*_{F,1}[k]y_F[k] - c'_{F,2}[k]y^*_F[k]) \\ &\quad = (|c'_{F,1}[k]|^2 - |c'_{F,2}[k]|^2) \\ &\qquad \cdot y'_{F,h}[k] + (c'^*_{F,1}[k]\tilde{w}_F[k] - c'_{F,2}[k]\tilde{w}^*_F[k]).\end{aligned}$$

Based on the LMMSE criterion, $y'_{F,h}[k]$ is calculated by

$$\hat{y}'_{F,h}[k] = \frac{c^*_k}{|c_k|^2 + \frac{\tilde{\sigma}^2_F}{\sigma^2_{F,y}}}(c'^*_{F,1}[k]y_F[k] - c'_{F,2}[k]y^*_F[k]), \tag{33}$$

where $c_k = |c'_{F,1}[k]|^2 - |c'_{F,2}[k]|^2$. It can be seen that all the elements $\hat{y}'_{F,h}[k]$ for $k = 0, \ldots, N-1$ are derived, which form the recovered vector $\hat{\boldsymbol{y}}'_{F,h} = [\hat{y}'_{F,h}[0], \ldots, \hat{y}'_{F,h}[N-1]]^T$.

B. CIR Equalization

Next, we equalize the CIR $\boldsymbol{h}'$ in $\hat{\boldsymbol{y}}'_{F,h}$, and the expected Rx signal $\boldsymbol{y}'_{F,p}$ after equalization is defined as $\boldsymbol{y}'_{F,p} = e^{j\theta}\boldsymbol{P}_R\boldsymbol{P}_T\boldsymbol{x}_F$. Still, we employ the LMMSE criterion to recover $\boldsymbol{y}'_{F,p}$ as

$$\hat{\boldsymbol{y}}'_{F,p} = \boldsymbol{A}\hat{\boldsymbol{y}}'_{F,h}, \tag{34}$$

where $\boldsymbol{A}$ is the coefficient matrix to be determined.

We denote that $\hat{\boldsymbol{y}}'_{F,h} = \boldsymbol{y}'_{F,h} + \tilde{\boldsymbol{w}}_{F,h}$. Since $|c'_{F,1}[k]|$ is close to 1 and $|c'_{F,2}[k]|$ is small, we assume that the noise power after IQ imbalance compensation is unchanged. So, we still consider the noise power as $\tilde{\sigma}^2_F \approx N\sigma^2$ in the noise term $\tilde{\boldsymbol{w}}_{F,h}$. Besides, the time-domain data symbols in $\boldsymbol{x}$ is considered to be independent of each other with average power σ^2_x, i.e. $\text{E}[\boldsymbol{x}\boldsymbol{x}^H] = \sigma^2_x\boldsymbol{I}_N$, leading to $\text{E}[\boldsymbol{x}_F\boldsymbol{x}^H_F] = \sigma^2_{F,x}\boldsymbol{I}_N$ with $\sigma^2_{F,x} = N\sigma^2_x$. Then, $\boldsymbol{A}$ is derived as

$$\begin{aligned}\boldsymbol{A} &= \arg\min_{\boldsymbol{A}} \text{E}\left[||\boldsymbol{A}\hat{\boldsymbol{y}}'_{F,h} - \boldsymbol{y}'_{F,p}||^2\right] \\ &= \arg\min_{\boldsymbol{A}} \text{E}\big[||\boldsymbol{A}(\alpha\boldsymbol{P}_R\text{diag}\{\boldsymbol{h}_F\}\boldsymbol{P}_T\boldsymbol{x}_F + \tilde{\boldsymbol{w}}_{F,h}) \\ &\quad - e^{j\theta}\boldsymbol{P}_R\boldsymbol{P}_T\boldsymbol{x}_F||^2\big].\end{aligned} \tag{35}$$

Although (35) has a relatively complicated form, $\boldsymbol{A}$ can be approximated as $\boldsymbol{A} \approx \text{diag}\{\boldsymbol{e}_F\}$, where the kth element of $\boldsymbol{e}_F$ is given by

$$e_F[k] = \frac{\sigma^2_{F,x}h'^*_F[k]}{\tilde{\sigma}^2_F + \sigma^2_{F,x}|h'_F[k]|^2}, \quad k = 0, \ldots, N-1, \tag{36}$$

which is independent of the PN. The derivation of $\boldsymbol{A}$ is given in Appendix D.

C. Phase Noise Compensation

After the CIR equalization, a time-domain decision-directed PN compensation method is utilized to jointly detect the data symbols and compensate the PN with the aid of phase-tracking (PT) pilots. Note that $\boldsymbol{y}'_{F,p}$ can be expressed as

$$\boldsymbol{y}'_{F,p} = e^{j\theta}\boldsymbol{P}_R\boldsymbol{P}_T\boldsymbol{x}_F = e^{j\theta}\left(\frac{1}{N}\boldsymbol{F}\text{diag}\{\boldsymbol{p}_R\}\boldsymbol{F}^H\right) \cdot \left(\frac{1}{N}\boldsymbol{F}\text{diag}\{\boldsymbol{p}_T\}\boldsymbol{F}^H\right)\boldsymbol{x}_F = \boldsymbol{F}\text{diag}\{\boldsymbol{p}'\}\boldsymbol{x},$$

where $\boldsymbol{p}' = [e^{j\theta'[0]}, \ldots, e^{j\theta'[N-1]}]^T$ with $\theta'[n] = \theta + \theta_T[n] - \theta_R[n]$. Replacing $\boldsymbol{y}'_{F,p}$ by $\hat{\boldsymbol{y}}'_{F,p}$, the corresponding time-domain signal after IDFT is

$$\hat{\boldsymbol{y}}_p = \frac{1}{N}\boldsymbol{F}^H\hat{\boldsymbol{y}}'_{F,p} = \text{diag}\{\boldsymbol{p}'\}\boldsymbol{x} + \boldsymbol{w}_p,$$

where $\boldsymbol{w}_p$ denotes the noise term in $\hat{\boldsymbol{y}}_p$. It can be seen that the phase of the time-domain Tx vector $\boldsymbol{x}$ is rotated by $\boldsymbol{p}'$, and the phase ambiguity θ is contained in $\boldsymbol{p}'$. Therefore, to recover $\boldsymbol{x}$, we need to estimate the phase rotation $\theta'[n]$ for $n = 0, \ldots, N-1$.

We assume that there are N_{pt} uniformly inserted PT pilots in $\boldsymbol{x}$ with interval $T = \frac{N}{N_{pt}}$. Hence, the indices of all PT pilots can be denoted as $\{n_0 + gT | g = 0, \ldots, N_{pt} - 1\}$, where $n_0 < T$ denotes the index of the first PT pilot. Then, the initial estimation of $\theta'[n]$ is given by

$$\begin{aligned}\hat{\theta}'[n] &= \arg\min_{\theta'[n_0+gT]} |x[n_0+gT]e^{j\theta'[n_0+gT]} - \hat{y}_p[n_0+gT]|^2 \\ &= \angle(x^*[n_0+gT]\hat{y}_p[n_0+gT]), \\ n &= gT, \ldots, (g+1)T-1, \end{aligned} \tag{37}$$

which form the initial estimation of $\boldsymbol{p}'$ as $\hat{\boldsymbol{p}}' = [e^{j\hat{\theta}'[0]}, \ldots, e^{j\hat{\theta}'[N-1]}]^T$. Alternatively, a CPE of the entire data SC block can be estimated via PT pilots as the initial $\hat{\theta}'[n]$ in low-SNR scenario. Then, $\boldsymbol{x}$ is recovered by

$$\hat{\boldsymbol{x}} = \text{diag}\{\hat{\boldsymbol{p}}'^*\}\hat{\boldsymbol{y}}_p, \tag{38}$$

and we denote the Tx vector after decision as $\bar{\boldsymbol{x}}$ whose last N_{cp} symbols are set to be 0 due to the ZP SC block.

Next, the decision data symbols $\bar{\boldsymbol{x}}$ can be utilized to improve the phase estimation $\hat{\theta}'[n]$. We split $\hat{\theta}'[0], \ldots, \hat{\theta}'[N-1]$ into N_θ short segments with length of $T_\theta = \frac{N}{N_\theta}$, where the $\hat{\theta}'[n]$ in one segment is considered as constant. Therefore, $\hat{\theta}'[n]$ for the gth segment is updated by

$$\begin{aligned}\hat{\theta}'[n] &= \arg\min_{\theta_g} ||\bar{\boldsymbol{x}}_g e^{j\theta_g} - \hat{\boldsymbol{y}}_{p,g}||^2 \\ &= \angle(\bar{\boldsymbol{x}}_g^H\hat{\boldsymbol{y}}_{p,g}), \quad n = gT_\theta, \ldots, (g+1)T_\theta - 1, \end{aligned} \tag{39}$$

where $\bar{\boldsymbol{x}}_g = [\bar{x}[gT_\theta], \ldots, \bar{x}[(g+1)T_\theta - 1]]^T$, $\hat{\boldsymbol{y}}_{p,g} = [\hat{y}_p[gT_\theta], \ldots, \hat{y}_p[(g+1)T_\theta - 1]]^T$. It can be seen that an iteration between (38) and (39) can be performed alternately to improve the data symbol detection $\hat{\boldsymbol{x}}$ and the phase estimation $\hat{\theta}'[n]$.

As a summary for the proposed methods in Section V-A, V-B, V-C, the entire channel equalization procedure is presented in Algorithm 2. The number of iterations in PN compensation is denoted as n_{it}. The computational complexity of Algorithm 2 is estimated in terms of the number of complex multiplications required. Firstly, a FFT is required to derive $\boldsymbol{y}_F$ before the algorithm starts, which induces $\frac{N}{2}\log_2(N)$ multiplications. Then, we consider the IQ imbalance compensation step. In (32), the multiplication and inverse of 2×2 matrix induce 26 multiplications, which are repeated for $k = 1, \ldots, \frac{N}{2} - 1$ and results in $13N - 26$ multiplications. In (33), 7 multiplications are repeated for $k = 0, \frac{N}{2}$, leading to 14 multiplications. For CIR equalization, according to (36), the computation of $e_F[k], k = 0, \ldots, N-1$ requires $4N$ multiplications, and the equalization in (34) requires N multiplications. Next, a IFFT is conducted to derive $\hat{\boldsymbol{y}}_p$ and induces $\frac{N}{2}\log_2(N)$ multiplications. The initial estimation of PN in (37) needs $2N_{pt}$ multiplications. Finally, for each iteration between data symbol and PN estimation, the derivation of $\hat{\boldsymbol{x}}$ in (38) requires N multiplications, and the symbol decision needs $N|\chi|$ multiplications, and the update of $\hat{\theta}'[n]$ in (39) induces $N + N_\theta$ multiplications, where $|\chi|$ represents the size of constellation map. Therefore, the total number of complex multiplications is approximately

$$\begin{aligned}&(13N-12) + (5N) \\ &\quad + (2N_{pt} + n_{it}((|\chi|+2)N + N_\theta)) + N\log_2(N).\end{aligned}$$

Since n_{it} is small, it can be seen that the overall complexity is in the order of $O(N\log_2(N))$, which has the same order as FFT. Therefore, the total complexity of our proposed channel equalization scheme is relatively low, which is preferred in ultra-high data rate THz communications.

Algorithm 2 Channel Equalization Algorithm for Signal Model (8)

Require: $\boldsymbol{y}_F$, $\boldsymbol{h}'$, $\boldsymbol{c}'_1$, $\boldsymbol{c}'_2$;
1: Compensate IQ imbalance according to (32), (33) and derive $\hat{\boldsymbol{y}}'_{F,h}$;
2: Equalize CIR according to (34), (36) and derive $\hat{\boldsymbol{y}}'_{F,p}$;
3: Derive the initial phase estimation $\hat{\theta}'[n]$ according to (37) via PT pilots;
4: **for** $i = 1 : n_{it}$ **do**
5: Recover the data symbol vector $\boldsymbol{x}$ according to (38) and derive $\bar{\boldsymbol{x}}$ after symbol decision;
6: Update the estimation $\hat{\theta}'[n]$ according to (39) via $\bar{\boldsymbol{x}}$;
7: **end for**

VI. Simulation Results

The performance of our proposed CE and equalization method are evaluated via Monte Carlo simulation in this section. In our simulations, the SC block lengths for pilot block and data block are $N = 32$, $N = 512$, respectively, and $N_{cp} = 16$. The CIR $h[n]$ is randomly generated by complex Gaussian distribution with zero mean and exponentially decaying variance with decaying factor 0.5 [22], i.e. $h[n] \sim \mathcal{CN}(0, 0.5^n)$, and the length of CIR is set to be $L = 8$. The IQ imbalance parameters are $\gamma = 0.85$ corresponding to a 1.41dB imbalance on amplitude, and $\phi = 5^\circ$ [22]. The equivalent baseband impulse response of the LPFs $g_I[n]$, $g_Q[n]$ are $[0.1, 1, 0.01]$, $[0.01, 1, 0.2]$ [24], respectively. The maximum length of $h[n]$

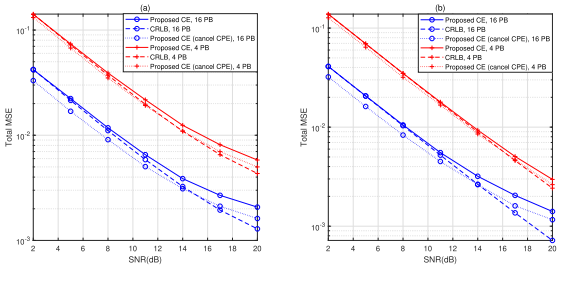

Fig. 4. MSE of Stage-1 CE. (a) $\sigma^2_{\theta_T} = \sigma^2_{\theta_R} = 5\times10^{-4}$, (b) $\sigma^2_{\theta_T} = \sigma^2_{\theta_R} = 5 \times 10^{-5}$. Since the CRLB is derived for the MSE before cancelling CPE, the MSE after cancelling CPE is slightly lower than the CRLB in low-SNR region.

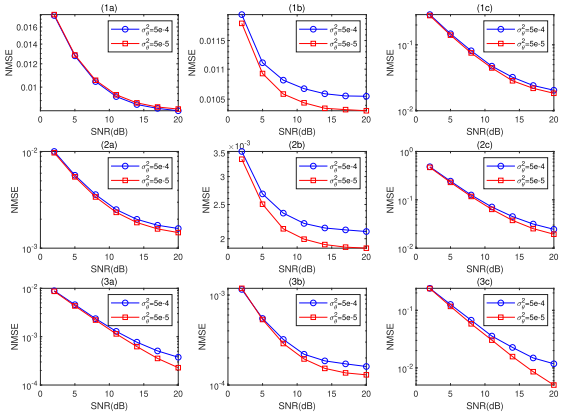

Fig. 5. NMSE of Stage-2 CE with 16 PBs, $\sigma^2_{\theta_T} = \sigma^2_{\theta_R} = \sigma^2_{\theta} = 5 \times 10^{-4}, 5 \times 10^{-5}$. (1) 1st set of $g_I[n]$, $g_Q[n]$, $\hat{g}[n]$ [24]: $[0.1, 1, 0.01]$, $[0.01, 1, 0.2]$, $[0, 1, 0]$, (2) 2nd set of $g_I[n]$, $g_Q[n]$, $\hat{g}[n]$ [24]: $[0.03, 1, 0.01]$, $[0.02, 1, 0.08]$, $[0, 1, 0]$, (3) 3rd set of $g_I[n]$, $g_Q[n]$, $\hat{g}[n]$ [23]: $[1, 0.1]$, $[1, -0.1]$, $[1, 0]$. (a) NMSE of $\hat{\boldsymbol{h}}'$, (b) NMSE of $\hat{\boldsymbol{c}}'_1$, (c) NMSE of $\hat{\boldsymbol{c}}'_2$.

and $c_1[n](c_2[n])$ are set to be $L_h = 12$, $L_c = 4$ for the two-stage CE algorithm. The PN parameters $\sigma^2_{\theta_T}$, $\sigma^2_{\theta_R}$ are set to be $\sigma^2_{\theta_T} = \sigma^2_{\theta_R} = 5 \times 10^{-4}$, 5×10^{-5}, corresponding to a strong PN (e.g. $T_s = 10^{-10}$s, $f_{3\text{dB}} = 795.8$kHz) and a relatively weak PN (e.g. $T_s = 10^{-10}$s, $f_{3\text{dB}} = 79.58$kHz). Two modulation schemes QPSK, 16QAM are simulated. The time-domain Tx power σ_x^2 is normalized to 1 for both pilot transmission and data transmission, and the signal-to-noise ratio (SNR) is defined as SNR $= \frac{1}{\sigma^2}$.

A. Stage-1 Channel Estimation

Firstly, we investigate the total MSE of $\hat{\boldsymbol{h}}'_1$ and $\hat{\boldsymbol{h}}'_2$ in the Stage-1 CE, which is expressed as $||\hat{\boldsymbol{h}}'_1 - \boldsymbol{h}'_1||^2 + ||\hat{\boldsymbol{h}}'_2 - \boldsymbol{h}'_2||^2$. Moreover, since a CPE (or phase ambiguity) θ [1] in the estimation $\hat{\boldsymbol{h}}'_1$, $\hat{\boldsymbol{h}}'_2$ is insignificant, we also consider the total MSE after cancelling the CPE θ which is computed by $\theta = \arg\min_{\theta} ||\boldsymbol{h}_1 - e^{j\theta}\hat{\boldsymbol{h}}'_1||^2 = \angle(\hat{\boldsymbol{h}}'^H_1 \boldsymbol{h}_1)$. The estimation after cancelling CPE is $e^{j\theta}\hat{\boldsymbol{h}}'_1$ and $e^{-j\theta}\hat{\boldsymbol{h}}'_2$, and the corresponding total MSE is expressed as $||e^{j\theta}\hat{\boldsymbol{h}}'_1 - \boldsymbol{h}_1||^2 + ||e^{-j\theta}\hat{\boldsymbol{h}}'_2 - \boldsymbol{h}_2||^2$. Besides, the approximate CRLB derived in (21) is also computed. It should be noted that the CRLB is derived for the total MSE before cancelling CPE.

The Stage-1 CE performance with $q = 4$, $M = 4, 16$ pilot blocks (PBs) is evaluated in Fig. 4. In the low-SNR region, the CE performance is mainly limited by the high noise power. Since the Stage-1 CE can efficiently mitigate the AWGN by combining multiple pilot blocks via LS criterion, its performance is close to the CRLB which decreases rapidly with the growth of SNR. In this region, cancelling the CPE results in a further decrease of the total MSE, so the total MSE after cancelling CPE becomes smaller than the CRLB.

In the high-SNR region, the performance of Stage-1 CE is limited by phase tracking error resulting from the approximations adopted in Stage-1 CE. Specifically, the phase of the second pilot block is approximated as 0 in the initial estimation and the phase variation between two neighboring pilot blocks is approximated by first-order Taylor expansion in EKF. These approximations lead to small phase estimation error of each pilot block, which makes the total MSE before cancelling CPE become higher than the CRLB. In this case, cancelling the CPE can only rule out a common phase estimation error over all pilot blocks, so the total MSE after cancelling CPE is larger than the CRLB in high-SNR region. However, the total MSE of Stage-1 CE can still decrease to a relatively low level with the increase of SNR, providing a sufficiently high CE accuracy.

On the other hand, it can be seen that the total MSE decreases drastically when the number of PBs M increases from 4 to 16. This demonstrates that the proposed Stage-1 CE method can efficiently combine multiple PBs under PN, and indicates that the estimation accuracy can be significantly improved when more PBs are transmitted.

B. Stage-2 Channel Estimation

Next, the normalized MSE (NMSE) of $\hat{\boldsymbol{h}}'$, $\hat{\boldsymbol{c}}'_1$, $\hat{\boldsymbol{c}}'_2$ is studied. As indicated by (28), there are ambiguities α and θ in $\boldsymbol{h}'$, $\boldsymbol{c}'_1$, $\boldsymbol{c}'_2$, so we need to cancel the ambiguity before computing NMSE. We replace $\hat{\boldsymbol{h}}'$ by $\beta\hat{\boldsymbol{h}}'$, where $\beta = \arg\min_{\beta} ||\boldsymbol{h} - \beta\hat{\boldsymbol{h}}'||^2 = \frac{\hat{\boldsymbol{h}}'^H \boldsymbol{h}}{||\hat{\boldsymbol{h}}'||^2}$, and meanwhile $\hat{\boldsymbol{c}}'_1$, $\hat{\boldsymbol{c}}'_2$ are replaced by $\beta^{-1}\hat{\boldsymbol{c}}'_1$, $(\beta^*)^{-1}\hat{\boldsymbol{c}}'_2$. Note that there is still a phase ambiguity in $\beta^{-1}\hat{\boldsymbol{c}}'_1$, $(\beta^*)^{-1}\hat{\boldsymbol{c}}'_2$, which is computed by $\theta = \arg\min_{\theta} ||\boldsymbol{c}_1 - e^{j\theta}\beta^{-1}\hat{\boldsymbol{c}}'_1||^2 = \angle((\beta^{-1}\hat{\boldsymbol{c}}'_1)^H \boldsymbol{c}_1)$. So, $\beta^{-1}\hat{\boldsymbol{c}}'_1$, $(\beta^*)^{-1}\hat{\boldsymbol{c}}'_2$ are further replaced by $e^{j\theta}\beta^{-1}\hat{\boldsymbol{c}}'_1$, $e^{-j\theta}(\beta^*)^{-1}\hat{\boldsymbol{c}}'_2$. So, the NMSE of $\hat{\boldsymbol{h}}'$, $\hat{\boldsymbol{c}}'_1$, $\hat{\boldsymbol{c}}'_2$ is computed by

$$\frac{||\beta\hat{\boldsymbol{h}}' - \boldsymbol{h}||^2}{||\boldsymbol{h}||^2}, \quad \frac{||e^{j\theta}\beta^{-1}\hat{\boldsymbol{c}}'_1 - \boldsymbol{c}_1||^2}{||\boldsymbol{c}_1||^2}, \quad \frac{||e^{-j\theta}(\beta^*)^{-1}\hat{\boldsymbol{c}}'_2 - \boldsymbol{c}_2||^2}{||\boldsymbol{c}_2||^2}.$$

Apart from the $g_I[n]$, $g_Q[n]$ given at the beginning of this Section, the NMSE of another two sets of $g_I[n]$, $g_Q[n]$ are also investigated, as demonstrated in Fig. 5. Note that the NMSE of different sets of $g_I[n]$, $g_Q[n]$ can be very different, and a NMSE floor would appear with the increase of SNR. As mentioned in Section IV, this phenomenon mainly results from the uncertainty of the solution to a deconvolution

[1]Note that θ is the CPE of $\hat{\boldsymbol{h}}'_1$, $\hat{\boldsymbol{h}}'_2$, which is different from the CPE of SC block defined in Section III-A.

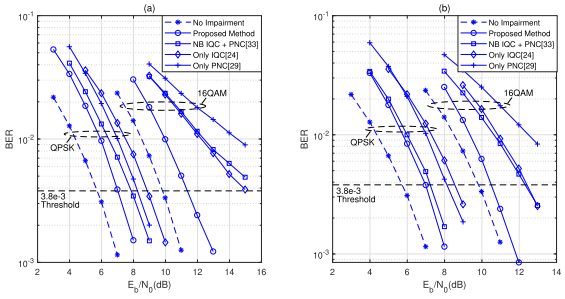

Fig. 6. BER performance with perfect CSI, $N_{pt} = 8$, $n_{it} = 4$. PNC represents PN compensation, and IQC represents IQ imbalance compensation. (a) $\sigma^2_{\theta_T} = \sigma^2_{\theta_R} = 5 \times 10^{-4}$, (b) $\sigma^2_{\theta_T} = \sigma^2_{\theta_R} = 5 \times 10^{-5}$.

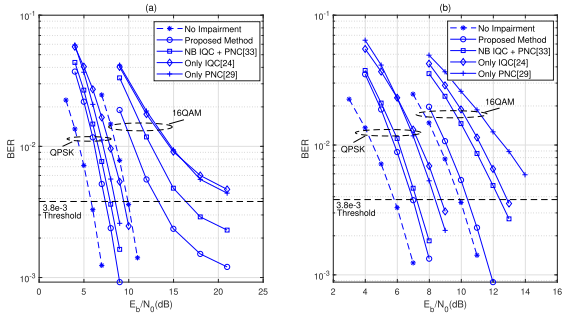

Fig. 7. BER performance with the estimated CSI, $N_{pt} = 8$, $n_{it} = 4$ and 16 PBs for CE. PNC represents PN compensation, and IQC represents IQ imbalance compensation. (a) $\sigma^2_{\theta_T} = \sigma^2_{\theta_R} = 5 \times 10^{-4}$, (b) $\sigma^2_{\theta_T} = \sigma^2_{\theta_R} = 5 \times 10^{-5}$.

problem, i.e. a deconvolution could have multiple feasible solutions (not counting the solutions caused by scaling factor ambiguity). Specifically, we could approach the true value of $\boldsymbol{h}'$, $\boldsymbol{c}'_1$, $\boldsymbol{c}'_2$ by utilizing gradient descent demixing with a coarse estimation as its start point. However, the gradient descent may converge to a feasible solution to the deconvolution which is close to the true solution, leading to an error floor of NMSE. Nevertheless, according to the results shown in Fig. 5, we could still acquire a relatively accurate CSI which can be used for channel equalization. In Section VI-C below, we choose the first set of $g_I[n]$, $g_Q[n]$ for the simulations on channel equalization, which has relatively larger NMSE among the three sets of $g_I[n]$, $g_Q[n]$.

C. Channel Equalization

The uncoded BER of our proposed channel equalization scheme is investigated under perfect CSI and our estimated CSI respectively. For comparison, the BER without RF impairments is simulated as a lower bound of BER performance. Moreover, the BER with only WB IQ imbalance compensation [24], with only phase noise compensation [29], with NB IQ imbalance and PN compensation [33] are also simulated. Since the PN compensation method in [33] is for OFDM systems, we replace its PN compensation method by our PN compensation method.

The BER performance with perfect CSI is simulated in Fig. 6. We compare the performance of the proposed equalization scheme and the state-of-the-art equalization schemes at the 3.8×10^{-3} BER threshold for the forward error correction code with 7% overhead [41]. For QPSK modulation, the proposed equalization scheme has 0.8dB, 0.3dB gain compared to the state-of-the-art schemes in the strong-PN scenario and the weak-PN scenario, respectively. Since QPSK is a low-order modulation scheme which is robust to the distortions caused by RF impairments, the performance gain of eliminating RF impairments is relatively low for QPSK. On the other hand, for high-order 16QAM modulation transmitted under higher SNR, the BER is mainly limited by RF impairments instead of AWGN. Therefore, the performance gain of our equalization scheme becomes prominent and increases to 3.7dB, 1.8dB in the two scenarios with strong and relatively weak PN, respectively, which indicates that our proposed equalization scheme can improve the BER performance significantly for high-order modulation.

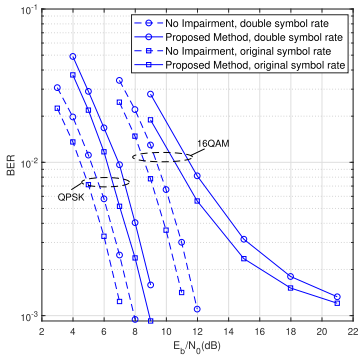

Fig. 8. BER performance comparison between the original symbol rate and double symbol rate, with the estimated CSI, $N_{pt} = 8$, $n_{it} = 4$ and 16 PBs for CE.

The BER performance with the estimated CSI is simulated in Fig. 7. Because of the imperfection of CSI, the BER performance of our proposed method degrades. Nevertheless, at the 3.8×10^{-3} BER threshold, our proposed equalization scheme obtains 0.5dB, 0.2dB performance gain for QPSK transmission in the two scenarios, which increase to 3.2dB, 2.0dB for 16QAM transmission. This verifies that the CSI derived by the proposed two-stage CE can be used for our equalization scheme to detect data symbols efficiently.

Furthermore, the BER performance comparison between the original symbol rate and double symbol rate is shown in Fig. 8. To model the CIR with double symbol rate, we use sinc-interpolation to derive the double-length CIR $h[n]$. And the sample rate of $g_I[n], g_Q[n]$ is doubled by inserting zeros, which ensures that the new frequency responses of $g_I[n], g_Q[n]$ in the lower half bandwidth are same as their frequency responses in the original bandwidth. Then, a random Gaussian disturbance is added to both $g_I[n]$ and $g_Q[n]$, so $g_I[n], g_Q[n]$ are modeled as $[0.01, 0, 1, 0, 0.01] + \boldsymbol{n}_I^T$, $[0.01, 0, 1, 0, 0.2] + \boldsymbol{n}_Q^T$, respectively, where $\boldsymbol{n}_I, \boldsymbol{n}_Q \in \mathbb{R}^{5\times 1}$ whose elements are i.i.d. Gaussian random variables with distribution $N(0, (0.1)^2)$.

It is evident that double symbol rate degrades the BER performance of our equalization for both QPSK and 16QAM

due to the more severe inter-symbol interference (ISI) caused by the longer CIR and WB IQ imbalance parameters. Even for the ideal case without RF impairments, the performance also deteriorates because the longer CIR would increase the ISI level as well. On the other hand, the performance loss in our proposed method becomes only slightly worse than the ideal scenario for 16QAM modulation, which indicates that the more severe IQ imbalance under double symbol rate can be mitigated efficiently by using our proposed equalization.

VII. Conclusion

In this paper, we investigate the channel estimation and equalization for Terahertz receiver under WB IQ imbalance and PN. Since the CIR and the WB IQ imbalance parameters need to be estimated separately, a two-stage CE method is proposed, which manages to estimate the CIR and WB IQ imbalance parameters under PN for the first time. In Stage-1 CE, the total CSI is estimated firstly. We employ EKF to track the PN of each pilot block so that we can combine multiple consecutive pilot blocks to improve the estimation accuracy. In Stage-2 CE, the CIR and IQ imbalance parameters are separated from the total CSI via a deconvolution approach. Finally, based on the CSI derived in the Stage-2 CE, we propose a low-complexity channel equalization scheme, which compensates the WB IQ imbalance and PN, and equalizes the CIR as well.

Appendix A

Derivation of (3). The corresponding bandpass signal $r_{RF}(t)$ of $r(t)$ is expressed as

$$\begin{aligned} r_{RF}(t) &= 2\mathrm{Re}\{r(t)e^{j2\pi f_c t}\} \\ &= 2(r_I(t)cos(2\pi f_c t) - r_Q(t)sin(2\pi f_c t)), \end{aligned} \quad (40)$$

where $r_I(t) = \mathrm{Re}\{r(t)\}$, $r_Q(t) = \mathrm{Im}\{r(t)\}$. The in-phase, quadrature signals are given by

$$y_I(t) = (r_{RF}(t)cos(2\pi f_c t + \theta_R(t))) * g_I(t), \quad (41)$$

$$y_Q(t) = (-r_{RF}(t)\gamma sin(2\pi f_c t + \theta_R(t) + \phi)) * g_Q(t), \quad (42)$$

respectively. Note that the high-frequency components in $y_I(t)$, $y_Q(t)$ would be filtered by $g_I(t)$, $g_Q(t)$. Combining (40)-(42) and $y(t) = y_I(t) + jy_Q(t)$, we arrive at

$$\begin{aligned} y(t) = (r(t)e^{-j\theta_R(t)}) * \frac{1}{2}(g_I(t) + \gamma e^{-j\phi} g_Q(t)) \\ + (r^*(t)e^{j\theta_R(t)}) * \frac{1}{2}(g_I(t) - \gamma e^{j\phi} g_Q(t)). \end{aligned}$$

Appendix B

Derivation of the properties of $\boldsymbol{x}_{F,m}$, i.e. $x_{F,m}[k] = x_m^*[k]x_{F,m}[0]$ and $x_{F,m}[0] = e^{j\pi\frac{(qm)^2}{N}}x_{F,0}[0]$.

According to (17) and N is an even number, we have

$$\begin{aligned} x_{F,m}[k] &= \sum_{n=0}^{N-1} e^{-j\pi\frac{n(n+2qm)}{N}} \cdot e^{-j2\pi\frac{nk}{N}} \\ &= e^{j\pi\frac{k(k+2qm)}{N}} \sum_{n=0}^{N-1} e^{-j\pi\frac{(n+k)(n+k+2qm)}{N}} \\ &= e^{j\pi\frac{k(k+2qm)}{N}} \sum_{n=0}^{N-1} e^{-j\pi\frac{n(n+2qm)}{N}} \\ &= e^{j\pi\frac{k(k+2qm)}{N}} x_{F,m}[0] = x_m^*[k]x_{F,m}[0]. \end{aligned} \quad (43)$$

Therefore, the property $x_{F,m}[k] = x_m^*[k]x_{F,m}[0]$ is derived.

Similar to (43), the property $x_{F,m}[0] = e^{j\pi\frac{(qm)^2}{N}}x_{F,0}[0]$ can be derived by

$$\begin{aligned} x_{F,m}[0] &= \sum_{n=0}^{N-1} e^{-j\pi\frac{n(n+2qm)}{N}} = e^{j\pi\frac{(qm)^2}{N}} \sum_{n=0}^{N-1} e^{-j\pi\frac{(n+qm)^2}{N}} \\ &= e^{j\pi\frac{(qm)^2}{N}} \sum_{n=0}^{N-1} e^{-j\pi\frac{n^2}{N}} = e^{j\pi\frac{(qm)^2}{N}} x_{F,0}[0]. \end{aligned}$$

Appendix C

Derivation of P_E. We use $\boldsymbol{E}_R$ to demonstrate the computation of P_E, and it could be easily seen that the P_E for $\boldsymbol{E}_T$, $\tilde{\boldsymbol{E}}_T$, $\tilde{\boldsymbol{E}}_R$ are the same.

We denote the Rx PN for the mth pilot block as $\theta_R[0], \ldots, \theta_R[N-1]$. Since $\boldsymbol{E}_R$ is a circulant matrix, the average power of the elements in its first row is same as the average power of the elements in its first column. As we mentioned at the end of Section II-C, the first column of $\boldsymbol{E}_R$ is the vector $\boldsymbol{p}_F = \frac{1}{N}\boldsymbol{F}\boldsymbol{p}_{R,m}$ whose first element is set to zero, where $\boldsymbol{p}_{R,m} = [e^{-j\theta_R[0]}, \ldots, e^{-j\theta_R[N-1]}]^T$. Therefore, we actually need to compute the DFT of $\boldsymbol{p}_{R,m}$ with a scaling factor $1/N$.

According to the random-walk PN model (5), we have $e^{-j\theta_R[n]} = e^{-j\theta_R[0]} \cdot e^{-j\sum_{n_0=1}^{n}\Delta\theta_R[n_0]}$, where $\Delta\theta_R[n]$ have i.i.d. distribution $\mathcal{N}(0, \sigma_{\theta_R}^2)$. Then, the kth element of $\boldsymbol{p}_F$ is

$$p_F[k] = \frac{1}{N}e^{-j\theta_R[0]} \sum_{n=0}^{N-1} \left(e^{-j\sum_{n_0=1}^{n}\Delta\theta_R[n_0]} \cdot e^{-j2\pi\frac{nk}{N}}\right). \quad (44)$$

To compute the power of $p_F[k]$, we define $p_{i,k}$ as

$$p_{i,k} = e^{-j2\pi\frac{ik}{N}} + \sum_{n=i+1}^{N-1} \left(e^{-j\sum_{n_0=i+1}^{n}\Delta\theta_R[n_0]} \cdot e^{-j2\pi\frac{nk}{N}}\right),$$

where $i = 0, \ldots, N-1$, $p_{N-1,k} = e^{-j2\pi\frac{(N-1)k}{N}}$, and we have $p_F[k] = \frac{1}{N}e^{-j\theta_R[0]}p_{0,k}$. Moreover, we have

$$p_{i,k} = e^{-j2\pi\frac{ik}{N}} + e^{-j\Delta\theta_R[i+1]}p_{i+1,k}. \quad (45)$$

Note that $e^{-j\Delta\theta_R[i+1]}$ is independent of $p_{i+1,k}$, so the expectation of $p_{i,k}$ can be given by

$$\mathrm{E}[p_{i,k}] = e^{-j2\pi\frac{ik}{N}} + \mathrm{E}[e^{-j\Delta\theta_R[i+1]}]\mathrm{E}[p_{i+1,k}]. \quad (46)$$

Using the characteristic function of Gaussian distribution, we have $\mathrm{E}[e^{-j\Delta\theta_R[i+1]}] = e^{-\frac{1}{2}\sigma_{\theta_R}^2}$. Then, from $p_{N-1,k} = e^{-j2\pi\frac{(N-1)k}{N}}$, we can derive $\mathrm{E}[p_{i,k}]$ for $i = 0, \ldots, N-2$ according to (46).

On the other hand, from (45), we have

$$\begin{aligned} \mathrm{E}[|p_{i,k}|^2] &= \mathrm{E}[|e^{-j2\pi\frac{ik}{N}} + e^{-j\Delta\theta_R[i+1]}p_{i+1,k}|^2] \\ &= 1 + e^{j2\pi\frac{ik}{N}}\mathrm{E}[e^{-j\Delta\theta_R[i+1]}]\mathrm{E}[p_{i+1,k}] \\ &\quad + e^{-j2\pi\frac{ik}{N}}\mathrm{E}[e^{j\Delta\theta_R[i+1]}]\mathrm{E}[p_{i+1,k}]^* + \mathrm{E}[|p_{i+1,k}|^2]. \end{aligned} \quad (47)$$

From $\mathrm{E}[|p_{N-1,k}|^2]=1$, $\mathrm{E}[|p_{i,k}|^2]$ for $i=0,\ldots,N-2$ can be computed according to (47), and then the power of $p_F[k]$ is computed by $\mathrm{E}[|p_F[k]|^2]=\frac{1}{N^2}\mathrm{E}[|p_{i,0}|^2]$.

Finally, we can derive P_E as $P_E=\sum_{k=1}^{N-1}\mathrm{E}[|p_F[k]|^2]$.

APPENDIX D

Derivation of $\boldsymbol{A}$ from (35). For convenience, we denote $J_0(\boldsymbol{A})$ as

$$J_0(\boldsymbol{A}) = \mathrm{E}\left[||\boldsymbol{A}(\alpha\boldsymbol{P}_R\mathrm{diag}\{\boldsymbol{h}_F\}\boldsymbol{P}_T\boldsymbol{x}_F+\tilde{\boldsymbol{w}}_{F,h})-e^{j\theta}\boldsymbol{P}_R\boldsymbol{P}_T\boldsymbol{x}_F||^2\right].$$

Then, we have

$$\begin{aligned}J_0(\boldsymbol{A}) &= \mathrm{E}\left[||e^{j\theta}(\boldsymbol{A}\boldsymbol{P}_R\mathrm{diag}\{\boldsymbol{h}'_F\}-\boldsymbol{P}_R)\boldsymbol{P}_T\boldsymbol{x}_F+\boldsymbol{A}\tilde{\boldsymbol{w}}_{F,h}||^2\right]\\ &\overset{(*)}{=}\sigma_{F,x}^2\mathrm{tr}\left((\boldsymbol{A}\boldsymbol{P}_R\mathrm{diag}\{\boldsymbol{h}'_F\}-\boldsymbol{P}_R)^H\right.\\ &\quad\left.\times\ (\boldsymbol{A}\boldsymbol{P}_R\mathrm{diag}\{\boldsymbol{h}'_F\}-\boldsymbol{P}_R)\right)+\tilde{\sigma}_F^2\mathrm{tr}(\boldsymbol{A}^H\boldsymbol{A}),\end{aligned}$$

where (*) is derived according to $\boldsymbol{P}_T=\frac{1}{N}\boldsymbol{F}\mathrm{diag}\{\boldsymbol{p}_T\}\boldsymbol{F}^H$ is a unitary matrix.

The derivative of $J_0(\boldsymbol{A})$ with respect to $\boldsymbol{A}^*$ is given by

$$\frac{\partial J_0}{\partial \boldsymbol{A}^*}=\sigma_{F,x}^2(\boldsymbol{A}\boldsymbol{P}_R\mathrm{diag}\{\boldsymbol{h}'_F\}-\boldsymbol{P}_R)\mathrm{diag}\{\boldsymbol{h}'^*_F\}\boldsymbol{P}_R^H+\tilde{\sigma}_F^2\boldsymbol{A}.$$

By letting $\frac{\partial J_0}{\partial \boldsymbol{A}^*}=\boldsymbol{0}$, we have

$$\boldsymbol{A}=\left(\boldsymbol{P}_R\mathrm{diag}\{|\boldsymbol{h}'_F|^2\}\boldsymbol{P}_R^H+\frac{\tilde{\sigma}_F^2}{\sigma_{F,x}^2}\boldsymbol{I}\right)^{-1}\left(\boldsymbol{P}_R\mathrm{diag}\{\boldsymbol{h}'^*_F\}\boldsymbol{P}_R^H\right),$$

where we denote $|\boldsymbol{h}'_F|^2=[|h'_F[0]|^2,\ldots,|h'_F[N-1]|^2]^T$ for convenience. Note that PN is slow time-varying and $\boldsymbol{F}^H\mathrm{diag}\{|\boldsymbol{h}'_F|^2\}\boldsymbol{F}$ is a diagonally dominant matrix, so we have

$$\begin{aligned}&\boldsymbol{P}_R\mathrm{diag}\{|\boldsymbol{h}'_F|^2\}\boldsymbol{P}_R^H\\ &\quad=\frac{1}{N^2}\boldsymbol{F}\left(\mathrm{diag}\{\boldsymbol{p}_R\}\boldsymbol{F}^H\mathrm{diag}\{|\boldsymbol{h}'_F|^2\}\boldsymbol{F}\mathrm{diag}\{\boldsymbol{p}_R^*\}\right)\boldsymbol{F}^H\\ &\quad\approx\mathrm{diag}\{|\boldsymbol{h}'_F|^2\},\end{aligned}$$

and similarly, we also have $\boldsymbol{P}_R\mathrm{diag}\{\boldsymbol{h}'^*_F\}\boldsymbol{P}_R^H\approx\mathrm{diag}\{\boldsymbol{h}'^*_F\}$. Therefore, we arrive at

$$\boldsymbol{A}\approx\left(\mathrm{diag}\{|\boldsymbol{h}'_F|^2\}+\frac{\tilde{\sigma}_F^2}{\sigma_{F,x}^2}\boldsymbol{I}\right)^{-1}\mathrm{diag}\{\boldsymbol{h}'^*_F\}.$$

REFERENCES

[1] H.-J. Song and T. Nagatsuma, "Present and future of terahertz communications," *IEEE Trans. THz Sci. Technol.*, vol. 1, no. 1, pp. 256–263, Sep. 2011.

[2] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Phys. Commun.*, vol. 12, pp. 16–32, Sep. 2014.

[3] H. Elayan, O. Amin, B. Shihada, R. M. Shubair, and M.-S. Alouini, "Terahertz band: The last piece of RF spectrum puzzle for communication systems," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1–32, 2020.

[4] A. Hirata *et al.*, "10-Gbit/s wireless link using InP HEMT MMICs for generating 120-GHz-band millimeter-wave signal," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 5, pp. 1102–1109, May 2009.

[5] S. Koenig *et al.*, "Wireless sub-THz communication system with high data rate," *Nature Photon.*, vol. 7, no. 12, pp. 977–981, Oct. 2013.

[6] X. Yu *et al.*, "400-GHz wireless transmission of 60-Gb/s Nyquist-QPSK signals using UTC-PD and heterodyne mixer," *IEEE Trans. THz Sci. Technol.*, vol. 6, no. 6, pp. 765–770, Nov. 2016.

[7] S. Jia *et al.*, "120 Gb/s multi-channel THz wireless transmission and THz receiver performance analysis," *IEEE Photon. Technol. Lett.*, vol. 29, no. 3, pp. 310–313, Feb. 1, 2017.

[8] F. Boes *et al.*, "Ultra-broadband MMIC-based wireless link at 240 GHz enabled by 64 GS/s DAC," in *Proc. 39th Int. Conf. Infr., Millim., THz Waves (IRMMW-THz)*, Sep. 2014, pp. 1–2.

[9] I. Kallfass *et al.*, "64 Gbit/s transmission over 850 m fixed wireless link at 240 GHz carrier frequency," *J. Infr., Millim., Terahertz Waves*, vol. 36, no. 2, pp. 221–233, Feb. 2015.

[10] I. Kallfass *et al.*, "Towards MMIC-based 300 GHz indoor wireless communication systems," *IEICE Trans. Electron.*, vol. E98.C, no. 12, pp. 1081–1090, Dec. 2015.

[11] J. Grzyb, P. R. Vazquez, B. Heinemann, and U. R. Pfeiffer, "A high-speed QPSK/16-QAM 1-m wireless link with a tunable 220–260 GHz LO carrier in SiGe HBT technology," in *Proc. 43rd Int. Conf. Infr., Millim., Terahertz Waves (IRMMW-THz)*, Sep. 2018, pp. 1–2.

[12] P. Rodriguez-Vazquez, J. Grzyb, B. Heinemann, and U. R. Pfeiffer, "A 16-QAM 100-Gb/s 1-M wireless link with an EVM of 17% at 230 GHz in an SiGe technology," *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 4, pp. 297–299, Apr. 2019.

[13] I. F. Akyildiz and J. M. Jornet, "Realizing ultra-massive MIMO (1024×1024) communication in the (0.06–10) terahertz band," *Nano Commun. Netw.*, vol. 8, pp. 46–54, Jun. 2016.

[14] Z. Zhu, X. Huang, and H. Leung, "Joint I/Q mismatch and distortion compensation in direct conversion transmitters," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2941–2951, Jun. 2013.

[15] K. S. Lorenz, J. Goodman, G. Stantchev, and N. A. Pendergrass, "Generalized transmitter compensation of frequency dependent I/Q imbalance," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2220–2231, May 2016.

[16] Z. A. Khan, E. Zenteno, P. Handel, and M. Isaksson, "Digital predistortion for joint mitigation of I/Q imbalance and MIMO power amplifier distortion," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 1, pp. 322–333, Jan. 2017.

[17] Y. R. Ramadan, H. Minn, and M. E. Abdelgelil, "Precompensation and system parameters estimation for low-cost nonlinear tera-hertz transmitters in the presence of I/Q imbalance," *IEEE Access*, vol. 6, pp. 51814–51833, Oct. 2018.

[18] 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.

[19] K.-Y. Sung and C.-C. Chao, "Estimation and compensation of I/Q imbalance in OFDM direct-conversion receivers," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 3, pp. 438–453, Jun. 2009.

[20] A.-A.-A. Boulogeorgos, V. M. Kapinas, R. Schober, and G. K. Karagiannidis, "I/Q-imbalance self-interference coordination," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4157–4170, Jun. 2016.

[21] B. Narasimhan, D. Wang, S. Narayanan, H. Minn, and N. Al-Dhahir, "Digital compensation of frequency-dependent joint Tx/Rx I/Q imbalance in OFDM systems under high mobility," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 3, pp. 405–417, Jun. 2009.

[22] Y.-C. Pan and S.-M. Phoong, "A time-domain joint estimation algorithm for CFO and I/Q imbalance in wideband direct-conversion receivers," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2353–2361, Jul. 2012.

[23] C. Zhang, Z. Xiao, B. Gao, L. Su, and D. Jin, "Three-stage treatment of TX/RX IQ imbalance and channel with CFO for SC-FDE systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 297–300, Feb. 2014.

[24] X. Zhang, H. Li, W. Liu, and J. Qiao, "Iterative IQ imbalance compensation receiver for single carrier transmission," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8238–8248, Sep. 2017.

[25] B. F. Beidas, "Radio-frequency impairments compensation in ultra high-throughput satellite systems," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6025–6038, Sep. 2019.

[26] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sep. 2012.

[27] C. Zhang, Z. Xiao, B. Gao, L. Su, and D. Jin, "Iterative Tx and Rx phase noise compensation for 60 GHz systems with SC-FDE transmission," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 5158–5162.

[28] P. Pedrosa, R. Dinis, F. Nunes, and A. Rodrigues, "Joint frequency domain equalisation and phase noise estimation for single-carrier modulations in doubly-selective channels," *IET Commun.*, vol. 9, no. 8, pp. 1138–1146, May 2015.
[29] X. Cheng, N. Lou, and B. Yuan, "Iterative decision-aided compensation of phase noise in millimeter-wave SC-FDE systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1030–1033, May 2016.
[30] N. Iqbal and A. Zerguine, "AFD-DFE using constraint-based RLS and phase noise compensation for uplink SC-FDMA," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4435–4443, May 2017.
[31] X. Cheng, K. Xu, and S. Li, "Compensation of phase noise in uplink massive MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1764–1778, Mar. 2019.
[32] J. Tubbax *et al.*, "Compensation of IQ imbalance and phase noise in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 872–877, May 2005.
[33] Q. Zou, A. Tarighat, and A. Sayed, "Joint compensation of IQ imbalance and phase noise in OFDM wireless systems," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 404–414, Feb. 2009.
[34] P. Rabiei, W. Namgoong, and N. Al-Dhahir, "Reduced-complexity joint baseband compensation of phase noise and I/Q imbalance for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3450–3460, Nov. 2010.
[35] R. Hamila, O. Ozdemir, and N. Al-Dhahir, "Beamforming OFDM performance under joint phase noise and I/Q imbalance," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2978–2989, May 2016.
[36] H. Minn, Q. Zhan, N. Al-Dhahir, and H. Huang, "In-phase and quadrature timing mismatch estimation and compensation in millimeter-wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4317–4331, Jul. 2017.
[37] J. Dong and Y. Shi, "Nonconvex demixing from bilinear measurements," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5152–5166, Oct. 2018.
[38] C. Han, A. O. Bicen, and I. F. Akyildiz, "Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2402–2412, May 2015.
[39] S. Kim and A. Zajic, "Statistical modeling and simulation of short-range device-to-device communication channels at sub-THz frequencies," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6423–6433, Sep. 2016.
[40] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, Oct. 2011.
[41] *Forward Error Correction for High Bit-Rate DWDM Submarine Systems*, document ITU-T Rec. G.975.1 (02/2004), I.-T. S. Group, 2005.
[42] J. C. Scheytt, D. Wrana, M. Bahmanian, and I. Kallfass, "Ultra-low phase noise frequency synthesis for THz communications using optoelectronic PLLs," in *Proc. 3rd Int. Workshop Mobile THz Syst. (IWMTS)*, Jul. 2020, pp. 1–4.
[43] F. Herzel, S. A. Osmany, and J. C. Scheytt, "Analytical phase-noise modeling and charge pump optimization for fractional-N PLLs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 8, pp. 1914–1924, Aug. 2010.

Ziyuan Sha received the B.S. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2017, where he is currently pursuing the Ph.D. degree. His main research interests include signal processing and resource management techniques in millimeter wave and terahertz communications.

Zhaocheng Wang (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees from Tsinghua University, in 1991, 1993, and 1996, respectively.

From 1996 to 1997, he was a Post-Doctoral Fellow with Nanyang Technological University, Singapore. From 1997 to 1999, he was a Research Engineer/Senior Engineer with the OKI Techno Centre (Singapore) Pte. Ltd., Singapore. From 1999 to 2009, he was a Senior Engineer/Principal Engineer with Sony Deutschland GmbH, Germany. Since 2009, he has been a Professor with the Department of Electronic Engineering, Tsinghua University, where he is currently the Director of the Broadband Communication Key Laboratory, Beijing National Research Center for Information Science and Technology (BNRist). He has authored or coauthored two books, which have been selected by *IEEE Press Series on Digital and Mobile Communication* and published by (Wiley-IEEE Press). He has authored/coauthored more than 160 peer-reviewed journal articles. He holds 46 U.S./EU granted patents (23 of them as the first inventor). His research interests include wireless communications, millimeter wave communications, and optical wireless communications. He is a fellow of the Institution of Engineering and Technology. He was a recipient of the ICC2013 Best Paper Award, the OECC2015 Best Student Paper Award, the 2016 IEEE Scott Helt Memorial Award, the 2016 IET Premium Award, the 2016 National Award for Science and Technology Progress (First Prize), the ICC2017 Best Paper Award, the 2018 IEEE ComSoc Asia-Pacific Outstanding Paper Award, and the 2020 IEEE ComSoc Leonard G. Abraham Prize. He was an Associate Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2011 to 2015 and IEEE COMMUNICATIONS LETTERS from 2013 to 2016. He is currently an Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE SYSTEMS JOURNAL, and IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY.