

Joint channel and clipping amplitude estimation and signal detection for clipped OTFS

Yufan Chen, Dongxuan He, *Member, IEEE*, Hua Wang, *Member, IEEE*, Weijie Yuan, *Senior Member, IEEE*, and Tony Q. S. Quek, *Fellow, IEEE*

Abstract—This paper investigates the receiver design for clipped orthogonal time frequency space (OTFS) systems, where the user devices are equipped with power amplifiers (PAs) with low dynamic range. To improve power efficiency, the PAs have to work near the saturation points, which leads to unknown nonlinear distortions, thus making the signal detection more challenging. To solve this problem, techniques like intentional clipping or pre-distortion are adopted, thus approximating the outputs of the PAs as clipped signals. To further compensate for the unknown time-varying multipath channel and the clipping distortion at the receiver, the channel and clipping amplitude (CA) estimation, channel tracking, and signal detection are studied in this paper. Firstly, a receiver framework is developed for clipped OTFS. Secondly, by adopting the sparsity of the delay-Doppler (DD) domain channel and the piecewise linearized signal model with respect to CA, a novel sparse Bayesian learning (SBL) based joint channel and CA estimation scheme is proposed. Then, to further reduce the estimation error and bit error rate, a Kalman filter (KF) based channel tracking scheme and a minimum mean square error decision feedback blockwise equalization (MMSE-DFBE) based detection scheme are proposed. These two schemes are integrated in an expectation maximization (EM) based iterative tracking and detection algorithm. Finally, numerical simulations are conducted to demonstrate the superiority of the proposed schemes in terms of both estimation error and bit error rate.

Index Terms—Orthogonal time frequency space (OTFS), channel estimation, clipping amplitude estimation, channel tracking, iterative detection.

I. INTRODUCTION

FUTURE wideband communication networks are expected to support higher user mobility speeds, which suffers from higher Doppler shifts, thus bringing great challenges to the waveform and receiver design. Specifically, the conventional orthogonal frequency division multiplexing (OFDM) waveform with pointwise channel estimation and detection will suffer from severe performance loss due to the intercarrier interference (ICI) caused by the high Doppler effects. The high peak-to-average power ratio (PAPR) of OFDM is another demerit. These issues pose an urgent need for new waveforms.

This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFE0200404, and in part by the National Natural Science Foundation of China under Grant 62101306, and also in part by the National Research Foundation, Singapore and Infocomm Media Development Authority under its Future Communications Research & Development Programme. (*Corresponding author: Dongxuan He*)

Y. Chen, D. He, and H. Wang are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mails: 3120185428@bit.edu.cn, dongxuan_he@bit.edu.cn, wanghua@bit.edu.cn).

W. Yuan is with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: yuanwj@sustech.edu.cn).

T. Q. S. Quek is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore 487372 (e-mail: tonyquek@sutd.edu.sg).

The orthogonal time frequency space (OTFS) was proposed in [1], where the symbols are modulated in the delay-Doppler (DD) domain. OTFS exploits both time and frequency diversity and significantly outperforms OFDM in the time-varying multipath fading channel [2]. Thanks to the sparsity and low variability of the DD domain channel, the detection and channel estimation for OTFS are also simple and robust [3]. Due to its excellent performance in high mobility systems, many studies on OTFS-based transceiver design have been carried out in recent years. The OTFS with pulse shaping, windowing, and precoding were studied in [4], [5], and [6], respectively. In [7], a channel estimation scheme was proposed, which used a DD domain embedded pilot with a guard interval. Sparse channel estimation schemes were studied in [8]–[10], which exploited the sparsity of the DD domain channel to further improve the estimation accuracy. Other works [11]–[13] have proposed channel estimation schemes using the time-frequency (TF) domain pilots to be compatible with the OFDM systems. Signal detection schemes for OTFS have also been well studied by the literature, where both linear detectors [14]–[16] and iterative detectors [17]–[18] were proposed.

Although the OTFS transmitter and receiver schemes have been widely studied, the application of OTFS in future networks is still challenging [19]. One important challenge is the nonlinearity of the power amplifier (PA). Similar to OFDM, OTFS is a multicarrier waveform suffering from the high PAPR problem. In scenarios such as the uplink of non-terrestrial networks (NTN) or high mobility machine type communications (MTC), many devices are equipped with PAs with low dynamic range. Therefore, the PAs have to work near their saturation points to enhance power efficiency. This inevitably results in distortions due to the nonlinearity of PAs.

Nevertheless, the impacts of the nonlinear PAs were only considered in a few studies on OTFS. Some of these works focused on evaluating the error performance [20]–[23], while using the same receivers as those in linear systems and assuming known channel response. The authors in [24] proposed a novel nonlinear detection and a novel embedded pilot pattern for channel estimation while assuming the nonlinear parameters known at the receiver. Nevertheless, considering the environmental changes and the aging of electronic components, the output characteristics of PAs are slowly time-varying. It is costly to measure and calibrate each device beforehand. Therefore, estimating the overall nonlinearity at the receiver is important, which affects the recovery of the distorted pilot symbols, thus further affecting the channel estimation and signal detection.

It is difficult to design a common parameter estimator for various types of nonlinear PA models. To simplify the

estimation and detection problem, techniques like intentional clipping or pre-distortion are typically applied at the transmitter [25], [26], [27]. Then the overall nonlinearity can be well approximated by a soft limiter (clipper) model [28], [29], where the clipping amplitude (CA) is assumed unknown to the receiver. Some existing works have studied the receiver schemes based on the soft limiter model. In [30], the authors proposed both blind and pilot based CA estimation schemes under the additive white Gaussian noise (AWGN) channel. The authors in [31] further proposed an alternating optimization algorithm to jointly estimate the CA and the multipath fading channel. The detection of clipped OFDM with known channel response and CA was studied in [32], where a maximum-likelihood (ML) detector was proposed. The authors in [33] proposed blind corrections of clipped OFDM with unknown CA. Other works have proposed clipping noise cancellation (CNC) algorithms for single-input-single-output (SISO) OFDM [34] and multiple-input-multiple-output (MIMO) OFDM [35]. Nevertheless, all these methods are only applicable for OFDM under static or slow time-varying channels. For clipped OTFS, estimating and tracking the DD domain channel, as well as signal detection, are more challenging tasks. Specifically, the increased dimension of unknown channel parameters makes the estimation more difficult when compared to OFDM. The fast time-varying channel response also raises a trade-off issue between estimation accuracy and the pilot/guard interval/computation overhead. Moreover, in addition to distortion cancellation, the residual intersymbol/intercarrier interference issue should also be handled for the detection of clipped OTFS symbols.

In this paper, we have proposed novel receiver schemes for joint DD channel and clipping amplitude estimation, channel tracking, as well as iterative signal detection of clipped OTFS. Specifically, our contributions are summarized as follows:

1) The receiver design problem for clipped OTFS is formulated, where the linearized sparse representation of the received OTFS signal is derived. Accordingly, a framework for joint estimation, tracking, and detection is developed.

2) A sparse Bayesian learning based joint DD domain channel and CA estimation is proposed. A specialized DD domain pilot generating criterion is also provided to reduce the complexity. The performance bounds for the mean square errors (MSEs) of the joint estimations are also derived.

3) A Kalman filter (KF) based scheme is derived for tracking the channel variations at the data blocks, which further reduces the estimation error by exploiting the observations from all data blocks.

4) A minimum mean square error decision feedback block-wise equalization (MMSE-DFBE) scheme is proposed for signal detection at the data blocks, which improves the bit error performance by minimizing the MSE using the correlation of the received and detected symbols.

5) An expectation maximization (EM) based iterative tracking and detection algorithm is also developed, which integrates the proposed tracking and detection approaches. Numerical simulations are conducted to show the benefits of our approach.

The rest of this paper is organized as follows. Section II

introduces the system model and the receiver framework for clipped OTFS. In Section III, the proposed SBL based joint channel and CA estimation algorithm is introduced. The lower bounds for MSE of the channel and CA estimation are also derived. Section IV introduces the KF based channel tracking scheme and the MMSE-DFBE based signal detection scheme, as well as the EM based algorithm for iterative tracking and detection. In Section V, the simulation results of the proposed and existing schemes are presented and discussed. In Section VI, we conclude this paper.

Notations: The superscripts $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^n$ represent the transpose, the Hermitian transpose, and the n -th power operation, respectively. $\arg(\cdot)$, $\sum(\cdot)$ and $\mathbb{E}(\cdot)$ are the argument, the summation, and the expectation operators, respectively. $\delta(\cdot)$ denotes the Dirac delta function. $\|\cdot\|_n$ denotes the ℓ_n -norm. The notations \odot and \otimes denote the Hadamard product and the Kronecker product, respectively. $\text{vec}(\cdot)$ represents the operation of vectorizing a matrix into a column vector. $\text{diag}(\cdot)$ represents the operation of constructing a diagonal matrix using a vector or extracting diagonal elements from a matrix. $\langle \cdot \rangle$ denotes the hard decision to the nearest constellation points. \mathbf{I}_n denotes the n -dimensional identity matrix. The normalized n -point discrete Fourier transform (DFT) and inverse DFT (IDFT) matrices are denoted by \mathbf{F}_n and \mathbf{F}_n^H , respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System model

In this paper, a SISO OTFS system is considered. The input data frame is segmented and mapped into K OTFS data blocks in the DD domain with M delay bins and N Doppler bins for each block.

The k -th DD domain OTFS block is denoted by $\mathbf{X}_k \in \mathbb{C}^{M \times N}$, $k = 0, \dots, K$, which is first transformed into the TF domain using the inverse symplectic finite Fourier transform (ISFFT) as [4]

$$\mathbf{S}_k = \mathbf{F}_M \mathbf{X}_k \mathbf{F}_N^H. \quad (1)$$

Next, \mathbf{S}_k is transformed into the time domain vector $\mathbf{s}_k \in \mathbb{C}^{MN \times 1}$ through the Heisenberg transform as follows

$$\mathbf{s}_k = \text{vec}(\mathbf{P}_{\text{tx}} \mathbf{F}_M^H \mathbf{S}_k) = (\mathbf{F}_N^H \otimes \mathbf{P}_{\text{tx}}) \mathbf{x}_k, \quad (2)$$

where $\mathbf{x}_k = \text{vec}(\mathbf{X}_k)$, $\mathbf{P}_{\text{tx}} \in \mathbb{C}^{M \times M}$ is a diagonal matrix of transmit pulse shaping.

Then, a cyclic prefix (CP) is added to each block. Further, a frame is formed with the blocks. In this paper, an efficient frame structure using the block-type pilot is adopted [12], which is shown in Fig.1. This structure only needs one pilot block over all the $(K + 1)$ blocks and does not need guard intervals. Therefore when K is large, it is more efficient than the frequency domain pilot [13] or DD domain embedded pilot [7]. Furthermore, the superposition of the pilot and the data on each time domain sample is avoided by using this structure. As such, the extracted pilot block does not contain interference and additional distortion from the unknown data, which makes the estimation of the channel and CA more accurate and easier.

The signal frame then passes through the cascaded combination of a clipper/pre-distorter and a power amplifier, which is well approximated as a soft limiter. Let a scalar A denote the

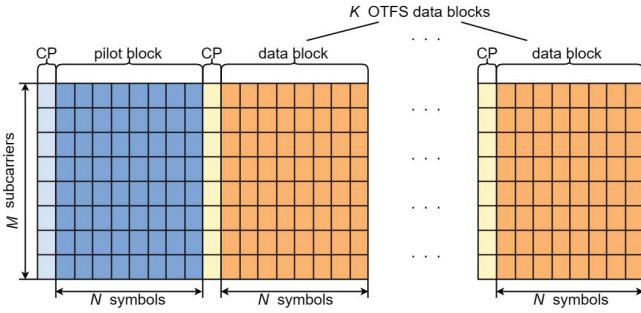


Fig. 1. Frame structure for OTFS with the block-type pilot.

clipping amplitude¹, the clipped output signal corresponding to the input signal $s(t) = a(t) e^{j\arg(s(t))}$ at time t is expressed as

$$g(s(t); A) = \begin{cases} s(t), & a(t) \leq A, \\ Ae^{j\arg(s(t))}, & a(t) > A. \end{cases} \quad (3)$$

After passing through the high mobility multipath channel, the received signal is given by

$$r(t) = \int \int h(\tau, \nu) g(s(t - \tau); A) e^{j2\pi\nu(t - \tau)} d\tau d\nu + v(t), \quad (4)$$

where $s(t)$ is the continuous-time domain transmitted signal, $v(t)$ is the additive white Gaussian noise, $h(\tau, \nu)$ is the delay-Doppler channel response, given by

$$h(\tau, \nu) = \sum_{p=1}^P h_p \delta(\tau - \tau_p) \delta(\nu - \nu_p), \quad (5)$$

where P is the number of paths, h_p , τ_p , and ν_p are the complex gain, delay, and Doppler shift of the p -th path, respectively.

Denoting NT and $M\Delta f$ the duration and bandwidth of one OTFS block, respectively, the normalized delay and Doppler of the p -th path are given by

$$\tau_p = \tau_p M\Delta f, o_p = \nu_p NT. \quad (6)$$

We assume that the bandwidth is sufficient to approximate the path delays to the nearest sampling points, thus we can ignore non-integer l_p values [36]. The received n -th discrete time sample is thus given by

$$r(n) = \sum_{p=1}^P h_p e^{j2\pi \frac{o_p(n-l_p)}{MN}} s([n - l_p]_{MN}) + v(n). \quad (7)$$

Then, after deframing and removing CP, the k -th block can be denoted in vector form, given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{g}(\mathbf{s}_k; A) + \mathbf{v}_k, \quad (8)$$

where $\mathbf{g}(\mathbf{s}_k; A) = [g(\mathbf{s}_k(1); A), \dots, g(\mathbf{s}_k(MN); A)]^T$ is the clipped transmitted vector, $\mathbf{v}_k \in \mathbb{C}^{MN \times 1}$ is the noise vector, $\mathbf{H}_k \in \mathbb{C}^{MN \times MN}$ is the time-delay domain channel response, given by

$$\mathbf{H}_k = \sum_{p=1}^P h_k^p \boldsymbol{\Pi}^{l_p} \boldsymbol{\Delta}_p, \quad (9)$$

¹For a specific device, A may vary with the operating point and the maximum output power of the PA, so it is unknown to the receiver beforehand.

in which $\boldsymbol{\Pi} \in \mathbb{Z}^{MN \times MN}$ is a circulant permutation matrix given by

$$\boldsymbol{\Pi} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}_{MN \times MN}, \quad (10)$$

$\boldsymbol{\Delta}_p$ is a diagonal matrix given by

$$\boldsymbol{\Delta}_p = \begin{cases} \text{diag}(1, \omega_p, \dots, \omega_p^{MN-1}), & \text{if } l_p = 0, \\ \text{diag}(1, \dots, \omega_p^{MN-l_p-1}, \omega_p^{-l_p}, \dots, \omega_p^{-1}), & \text{if } l_p \neq 0. \end{cases} \quad (11)$$

where $\omega_p = e^{j\frac{2\pi o_p}{MN}}$.

B. Problem formulation

Since (8) is nonlinear with respect to A , direct estimation of A presents challenges. To tackle this, the estimation and detection problem is reformulated in this section. The transmitted vector can be denoted as

$$\mathbf{s}_k = \mathbf{a}_k \odot e^{j\phi_k}, \quad (12)$$

where $\mathbf{a}_k = [a_k(1), \dots, a_k(MN)]^T$ and $e^{j\phi_k} = [e^{j\phi_k(1)}, \dots, e^{j\phi_k(MN)}]^T$ are the amplitude vector and the phase vector, respectively. Next, a binary vector $\mathbf{c}_k = [c_k(1), \dots, c_k(MN)]^T$ is introduced [31], [37], where $c_k(n)$ indicates whether the n -th sample of \mathbf{s}_k is clipped or not, given by

$$c_k(n) = \begin{cases} 1, & a_k(n) > A \\ 0, & a_k(n) \leq A \end{cases}. \quad (13)$$

In this way, a linearized representation of the clipped vector can be derived as

$$\mathbf{z}_k = (1 - \mathbf{c}_k) \odot \mathbf{s}_k + A \mathbf{c}_k \odot e^{j\phi_k}, \quad (14)$$

and therefore we have

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{z}_k + \mathbf{v}_k. \quad (15)$$

We then note that \mathbf{H}_k is a matrix with more variables than \mathbf{z}_k , which makes directly reconstructing \mathbf{H}_k from the model of (15) still infeasible. Fortunately, (15) can be further reformulated by adopting the sparsity of the DD domain channel. Specifically, let M_τ and N_ν be the delay and Doppler grid sizes larger than the maximum normalized delay and Doppler taps i.e. $M_\tau > l_p, N_\nu > o_p, \forall p$. We then consider a new DD grid of G_τ delay bins and G_ν Doppler bins for a sparse representation of the DD channel. Since fractional normalized delays are not considered, the delay axis size is set as $G_\tau = M_\tau$. The Doppler axis sizes are set as $G_\nu = N_\nu$ and $G_\nu > N_\nu$ for integer and fractional normalized Dopplers, respectively [12]. Then the DD channel response can be re-expressed on the new grid, given by

$$h(\tau, \nu) = \sum_{m=0}^{G_\tau-1} \sum_{n=0}^{G_\nu-1} h_k^{(m,n)} \delta(\tau - \frac{m}{M\Delta f}) \delta(\nu - \frac{nN_\nu}{G_\nu NT}), \quad (16)$$

where $h_k^{(m,n)}$ is the complex gain of the path corresponding to m delay units and n Doppler units. The time-delay domain channel matrix is thus reformulated as

$$\mathbf{H}_k = \sum_{m=0}^{G_\tau-1} \sum_{n=0}^{G_\nu-1} h_k^{(m,n)} \boldsymbol{\Pi}^m \boldsymbol{\Delta}_{m,n}, \quad (17)$$

where $\boldsymbol{\Pi}$ is the same permutation matrix as (10) and

$$\boldsymbol{\Delta}_{m,n} = \begin{cases} \text{diag} \left(1, \omega_n, \dots, \omega_n^{MN-1} \right), & \text{if } m = 0, \\ \text{diag} \left(1, \dots, \omega_n^{MN-m-1}, \omega_n^{-m}, \dots, \omega_n^{-1} \right), & \text{if } m \neq 0, \end{cases} \quad (18)$$

where $\omega_n = e^{j\frac{2\pi n}{MN}}$. It is worth noting that in (17) only P out of the $G_\tau G_\nu$ delay-Doppler taps have a non-zero path gain coefficient $h_k^{(m,n)}$. This property further enables a sparse representation of the channel reconstruction model. Specifically, by substituting (17) into (15) and extracting all the $h_k^{(m,n)}$ coefficients into a vector, we obtain

$$\mathbf{y}_k = \mathbf{W} (\mathbf{I}_{G_\tau G_\nu} \otimes \mathbf{z}_k) \mathbf{h}_k + \mathbf{v}_k, \quad (19)$$

where $\mathbf{W} \in \mathbb{C}^{MN \times MNG_\tau G_\nu}$ is a fixed matrix given by

$$\mathbf{W} = [\boldsymbol{\Pi}^0 \boldsymbol{\Delta}_{0,0}, \dots, \boldsymbol{\Pi}^0 \boldsymbol{\Delta}_{0,G_\nu-1}, \boldsymbol{\Pi}^1 \boldsymbol{\Delta}_{1,0}, \dots, \boldsymbol{\Pi}^{G_\tau-1} \boldsymbol{\Delta}_{G_\tau-1,G_\nu-1}], \quad (20)$$

and \mathbf{h}_k is a sparse vector of path gains, given by

$$\mathbf{h}_k = [h_k^{(0,0)}, \dots, h_k^{(0,G_\nu-1)}, h_k^{(1,0)}, \dots, h_k^{(G_\tau-1,G_\nu-1)}]^T. \quad (21)$$

Denoting $\Phi(\mathbf{s}_k, A) = \mathbf{W} (\mathbf{I}_{G_\tau G_\nu} \otimes \mathbf{z}_k)$, we found that $\Phi(\mathbf{s}_k, A)$ is linearized in A if \mathbf{s}_k is fixed. Then, (19) can be reformulated as

$$\mathbf{y}_k = \Phi(\mathbf{s}_k, A) \mathbf{h}_k + \mathbf{v}_k. \quad (22)$$

For the pilot block where $k = 0$, the joint estimation problem can be formulated as solving \mathbf{h}_0 and A from the received pilot block \mathbf{y}_0 , with the known pilot symbols \mathbf{s}_0 . Note that if $\Phi(\mathbf{s}_0, A)$ is fixed, the channel estimation problem further turns to a sparse recovery problem, where the sparsity of \mathbf{h}_0 can be exploited to obtain sparse solutions for the ill-conditioned or underdetermined equations that cannot be well solved by conventional linear estimations.

In order to perform efficient joint estimation, tracking and detection based on (22), a receiver framework is further developed. In particular, the received pilot block \mathbf{y}_0 is sent to the Channel & CA estimation module, where the initial estimations of the channel vector $\hat{\mathbf{h}}_0$ and the clipping amplitude \hat{A} are obtained. Then, the Channel & CA tracking module is utilized to update the channel coefficients $\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_K$, while the data blocks $\hat{\mathbf{X}}_1, \dots, \hat{\mathbf{X}}_K$ are detected by the signal detection module. These two modules work iteratively. The proposed transmitter and receiver are shown schematically in Fig.2.

III. JOINT ESTIMATION OF DELAY-DOPPLER CHANNEL AND CLIPPING AMPLITUDE

In this section, a sparse Bayesian learning (SBL) based joint channel and CA estimation scheme is proposed, as well as a corresponding DD domain pilot design criterion. The flowchart of our proposed SBL based joint estimation is shown in Fig.3.

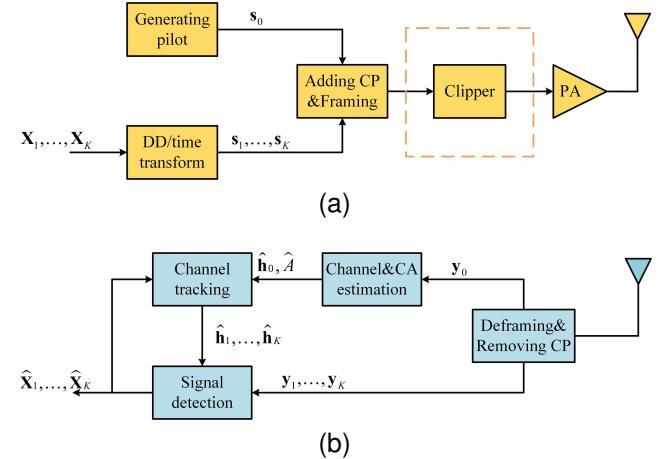


Fig. 2. (a) Transmitter and (b) receiver for clipped OTFS.

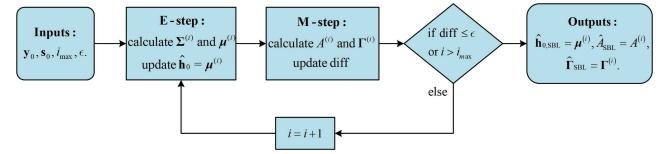


Fig. 3. Flowchart of the proposed SBL based joint estimation algorithm.

A. SBL based joint channel and CA estimation

Let a single index $u = mG_\nu + n$ represent the vectorized delay-Doppler indices (m, n) in (21), then the channel coefficient vector is represented as $\mathbf{h}_0 = [h_0^{(0)}, \dots, h_0^{(u)}, \dots, h_0^{(G_\tau G_\nu - 1)}]^T$. In the SBL framework [38], in order to encourage sparsity, a vector of hyper-parameters is defined to govern the prior distribution of the channel vector to be reconstructed. Since the noise is Gaussian distributed, a zero-mean Gaussian prior is assumed to ensure conjugation with the Gaussian likelihood. Therefore we assume $\mathbf{h}_0 \sim \mathcal{CN}(0, \gamma)$, where $\gamma = [\gamma(1), \dots, \gamma(u), \dots, \gamma(G_\tau G_\nu - 1)]^T$ is a vector of hyper-parameters. The parametrized prior distribution is given by

$$p(\mathbf{h}_0; \gamma) = \prod_{u=1}^{G_\tau G_\nu} \frac{1}{(\pi \gamma(u))} \exp \left(-\frac{|h_0^{(u)}|^2}{\gamma(u)} \right). \quad (23)$$

Given the prior distribution of the channel, the pilot based joint channel and CA estimation problem is formulated as

$$[\hat{\mathbf{h}}_0, \hat{A}, \Gamma] = \arg \max_{\mathbf{h}_0, A, \Gamma} p(\mathbf{y}_0; \mathbf{h}_0, A, \Gamma), \quad (24)$$

where $\Gamma = \text{diag}(\gamma)$. The above objective function is not convex, which makes the closed form solution hard to obtain. To this end, an EM based SBL [39] algorithm is adopted to solve (24). In the EM algorithm, \mathbf{h}_0 is treated as a hidden variable, while A and γ are the parameters to be estimated. The expectation of the log-likelihood function is calculated

in the E-step and maximized in the M-step. Specifically, the E-step and the M-step of the i -th iteration are given as

$$\begin{aligned} \text{E - step : } & \mathcal{L}(A, \gamma | A^{(i-1)}, \gamma^{(i-1)}) \\ & = \mathbb{E}_{\mathbf{h}_0 | \mathbf{y}_0; A^{(i-1)}, \gamma^{(i-1)}} [\log p(\mathbf{y}_0, \mathbf{h}_0; A, \gamma)], \end{aligned} \quad (25)$$

$$\begin{aligned} \text{M - step : } & (A^{(i)}, \gamma^{(i)}) \\ & = \arg \max_{A, \gamma} \mathcal{L}(A, \gamma | A^{(i-1)}, \gamma^{(i-1)}). \end{aligned} \quad (26)$$

The E-step requires the posterior density of the sparse vector with respect to $\gamma^{(i-1)}$ and $A^{(i-1)}$ from the last iteration, given by $p(\mathbf{h}_0 | \mathbf{y}_0; A^{(i-1)}, \gamma^{(i-1)}) = \mathcal{CN}(\mathbf{h}_0; \boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma}^{(i)})$, where

$$\boldsymbol{\Sigma}^{(i)} =$$

$$\left[\sigma_v^{-2} \boldsymbol{\Phi}(\mathbf{s}_0, A^{(i-1)})^H \boldsymbol{\Phi}(\mathbf{s}_0, A^{(i-1)}) + (\boldsymbol{\Gamma}^{(i-1)})^{-1} \right]^{-1}, \quad (27)$$

$$\boldsymbol{\mu}^{(i)} = \sigma_v^{-2} \boldsymbol{\Sigma} \boldsymbol{\Phi}(\mathbf{s}_0, A^{(i-1)})^H \mathbf{y}_0. \quad (28)$$

In the M-step, the objective function is maximized with respect to A and γ by setting $\mathbf{h}_0 = \boldsymbol{\mu}^{(i)}$. According to the Bayesian rule, the objective function can be denoted as

$$\begin{aligned} & \mathcal{L}(A, \gamma | A^{(i-1)}, \gamma^{(i-1)}) \\ & = \mathbb{E}_{\mathbf{h}_0 | \mathbf{y}_0; A^{(i-1)}, \gamma^{(i-1)}} [\log p(\mathbf{y}_0, \mathbf{h}_0; A, \gamma)] \\ & = \mathbb{E}_{\mathbf{h}_0 | \mathbf{y}_0; A^{(i-1)}, \gamma^{(i-1)}} [\log p(\mathbf{y}_0 | \mathbf{h}_0; A)] \\ & + \mathbb{E}_{\mathbf{h}_0 | \mathbf{y}_0; A^{(i-1)}, \gamma^{(i-1)}} [\log p(\mathbf{h}_0; \gamma)] + c, \end{aligned} \quad (29)$$

where c is a constant independent of A and γ . We note that the first term of (29) is independent of γ , while the second term is independent of A , which can be maximized separately.

The maximization of the first term of (29) is expressed as

$$A^{(i)} = \arg \min_{A > 0} \|\mathbf{y}_0 - \boldsymbol{\Phi}(\mathbf{s}_0, A) \mathbf{h}_0\|^2. \quad (30)$$

Let $\overset{\circ}{\mathbf{a}}_0 = \mathbf{P}_s \mathbf{a}_0$ denote the pilot amplitude that is sorted from the minimum to the maximum, where $\mathbf{P}_s \in \mathbb{C}^{MN \times MN}$ is the sorting permutation matrix. Similarly, the sorted binary vector, pilot, and phase vector are denoted as $\overset{\circ}{\mathbf{c}}_0 = \mathbf{P}_s \mathbf{c}_0$, $\overset{\circ}{\mathbf{s}}_0 = \mathbf{P}_s \mathbf{s}_0$, and $\overset{\circ}{\phi}_0 = \mathbf{P}_s \phi_0$, respectively. Then, (30) can be reformulated as

$$A^{(i)} = \arg \min_{A_0 > 0} \left\| \mathbf{y}_0 - \boldsymbol{\Phi}_u(\overset{\circ}{\mathbf{c}}_0) \mathbf{h}_0 - \boldsymbol{\Phi}_c(\overset{\circ}{\mathbf{c}}_0, A) \mathbf{h}_0 \right\|^2, \quad (31)$$

where $\boldsymbol{\Phi}_u(\overset{\circ}{\mathbf{c}}_0) = \mathbf{W} \left(\mathbf{I}_{G_\tau G_\nu} \otimes \left(\mathbf{P}_s^T \left((1 - \overset{\circ}{\mathbf{c}}_0) \odot \overset{\circ}{\mathbf{s}}_0 \right) \right) \right)$ and $\boldsymbol{\Phi}_c(\overset{\circ}{\mathbf{c}}_0, A) = \mathbf{W} \left(\mathbf{I}_{G_\tau G_\nu} \otimes \left(\mathbf{P}_s^T \left(A \overset{\circ}{\mathbf{c}}_0 \odot e^{j \overset{\circ}{\phi}_0} \right) \right) \right)$.

The objective function (31) is linear with respect to A if $\overset{\circ}{\mathbf{c}}_0$ is fixed. Therefore, the global minimum can be obtained by calculating the minimum values in each fixed interval. From (13), we note that $\overset{\circ}{\mathbf{c}}_0$ is a binary vector, in which the first consecutive elements are all zeros and the others are all ones. Regarding the set of elements in $\overset{\circ}{\mathbf{a}}_0$ with the same value as a

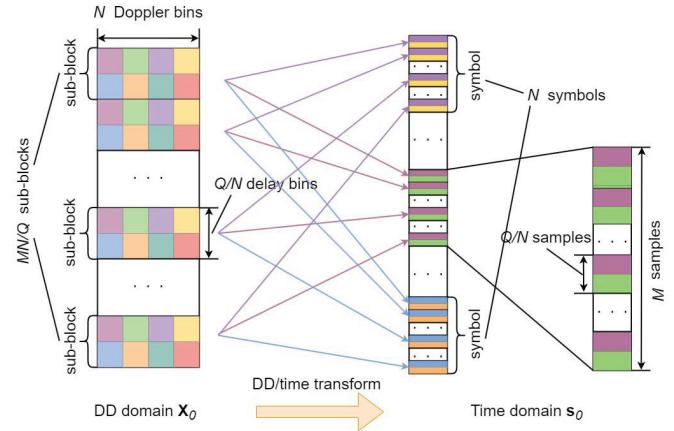


Fig. 4. OTFS pilot structure for joint channel&CA estimation.

group, the total number of intervals N_0 depends on the number of groups. We now consider two special cases for N_0 values:

- 1) All the pilots have the same amplitude, where $\overset{\circ}{\mathbf{a}}_0(1) = \overset{\circ}{\mathbf{a}}_0(2) \dots = \overset{\circ}{\mathbf{a}}_0(MN)$ and $N_0 = 1$, $\overset{\circ}{\mathbf{c}}_0$ can only be all zero or all one. Accurate estimation of A and \mathbf{h}_0 cannot be obtained in both occasions. That is because when N_0 is very small, different combinations of A s and \mathbf{h}_0 s may result in highly correlated or even identical \mathbf{y}_0 .
- 2) The DD domain pilot symbols are randomly generated with all different time domain samples where $\overset{\circ}{\mathbf{a}}_0(1) < \overset{\circ}{\mathbf{a}}_0(2) \dots < \overset{\circ}{\mathbf{a}}_0(MN)$ and therefore $N_0 = MN$. In this case, the number of intervals to be traversed is very large, which will lead to high computational complexity.

In conclusion, for joint estimation of channel and CA, N_0 should not be set too small or too large. Accordingly, a specialized pilot structure is developed, where the value of N_0 can be flexibly selected. Firstly, the DD domain pilot is segmented into $\frac{MN}{Q}$ OTFS sub-blocks along the delay axis, given as

$$\mathbf{X}_0 = \begin{bmatrix} \mathbf{X}_{0,1} \\ \mathbf{X}_{0,2} \\ \vdots \\ \mathbf{X}_{0,\frac{MN}{Q}} \end{bmatrix}, \quad (32)$$

where $\mathbf{X}_{P,i} \in \mathbb{C}^{\frac{Q}{N} \times N}$, $i = 1, \dots, \frac{MN}{Q}$ are the known pilot sub-blocks of length Q . After the ISFFT and the Heisenberg transform, the time domain pilot is denoted as

$$\mathbf{s}_0 = \text{vec}(\mathbf{P}_{tx} \mathbf{X}_0 \mathbf{F}_N^H) = \text{vec} \begin{bmatrix} \mathbf{P}_{tx,1} \mathbf{X}_{0,1} \mathbf{F}_N^H \\ \mathbf{P}_{tx,2} \mathbf{X}_{0,2} \mathbf{F}_N^H \\ \vdots \\ \mathbf{P}_{tx,\frac{MN}{Q}} \mathbf{X}_{0,\frac{MN}{Q}} \mathbf{F}_N^H \end{bmatrix}, \quad (33)$$

where $\mathbf{P}_{tx,j} \in \mathbb{C}^{\frac{Q}{N} \times \frac{Q}{N}}$ is the j -th block on the diagonal of \mathbf{P}_{tx} . We further let all the segments in (33) to be identical, i.e., $\mathbf{P}_{tx,1} \mathbf{X}_{0,1} \mathbf{F}_N^H = \mathbf{P}_{tx,2} \mathbf{X}_{0,2} \mathbf{F}_N^H = \dots = \mathbf{P}_{tx,\frac{MN}{Q}} \mathbf{X}_{0,\frac{MN}{Q}} \mathbf{F}_N^H$, the number of different values in \mathbf{s}_0 is thus upper bounded by that in each part. Take the general rectangular pulse where $\mathbf{P}_{tx} = \mathbf{I}_M$ as an example, we obtain $\mathbf{X}_{0,1} = \mathbf{X}_{0,2} = \dots = \mathbf{X}_{0,\frac{MN}{Q}}$. We further assume that the $\mathbf{X}_{0,1}$ is randomly

generated and all its Q samples are different, therefore the number of constant intervals of $\hat{\mathbf{c}}_0$ is $N_0 = Q$. In this way, we have generated a pilot block such that the sorted amplitudes $\hat{a}_0 \left(q \frac{MN}{Q} - \frac{MN}{Q} + 1 \right) = \dots = \hat{a}_0 \left(q \frac{MN}{Q} \right) = \hat{a}_0^{(q)}, q = 1, \dots, Q$. The above pilot structure can also be schematically explained by Fig.4. In the DD domain, The j -th sub-block occupies the $\left(\frac{Q}{N}(j-1) + 1 \right)$ -th to the $\frac{Q}{N}j$ -th delay bin. Meanwhile, it occupies the $\left(\frac{Q}{N}(j-1) + 1 \right)$ -th to the $\frac{Q}{N}j$ -th sample of each block in the time domain. If the DD domain sub-blocks are identical, the time domain samples are also identical. By adopting this pilot structure, the highly correlated combinations of the channel and the clipped pilot are effectively avoided, while the complexity of CA estimation is also reduced.

With the proposed pilot, the objective function in the q -th constant interval is denoted as

$$\begin{aligned} J^{(q)}(A) &= \left\| \mathbf{y}_0 - \Phi_u \left(\hat{\mathbf{c}}_0^{(q)} \right) \mathbf{h}_0 - \Phi_c \left(\hat{\mathbf{c}}_0^{(q)}, A \right) \mathbf{h}_0 \right\|^2 \\ &= \left\| \mathbf{y}_0 - \Phi_u \left(\hat{\mathbf{c}}_0^{(q)} \right) \mathbf{h}_0 \right\|^2 + \left\| \widehat{\Phi}_c \left(\hat{\mathbf{c}}_0^{(q)} \right) \mathbf{h}_0 \right\|^2 A^2 \\ &\quad - 2\operatorname{Re} \left\{ \left[\mathbf{y}_0 - \Phi_u \left(\hat{\mathbf{c}}_0^{(q)} \right) \mathbf{h}_0 \right]^H \widehat{\Phi}_c \left(\hat{\mathbf{c}}_0^{(q)} \right) \mathbf{h}_0 \right\} A, \end{aligned} \quad (34)$$

where $\hat{\mathbf{c}}_0^{(q)}$ denotes the known constant vector $\hat{\mathbf{c}}_0$ when $\hat{a}_0 \left(q \frac{MN}{Q} \right) \leq A < \hat{a}_0 \left((q+1) \frac{MN}{Q} \right)$, and $\widehat{\Phi}_c \left(\hat{\mathbf{c}}_0^{(q)} \right) = \mathbf{W} \left(\mathbf{I}_{G_\tau G_\nu} \otimes \left(\mathbf{P}_s^T \left(\hat{\mathbf{c}}_0^{(q)} \odot e^{j\phi_0} \right) \right) \right)$ is also a constant vector independent of A .

Taking the derivative of (34) with respect to A and setting it to zero, the minimizer is obtained as

$$\tilde{A}_q = \frac{\operatorname{Re} \left\{ \left[\mathbf{y}_0 - \Phi_u \left(\hat{\mathbf{c}}_0^{(q)} \right) \mathbf{h}_0 \right]^H \widehat{\Phi}_c \left(\hat{\mathbf{c}}_0^{(q)} \right) \mathbf{h}_0 \right\}}{\left\| \widehat{\Phi}_c \left(\hat{\mathbf{c}}_0^{(q)} \right) \mathbf{h}_0 \right\|^2}. \quad (35)$$

Then the CA estimation of the q -th interval is given as

$$\hat{A}_q = \begin{cases} \hat{a}_0 \left((q+1) \frac{MN}{Q} \right), & \text{if } \tilde{A}_q \geq \hat{a}_0 \left((q+1) \frac{MN}{Q} \right), \\ \hat{a}_0 \left(q \frac{MN}{Q} \right), & \text{if } \tilde{A}_q < \hat{a}_0 \left(q \frac{MN}{Q} \right), \\ \tilde{A}_q, & \text{else.} \end{cases} \quad (36)$$

Finally, (30) is solved by comparing the objective functions in all Q intervals and finding the minimum, given as

$$A^{(i)} = \arg \min_{q=1, \dots, Q} J^{(q)}(\hat{A}_q). \quad (37)$$

We then consider the maximization of the second term of (29) with respect to γ . By substituting $p(\mathbf{h}_0; \gamma)$ and taking the derivative with respect to γ , we obtain the maximizer, which is given by

$$\begin{aligned} \gamma^{(i)} &= \arg \max_{\gamma} \mathbb{E}_{\mathbf{h}_0 | \mathbf{y}_0; \gamma^{(i-1)}} [\log p(\mathbf{h}_0; \gamma)] \\ &= \mathbb{E}_{\mathbf{h}_0 | \mathbf{y}_0; \gamma^{(i-1)}} [|\mathbf{h}_0|^2] \\ &= \operatorname{diag} (\Sigma^{(i)}) + |\mu^{(i)}|^2, \end{aligned} \quad (38)$$

where $\operatorname{diag} (\Sigma^{(i)}) \in \mathbb{C}^{G_\tau G_\nu \times 1}$ is a vector containing the diagonal elements of $\Sigma^{(i)}$.

Based on the above estimation methods, an SBL based joint channel and CA estimation (SBL-JE) algorithm is proposed, which is detailed as **Algorithm 1**. With the maximizations in M-step, we have $\mathcal{L}(A^{(i+1)}, \gamma^{(i+1)} | A^{(i)}, \gamma^{(i)}) \geq \mathcal{L}(A^{(i)}, \gamma^{(i)} | A^{(i-1)}, \gamma^{(i-1)})$ therefore the objective function $\mathcal{L}(A, \gamma | A^{(i)}, \gamma^{(i)})$ increases monotonically. This property guarantees the convergence of the proposed algorithm to approach the maximum likelihood estimations [40].

Algorithm 1 SBL based joint estimation of channel and CA

Input: Received pilot block \mathbf{y}_0 , original pilot block \mathbf{s}_0 , maximum iteration time i_{\max} , stopping accuracy ϵ .

Output: DD channel estimation $\hat{\mathbf{h}}_{0,\text{SBL}}$, clipping amplitude estimation \hat{A}_{SBL} , hyperparameter matrix $\hat{\Gamma}_{\text{SBL}}$.

- 1: Calculate $\mathbf{P}_s, \hat{\mathbf{c}}_0^{(1)}, \dots, \hat{\mathbf{c}}_0^{(Q)}, \hat{a}_0^{(1)}, \dots, \hat{a}_0^{(Q)}$, and ϕ_0 .
- 2: Initialize $\Gamma^{(0)} = \mathbf{I}_{G_\tau G_\nu}$, $\mu^{(0)} = \mathbf{0}_{G_\tau G_\nu}$, and $A^{(0)}$ using the procedure of (35)-(37) by setting $\mathbf{h}_0 = \hat{\mathbf{h}}_{0,\text{MMSE}}$.
- 3: Set $\text{diff} = 1, i = 1$.
- 4: **while** ($\text{diff} > \epsilon$ & $i \leq i_{\max}$) **do**
- 5: **E-step:**
- 6: Update $\Sigma^{(i)}$ and $\mu^{(i)}$ using (27) and (28).
- 7: Setting $\mathbf{h}_0 = \mu^{(i)}$.
- 8: **M-step:**
- 9: Update $A^{(i)}$ using the procedure of (35)-(37).
- 10: **for** $u = 1, \dots, G_\tau G_\nu$ **do**
- 11: $\Gamma^{(i)}(u, u) = \Sigma^{(i)}(u, u) + |\mu^{(i)}(u)|^2$.
- 12: **end for**
- 13: Calculate $\text{diff} = \frac{\|\mu^{(i)} - \mu^{(i-1)}\|^2}{\|\mu^{(i)}\|^2}$.
- 14: Set $i = i + 1$.
- 15: **end while**
- 16: Set outputs $\hat{\mathbf{h}}_{0,\text{SBL}} = \mu^{(i)}, \hat{A}_{\text{SBL}} = A^{(i)}, \hat{\Gamma}_{\text{SBL}} = \Gamma^{(i)}$.

The hyperparameter matrix is empirically initialized as $\Gamma^{(0)} = \mathbf{I}_{G_\tau G_\nu}$ to enable accurate estimations. The initial CA estimation $A^{(0)}$ is obtained using the procedure of (35)-(37) by setting $\mathbf{h}_0 = \hat{\mathbf{h}}_{0,\text{MMSE}}$, where $\hat{\mathbf{h}}_{0,\text{MMSE}}$ is calculated by MMSE estimation using the original pilots, given by

$$\begin{aligned} \hat{\mathbf{h}}_{0,\text{MMSE}} &= \\ \sigma_v^{-2} \left[\sigma_v^{-2} \Phi_u(\mathbf{0})^H \Phi_u(\mathbf{0}) + \mathbf{I}_{G_\tau G_\nu} \right]^{-1} \Phi_u(\mathbf{0})^H \mathbf{y}_0. \end{aligned} \quad (39)$$

The iteration is stopped when either $i \geq i_{\max}$, or the convergency criteria $\text{diff} = \frac{\|\mu^{(i)} - \mu^{(i-1)}\|^2}{\|\mu^{(i)}\|^2} \leq \epsilon$ is reached.

Finally, we obtain the MAP estimation of the sparse channel as $\hat{\mathbf{h}}_{0,\text{SBL}} = \mu^{(i)}$, and the CA estimation $\hat{A}_{\text{SBL}} = A^{(i)}$.

B. Computational complexity and performance bounds

The computational complexity of **Algorithm 1** is dominated by the EM iterations. In the E-step, The complexity order of calculating $\Sigma^{(i)}$ and $\mu^{(i)}$ are $\mathcal{O}(G_\tau^3 G_\nu^3)$ and $\mathcal{O}(MNG_\tau G_\nu + G_\tau^2 G_\nu^2)$ for each iteration, respectively. In the M-step, the complexity order of calculating $\Gamma^{(i)}$ and $A^{(i)}$ are $\mathcal{O}(G_\tau G_\nu)$ and $\mathcal{O}(QMNG_\tau G_\nu)$ for each iteration, respectively. When the grid size is large, calculating $G_\tau G_\nu$

$$\begin{aligned} J_D^{22} &= \mathbb{E} \left[\frac{\partial \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)}{\partial A} \frac{\partial \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)^T}{\partial A} \right] \\ &= \frac{1}{\sigma_v^2} \left\{ \mathbf{h}_0^H [\mathbf{I}_{G_\tau G_\nu} \otimes (\mathbf{c}_0 \odot e^{j\phi_0})]^H \mathbf{W}^H \mathbf{W} [\mathbf{I}_{G_\tau G_\nu} \otimes (\mathbf{c}_0 \odot e^{j\phi_0})] \mathbf{h}_0 \right. \\ &\quad \left. + \mathbf{h}_0^T [\mathbf{I}_{G_\tau G_\nu} \otimes (\mathbf{c}_0 \odot e^{j\phi_0})]^T \mathbf{W}^T \mathbf{W}^* [\mathbf{I}_{G_\tau G_\nu} \otimes (\mathbf{c}_0 \odot e^{j\phi_0})]^* \mathbf{h}_0^* \right\}, \end{aligned} \quad (42)$$

dimension matrix inversion in the E-step will be challenging. Fortunately, existing studies have proposed low-complexity versions of SBL that avoid high-dimensional matrix inversions [41]– [43]. As a general solution, the proposed EM based algorithm is compatible to these methods, and the order of complexity can be reduced to $\mathcal{O}(G_\tau G_\nu \log(G_\tau G_\nu))$.

In order to evaluate the MSE performance of the proposed joint estimation method, we also derive the Bayesian Cramér-Rao bound (BCRB) [44]. The unknown complex and real parameters are denoted by a vector $\boldsymbol{\theta}_0 = [\mathbf{h}_0^T, \mathbf{h}_0^H, A]^T$. The Bayesian Fisher information matrix (FIM) is given by $\mathbf{J}_B = \mathbf{J}_D + \mathbf{J}_P$, where $\mathbf{J}_D = \mathbb{E} \left[\frac{\partial^2 \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0^H} \right]$ is the data information matrix, $\mathbf{J}_P = \mathbb{E} \left[\frac{\partial^2 \log p(\mathbf{y}_0; \boldsymbol{\gamma})}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0^H} \right]$ is the prior information matrix. The log probability density function (PDF) of the data is given by

$$\begin{aligned} \log p(\mathbf{y}_0; \boldsymbol{\theta}_0) &= \\ &- MN \log(\pi \sigma_v^2) - \frac{\|\mathbf{y}_0 - \Phi(\mathbf{s}_0, A) \mathbf{h}_0\|^2}{\sigma_v^2}. \end{aligned} \quad (40)$$

According to [45], the full data information matrix is derived as

$$\mathbf{J}_D = \left[\begin{array}{c|c} \mathbf{J}_D^{11} & \mathbf{J}_D^{12} \\ \hline \mathbf{J}_D^{21} & J_D^{22} \end{array} \right], \quad (41)$$

where J_D^{22} is a scalar given by (42) at the top of this page. The other parts are denoted as $\mathbf{J}_D^{11} = [\mathbf{A}_1 \quad \mathbf{A}_2^*], \mathbf{J}_D^{21} = [\mathbf{P}_D \quad \mathbf{P}_D^*], \mathbf{J}_D^{12} = \begin{bmatrix} \mathbf{P}_D^H \\ \mathbf{P}_D^T \end{bmatrix} = (\mathbf{J}_D^{21})^H$, where

$$\begin{aligned} \mathbf{A}_1 &= \mathbb{E} \left[\frac{\partial \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)}{\partial \mathbf{h}_0^*} \frac{\partial \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)^H}{\partial \mathbf{h}_0^*} \right] \\ &= \frac{1}{\sigma_v^2} (\mathbf{I}_{G_\tau G_\nu} \otimes \mathbf{z}_0)^H \mathbf{W}^H \mathbf{W} (\mathbf{I}_{G_\tau G_\nu} \otimes \mathbf{z}_0), \end{aligned} \quad (43)$$

$$\begin{aligned} \mathbf{A}_2 &= \mathbb{E} \left[\frac{\partial \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)}{\partial \mathbf{h}_0} \frac{\partial \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)^T}{\partial \mathbf{h}_0} \right] \\ &= \mathbf{0}_{G_\tau G_\nu \times G_\tau G_\nu}, \end{aligned} \quad (44)$$

$$\begin{aligned} \mathbf{P}_D &= \mathbb{E} \left[\frac{\partial \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)}{\partial A} \frac{\partial \log p(\mathbf{y}_0; \boldsymbol{\theta}_0)^H}{\partial \mathbf{h}_0^*} \right] \\ &= \frac{1}{\sigma_v^2} \mathbf{h}_0^H [\mathbf{I}_{G_\tau G_\nu} \otimes (\mathbf{c}_0 \odot e^{j\phi_0})]^H \mathbf{W}^H \mathbf{W} (\mathbf{I}_{G_\tau G_\nu} \otimes \mathbf{z}_0). \end{aligned} \quad (45)$$

Similarly, the prior information matrix is derived as

$$\mathbf{J}_P = \left[\begin{array}{c|c} \mathbf{J}_P^{11} & \mathbf{J}_P^{12} \\ \hline \mathbf{J}_P^{21} & J_P^{22} \end{array} \right], \quad (46)$$

where $\mathbf{J}_P^{11} = \begin{bmatrix} \boldsymbol{\Gamma}^{-1} & \mathbf{0}_{G_\tau G_\nu \times G_\tau G_\nu} \\ \mathbf{0}_{G_\tau G_\nu \times G_\tau G_\nu} & \boldsymbol{\Gamma}^{-1} \end{bmatrix}, \mathbf{J}_P^{21} = \mathbf{0}_{1 \times G_\tau G_\nu}, \mathbf{J}_P^{12} = \mathbf{0}_{G_\tau G_\nu \times 1}, J_P^{22} = 0$. Finally, the BCRB of CA and DD channel are derived as

$$\text{BCRB}(A) = [J_D^{22} - 2\text{Re}\{\mathbf{P}_D \mathbf{D}_1 \mathbf{P}_D^H + \mathbf{P}_D^* \mathbf{D}_2 \mathbf{P}_D^H\}]^{-1}, \quad (47)$$

$$\text{BCRB}(\mathbf{h}_0) =$$

$$\mathbf{D}_1 + (\mathbf{D}_1 \mathbf{P}_D^H + \mathbf{D}_2^* \mathbf{P}_D^T) \text{BCRB}(A) (\mathbf{P}_D \mathbf{D}_1^H + \mathbf{P}_D^* \mathbf{D}_2^T), \quad (48)$$

where

$$\begin{aligned} \mathbf{D}_1 &= [\mathbf{A}_1 + \boldsymbol{\Gamma}^{-1} - \mathbf{A}_2^* (\mathbf{A}_1^* + \boldsymbol{\Gamma}^{-1})^{-1} \mathbf{A}_2]^{-1} \\ &= (\mathbf{A}_1 + \boldsymbol{\Gamma}^{-1})^{-1}, \end{aligned} \quad (49)$$

and

$$\mathbf{D}_2 = -(\mathbf{A}_1^* + \boldsymbol{\Gamma}^{-1}) \mathbf{A}_2 \mathbf{D}_1 = \mathbf{0}_{G_\tau G_\nu \times G_\tau G_\nu}. \quad (50)$$

IV. CHANNEL TRACKING AND ITERATIVE DETECTION FOR CLIPPED OTFS

In Section III, an estimation scheme of channel and CA for the pilot block was proposed, of which the estimation error is bounded by the power of the clipped pilot block. To avoid the error accumulating over time, the estimated channel should be updated at each data block. Since the data blocks do not contain pilot symbols, tracking the time-varying channel also faces challenges. A straightforward approach is to repeat the procedure of joint channel and CA estimation in **Algorithm 1**, where the detected data symbols using the initial estimations are treated as new known symbols. However, this approach is computationally expensive and only exploits the observations from the current data block, which may lead to lower accuracy if the decisions are unreliable. On the other hand, linear equalization based iterative detection for clipped signal ignores the knowledge of the detected symbols, therefore the error performance can be improved.

In this section, we first propose to adopt a Kalman filter based approach to track the channel variations using the observations from all the blocks. Then, a minimum mean square error decision feedback blockwise equalization (MMSE-DFBE) based detector for clipped OTFS is proposed. Finally, by integrating both approaches, an EM based iterative tracking and detection (EM-ITD) algorithm is developed. The flowchart of our proposed algorithm is shown in Fig.5.

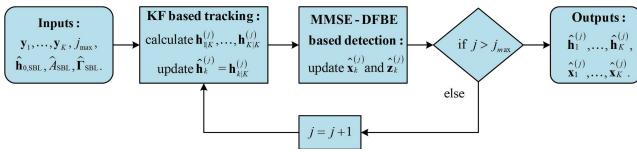


Fig. 5. Flowchart of the proposed iterative tracking and detection algorithm.

A. KF based channel tracking

In a practical high mobility multipath channel, the amplitude, delay, and Doppler shift of each path, as well as the CA, are approximately constant during a very short period of time, e.g. a frame duration in the order of milliseconds. Therefore, A is assumed to be identical for different blocks in a frame, and we consider a time correlation model of the channel where only the phase of each complex path varies with a rate depending on the Doppler shift. Specifically, under the frame structure shown in Fig.1, the complex path gain of the (m, n) -th bin of the k -th and the $(k+1)$ -th OTFS block have the following relation, given by

$$h_{k+1}^{(m,n)} = h_k^{(m,n)} e^{j \frac{2\pi N_\nu (MN + N_{CP}) \cdot n}{G_\nu MN}}. \quad (51)$$

In a Kalman filter framework, the arguments to be tracked is denoted by the state vector \mathbf{h}_k . Then the process model is given by the following difference equation as

$$\mathbf{h}_{k+1} = \mathbf{T}\mathbf{h}_k, \quad (52)$$

where \mathbf{T} is a diagonal matrix that relates the state at the k -th block to the state at the $(k+1)$ -th block. The expression of \mathbf{T} is given by (53) at the top of the next page. The measurement model is given as

$$\mathbf{y}_k = \mathbf{y}_{k|k-1} + \mathbf{E}_k (\mathbf{h}_k - \mathbf{h}_{k|k-1}) + \mathbf{v}_k, \quad (54)$$

where $\mathbf{y}_{k|k-1}$ and $\mathbf{h}_{k|k-1}$ are the priori mean of the measurement and the state vector, respectively. In the j -th iteration, \mathbf{E}_k is calculated by

$$\begin{aligned} \mathbf{E}_k &= \Phi \left(\hat{\mathbf{s}}_k^{(j-1)}, \hat{A}_{SBL} \right) \\ &= \mathbf{W} \left(\mathbf{I}_{G_\tau G_\nu} \otimes \hat{\mathbf{z}}_k^{(j-1)} \right). \end{aligned} \quad (55)$$

Given (52) and (54), the posterior mean and covariance using the observations from all the K blocks can be calculated recursively by applying the Kalman filter and smoother (KFS) update equations [46]. Firstly, the prediction equations for the k -th block are given as

$$\mathbf{h}_{k|k-1} = \mathbf{T}\mathbf{h}_{k-1|k-1}, \quad (56)$$

$$\mathbf{P}_{k|k-1} = \mathbf{T}\mathbf{P}_{k-1|k-1}\mathbf{T}^H, \quad (57)$$

where $\mathbf{P}_{k|k-1}$ is the priori covariance of the state, $\mathbf{h}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ are the posterior mean and covariance from the last block, respectively.

Next, the Kalman gain and filtering equations at the k -th block are given as

$$\mathbf{G}_k = \mathbf{P}_{k|k-1} \mathbf{E}_k^H (\mathbf{E}_k \mathbf{P}_{k|k-1} \mathbf{E}_k^H + \sigma_v^2 \mathbf{I}_{MN})^{-1}, \quad (58)$$

$$\mathbf{h}_{k|k} = \mathbf{h}_{k|k-1} + \mathbf{G}_k (\mathbf{y}_k - \mathbf{E}_k \mathbf{h}_{k|k-1}), \quad (59)$$

$$\mathbf{P}_{k|k} = (\mathbf{I}_{G_\tau G_\nu} - \mathbf{G}_k \mathbf{E}_k) \mathbf{P}_{k|k-1}. \quad (60)$$

Upon prediction and filtering from $k = 1$ to $k = K$, the posterior mean $\mathbf{h}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ from all the K blocks can be obtained. Finally, a smoother is employed to ensure that the observations from all the K blocks are included for each block, given as

$$\begin{cases} \mathbf{h}_{k-1|K} = \mathbf{T}^{-1} \mathbf{h}_{k|K}, \\ \mathbf{P}_{k-1|K} = \mathbf{P}_{k|K}. \end{cases} \quad (61)$$

B. MMSE-DFBE based iterative detection

For the detection of clipped OTFS signal under fast fading channel, a linear equalizer ignores using the decision for further interference and distortion cancellation. In this section, a novel MMSE-DFBE based detector for clipped OTFS is proposed, which minimizes the MSE by exploiting the decision feedback of the last iteration.

The general architecture of the proposed MMSE-DFBE detector during the j -th iteration is shown in Fig.6, which consists of an equalization part that cancels the interference and a decision part that cancels the distortion and feeds back the decisions.

The equalization part includes a block feedforward (BFF) equalizer $\mathbf{C}_k^{(j)} \in \mathbb{C}^{MN \times MN}$ and a block feedback (BFB) equalizer $\mathbf{B}_k^{(j)} \in \mathbb{C}^{MN \times MN}$. The equalization output is given by

$$\hat{\mathbf{z}}_k^{(j)} = \mathbf{C}_k^{(j)} \mathbf{y}_k + \mathbf{B}_k^{(j)} \hat{\mathbf{z}}_k^{(j-1)}. \quad (62)$$

The principle for equalizers design is to minimize the MSE with the transmit signal \mathbf{z}_k , given by

$$\begin{aligned} [\mathbf{C}_k^{(j)}, \mathbf{B}_k^{(j)}] &= \arg \min_{\mathbf{C}_k, \mathbf{B}_k} \mathbb{E} \left[\left| \hat{\mathbf{z}}_k^{(j)} - \mathbf{z}_k \right|^2 \right] \\ &= \arg \min_{\mathbf{C}_k, \mathbf{B}_k} \mathbb{E} \left[\left| \mathbf{C}_k^{(j)} \mathbf{y}_k + \mathbf{B}_k^{(j)} \hat{\mathbf{z}}_k^{(j-1)} - \mathbf{z}_k \right|^2 \right]. \end{aligned} \quad (63)$$

Denoting the powers of the transmitted and the detected signal from the last iteration as $\mathcal{P}_{\mathbf{z}_k} = \mathbb{E} \left[|\mathbf{z}_k|^2 \right]$ and $\mathcal{P}_{\hat{\mathbf{z}}_k^{(j-1)}} = \mathbb{E} \left[|\hat{\mathbf{z}}_k^{(j-1)}|^2 \right]$, respectively. Their correlation is denoted as $\mathcal{R}_{\mathbf{z}_k \hat{\mathbf{z}}_k^{(j-1)}} = \mathbb{E} \left[\mathbf{z}_k^* \hat{\mathbf{z}}_k^{(j-1)} \right]$. It should be noted that the exact value of $\mathcal{R}_{\mathbf{z}_k \hat{\mathbf{z}}_k^{(j-1)}}$ is unavailable since \mathbf{z}_k is unknown. In the practical iterations, a scaled correlation between ZF equalized signal and $\hat{\mathbf{z}}_k^{(j-1)}$ is used as an approximation, given by

$$\hat{\mathcal{R}}_{\mathbf{z}_k \hat{\mathbf{z}}_k^{(j-1)}} = \frac{\eta}{MN} \left[(\mathbf{H}_k^H \mathbf{H}_k)^{-1} \mathbf{H}_k^H \mathbf{y}_k \right]^H \hat{\mathbf{z}}_k^{(j-1)}, \quad (64)$$

where η is a factor that avoids an overestimate of $|\mathcal{R}_{\mathbf{z}_k \hat{\mathbf{z}}_k^{(j-1)}}|$. An overestimation may introduce more error and lead to worse equalization performance [47].

Upon introducing these expressions, the MSE is further denoted by (65) at the top of the next page.

Furthermore, we want to design $\mathbf{C}_k^{(j)}$ that preserves the desired part of the signal while $\mathbf{B}_k^{(j)}$ cancels the interference part. Therefore, the diagonal elements of $\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}$ and $\mathbf{B}_k^{(j)}$ must satisfy the zero-sum constraints, i.e.

$$\mathbf{T} = \mathbf{I}_{G_\tau} \otimes \text{diag} \left(e^{j \frac{2\pi N_\nu (MN + N_{CP}) \cdot 0}{G_\nu MN}}, \dots, e^{j \frac{2\pi N_\nu (MN + N_{CP}) \cdot (G_\nu - 1)}{G_\nu MN}} \right). \quad (53)$$

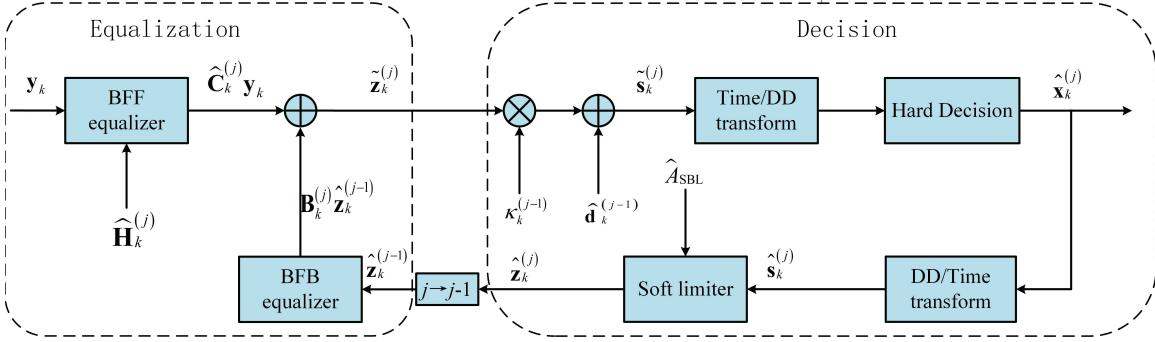


Fig. 6. Architecture of the MMSE-DFBE detector.

$$\begin{aligned} \mathcal{M}(\mathbf{B}_k^{(j)}, \mathbf{C}_k^{(j)}) &= \frac{1}{MN} \text{tr} \left\{ (\mathbf{C}_k^{(j)})^H \mathbf{C}_k^{(j)} \sigma_v^2 + (\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN})^H (\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}) \mathcal{P}_{\mathbf{z}_k} \right. \\ &\quad \left. + 2\text{Re} \left[(\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}) \mathbf{B}_k^{(j)} \mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}} \right] + (\mathbf{B}_k^{(j)})^H \mathbf{B}_k^{(j)} \mathcal{P}_{\widehat{\mathbf{z}}_k^{(j-1)}} \right\}. \end{aligned} \quad (65)$$

$\text{tr} [\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}] = 0$ and $\text{tr} [\mathbf{B}_k^{(j)}] = 0$. Then, we have the following theorem for the design of equalizers:

Theorem 1: The BFF and BFB equalizers that minimize $\mathcal{M}(\mathbf{B}_k^{(j)}, \mathbf{C}_k^{(j)})$ are given by

$$\mathbf{C}_k^{(j)} = \frac{\widehat{\mathbf{C}}_k^{(j)}}{\beta_k^{(j)}}, \quad (66)$$

$$\mathbf{B}_k^{(j)} = \frac{\mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}}}{\mathcal{P}_{\widehat{\mathbf{z}}_k^{(j-1)}}} (\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}), \quad (67)$$

where $\beta_k^{(j)} = \frac{1}{MN} \text{tr} [\widehat{\mathbf{C}}_k^{(j)} \mathbf{H}_k]$, and

$$\widehat{\mathbf{C}}_k^{(j)} = \left[\mathbf{H}_k^H \mathbf{H}_k \left(\mathcal{P}_{\mathbf{z}_k} - \frac{|\mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}}|^2}{\mathcal{P}_{\widehat{\mathbf{z}}_k^{(j-1)}}} \right) + \sigma_v^2 \mathbf{I}_{MN} \right]^{-1} \mathbf{H}_k^H. \quad (68)$$

Proof: Please refer to Appendix A. ■

In the first iteration when $j = 1$, the detection is unavailable and we set $\widehat{\mathcal{R}}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(0)}} = 0$. Thus we have $\mathbf{B}_k^{(1)} = \mathbf{0}_{MN}$, and

$$\mathbf{C}_k^{(1)} = \left[\mathbf{H}_k^H \mathbf{H}_k + \frac{\sigma_v^2}{\mathcal{P}_{\mathbf{z}_k}} \mathbf{I}_{MN} \right]^{-1} \mathbf{H}_k^H, \quad (69)$$

which is equivalent to the MMSE linear blockwise equalizer (MMSE-LBE).

In the decision part, we first use a distortion canceller similar to the approach in [32], where the output signal is given as

$$\widetilde{\mathbf{s}}_k^{(j)} = \frac{1}{\kappa_k^{(j-1)}} (\widetilde{\mathbf{z}}_k^{(j)} - \widehat{\mathbf{d}}_k^{(j-1)}), \quad (70)$$

where $\kappa_k^{(j-1)}$ is the correlation factor between distorted and undistorted signal, and $\widehat{\mathbf{d}}_k^{(j-1)}$ is the uncorrelated distortion

vector, which can be approximated using the decision feedbacks from the last iteration as

$$\kappa_k^{(j-1)} = \left| \frac{(\widehat{\mathbf{z}}_k^{(j-1)})^H \widehat{\mathbf{s}}_k^{(j-1)}}{(\widehat{\mathbf{s}}_k^{(j-1)})^H \widehat{\mathbf{s}}_k^{(j-1)}} \right|, \quad (71)$$

$$\widehat{\mathbf{d}}_k^{(j-1)} = \widehat{\mathbf{z}}_k^{(j-1)} - \kappa_k^{(j-1)} \widehat{\mathbf{s}}_k^{(j-1)}. \quad (72)$$

Next, the time domain signal is transformed into the DD domain, where the hard decision is performed, given as

$$\widehat{\mathbf{x}}_k^{(j)} = \langle (\mathbf{F}_N \otimes \mathbf{P}_{\text{rx}}) \widetilde{\mathbf{s}}_k^{(j)} \rangle, \quad (73)$$

where $\mathbf{P}_{\text{rx}} \in \mathbb{C}^{M \times M}$ is a diagonal matrix of receive pulse shaping.

Then, the unclipped version of the decision feedback is calculated by a DD/time transform as

$$\widehat{\mathbf{s}}_k^{(j)} = (\mathbf{F}_N^H \otimes \mathbf{P}_{\text{tx}}) \widehat{\mathbf{x}}_k^{(j)}. \quad (74)$$

Finally, the clipped version of the decision feedback is calculated using a soft limiter with estimated \widehat{A}_{SBL} as

$$\widehat{\mathbf{z}}_k^{(j)} = \mathbf{g}(\widehat{\mathbf{s}}_k^{(j)}; \widehat{A}_{\text{SBL}}). \quad (75)$$

During the first iteration, there is no feedback and we set $\kappa_k^{(0)} = 1$, and $\widehat{\mathbf{d}}_k^{(0)} = \mathbf{0}_{MN \times 1}$. The procedure can be simplified as a normal DD domain decision without distortion cancellation.

C. Iterative tracking and detection

We have proposed a channel tracking scheme and an iterative detection scheme. In order to take advantages of both of them, we then integrate them into an EM-based framework, in which the channel is updated using the newly detected

Algorithm 2 EM based iterative tracking and detection

Input: Received data block $\mathbf{y}_1, \dots, \mathbf{y}_K$, maximum iteration time j_{\max} .

Output: DD channel estimation at data blocks $\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_K$, detected data symbols $\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_K$.

- 1: Calculate \hat{A}_{SBL} , $\hat{\mathbf{h}}_{0,SBL}$, $\hat{\Gamma}_{SBL}$ using Algorithm 1.
- 2: Initialize $\mathbf{h}_{0|0}^{(1)} = \hat{\mathbf{h}}_{0,SBL}$, $\mathbf{P}_{0|0}^{(1)} = \hat{\Gamma}_{SBL}$, $\hat{\mathbf{z}}_1^{(0)} = \dots = \hat{\mathbf{z}}_K^{(0)} = \mathbf{0}_{MN \times 1}$, $\hat{\mathbf{h}}_1^{(0)} = \dots = \hat{\mathbf{h}}_K^{(0)} = \mathbf{0}_{G_\tau G_\nu \times 1}$.
- 3: Set $j = 1$.
- 4: **while** ($j \leq j_{\max}$) **do**
- 5: **E-step:**
- 6: **for** $k = 1, \dots, K$ **do**
- 7: Prediction:
- 8: $\mathbf{h}_{k|k-1}^{(j)} = \mathbf{T}\mathbf{h}_{k-1|k-1}^{(j)}$,
- 9: $\mathbf{P}_{k|k-1}^{(j)} = \mathbf{T}\mathbf{P}_{k-1|k-1}^{(j)}\mathbf{T}^H$.
- 10: Filtering:
- 11: $\mathbf{E}_k = \mathbf{W} \left(\mathbf{I}_{G_\tau G_\nu} \otimes \hat{\mathbf{z}}_k^{(j-1)} \right)$,
- 12: $\mathbf{G}_k = \mathbf{P}_{k|k-1}^{(j)} \mathbf{E}_k^H \left(\mathbf{E}_k \mathbf{P}_{k|k-1}^{(j)} \mathbf{E}_k^H + \sigma_v^2 \mathbf{I}_{MN} \right)^{-1}$,
- 13: $\mathbf{h}_{k|k}^{(j)} = \mathbf{h}_{k|k-1}^{(j)} + \mathbf{G}_k \left(\mathbf{y}_k - \mathbf{E}_k \mathbf{h}_{k|k-1}^{(j)} \right)$,
- 14: $\mathbf{P}_{k|k}^{(j)} = (\mathbf{I}_{G_\tau G_\nu} - \mathbf{G}_k \mathbf{E}_k) \mathbf{P}_{k|k-1}^{(j)}$.
- 15: **end for**
- 16: **for** $k = K, K-1, \dots, 1$ **do**
- 17: Smoothing:
- 18: $\mathbf{h}_{k-1|K}^{(j)} = \mathbf{T}^{-1} \mathbf{h}_{k|K}^{(j)}$,
- 19: $\mathbf{P}_{k-1|K}^{(j)} = \mathbf{P}_{k|K}^{(j)}$.
- 20: Update $\hat{\mathbf{H}}_k^{(j)}$ from $\mathbf{h}_{k|K}^{(j)}$.
- 21: **end for**
- 22: **M-step:**
- 23: **for** $k = 1, \dots, K$ **do**
- 24: Equalization:
- 25: calculate $\mathbf{C}_k^{(j)}$ and $\mathbf{B}_k^{(j)}$, and $\hat{\mathbf{z}}_k^{(j)}$ using (66), (67), and (62).
- 26: Decision: calculate $\hat{\mathbf{x}}_k^{(j)}$ and $\hat{\mathbf{z}}_k^{(j)}$ using the procedure of (70)-(75).
- 27: **end for**
- 28: Set $j = j + 1$.
- 29: **end while**
- 30: **for** $k = 1, \dots, K$ **do**
- 31: Set outputs $\hat{\mathbf{h}}_k = \hat{\mathbf{h}}_k^{(j)}$, $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^{(j)}$.
- 32: **end for**

symbols in the E-step, while the symbol detection is performed using the updated channel in the M-step. An initial update is obtained by the pilot based estimation of **Algorithm 1**. In this way, an EM-ITD approach is further developed, which is detailed as **Algorithm 2**, where j_{\max} is the maximum iteration time. Since the hyperparameters are identical for all blocks and have reached convergence in the initial estimation, very few iterations are required before realizing a good performance.

The complexity of the E-step of **Algorithm 2** is dominated by calculating the Kalman gain \mathbf{G}_k , where MN -dimensional matrix inversion is costly when MN is large. However, the sparsity of the prior covariance matrix $\mathbf{P}_{k|k-1}$ can be further

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	6 GHz
Maximum speed	1000 km/h
Maximum Doppler shift	5.54 kHz
Number of OTFS data blocks in each frame K	13
Subcarrier spacing Δf	120 kHz
Number of subcarriers M	32
Number of symbols in each block N	8
Number of samples in each pilot sub-block Q	32
Number of samples in CP N_{CP}	16
Modulation	QPSK
Channel fading model	TDL-A [49]
Desired delay spread $DS_{desired}$	363ns
Maximum normalized delay	14
Delay grid size G_τ	16
Maximum normalized Doppler	0.37
Doppler grid size G_ν	8

exploited, where the diagonal elements that approach zero can be ignored and only P elements are preserved, then the matrix inversion lemma can be applied to reduce the dimension of inversion to P [48]. In this way, the dominant complexity order of the Kalman filtering for each block can be reduced to $\mathcal{O}(M^2N^2)$. The complexity of the M-step is dominated by calculating the equalizers. Still, we can preserve the P dominant elements in each column of $\hat{\mathbf{H}}_k^{(j)}$, then the computational complexity of $\mathbf{B}_k^{(j)}$ and $\mathbf{C}_k^{(j)}$ are both $\mathcal{O}(PM^2N^2)$ for each iteration.

V. SIMULATION RESULTS

In this section, numerical simulations are conducted to validate the effectiveness of the proposed approaches. The performance using the MMSE estimator and the SBL estimator in [12] using both clipped and linear received signals are also included for comparison. The estimation errors of channel and CA are demonstrated in terms of NMSE, which are defined as $\text{NMSE}(\hat{\mathbf{h}}_k) = \frac{\|\mathbf{h}_k - \hat{\mathbf{h}}_k\|_2^2}{\|\mathbf{h}_k\|_2^2}$ and $\text{NMSE}(\hat{A}) = \frac{\|A - \hat{A}\|^2}{\|A\|^2}$, respectively. The results are obtained by averaging the NMSEs and BERs of 10000 random frames. Moreover, the estimation errors are benchmarked with respect to the BCRBs derived in Sec.III, while the bit errors are benchmarked with respect to those using known channels and CAs. Without loss of generality, quadrature phase shift keying (QPSK) modulation and rectangular pulse are uniformly used. The tapped delay line (TDL) channel model suggested by 3GPP [49] is used, with a desired delay spread of 363ns. The maximum moving speed is 1000km/h. More details for simulation parameters are shown in Table.I. The stopping parameters for SBL-based estimations and iterative detection are set as $\epsilon = 10^{-6}$, $i_{\max} = 50$, and $j_{\max} = 4$. The factor η is set as 0.7. These parameters are selected empirically to balance performance and complexity, and are consistent with common settings in related studies on SBL and DFE [12], [47]. In all simulations, the normalized average power of the input signal to the clipper is set equal to 1, the clipping amplitude and noise power are also normalized. Therefore, the CA and the signal to noise power ratio (SNR) in decibels are denoted as $20\log_{10}(A)[\text{dB}]$ and $-10\log_{10}(\sigma_v^2)[\text{dB}]$, respectively. Three different scenarios

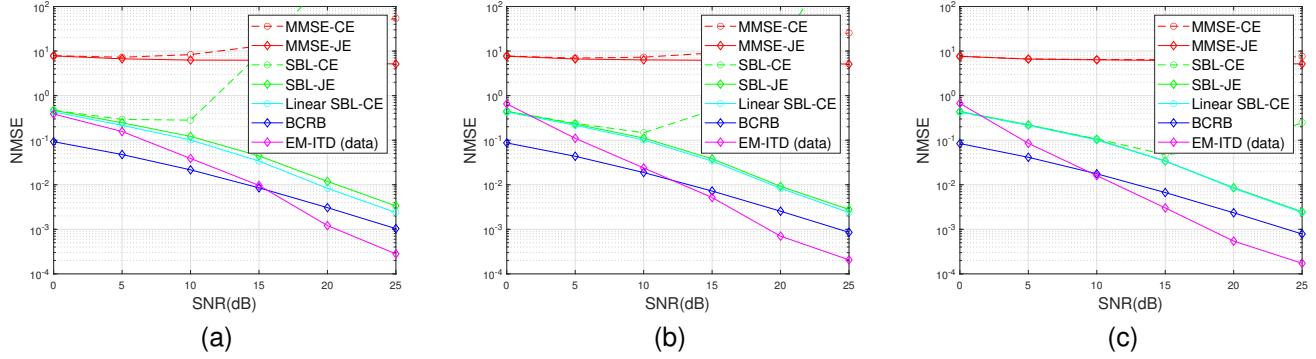


Fig. 7. NMSE of channel estimation versus SNR for different schemes when (a) CA=1 dB, (b) CA =3 dB, and (c) CA = 5 dB.

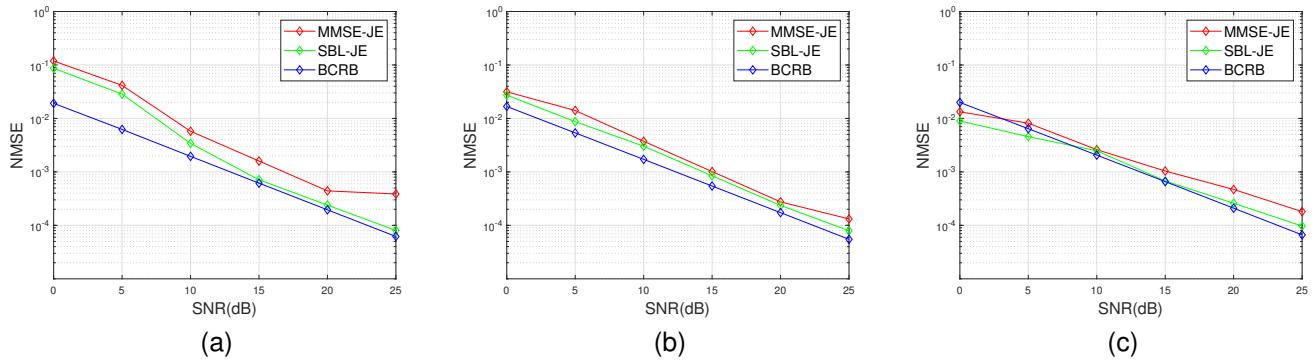


Fig. 8. NMSE of CA estimation versus SNR for different schemes when (a) CA=1 dB, (b) CA =3 dB, and (c) CA = 5 dB.

are considered in our simulations, where CA are set equal to 1dB, 3dB, and 5dB. The normalized peak amplitude of the transmit signal is around 9dB, therefore the degradations by using additional input backoff (IBO) for unclipped linear PA outputs are around 8dB, 6dB, and 4dB for the above scenarios.

Fig.7 demonstrates the NMSE of DD channel estimation versus SNR when CA = 1dB, 3dB, and 5dB. The curves "SBL-JE" and "EM-ITD" represent the proposed approaches using **Algorithm 1** and **Algorithm 2**, respectively. The curves "MMSE-CE" and "SBL-CE" represent the MMSE and SBL based DD channel estimation in [12] without CA estimation, while "Linear SBL-CE" represents the same scheme under linear PA outputs. The curve "MMSE-JE" represents the MMSE-based channel estimation with CA estimation using an alternating optimization method similar to [31]. It is observed that the MMSE-based approaches perform poor NMSE at all SNRs, since they ignore the sparsity of the DD-domain channel. For the "SBL-CE" curve, it is worth noting that the NMSE increases as SNR increases at high SNR. This is because the clipped pilot can be regarded as adding a fixed distortion to the original pilot. At high SNR, the distortion part dominates the total power of distortion plus noise. However, the SBL based channel estimation only brings the noise power into calculation and cannot handle the distortion. In that case, a higher SNR leads to more severe overestimates of the signal to noise plus distortion power ratio (SNDR) in the E-step. SBL is susceptible to this error and fails to converge to the

correct estimates. From Fig.7(a) to Fig.7(c), we note that as CA increases, the turning point of the NMSE curve comes at higher SNR due to the lower distortion part power. This issue is fixed in the case of the proposed joint estimation by using the estimated clipped pilots, where the effect of distortion is greatly reduced. The NMSE curves of the proposed SBL-JE are shown to be close to the SBL estimations with linear PAs, and are also shown to have smaller gaps with the BCRBs. Furthermore, lower NMSEs can be obtained by using the proposed EM-ITD, since the channel tracking exploits the observations from all the data blocks to further eliminate the influence of noise. Specifically, by using EM-ITD, the required SNR for a target channel estimation NMSE of 10^{-3} are reduced by 6dB, 8dB, and 10dB for CA = 1dB, 3dB, and 5dB scenarios, respectively.

The NMSE of CA estimation versus SNR when CA = 1dB, 3dB, and 5dB scenarios are shown in Fig.8. It is observed that in all scenarios the SBL-JE curves approach the BCRB. From Fig.8(c), we note that when CA= 5dB, the NMSE of the proposed JE is lower than the calculated BCRB at low SNR. In fact, the estimated \hat{A}_q is restricted to a fixed interval as shown by (36). The BCRB calculates the lower bound NMSE for \tilde{A}_q in (35) that is not restricted, which may be higher than that of the restricted CA estimation error when the SNR is low. On the other hand, the SBL-JE shows lower NMSE than MMSE-JE in all scenarios. This is because the MMSE based channel estimation is not available, which further affects the

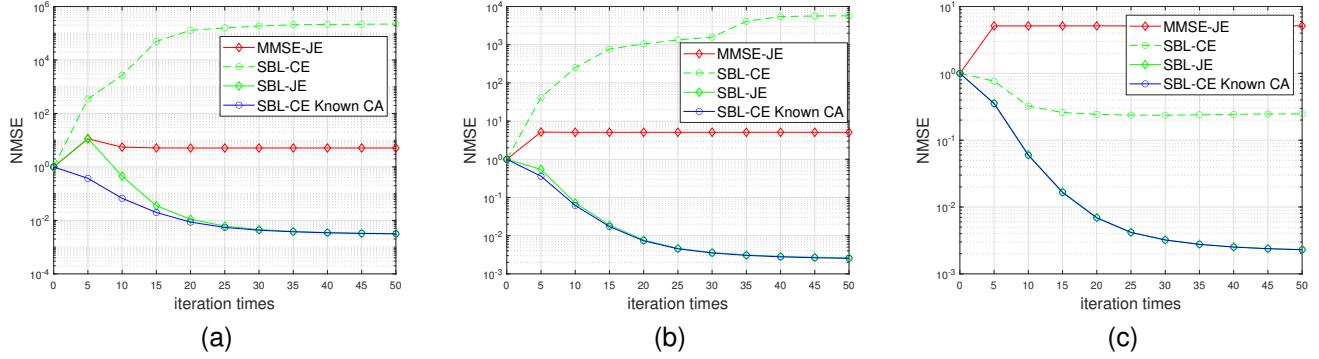


Fig. 9. NMSE of channel estimation versus iteration times for different schemes when (a) CA=1 dB, (b) CA =3 dB, and (c) CA = 5 dB.

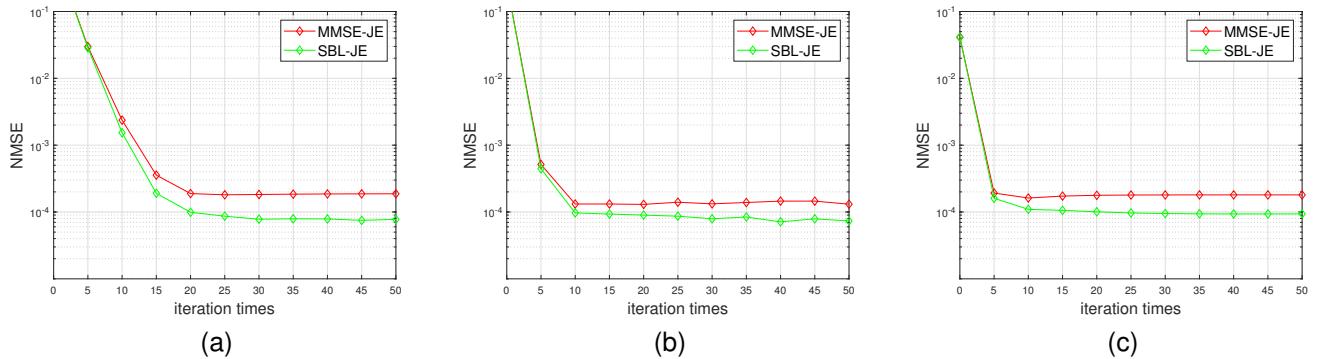


Fig. 10. NMSE of CA estimation versus iteration times for different schemes when (a) CA=1 dB, (b) CA =3 dB, and (c) CA = 5 dB.

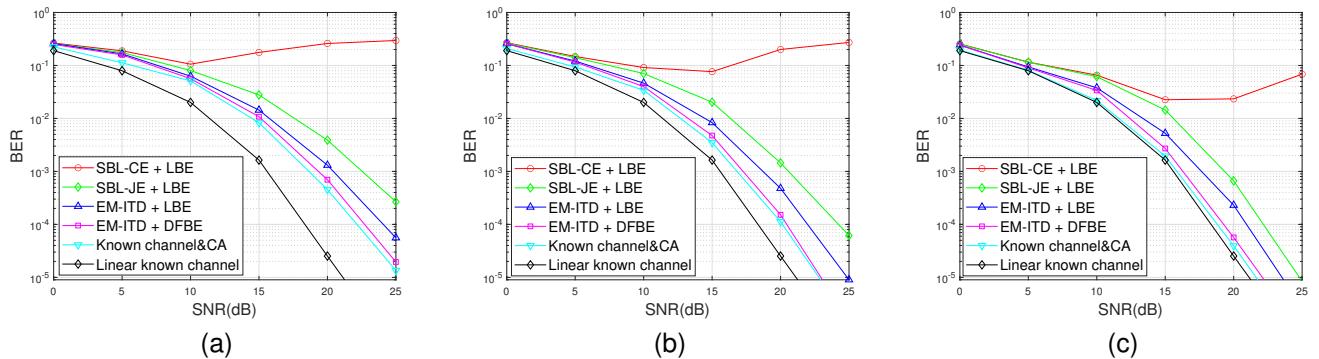


Fig. 11. BER versus SNR for different schemes when (a) CA=1 dB, (b) CA =3 dB, and (c) CA = 5 dB.

accuracy of the alternating optimization method.

Fig.9 and Fig.10 show the NMSEs of channel and CA estimations with respect to iteration times i for different scenarios. The results are obtained by averaging 10000 different channel realizations at SNR = 25 dB. The NMSE of DD channel using SBL-CE while assuming known CA is also included. It is noted that the NMSEs of the proposed SBL-JE decrease with the number of iterations for all scenarios, therefore verifying its convergence. The convergence rate of channel estimation is almost the same as SBL-CE with known CA, which converges after about 30 iterations for all scenarios. Furthermore, existing iterative schemes are shown unable to converge to correct

channel estimates. On the other hand, the CA estimations converge after 20, 15, and 5 iterations for CA= 1dB, 3dB, and 5dB scenarios, respectively.

Fig.11 demonstrates the BERs by using different receivers for the three scenarios. The BER curves of both linear and nonlinear PA outputs using the proposed DFBE based detector with known channel and CA are also plotted for comparison. It is found that the BER by using SBL-CE degrades severely in all scenarios. This is mainly due to its large channel estimation error without CA estimation as we have presented earlier. Therefore, it is necessary to conduct CA estimation, even under a mild distortion at CA= 5dB. It is also observed that

TABLE II
COMPLEXITY AND AVERAGE RUNNING TIME FOR DIFFERENT SCHEMES.

Receiver procedures	Methods	Complexity(per iteration)	Running time
Initial estimation at the pilot block	MMSE-CE	$\mathcal{O}(G_\tau^3 G_\nu^3)$ (linear)	0.006s
	SBL-CE	$\mathcal{O}(G_\tau^3 G_\nu^3 + G_\tau^2 G_\nu^2 + MNG_\tau G_\nu + G_\tau G_\nu) \rightarrow \mathcal{O}(G_\tau^3 G_\nu^3 + MNG_\tau G_\nu)$	0.006s
	MMSE-JE (alternating optimization)	$\mathcal{O}(G_\tau^3 G_\nu^3 + QMNG_\tau G_\nu)$	0.008s
	SBL-JE (proposed)	$\mathcal{O}(G_\tau^3 G_\nu^3 + G_\tau^2 G_\nu^2 + MNG_\tau G_\nu + G_\tau G_\nu + QMNG_\tau G_\nu) \rightarrow \mathcal{O}(G_\tau^3 G_\nu^3 + (Q+1)MNG_\tau G_\nu)$	0.008s
Iterative tracking and detection at the data blocks	KF based tracking (proposed)	dominated by $\mathcal{O}(M^2 N^2)$	0.001s
	MMSE-LBE based detector	dominated by $\mathcal{O}(PM^2 N^2)$	0.007s
	MMSE-DFBE based detector (proposed)	dominated by $\mathcal{O}(PM^2 N^2)$	0.009s

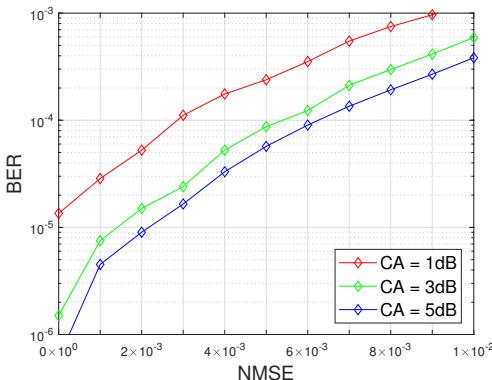


Fig. 12. BER versus estimation NMSE when SNR = 25 dB.

The EM-ITD holds lower BER than SBL-JE due to the more accurate channel estimations. Moreover, we note that DFBE also outperforms LBE, since it exploits the correlation of the received and the detected symbols for further interference cancellation. By approving the proposed iterative tracking and DFBE, the required SNR for “EM-ITD+DFBE” is 4 dB lower than that for “SBL-JE+LBE” at a target BER of 10^{-4} . The BER of “EM-ITD+DFBE” is very close to that of DFBE with known channel and CA, with an almost negligible gap. These results validate the superiority of our proposed approaches. Moreover, with the proposed detector, the performance gaps between clipped and linear PA output signals are around 4dB, 2.5dB, and 0.5dB for three scenarios, which are significantly lower than the required additional IBOs for linear PA outputs. Therefore, the proposed receiver can improve the efficiency of the PA by decreasing IBOs.

To verify the influence of initial estimation accuracy, simulation results of BER of the proposed EM-ITD under varying channel estimation NMSEs are shown in Fig. 12. The SNR is set equal to 25dB. It can be observed that the BERs increase quickly as the NMSE increases from 0 to 10^{-2} . This is because

inaccurate initial estimation may lead to unreliable decision feedback on data blocks. As a result, the iterative channel tracking and DFBE cannot guarantee the system performance due to error propagation. Therefore, accurate initial estimation is crucial for the receiver of clipped OTFS.

The computational complexity and CPU running times for the proposed and relevant schemes are summarized in Table. II. The running times are averaged from 1000 runs, which are performed in MATLAB R2022b using a 2.10 GHz Intel Core i7 CPU with 8 GB of memory storage. For the pilot based initial estimation, each EM iteration in the proposed SBL-JE has the same complexity order as the MMSE-JE. When compared to SBL-CE, the proposed method introduces an additional estimation of CA in the M-step, which slightly increases complexity, and the average run time for each iteration is increased by approximately 30 percent. In each frame, the iterative initial estimation is only executed at the pilot block, then the proposed KF based tracking provides efficient acquisition of channel response at the data blocks and has little impact on the total running time. Therefore the total processing time for each frame is not significantly increased when compared to non-iterative estimation. For the detection of data blocks, the proposed MMSE-DFBE has the same complexity order as the conventional MMSE-LBE, and the running time is slightly increased by additionally calculating $B_k^{(j)}$ in each iteration. In conclusion, the proposed receiver schemes achieve better performance while maintaining the complexity and runtime at a similar level as conventional schemes. Therefore, we believe applying the proposed schemes can be efficient and beneficial in real-world deployment.

VI. CONCLUSION

In this paper, we studied the channel and CA estimation, channel tracking, and signal detection techniques for clipped OTFS systems. Based on an efficient frame structure, a framework for joint channel and CA estimation, channel tracking,

$$\begin{aligned} \mathcal{G}\left(\mathbf{B}_k^{(j)}, \mathbf{C}_k^{(j)}, \lambda^{(j)}, \mu^{(j)}\right) &= \frac{1}{MN} \operatorname{tr} \left\{ \left(\mathbf{C}_k^{(j)}\right)^H \mathbf{C}_k^{(j)} \sigma_v^2 + \left(\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}\right)^H \left(\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}\right) \mathcal{P}_{\mathbf{z}_k} \right. \\ &\quad \left. + 2\operatorname{Re} \left[\left(\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}\right) \mathbf{B}_k^{(j)} \mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}} \right] + \left(\mathbf{B}_k^{(j)}\right)^H \mathbf{B}_k^{(j)} \mathcal{P}_{\widehat{\mathbf{z}}_k^{(j-1)}} + \lambda^{(j)} \mathbf{B}_k^{(j)} + \mu^{(j)} \left(\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}\right) \right\}, \end{aligned} \quad (76)$$

and symbol detection was developed. Additionally, an SBL-based joint channel and CA estimation algorithm for the pilot block was proposed, which exploits the sparsity of the DD domain channel to estimate the channel and CA iteratively under the SBL framework. A specialized DD domain pilot was also provided to reduce the complexity of the proposed approach. We also derived the BCRB of the joint estimation algorithm as a performance benchmark. Furthermore, we proposed an iterative tracking and detection algorithm for the data blocks, where an KF based channel tracking and an MMSE-DFBE based signal detection were derived and integrated into an EM-based framework. Our simulation results on estimation NMSE and BER demonstrated the benefits of our proposed approaches over the existing approaches.

APPENDIX A PROOF OF THE THEOREM-1

To realize the equalizer design that minimizes the MSE under the given constraints, the Lagrange multiplier method is adopted. The Lagrangian is written as (76) at the top of this page, where $\lambda^{(j)}$ and $\zeta^{(j)}$ are the Lagrange multipliers. By setting to zero the derivatives of $\mathcal{G}\left(\mathbf{B}_k^{(j)}, \mathbf{C}_k^{(j)}, \lambda^{(j)}, \zeta^{(j)}\right)$ with respect to $\mathbf{B}_k^{(j)}$, $\mathbf{C}_k^{(j)}$, $\lambda^{(j)}$, and $\zeta^{(j)}$, the Karush-Kuhn-Tucker (KKT) conditions are obtained as [50]

$$2\left(\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}\right) \mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}} + 2\mathbf{B}_k^{(j)} \mathcal{P}_{\widehat{\mathbf{z}}_k^{(j-1)}} = -\lambda^{(j)} \mathbf{I}_{MN}, \quad (77)$$

$$\begin{aligned} \mathbf{C}_k^{(j)} \sigma_v^2 + 2\left(\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN}\right) \mathbf{H}_k^H \mathcal{P}_{\mathbf{z}_k} + 2\mathbf{B}_k^{(j)} \mathbf{H}_k^H \mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}} \\ = -\zeta^{(j)} \mathbf{H}_k^H, \end{aligned} \quad (78)$$

$$\operatorname{tr} \left[\mathbf{B}_k^{(j)} \right] = 0, \quad (79)$$

$$\operatorname{tr} \left[\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN} \right] = 0. \quad (80)$$

From (77) we obtain

$$\mathbf{B}_k^{(j)} = \frac{\mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}}}{\mathcal{P}_{\widehat{\mathbf{z}}_k^{(j-1)}}} \left(\mathbf{C}_k^{(j)} \mathbf{H}_k - \alpha_k^{(j)} \mathbf{I}_{MN} \right), \quad (81)$$

where $\alpha_k^{(j)}$ is a scalar. We then search for a suitable $\alpha_k^{(j)}$ so that (79) and (80) hold. By taking the trace of (81) and substituting (79) and (80) into both sides, we obtain $\alpha_k^{(j)} = 1$ and

$$\mathbf{B}_k^{(j)} = \frac{\mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}}}{\mathcal{P}_{\widehat{\mathbf{z}}_k^{(j-1)}}} \left(\mathbf{C}_k^{(j)} \mathbf{H}_k - \mathbf{I}_{MN} \right). \quad (82)$$

By substituting (81) into (78), we obtain

$$\mathbf{C}_k^{(j)} = \frac{\widehat{\mathbf{C}}_k^{(j)}}{\beta_k^{(j)}}, \quad (83)$$

where $\beta_k^{(j)}$ is also a scalar and

$$\widehat{\mathbf{C}}_k^{(j)} = \left[\mathbf{H}_k^H \mathbf{H}_k \left(\mathcal{P}_{\mathbf{z}_k} - \frac{\left| \mathcal{R}_{\mathbf{z}_k \widehat{\mathbf{z}}_k^{(j-1)}} \right|^2}{\mathcal{P}_{\widehat{\mathbf{z}}_k^{(j-1)}}} \right) + \sigma_v^2 \mathbf{I}_{MN} \right]^{-1} \mathbf{H}_k^H. \quad (84)$$

Finally we search for a suitable $\beta_k^{(j)}$ so that (80) holds. Upon substituting (83) into (80) we obtain

$$\beta_k^{(j)} = \frac{1}{MN} \operatorname{tr} \left[\widehat{\mathbf{C}}_k^{(j)} \mathbf{H}_k \right]. \quad (85)$$

REFERENCES

- [1] R. Hadani et al., "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, 2017, pp. 1–6.
- [2] K. R. Murali and A. Chockalingam, "On OTFS modulation for high-Doppler fading channels," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2018, pp. 1–10.
- [3] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501–6515, Oct. 2018.
- [4] P. Raviteja, Y. Hong, et al., "Practical pulse-shaping waveforms for reduced-cyclic-prefix otfs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 957–961, Jan. 2019.
- [5] Z. Wei, W. Yuan, et al., "Transmitter and receiver window designs for orthogonal time-frequency space modulation," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2207–2223, Apr. 2021.
- [6] H. Zhang, X. Huang, et al., "Adaptive transmission with frequency-domain precoding and linear equalization over fast fading channels," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7420–7430, Nov. 2021.
- [7] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay-Doppler channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906–4917, May. 2019.
- [8] L. Zhao, W. Gao, and W. Guo, "Sparse Bayesian learning of delay-Doppler channel for OTFS system," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2766–2769, Dec. 2020.
- [9] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, "Off-grid channel estimation with sparse Bayesian learning for OTFS systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7407–7426, Sep. 2022.
- [10] Q. Wang, Y. Liang, Z. Zhang, and P. Fan, "2D off-grid decomposition and SBL combination for OTFS channel estimation," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3084–3098, May. 2023.
- [11] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Bayesian learning aided sparse channel estimation for orthogonal time frequency space modulated systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8343–8348, Aug. 2021.
- [12] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Bayesian learning aided simultaneous row and group sparse channel estimation in orthogonal time frequency space modulated MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 635–648, Jan. 2022.
- [13] H.-T. Sheng and W.-R. Wu, "Time-frequency domain channel estimation for OTFS systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 937–948, Feb. 2024.
- [14] S. Tiwari, S. Das, et al., "Low complexity Immse receiver for otfs," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2205–2209, Dec. 2019.
- [15] G. Surabhi and A. Chockalingam, "Low-complexity linear equalization for otfs modulation," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 330–334, Feb. 2020.
- [16] C. Jin, Z. Bie, et al., "A simple two-stage equalizer for otfs with rectangular windows," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1158–1162, Apr. 2021.

- [17] S. Li, W. Yuan, et al., "Cross domain iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2227–2242, Apr. 2022.
- [18] H. Wen, W. Yuan, et al., "Mf-oamp-based joint channel estimation and data detection for ofts systems," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2948–2953, Feb. 2024.
- [19] L. Xiao, S. Li, Y. Qian, D. Chen, and T. Jiang, "An overview of OTFS for Internet of Things: Concepts, benefits, and challenges," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7596–7618, May. 2022.
- [20] R. Marsalek, J. Blumenstein, D. Schützenhöfer, and M. Pospisil, "OTFS modulation and influence of wideband RF impairments measured on a 60 GHz testbed," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun.*, Atlanta, GA, USA, 2020, pp. 1–5.
- [21] S. Sharma, A. Singh, K. Deka and C. Adjih, "Impact of nonlinear power amplifier on BER performance of OTFS modulation," in *2023 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, Jaipur, India, 2023, pp. 627–632.
- [22] S. G. Neelam and P. R. Sahu, "Error performance of OTFS in the presence of IQI and PA nonlinearity," in *Proc. Nat. Conf. Commun.*, 2020, pp. 1–6.
- [23] Y. Chen, H. Liu, Z. Lu and H. Wang, "Block scalable OFDM-Based waveform for NTN uplink—A comparative study of multiple candidate waveforms over NTN scenarios using NR numerology," *IEEE Access.*, vol. 12, pp. 31262–31278, 2024.
- [24] P. Priya, C. S. Reddy, and D. Sen, "Channel estimator and nonlinear detector for Mmwave beamformed OTFS systems in high mobility scenarios," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 11 698– 11 713, Sep. 2023.
- [25] A. N. D'Andrea, V. Lottici, and R. Reggiannini, "RF power amplifier linearization through amplitude and phase predistortion," *IEEE Trans. Commun.*, vol. 44, pp. 1477–1484, Nov. 1996.
- [26] W. G. Jeon, K. H. Chang, and Y. S. Cho, "An adaptive data predistorter for compensation of nonlinear distortion in OFDM systems," *IEEE Trans. Commun.*, vol. 45, pp. 1167–1171, Oct. 1997.
- [27] H. Ochiai and H. Imai, "Performance analysis of deliberately clipped OFDM signals," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2270–2277, Nov. 2000.
- [28] R. Raich, H. Qian, and G. T. Zhou, "Optimization of SNDR for amplitude-limited nonlinearities," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1964–1972, Nov. 2005.
- [29] T. Jiang, C. Li, and C. Ni, "Effect of PAPR reduction on spectrum and energy efficiencies in OFDM systems with class-A HPA over AWGN channel," *IEEE Trans. Broadcast.*, vol. 59, no. 3, pp. 513–519, Sep. 2013.
- [30] E. Olfat and M. Bengtsson, "Estimation of the clipping level in OFDM systems," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, 2015, pp. 1169–1173.
- [31] E. Olfat and M. Bengtsson, "Joint channel and clipping level estimation for OFDM in IoT-based networks," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4902–4911, 2017.
- [32] J. Tellado, L. M. C. Hoo, and J.M. Cioffi, "Maximum-likelihood detection of nonlinearly distorted multicarrier symbols by iterative decoding," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 218–228, Feb. 2003.
- [33] W. Wellisch, I. Ulian, and A. Barreto, "Iterative correction of clipped OFDM signals with unknown clipping levels," in *Vehicular Technology Conference (VTC 2013-Spring)*, no. 3, 2013.
- [34] H. Chen and A. M. Haimovich, "Iterative estimation and cancellation of clipping noise for OFDM signals," *IEEE Commun. Lett.*, vol. 7, no. 7, pp. 305–307, Jul. 2003.
- [35] M. Wachowiak and P. Kryszkiewicz, "Clipping noise cancellation receiver for the downlink of massive MIMO OFDM system," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 6061–6073, Oct. 2023.
- [36] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [37] D. Declercq and G. B. Giannakis, "Recovering clipped OFDM symbols with Bayesian inference," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2000, vol. 1, pp. 157–160.
- [38] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, no. 3, pp. 211–244, Sep. 2001.
- [39] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
- [40] C. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.
- [41] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "A GAMP-based low complexity sparse Bayesian learning algorithm," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 294–308, Jan. 2018.
- [42] H. Duan, L. Yang, J. Fang, and H. Li, "Fast inverse-free sparse Bayesian learning via relaxed evidence lower bound maximization," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 774–778, Jun. 2017.
- [43] A. Lin, A. H. Song, B. Bilgic, and D. Ba, "Bayesian bounds for parameter estimation and nonlinear filtering/tracking," Wiley-IEEE press New York, 2007.
- [44] Online:https://www.ele.uri.edu/faculty/kay/New%20web/downloadable%20files/Nageha_complex%20CRLB.pdf
- [45] Online:http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf
- [46] N. Benvenuto and S. Tomasin, "Iterative design and detection of a DFE in the frequency domain," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1867–1875, Nov. 2005.
- [47] M. Raitoharju and R. Piché, "On computational complexity reduction methods for Kalman filter extensions," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 10, pp. 2–19, Oct. 2019.
- [48] Study on channel model for frequencies from 0.5 to 100 GHz., document TS 38.901 v16.1.0. 3GPP, Dec, 2019.
- [49] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.:Cambridge Univ. Press, 2004.



Yufan Chen received his B.S. degree in information engineering from the Beijing Institute of Technology (BIT), Beijing, China, in 2018. He is currently pursuing the Ph.D. degree in information and communication systems from BIT. His current interests include high mobility communication systems, waveform design, channel estimation and signal detection.



Dongxuan He received the B.S. degree in Automation from Beijing Institute of Technology (BIT) in 2013, and received Ph.D. degree in Information and Communication Systems from BIT in 2019. From 2017 to 2018, he was a visiting student in Singapore University of Technology and Design (SUTD). From 2019 to 2022, he was a postdoctoral researcher in the Department of Electronic Engineering, Tsinghua University. He is currently an assistant professor in the School of Information and Electronics, BIT. His current interests include Integrated sensing and communication (ISAC), Terahertz communication, and AI empowered wireless communications. He is currently serving as a Guest Editor for the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, the ELECTRONICS, and the SPACE: SCIENCE & TECHNOLOGY. He was a recipient of the Best Paper Award from 2024 IEEE ICSDP, 2025 IEEE IWCMC. He was also an Exemplary Reviewer of IEEE Wireless Communications Letters.



Hua Wang received his Ph.D. degree from Beijing Institute of Technology (BIT), Beijing, China, in 1999. He is now a Professor with School of Information and Electronics, BIT. From Feb. 2009 to Jan. 2010, he was a visiting professor with Department of Electrical Engineering, Arizona State University, USA. His research interests are in the fields of communication theory and signal processing, wireless networking, modem design and implementation for satellite communication.



Weijie Yuan (Senior Member, IEEE) is now an Assistant Professor with the Southern University of Science and Technology. His research interests include Integrated Sensing and Communications (ISAC), Orthogonal Time Frequency Space (OTFS), and the Low-Altitude Wireless Networks (LAWN). He currently serves as an Editor for the IEEE Transactions on Communications, IEEE Transactions on Wireless Communications, IEEE Transactions on Mobile Computing, IEEE Communications Magazine, IEEE Communications Standards Magazine, IEEE Transactions on Green Communications and Networking, IEEE Communications Letters, and IEEE Open Journal of Communications Society, a Guest Editor for the IEEE Transactions on Vehicular Technology, IEEE Transactions on Network Science and Engineering, and IEEE Internet of Things Journal. He was the Track-Chair for IEEE ICC 2025 and IEEE VTC 2025-Spring. He served as an Organizer/the Chair of several workshops and special sessions in flagship IEEE and ACM conferences, including IEEE ICC, IEEE VTC, IEEE GlobeCom, IEEE/CIC ICCC, IEEE SPAWC, IEEE WCNC, IEEE ICASSP, and ACM MobiCom. He is the Founding Chair of the IEEE ComSoc Special Interest Group (SIG) on LAWN as well as the SIG on OTFS. He was listed among the World's Top 2% Scientists by Stanford University for citation impact from 2021 to 2024, and among the Elsevier Highly-Cited Chinese Researchers. He was a recipient of the Best Editor from IEEE CommL, the Best Paper Award from IEEE ICC 2023, IEEE/CIC ICCC 2023, and IEEE GlobeCom 2024, as well as the 2025 IEEE Communications Society & Information Theory Society Joint Paper Award.



Tony Q.S. Quek (S'98-M'08-SM'12-F'18) received the B.E. and M.E. degrees in electrical and electronics engineering from the Tokyo Institute of Technology in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2008. Currently, he is the Associate Provost (AI & Digital Innovation) and Cheng Tsang Man Chair Professor with Singapore University of Technology and Design (SUTD). He also serves as the Director of the Future Communications R&D

Programme, the ST Engineering Distinguished Professor, and the AI-on-RAN Working Group Chair in AI-RAN Alliance. His current research topics include wireless communications and networking, network intelligence, non-terrestrial networks, open radio access network, and 6G.

Dr. Quek was honored with the 2008 Philip Yeo Prize for Outstanding Achievement in Research, the 2012 IEEE William R. Bennett Prize, the 2015 SUTD Outstanding Education Awards – Excellence in Research, the 2016 IEEE Signal Processing Society Young Author Best Paper Award, the 2017 CTTC Early Achievement Award, the 2017 IEEE ComSoc AP Outstanding Paper Award, the 2020 IEEE Communications Society Young Author Best Paper Award, the 2020 IEEE Stephen O. Rice Prize, the 2020 Nokia Visiting Professor, the 2022 IEEE Signal Processing Society Best Paper Award, the 2024 IIT Bombay International Award For Excellence in Research in Engineering and Technology, and the IEEE Communications Society WTC Recognition Award 2024. He is an IEEE Fellow, a WWRF Fellow, and a Fellow of the Academy of Engineering Singapore.