Uplink Multi-User OTFS: Transmitter Design Based on Statistical Channel Information

Mingcheng Nie, *Graduate Student Member, IEEE*, Shuangyang Li, *Member, IEEE*, Deepak Mishra, *Senior Member, IEEE*, Jinhong Yuan, *Fellow, IEEE*, and Derrick Wing Kwan Ng, *Fellow, IEEE*

***Abstract*— Orthogonal time frequency space (OTFS) has been widely acknowledged as a promising wireless technology for challenging transmission scenarios, including high-mobility channels. In this paper, we investigate the uplink multi-user OTFS transmission designs based on statistical channel information. Specifically, we investigate the pilot power allocation based on the *a priori* statistical channel state information (CSI) only, where performance on channel estimation is considered. We first derive the *a posteriori* Cramér-Rao bound (PCRB) based on the *a priori* channel information of each user. We unveil that the PCRB only relates to the user's pilot signal-to-noise ratio (SNR) and the maximum of delay and Doppler shifts under the practical power-delay and power-Doppler profiles. Furthermore, a pilot power allocation scheme is proposed to minimize the average PCRB of different users, whose closed-form optimal allocation solution is derived. Moreover, we study the impact of statistical CSI on transmission rates, where a tight approximation of the sum-rate is derived. Particularly, the approximated sum-rate only relates to the user's symbol SNR and the maximum of delay and Doppler shifts. More importantly, we propose a power allocation for different users based only on the statistical CSI to maximize the achievable sum-rate while ensuring user fairness. The optimal power allocation solution is obtained by a fractional programming approach. Our numerical results verify the derived PCRB and the sum-rate analysis, where a roughly 3 dB improvement in terms of channel estimation accuracy and a significant rate improvement can be obtained.**

***Index Terms*— OTFS modulation, channel estimation, achievable rates, high-mobility, power allocation, imperfect CSI.**

Received 15 April 2024; revised 4 September 2024 and 9 November 2024; accepted 5 December 2024. Date of publication 12 December 2024; date of current version 17 July 2035. The work of S. Li is supported in part by the European Union's Horizon 2020 Research and Innovation Program under MSCA Grant No. 101105732 - DDComRad. The work of D. Mishra is supported in part by the Australian Research Council Discovery Early Career Researcher Award (DECRA) - DE230101391. The work of J. Yuan is supported in part by the Australian Research Council (ARC) Discovery Project under Grant DP220103596, and in part by the ARC Linkage Project under Grant LP200301482. The work of D. W. K. Ng is supported by the Australian Research Council's Discovery Projects DP230100603. An earlier version of this paper was presented in part at the 2023 IEEE International Conference on Communications Workshops [DOI: 10.1109/ICCWorkshops57953.2023.10283761]. The associate editor coordinating the review of this article and approving it for publication was Z. Hadzi-Velkov. *(Corresponding author: Shuangyang Li.)*

Mingcheng Nie, Deepak Mishra, Jinhong Yuan, and Derrick Wing Kwan Ng are with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: m.nie@student.unsw.edu.au; d.mishra@unsw.edu.au; j.yuan@unsw.edu.au; w.k.ng@unsw.edu.au).

Shuangyang Li is with the Department of Electrical Engineering and Computer Science, Technische Universität Berlin, 10587 Berlin, Germany (e-mail: shuangyang.li@tu-berlin.de).

Digital Object Identifier 10.1109/TCOMM.2024.3516473

I. Introduction

ONE of the core objectives of next-generation wireless communications is to provide robust and reliable services in challenging scenarios, including high-mobility channels. For such channels, the currently employed time-frequency domain (TF) modulation, i.e., orthogonal frequency division modulation (OFDM), could suffer from severe Doppler effects, due to the relative motion between transceivers and potential synchronization errors, leading to the orthogonality damage between sub-carriers. Consequently, non-negligible data rate degradation will appear. To tackle such challenges, delay-Doppler (DD) domain wireless technologies, such as orthogonal time frequency space (OTFS) modulation, have recently emerged as a promising technology to provide robust and high rate communications in high-mobility scenarios [1], [2], [3]. Specifically, the OTFS modulation multiplexes the data symbols in the DD domain and spreads each DD domain symbol to the whole TF domain in an orthogonal manner, where all DD domain symbols will theoretically experience the whole TF domain channel fluctuation. Thanks to this, OTFS has the potential of achieving full channel diversity and has shown a better error performance than conventional OFDM over high-mobility scenarios. More importantly, the DD domain effective channel matrix can be sparsely and compactly represented by only a few parameters, which are quasi-static. Therefore, DD communication scheme offers benefits of delay- and Doppler-resilience, facilitating simple channel estimation [4], [5] and transceiver designs [6], [7].

A. State-of-the-Art and Motivation

To perform reliable data detection for robust communication, accurate channel state information (CSI) is crucial. As a result, efficient pilot designs in the DD domain are highly desirable for general DD communications. An embedded pilot scheme was proposed in [8], where a sufficiently large guard space was adopted to eliminate the overlapping between pilot and information symbols at the receiver side. The channel response can be simply estimated by comparing the received symbols in the channel estimation region with a threshold, where the estimation accuracy depends on the pilot and channel power due to the convolution property of DD domain channels [9]. Furthermore, the authors in [10] proposed a superimposed pilot scheme, which removed the

guard space for improving spectral efficiency. Since the interference between pilot symbols and data symbols is no longer avoidable without the guard space, iterative processing among threshold-based channel estimation, interference cancellation, and data detection was also adopted in [10], where the threshold is modified per iteration based on the accuracy of detected symbols to refine the channel estimation. To further enhance the channel estimation accuracy for sparse DD domain channels, the authors in [11] extended the idea of [8] by providing a reliable estimation algorithm. More specifically, the shifts of the pilot symbol facilitate a coarse estimation of on-grid delay and Doppler shifts through threshold comparison, where a low-complexity maximum likelihood (ML) estimator was used to refine the estimation of the unknown parameters. Note that the aforementioned work and most of the DD domain pilot designs do not explicitly exploit the statistical CSI. Therefore, their performance may be suboptimal in the sense of maximum *a posteriori* (MAP) estimation.

To evaluate the potential estimation improvement by considering the statistical CSI, a sensible performance metric needs to be developed. As the most meaningful and relevant performance metric, the Cramér-Rao bound (CRB) serves as a fundamental mathematical tool for characterizing optimal estimation performance. It has been extensively applied across various fields, such as positioning [12] and sensing [13], [14] Its applications in OTFS transmissions have also been studied in [15], [16], [17], and [18]. For example, the CRB for the joint sensing and communication using OTFS signals was first derived in [15], which can be achieved by using the proposed approximated ML estimator. The CRB of channel coefficient and fractional Doppler estimation was derived in [16] as a benchmark to evaluate the proposed estimation algorithm. Furthermore, the author in [17] derived the CRBs for OTFS channel estimation under non-ideal pulse shaping and various cyclic prefix (CPs) configurations, where the influences of these configurations on channel estimation performance were compared through numerical analysis. Besides, in [18], the CRB for OTFS channel estimation with superimposed pilots was derived, where a noticeable error floor appears in the high signal-to-noise ratio (SNR) regime due to the overlapping between pilots and data. It was also indicated that applying successive interference cancellation (SIC) can effectively mitigate this error floor and thereby enhancing CRB performance. However, we observe that the actual CSI is needed in the derivation of the CRB [15], [16], [17], [18], and consequently CRB can only provide limited insights for practical designs, where CSI is typically unknown at the transmitter [19]. In these cases, the expected mean-square-error (MSE) performance instead of the channel-dependent MSE performance of channel estimation is more important and relevant. It should be highlighted that the expected MSE performance under optimal estimation is bounded by the *a posteriori* Cramér-Rao bound (PCRB), which relies on the *a priori* channel statistical information instead of the actual CSI of each channel realization. More importantly, the impact of the pilot power allocation can also be reflected in PCRB as we will show later, which provides valuable insights for practical system implementations. A most recent work [20] also derived the PCRB in OTFS systems, employing it solely as a performance metric to evaluate the proposed Bayesian learning-assisted channel estimation schemes, where no resource allocation based on the PCRB was considered and the underlying model is different from our work.

Precoding is a widely-adopted means for improving communication performance, which usually relies on the CSI at the transmitter (CSIT). In the context of DD communications, many effective precoding designs based on the CSIT have been proposed. For instance, a low complexity precoding design based on maximum ratio transmission (MRT) for multi-user OTFS systems was proposed in [21], alongside a corresponding low complexity detector at the user side, enabling separate demodulation of individual DD domain information symbol at the user side. Specifically, this work assumes that the CSIT is available in the time-division duplex system where channel reciprocity holds generally. The authors in [22] proposed a precoded OTFS transmission scheme to minimize the system bit error rate (BER), where the optimal precoder was obtained based on the eigenvalue decomposition of the DD domain effective channel matrix. Additionally, two precoding designs for multi-user multiple-input and multiple-output OTFS (MU-MIMO-OTFS) were proposed in [23] based on the approximation of channel inversion and the assumption of perfect CSIT, where the performance of the considered system was robust to different user speeds. Furthermore, a novel DD domain precoding design for MU-MIMO-OTFS systems was proposed in [24] to pre-cancel the known interference before transmission by assuming the CSIT is available. This approach employed Tomlinson-Harashima precoding (THP), demonstrating that the constant "shaping loss" is the only loss of the achievable rate compared to the optimal transmission scenario in a high signal-to-noise ratio (SNR) regime. More importantly, a low-complexity implementation of THP was proposed by inserting zero symbols at certain positions in the DD grid to break the DD domain interference cycles. This allows that all the interference can be cancelled out in a symbol-by-symbol manner.

In practice, the perfect CSIT is generally hard to obtain, especially in the high-mobility channels. Consequently, precoding design based on outdated CSIT may introduce severe performance degradation. Instead of the actual CSI, statistical CSI is relatively easier to obtain at the transmitter side in practice. However, relevant designs for DD communications using statistical CSI are still in their infancy. On the other hand, a DD domain predictive precoder was also proposed in the literature to bypass the need of explicit CSIT. The authors in [25] proposed a predictive precoder, which is achieved by exploiting the temporal dependency of the estimated historical channels obtained via the feedback link from the receiver to the transmitter. However, the precoder prediction relies on previous data streams introduces a potential risk of using outdated CSI. Moreover, the considered feedback link requires additional data burden and thus degrades the spectrum efficiency.

B. Contributions

In this paper, we investigate the pilot power allocation and data power allocation for uplink multi-user OTFS transmissions based only on the statistical CSI. The main contributions are listed as follows

- Based on the *a priori* channel statistical information, we derive the closed-form PCRB for uplink multi-user OTFS with practical power-delay and power-Doppler profiles, where we verify that the PCRB remains independent from the actual delay and Doppler indices for each channel realization. Based on the derived PCRB, a pilot design is proposed to improve the channel estimation accuracy by minimizing the average PCRB of different users, where we show that the pilot design can be reduced to a power allocation problem, thanks to the inherent sparsity and compactness properties of DD domain channel. The optimal solution of such a problem is then derived by using the Lagrangian method.
- We provide an in-depth analysis of the achievable rates for uplink multi-user OTFS, where we reveal that the instantaneous sum-rate can be enhanced through resource allocation under channel fading disparities in multi-user transmissions. Based on the derived instantaneous sum-rate, we further study the average rate aiming to eliminate the impact on instantaneous CSI from the channel according to the underlying power profiles. However, the instantaneous CSI is still required when calculating the average rate due to the existance of inter-user interference. Therefore, we propose to approximate the average sum-rate by using Taylor expansion for eliminating the impact of instantaneous CSI. Based on the proposed approximation, we show that the rate optimization can be simplified to a power allocation problem among different users, where user fairness can be considered. Particularly, the underlying power allocation problem is presented in the form of multiple ratios, which can be solved by the standard fractional programming approach.
- Our numerical results verify the effectiveness of the proposed designs in terms of average normalized-mean-square error (NMSE) of channel estimation accuracy and average sum-rate performance. In particular, our results show that the NMSE performance of the proposed pilot power allocation obtained a roughly 3 dB improvement compared to the conventional embedded pilot power allocation with the given power budget. Moreover, we show that the proposed approximation of the sum-rate, which solely relies on the statistical CSI, closely aligns with the actual sum-rate. As a result, a significant improvement in the average sum-rate performance over the conventional design is observed, which again justifies our approximation for the average sum-rate.

C. Notations

Vectors and matrices are denoted by boldface lowercase and uppercase letters, respectively. The openface letter $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex number fields, respectively. $\mathbb{E}$ and $\mathbb{E}_x[\cdot]$ represent the expectation with respect to (w.r.t.) all the randomness in the argument and x, respectively; $\frac{\partial f}{\partial x}$ denotes the partial derivative of the function f w.r.t. the variable x; $\widehat{x}$ represents the estimation of x. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts of a complex scalar/vector/matrix, respectively. The operators $\|\cdot\|$, $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote Euclidean norm, conjugate, transpose, and Hermitian operations, respectively; $\mathrm{vec}(\mathbf{X})$, $\det(\mathbf{X})$, and $\mathrm{Tr}(\mathbf{X})$ denote the vectorization, determinant, and trace operation on matrix $\mathbf{X}$, respectively; $[\cdot]_N$ represents the modulo operation w.r.t. N; $\otimes$ denotes the Kronecker product operator. $\mathrm{diag}\{x_1, \cdots, x_N\}$ denotes a diagonal square matrix whose diagonal consists of the elements of $x_1, \cdots, x_N$; $\mathbf{F}_N$ and $\mathbf{I}_M$ denote the normalized discrete Fourier transform (DFT) matrix of size $N \times N$ and the identity matrix of size $M \times M$, respectively; $\delta(\cdot)$ and $\exp(\cdot)$ denote the Delta function and the exponential function, respectively; $I(\cdot;\cdot)$ denotes the mutual information.

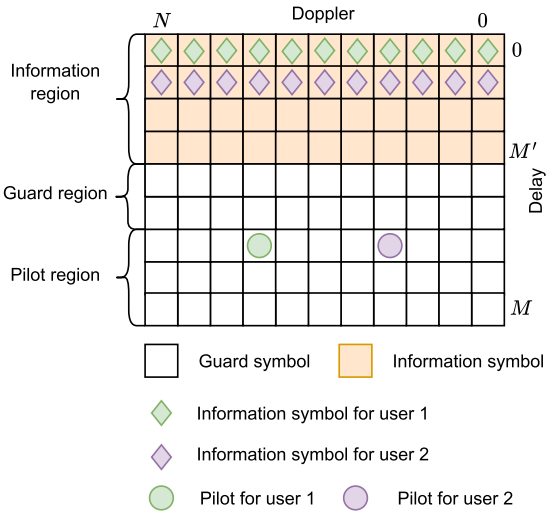

Fig. 1. Frame structure and the arrangement of pilot and information symbols.

II. System Model

We consider the uplink multi-user OTFS transmission with K_u users, where each user equips a single antenna and rectangular window[1] [26]. Specifically, we consider a two-dimensional grid Γ in the DD domain as $\Gamma = \{(\frac{l}{M\Delta f}, \frac{k}{NT}), l = 0, \cdots, M-1, k = 0, \cdots, N-1\}$, where M denotes the number of delay bins/the number of subcarriers, N denotes the number of Doppler bins/the number of time slots, Δf is the subcarrier spacing, and T is the time slots duration [2], respectively. Here, $\frac{1}{M\Delta f}$ and $\frac{1}{NT}$ are the sampling intervals along the delay and Doppler axis, respectively.

A. Frame Structure

We consider a DD domain frame structure for multi-user transmissions as shown in Fig. 1. The frame is segmented into three distinct regions: the information region, the pilot region, and a sufficient guard space in between. Such a structure not only facilitates more precise channel estimation for data detection within the same frame, but also ensures that the pilot

[1]In this paper, we employed a rectangular window but ignore the out-of-band emission [26].

power allocation and the data power allocation can be separated, as there will be no overlapping between corresponding symbols at the receiver. Moreover, in contrast to the multi-user frame structure in [8], placing all the users' information symbols within the information region allows an improved spectral efficiency, which is ensured by the application of inter-user interference cancellation. In the proposed frame structure, the pilot is represented by a single DD domain impulse, and the pilots of different users are sufficiently separated and placed in the pilot region. Additionally, our multi-user frame structure demands the same guard symbol overhead as the one in [8].

Let $\mathbf{X}_u \in \mathbb{C}^{M\times N}$ be the DD domain transmitted matrix of the u-th user, whose vector form is given by

$$\mathbf{x}_u \triangleq \mathrm{vec}(\mathbf{X}_u) = \mathbf{G}_u \mathbf{d}_u. \tag{1}$$

Here, $\mathbf{d}_u \in \mathbb{C}^{L\times 1}$, $1 \le L \le MN$, represents the DD domain information symbol vector of the u-th user and $\mathbf{G}_u \in \mathbb{C}^{MN\times L}$ denotes the corresponding DD domain mapping matrix of the u-th user, whose structure depends on the adopted multiple access (MA) scheme. Particularly, we require that adopted mapping matrices have special structures such that the aforementioned pilot and information symbol separation is guaranteed. To better explain the considered system model, let us consider an example. The delay division multiple access (DDMA) [27] allows different users to multiplex their information symbols along the whole Doppler axis with different delay indices.[2] The corresponding mapping matrix $\mathbf{G}_u$ is given by

$$\mathbf{G}_u = \mathbf{I}_N \otimes \mathbf{a}_u, \tag{2}$$

where $\mathbf{a}_u$ is a length-M column vector, containing a combination of one and zero entries. The entries marked as one denote the specific delay rows on which the information symbols are multiplexed. For example, if the m'-th entry of $\mathbf{a}_u$ is one and the remaining entries are zeros, it indicates that the information symbols are being multiplexed solely along the m'-th delay row. As shown in Fig. 1, the first user and second user multiplex their data along the first and second delay row, respectively, by setting $\mathbf{a}_1(1) = 1$ and $\mathbf{a}_2(2) = 1$ with all other entries being zero. More users can be added in this manner. Furthermore, by restricting the one entries for different users in $\mathbf{a}_u$ to only appear within the range of $0, \cdots, M'$, where $M' < M$, we establish an information region of dimensions $M' \times N$ as shown in Fig. 1(a).

Remark 1: Note that in this work, we assume the power of pilots and data are considered separately, which is a commonly adopted approach in the literature [8], [10], [28]. However, we acknowledge that considering the sum power of both pilots and data would reflect a more practical scenario, and this aspect will be addressed in future work.

[2]Note that DDMA is only one example of multiple access schemes; others, such as Doppler division multiple access, have also been exploited in [27]. In general, DD domain multiple access will suffer from multiuser interference due to the convolutional nature of the DD domain channel. Therefore, SIC receivers, or other type of equalizers, are necessary to enhance communication performance.

B. Multi-User OTFS Transmission

As shown in Fig. 2, the power of both information symbols and pilots for each user is first allocated, after which they are placed in the DD domain transmit vector $\mathbf{x}_u$. The time domain transmitted symbol vector for the u-th user, $\mathbf{s}_u \in \mathbb{C}^{MN\times 1}$, can be obtained by applying the inverse symplectic finite Fourier transform (ISFFT) and Heisenberg transform,[3] which is given by [29]

$$\mathbf{s}_u = \left(\mathbf{F}_N^H \otimes \mathbf{I}_M\right) \mathbf{x}_u. \tag{3}$$

After passing the channel, the time domain received symbol vector, $\mathbf{r} \in \mathbf{C}^{MN\times 1}$, is transformed to the DD domain by applying the Wigner transform and the symplectic finite Fourier transform (SFFT), which is given by

$$\mathbf{y} = \left(\mathbf{F}_N \otimes \mathbf{I}_M\right) \mathbf{r}, \tag{4}$$

where $\mathbf{y} \in \mathbb{C}^{MN\times 1}$ represents the DD domain received symbol vector. According to [30] and [31], the DD domain representation of a linear time-varying channel is given by

$$h_u(\tau,\nu) = \sum_{p=1}^{P} h_{p,u}\delta(\tau - \tau_{p,u})\delta(\nu - \nu_{p,u}), \tag{5}$$

where $h_{p,u}$, $\tau_{p,u} = (l_{p,u} + \imath_{p,u})\frac{1}{M\Delta f}$, and $\nu_{p,u} = (k_{p,u} + \kappa_{p,u})\frac{1}{NT}$ denote the channel coefficient, delay, and Doppler shifts of the p-th path of the u-th user, respectively. Here, $l_{p,u}$ and $k_{p,u}$ represent the integer delay and integer Doppler indices, respectively, while $-1/2 \le \imath_p \le 1/2$ and $-1/2 \le \kappa_p \le 1/2$ represent the fractional delay and fractional Doppler indices, respectively. In this paper, we consider the wireless channel with sufficient delay and Doppler resolutions, i.e., the fractional delay and Doppler indices are of zero values.[4] Thus, the time domain effective channel matrix between the u-th user and the base station (BS) with the rectangular pulse, $\mathbf{H}_u^{(\mathrm{T})}$, is given by [25]

$$\mathbf{H}_u^{(\mathrm{T})} = \sum_{p=1}^{P} h_{p,u} e^{-j2\pi\frac{k_{p,u}l_{p,u}}{MN}} \mathbf{\Delta}^{k_{p,u}} \mathbf{\Pi}^{l_{p,u}}, \tag{6}$$

where $\mathbf{\Delta} \in \mathbb{C}^{MN\times MN}$ denotes the phase rotating matrix as $\mathbf{\Delta} = \mathrm{diag}\left\{1, e^{j2\pi\frac{1}{MN}}, \cdots, e^{j2\pi\frac{MN-1}{MN}}\right\}$ and $\mathbf{\Pi} \in \mathbb{R}^{MN\times MN}$ denotes the forward cyclic shift permutation matrix as $\mathbf{\Pi} = [\boldsymbol{e}_2, \boldsymbol{e}_3, \cdots, \boldsymbol{e}_{MN}, \boldsymbol{e}_1]$, where $\boldsymbol{e}_i$, $1 \le i \le MN$, denotes the i-th column of the identity matrix $\mathbf{I}_{MN}$. The time domain input-output relationship in the vector representation can be written as

$$\mathbf{r} = \sum_{u=1}^{K_u} \mathbf{H}_u^{(\mathrm{T})} \mathbf{s}_u + \mathbf{w}, \tag{7}$$

[3]Note that a common choice for implementing the Heisenberg transform is the inverse fast Fourier transform (IFFT), i.e., the conventional OFDM modulator. Correspondingly, the fast Fourier transform (FFT), i.e., conventional OFDM demodulator, is used to implement the Wigner transform at the receiver side [2].

[4]Note that the considered integer delay and Doppler case is only valid with a sufficiently large M and N. As our main focus of this paper is the system design using statistical CSI, where the knowledge of fractional or integer delay and Doppler is not available, we restrict ourselves to only consider the integer delay and Doppler case. Nevertheless, we will examine our proposed scheme under fractional delay and Doppler cases in the numerical results.

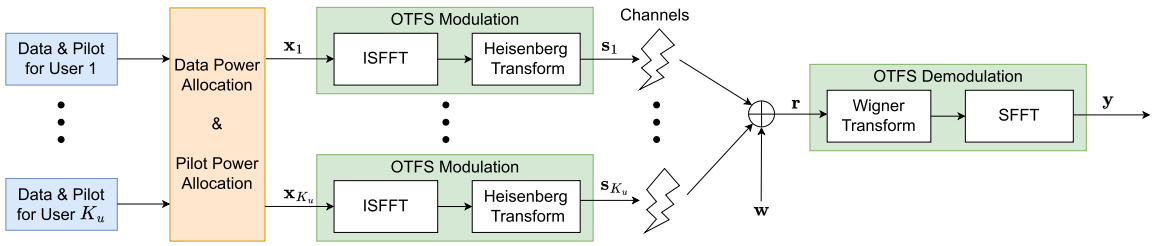

Fig. 2. Uplink multi-user OTFS transmission structure.

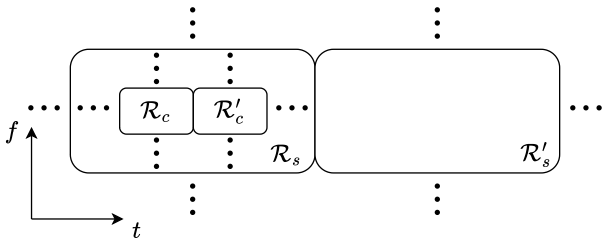

Fig. 3. Schematic illustration of coherence region $\mathcal{R}_c$ and stationarity region $\mathcal{R}_s$.

where $\mathbf{w} \in \mathbb{C}^{MN\times MN}$ represents the additive white Gaussian noise (AWGN) sample vector in the time domain with power spectral density (PSD) N_0 at the BS. By considering (1) – (7), the input-output relationship for uplink multi-user OTFS transmissions in the DD domain is given by[5]

$$\mathbf{y} = \sum_{u=1}^{K_u} \mathbf{H}_u \mathbf{x}_u + \mathbf{w}, \tag{8}$$

where $\mathbf{H}_u \in \mathbb{C}^{MN\times MN}$ denotes the DD domain effective channel matrix and is given by [25]

$$\begin{aligned}\mathbf{H}_u &= \left(\mathbf{F}_N \otimes \mathbf{I}_M\right) \mathbf{H}_u^{(\mathrm{T})} \left(\mathbf{F}_N^H \otimes \mathbf{I}_M\right) \\ &= \left(\mathbf{F}_N \otimes \mathbf{I}_M\right) \sum_{p=1}^{P} h_{p,u} e^{-j2\pi \frac{k_{p,u} l_{p,u}}{MN}} \mathbf{\Delta}^{k_{p,u}} \mathbf{\Pi}^{l_{p,u}} \left(\mathbf{F}_N^H \otimes \mathbf{I}_M\right).\end{aligned} \tag{9}$$

C. WSSUS and Averaging

It is worth highlighting that OTFS transmission occurs over a non-wide-sense stationary uncorrelated scattering (non-WSSUS) channel, where each OTFS frame corresponds to a stationarity region, within which the channel is approximately treated as WSSUS [32], [33], [34]. As a result, the dynamic channel fluctuations in the TF domain can be translated into a quasi-static response (5) in the DD domain. Moreover, our approach involves averaging over multiple OTFS frames to capture the statistical characterization of OTFS transmission, rather than focusing solely on the communication performance within one frame [35]. As shown in Fig. 3, the stationarity region $\mathcal{R}_s$ is significantly larger than the coherence region $\mathcal{R}_c$ [31], allowing for the existence of multiple coherence regions with different channel responses (where the conventional ergodic rate is defined [35]) within one stationarity region. While delay and Doppler responses remain constant within one stationarity region, they vary across different regions, forming the basis of our average definition. Furthermore, we assume that minimum delay, maximum delay, and maximum Doppler remain unchanged over several stationarity regions [36], during which our proposed allocation scheme based on statistical CSI remains valid following one selection. This assumption is grounded in the relatively unchanged physical geometry of the users over these regions under the given statistical information of the channel [36]. Additionally, we assume no feedback of instantaneous CSI in this work, due to the challenges of obtaining accurate CSIT in high-mobility scenarios.

[5]In (8), we use the same notation for the AWGN samples in both time domain and DD domain as they follow the same distribution.

D. A Priori Distribution of Channel Parameters

For the ease of derivation, we assume $l_{p,u}$ and $k_{p,u}$ are uniformly obtained from $[l_{\min,u}, l_{\max,u}]$ and $[-k_{\max,u}, k_{\max,u}]$, respectively, where $l_{\min,u}$, $l_{\max,u}$, and $k_{\max,u}$ represent the minimum delay, maximum delay, and maximum Doppler indices of the u-th user, respectively. The *a priori* probability density functions (PDFs) of $l_{p,u}$ and $k_{p,u}$ are given by [37]

$$p(l_{p,u}) = \begin{cases} \dfrac{1}{l_{\max,u} - l_{\min,u}}, & l_{\min,u} \le l_{p,u} \le l_{\max,u}, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

$$p(k_{p,u}) = \begin{cases} \dfrac{1}{2k_{\max,u}}, & -k_{\max,u} \le k_{p,u} \le k_{\max,u}, \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

and the joint PDF of delay and Doppler indices of the u-th user is obtained by $p(\mathbf{l}_u) = \prod_{p=1}^{P} p(l_{p,u})$ and $p(\mathbf{k}_u) = \prod_{p=1}^{P} p(k_{p,u})$ by assuming that the different resolvable paths are independent of each other. Assume that the channel coefficient of the u-th user's p-th path follows $h_{p,u} \sim \mathcal{CN}(0, \gamma_{p,u})$, where $\gamma_{p,u}$ is the variance of the corresponding channel coefficient that follows exponential power-delay and uniform power-Doppler profiles as [31]

$$\gamma_{p,u} = \frac{e^{-l_{p,u}}}{2k_{\max,u}}. \tag{12}$$

Following (12), the *a priori* PDF of $h_{p,u}$ is given by

$$p(h_{p,u}) = \frac{1}{\pi\gamma_{p,u}} \exp\left(-\frac{|h_{p,u}|^2}{\gamma_{p,u}}\right), \tag{13}$$

where $|\cdot|$ is the absolute value. Thus, the joint PDF of channel coefficient of u-th user is obtained by $p(\mathbf{h}_u) = \prod_{p=1}^{P} p(h_{p,u})$ due to independency assumption for different resolvable channel paths.

III. A Posteriori Cramér-Rao Bound Analysis for Channel Estimation

In this section, we present the PCRB analysis for DD domain channel estimation. According to [1] and [28], the input-output relationship (IOR) for uplink multi-user OTFS transmissions with rectangular pulses in the DD domain can be written by[6]

$$y[k,l] = z[k,l] + w[k,l], \tag{14}$$

where $y[k,l]$ denotes the received signal at the BS, i.e., the (k,l)-th entry of $\mathbf{y}$ in (8), $w[k,l]$ is the equivalent AWGN variable with variance N_0, and $z[k,l]$ is an intermediate variable defined by

$$z[k,l] = \sum_{u=1}^{K_u}\sum_{p=1}^{P} h_{p,u} e^{j2\pi\frac{(l-l_{p,u})k_{p,u}}{MN}} x_u[[k-k_{p,u}]_N, [l-l_{p,u}]_M], \tag{15}$$

where x_u denotes the pilot symbols of the u-th user and we omit the information symbols here as our current focus is on channel estimation. Note that the channel can be estimated by setting a threshold comparison, i.e., $y[k,l] = 0, \forall |y[k,l]| < \beta$ [8], then compensate the pilot and phase as shown in [28] for the practical rectangular pulse case.

A. A Posteriori Distribution of Channel Parameters

First, we define a vector that contains the channel parameters in (15). Note the channel coefficient is complex, and we decompose it into real and imaginary parts. Thus, we define a length-$4PK_u$ channel parameter vector $\boldsymbol{\theta} = [\mathbf{h}, \mathbf{l}, \mathbf{k}]$, where $\mathbf{h} = [\mathbf{h}_1, \cdots, \mathbf{h}_u, \cdots, \mathbf{h}_{K_u}]$ is a length-$2PK_u$ real vector contains the real and imaginary parts of channel coefficients, $\mathbf{l} = [\mathbf{l}_1, \cdots, \mathbf{l}_u, \cdots, \mathbf{l}_{K_u}]$ is a length-PK_u vector contains the delay indices, and $\mathbf{k} = [\mathbf{k}_1, \cdots, \mathbf{k}_u, \cdots, \mathbf{k}_{K_u}]$ is a length-PK_u vector contains the Doppler indices of all users. Specifically, $\mathbf{h}_u$, $\mathbf{l}_u$, and $\mathbf{k}_u$ for each user are given as

$$\mathbf{h}_u = [\Re(h_{1,u}), \cdots, \Re(h_{P,u}), \Im(h_{1,u}), \cdots, \Im(h_{P,u})], \tag{16}$$

$$\mathbf{l}_u = [l_{1,u}, \cdots, l_{P,u}], \text{and } \mathbf{k}_u = [\mathrm{k}_{1,u}, \cdots, \mathrm{k}_{P,u}], \tag{17}$$

where $\Re(h_{p,u})$ and $\Im(h_{p,u})$ represent the real and imaginary part of $h_{p,u}$. Based on the Bayesian theorem and the *a priori* information $p(\boldsymbol{\theta})$, the *a posteriori* PDF of $\boldsymbol{\theta}$ for given $\mathbf{y}$ is [40]

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta},\mathbf{y})}{p(\mathbf{y})} \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{18}$$

[6]The considered IOR is equivalent to the IORs described in [38] and [39] when employing a rectangular shaping pulse.

where $\mathbf{y}$ is the receive vector in (8). According to the power profiles in (12), we can obtain

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\mathbf{h}|\mathbf{k},\mathbf{l})p(\mathbf{k})p(\mathbf{l}). \tag{19}$$

B. A Posterior Cramér-Rao Bound Derivation

According to [41], the expected MSE matrix of $\boldsymbol{\theta}$ is bounded by the inverse of the *a posteriori* Fisher information matrix (FIM) $\mathbf{J}$ as $\mathbb{E}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right] \succeq \mathbf{J}^{-1} \triangleq \boldsymbol{\Gamma}$. Thus, we define the PCRBs of channel coefficients, delay indices, and Doppler indices by

$$\mathbb{E}\left[||\widehat{\mathbf{h}}-\mathbf{h}||^2\right] \geq \sum_{i=1}^{2PK_u} \boldsymbol{\Gamma}_{(i,i)} \triangleq \mathrm{PCRB}(\mathbf{h}), \tag{20}$$

$$\mathbb{E}\left[||\widehat{\mathbf{l}}-\mathbf{l}||^2\right] \geq \sum_{i=2PK_u+1}^{3PK_u} \boldsymbol{\Gamma}_{(i,i)} \triangleq \mathrm{PCRB}(\mathbf{l}), \tag{21}$$

$$\mathbb{E}\left[||\widehat{\mathbf{k}}-\mathbf{k}||^2\right] \geq \sum_{i=3PK_u+1}^{4PK_u} \boldsymbol{\Gamma}_{(i,i)} \triangleq \mathrm{PCRB}(\mathbf{k}), \tag{22}$$

where $\boldsymbol{\Gamma}_{(i,i)}$ denotes the i-th diagonal entry of $\boldsymbol{\Gamma}$ and the *a posteriori* FIM $\mathbf{J}$ can be obtained based on (19) by [41]

$$\begin{aligned}\mathbf{J}^{(i,j)} &= -\mathbb{E}_{\boldsymbol{\theta}\mathbf{y}}\left[\frac{\partial^2 \ln p(\boldsymbol{\theta}|\mathbf{z},\mathbf{y})}{\partial\theta_i\partial\theta_j}\right] \\ &= -\mathbb{E}_{\boldsymbol{\theta}\mathbf{y}}\left[\frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta},\mathbf{z})}{\partial\theta_i\partial\theta_j}\right] - \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln p(\mathbf{h}|\mathbf{k},\mathbf{l})}{\partial\theta_i\partial\theta_j}\right] \\ &= \mathbf{J}_r^{(i,j)} + \mathbf{J}_p^{(i,j)}.\end{aligned} \tag{23}$$

where $\mathbf{J}^{(i,j)}$ is the (i,j)-th entry of $\mathbf{J}$. Here, we denote $\mathbf{J}_r$ and $\mathbf{J}_p$ as the observed Fisher information matrix based on received uplink signals and the *a priori* information matrix based on prior known channel statistical information, respectively. Compared with conventional CRB's FIM $\mathbf{F} = -\mathbb{E}_{\mathbf{y}}\left[\frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta},\mathbf{z})}{\partial\theta_i\partial\theta_j}\right]$, the *a priori* information matrix $\mathbf{J}_p$ consist of the statistical information for the estimation, where the expected channel power in (12) is reflected. Note that the operation $\mathbb{E}_{\boldsymbol{\theta}}[\cdot]$ ensures that the PCRB is independent of the channel coefficient, delay and Doppler indices. Thus, the conditional logarithm PDF of $\mathbf{y}$ is given as follows

$$\ln p(\mathbf{y}|\boldsymbol{\theta},\mathbf{z}) = -MN\ln(\pi N_0) - \frac{1}{N_0}\sum_{k,l}|y[k,l]-z[k,l]|^2. \tag{24}$$

According to (23) and (24), we derive the following lemma.

Lemma 1: For multi-user OTFS transmission, the (i,j)-th entry of the observed Fisher information matrix $\mathbf{J}_r$ for the channel parameter vector $\boldsymbol{\theta}$ is obtained by

$$\mathbf{J}_r^{(i,j)} = \frac{2}{N_0}\mathbb{E}_{\boldsymbol{\theta}}\left[\Re\left\{\sum_{k,l}\left(\frac{\partial z[k,l]}{\partial\theta_i}\right)^*\left(\frac{\partial z[k,l]}{\partial\theta_j}\right)\right\}\right], \tag{25}$$

Proof: The proof is given in Appendix A.

By defining $k' = [k-k_{p,u}]_N$ and $l' = [l-l_{p,u}]_M$, the partial derivatives in (25) are given by

$$\frac{\partial z[k,l]}{\partial\Re(h_{p,u})} = e^{j2\pi\frac{(l-l_{p,u})k_{p,u}}{MN}} x_u[k',l'], \tag{26}$$

$$\frac{\partial z[k,l]}{\partial \Im(h_{p,u})} = j e^{j2\pi \frac{(l-l_{p,u})k_{p,u}}{MN}} x_u[k',l'], \tag{27}$$

$$\frac{\partial z[k,l]}{\partial l_{p,u}} = -h_{p,u} \frac{j2\pi k_{p,u}}{MN} e^{j2\pi \frac{(l-l_{p,u})k_{p,u}}{MN}} x_u[k',l'], \tag{28}$$

$$\frac{\partial z[k,l]}{\partial k_{p,u}} = h_{p,u} \frac{j2\pi (l-l_{p,u})}{MN} e^{j2\pi \frac{(l-l_{p,u})k_{p,u}}{MN}} x_u[k',l']. \tag{29}$$

By substituting the partial derivatives (26)-(29) into (25), we can observe that $\mathbf{J}_r$ is a diagonal matrix (see Appendix A). Thus, the i-th diagonal entry of $\mathbf{J}_r$ satisfies

$$\mathbf{J}_r^{(i,i)} = \begin{cases} \frac{2\mathbf{x}_u^H \mathbf{x}_u}{N_0}, & 1 \le i \le 2PK_u, \\ \frac{2\mathbf{x}_u^H \mathbf{x}_u}{N_0} A, & 2PK_u < i \le 3PK_u, \\ \frac{2\mathbf{x}_u^H \mathbf{x}_u}{N_0} B, & 3PK_u < i \le 4PK_u, \\ 0, & \text{otherwise}, \end{cases} \tag{30}$$

where $\mathbf{x}_u^H \mathbf{x}_u = \sum_{k',l'} x_u^*[k',l'] x_u[k',l']$. Note that $A = \mathbb{E}_{\boldsymbol{\theta}}\left[\left(|h_{p,u}| \frac{2\pi k_{p,u}}{MN}\right)^2\right]$ and $B = \mathbb{E}_{\boldsymbol{\theta}}\left[\left(|h_{p,u}| \frac{2\pi (l''-l_{p,u})}{MN}\right)^2\right]$, where $l'' = l_{p,u} + l_{\tau,u}$, and $l_{\tau,u}$ is the pilot delay index of u-th user. Based on (13) and (23), we have the following lemma.

Lemma 2: For multi-user OTFS transmission, the *a priori* information matrix $\mathbf{J}_p$ for the channel parameter vector $\boldsymbol{\theta}$ is a diagonal matrix, where the i-th diagonal entry of $\mathbf{J}_p$ is obtained by

$$\mathbf{J}_p^{(i,i)} = \begin{cases} \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{2}{\gamma_{p,u}}\right], & 1 \le i \le 2PK_u, \\ \mathbb{E}_{\boldsymbol{\theta}}\left[\left(\frac{1}{\gamma_{p,u}} \frac{\partial \gamma_{p,u}}{\partial l_{p,u}}\right)^2\right], & 2PK_u < i \le 3PK_u, \\ \mathbb{E}_{\boldsymbol{\theta}}\left[\left(\frac{1}{\gamma_{p,u}} \frac{\partial \gamma_{p,u}}{\partial k_{p,u}}\right)^2\right], & 3PK_u < i \le 4PK_u, \\ 0, & \text{otherwise}, \end{cases} \tag{31}$$

Proof: The proof is given in Appendix B.

From the above analysis, we can observe that the PCRB analysis relates to both the *a priori* information and the likelihood function $p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{z})$, whereas the conventional CRB only relates to the likelihood function [41]. As a result, the PCRB contains important information of the statistical channel characteristics rather than the temporal channel information that is only valid in a particular transmission, which can provide clear insights on the average MSE ergodically. More importantly, the actual CSI for each transmission is hard to obtain at the transmitter side, while the statistical channel information is generally available in the system design. Therefore, PCRB provides more implementation significance for communication designs. Our pilot power allocation considering PCRB will be investigated in the following subsection. Before that, we present the closed-form PCRB of delay and Doppler as

$$\mathrm{PCRB}(\mathbf{l}) = \sum_{u=1}^{K_u} \sum_{p=1}^{P} \left(\frac{2\mathbf{x}_u^H \mathbf{x}_u}{N_0} \mathbb{E}_{\boldsymbol{\theta}}\left[\left(|h_{p,u}| \frac{2\pi k_{p,u}}{MN}\right)^2\right] + \mathbb{E}_{\boldsymbol{\theta}}\left[\left(\frac{1}{\gamma_{p,u}} \frac{\partial \gamma_{p,u}}{\partial l_{p,u}}\right)^2\right] \right)^{-1}, \tag{32}$$

$$\mathrm{PCRB}(\mathbf{k}) = \sum_{u=1}^{K_u} \sum_{p=1}^{P} \left(\frac{2\mathbf{x}_u^H \mathbf{x}_u}{N_0} \mathbb{E}_{\boldsymbol{\theta}}\left[\left(|h_{p,u}| \frac{2\pi l_{\tau,u}}{MN}\right)^2\right] + \mathbb{E}_{\boldsymbol{\theta}}\left[\left(\frac{1}{\gamma_{p,u}} \frac{\partial \gamma_{p,u}}{\partial k_{p,u}}\right)^2\right] \right)^{-1}. \tag{33}$$

Next, we detail the closed-form PCRB of the channel coefficients in (20) based on (30) and (31) as follows

$$\mathrm{PCRB}(\mathbf{h}) = \sum_{u=1}^{K_u} \sum_{p=1}^{P} \left(\frac{\mathbf{x}_u^H \mathbf{x}_u}{N_0} + \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{1}{\gamma_{p,u}}\right] \right)^{-1} \tag{34}$$

$$= \sum_{u=1}^{K_u} P \left(\frac{\mathbf{x}_u^H \mathbf{x}_u}{N_0} + \underbrace{\frac{2k_{\max,u}\left(e^{l_{\max,u}} - e^{l_{\min,u}}\right)}{l_{\max,u} - l_{\min,u}}}_{\Phi_u} \right)^{-1}, \tag{35}$$

where (35) is obtained by substituting the exponential power-delay and uniform power-Doppler profiles in (12), and Φ_u denotes the u-th user's *a priori* channel statistical information that only relates to the minimum and maximum delay and maximum Doppler indices. The result in (34) demonstrates that the PCRB of the multi-user is the accumulation of each user's PCRB and the PCRB of a multi-path channel is the accumulation of each path's PCRB, which can be validated by the assumption that different users and paths of the channel are independent. Moreover, we can observe from (34) that the PCRB of each user relates to the corresponding pilot SNR and the channel statistic characteristic, i.e., the *expectation* of the inverse of the channel variance. Thanks to the expectation property of PCRB, we show that bounds in (35) only depend on the range of delay and Doppler indices even under the practical exponential power-delay and uniform power-Doppler profiles, which can be known at the transmitter side. Note that the detailed closed-form PCRBs of delay and Doppler indices are omitted but can be obtained in a similar approach by calculating the corresponding components in (30) and (31).

C. Pilot Power Allocation for Multi-User Transmissions Based on PCRB

In this section, we investigate the pilot power allocation for multi-user OTFS transmissions based on the derived PCRB. It should be noted that the classical threshold-based channel estimation performs reasonably well at high SNRs [8]. Therefore, we expect that optimizing the PCRB can further improve the estimation performance of the threshold-based channel estimation at high SNRs. Based on the previous discussions, we propose to allocate pilots' power based on the PCRB of the channel coefficient. Note that the delay and Doppler have sufficiently high resolution with large M and N values, and consequently they are unlikely to introduce estimation error in practical systems at high SNRs compared to the channel coefficient. Thus, we propose to allocate the pilots' power

by minimizing the average normalized PCRB (NPCRB) of channel coefficient estimation in (34) and formulate it as

$$\min_{\mathcal{P}_u} \ \mathrm{NPCRB}(\mathbf{h}) = \sum_{u=1}^{K_u} N_u P \left(\mathcal{P}_u + \Phi_u\right)^{-1} \tag{36a}$$

$$\text{subject to :} \quad \mathcal{P} = \sum_{u=1}^{K_u} \mathcal{P}_u, \tag{36b}$$

where $N_u = \mathbb{E}\left[\frac{1}{||\mathbf{h}_u||^2}\right]$ is the normalization factor[7] of u-th user's PCRB, P denotes the number of channel paths, $\mathcal{P}$ is the total pilot power of the cell, $\mathcal{P}_u \triangleq \mathbf{x}_u^H \mathbf{x}_u, \forall u = 1, \cdots, K_u$, is the pilot power of the u-th user, $\mathbf{x}_u$ denotes the pilot symbols of the u-th user, and Φ_u is shown in (35).

To solve the formulated optimization problem, considering λ as a Lagrangian multiplier for problem (36), we can write the underlying Lagrangian as follows

$$\mathcal{L}(\lambda, \mathcal{P}_1, \cdots, \mathcal{P}_u, \mathcal{P}) = P \sum_{u=1}^{K_u} N_u \left(\mathcal{P}_u + \Phi_u\right)^{-1} + \lambda \sum_{u=1}^{K_u} \mathcal{P}_u. \tag{37}$$

Then, we can find the optimal power allocation for (36) by solving the Karush-Kuhn-Tucker (KKT) conditions by $\frac{\partial \mathcal{L}}{\partial \mathcal{P}_u} = 0$, yields

$$\mathcal{P}_u = \sqrt{\frac{P N_u}{\lambda}} - \Phi_u. \tag{38}$$

After substituting (38) into the constraint, we can obtain the optimal value of λ as follows

$$\lambda = P \left(\frac{\sum_{i=1}^{K_u} \sqrt{N_i}}{\mathcal{P} + \sum_{i=1}^{K_u} \Phi_i} \right)^2, \tag{39}$$

where $\lambda > 0$. Finally, the optimal power allocation for each user can be obtained by substituting (39) into (38) as follows

$$\mathcal{P}_u = \frac{\sqrt{N_u}(\mathcal{P} + \sum_{i=1}^{K_u} \Phi_i)}{\sum_{i=1}^{K_u} \sqrt{N_i}} - \Phi_u. \tag{40}$$

We observe from (40) that a small value of maximum delay and Doppler indices will lead to a small Φ_u and N_u resulting in a small pilot power $\mathcal{P}_u$. This is because a small range of the delay and Doppler indices will statistically lead to a high average channel power according to the power profile (12), and therefore it does not need very high pilot power to have a good estimation performance.

Remark 2: Note that the channel model used for derivations considers the exponential power-delay profile and uniform power-Doppler profile. However, it should be noted that our derivations are always valid as long as the PDFs are square-summable [41]. Therefore, our derivations can be naturally extended to other models, e.g. the Jakes' model with proper truncation on the PDF. Additionally, the assumption of a uniform distribution for delay and Doppler shifts in (10) and (11) is not a prerequisite for our derivations, which remain valid across various distributions. However, it is important to emphasize that the performance gap between PCRB and CRLB is more relevant to the underlying power-delay and power-Doppler profiles and the values of minimum and maximum delay and Doppler indices, rather the specific PDFs of delay and Doppler shifts, as demonstrated in (23).

[7]Note that the normalization factor is an expected value of the channel coefficient, which can be computed before the transmission based on the given power profile and range of delay and Doppler shifts.

IV. Average Achievable Rate Optimization of Multi-User Transmissions

A. Sum-Rate of Uplink Multi-User OTFS

For the sake of presentation, we omit the pilot region in this section and focus on the data power allocation in this subsection. Therefore, we define $E_u \triangleq \mathbb{E}[\mathbf{x}_u], \forall u = 1, \cdots, K_u$ as the average symbol power of the u-th user and $E_{\text{tot}} = \sum_u E_u$ as the total symbol power budget. As commonly adopted for uplink transmissions, we perform successive interference cancellation (SIC) [42] at the receiver to recover the transmitted information symbols as shown in Fig. 4. Specifically, the multi-user interference (MUI) from all the previously decoded users is cancelled first when decodes the current user, while the MUI from the remaining undetected users and the residual MUI from previous decoded users are treated as additional AWGN. Furthermore, we consider a practical scenario where the receiver possesses only imperfect knowledge of the users' channels, i.e., the receiver is provided with partial information of the users' channels by pilot estimation. The resulting estimation error introduces additional interference, thereby impacting the achievable sum rate. Let $\widehat{\mathbf{H}}_u$ represents the estimated channel matrix of the u-th user, and $\Delta\mathbf{H}_u$ denotes the corresponding independent estimation error [43]. Then the actual channel for arbitrary users can be represented by

$$\mathbf{H}_u = \widehat{\mathbf{H}}_u + \Delta\mathbf{H}_u, \tag{41}$$

where the entries of $\Delta\mathbf{H}_u$ are assumed to be independent identically distributed (i.i.d.) and zero-mean circularly symmetric complex Gaussian variables with variance σ_e^2. It is noteworthy that the parameter σ_e^2 captures the quality of channel estimation and is assumed to be known at both the transmitter and receiver side [43]. Moreover, we assume that the distribution of channel coefficient $\mathbf{h}_u$, delay indices $\mathbf{l}_u$, and Doppler indices $\mathbf{k}_u$ within $\widehat{\mathbf{H}}_u$ remain consistent with those in $\mathbf{H}_u$, although the exact values may differ.

According to [27], [43], and [44], the sum-rate for the multi-user transmission with estimated channels can be defined by $I(\mathbf{y}; \mathbf{x}_1, \cdots, \mathbf{x}_{K_u} | \widehat{\mathbf{H}}_1, \cdots, \widehat{\mathbf{H}}_{K_u})$, whose closed-form expression is given in the following lemma.

Lemma 3: Considering an arbitrary multi-user transmission scheme, e.g., DDMA, the mapping matrix for the u-th user, $\mathbf{G}_u$, the corresponding estimated channel matrix, $\widehat{\mathbf{H}}_u$, and the variance of estimation error, σ_e^2, the closed-form expression of the achievable sum-rate under the Gaussian approximation is given by

$$\begin{aligned} &I(\mathbf{y}; \mathbf{x}_1, \cdots, \mathbf{x}_{K_u} | \widehat{\mathbf{H}}_1, \cdots, \widehat{\mathbf{H}}_{K_u}) \\ &= \sum_{u=1}^{K_u} I(\mathbf{y}; \mathbf{x}_u | \mathbf{x}_1, \cdots, \mathbf{x}_{u-1}, \widehat{\mathbf{H}}_1, \cdots, \widehat{\mathbf{H}}_{K_u}) \end{aligned} \tag{42}$$

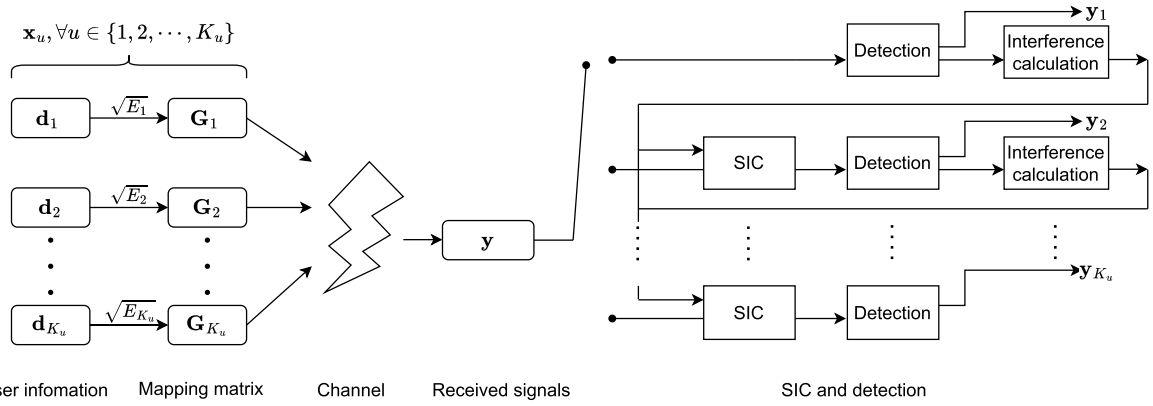

Fig. 4. DD domain uplink multiuser transmission with SIC detection.

$$= \sum_{u=1}^{K_u} \log_2 \det \left(\mathbf{I}_L + \frac{E_u}{\widetilde{N}_0 + \sigma_u^2} \mathbf{G}_u^H \widehat{\mathbf{H}}_u^H \widehat{\mathbf{H}}_u \mathbf{G}_u \right), \quad (43)$$

where $\widetilde{N}_0 = N_0 + E_{\text{tot}} L \sigma_e^2$ and σ_u^2 denotes the MUI power from the $(u+1)$-th to the K_u-th users and given by

$$\sigma_u^2 = \frac{1}{L} \text{Tr} \left(\sum_{u'=u+1}^{K_u} E_j \mathbf{G}_{u'}^H \widehat{\mathbf{H}}_{u'}^H \widehat{\mathbf{H}}_{u'} \mathbf{G}_{u'} \right). \quad (44)$$

Proof: The proof is given in Appendix C.

It should be noted that the achievable of the mutual information in (43) depends on the application of capacity-achieving codes and Gaussian input. In practical applications, it can be closely approached by adopting high-order constellations and near-optimal channel coding. Therefore, it is reasonable to consider this mutual information as the performance metric. According to [45], the Gaussian distribution represents the worst-case additive error for a given variance σ_e^2. This implies that if the rate in (43) is achievable when $\Delta \mathbf{H}_u$ follows a Gaussian distribution with variance σ_e^2, then it remains achievable for any other distribution of $\Delta \mathbf{H}_u$. Moreover, we assume that each entry in $\mathbf{H}_u$ is affected by i.i.d. Gaussian errors, which again represents the worst-case scenario. In practice, only a few entries in each row and column of $\mathbf{H}_u$ are non-zero, meaning that errors would only occur in these specific entries, rather than across the entire matrix $\mathbf{H}_u$. Consequently, our chosen estimation error model reflects the worst-case scenario to ensure the robustness of the achievable rate analysis.

Furthermore, it can be observed that the impact of imperfect CSI on the sum-rate performance in (43) depends on three parameters. Particularly, E_{tot} in $\widetilde{N}_0$ causes the system to be interference-limited in the high SNR regime, potentially leading to a performance floor, which is demonstrated numerically later. Additionally, with a larger number of transmit symbols L, the sum-rate performance can be degraded more severely under the same level of estimation error. Moreover, it can be observed from (43) that the achievable rate performance mainly relies on the symbol power, channel conditions, and mapping matrix of both current and undetected users, raising a variety of symbol power allocation schemes based on corresponding channel conditions and MA schemes to improve the achievable rate performance. However, although we recognize that the exact rate must be measured, acquiring accurate CSI at the transmitter for uplink transmission is generally impractical especially over time-varying channels. Consequently, we want a statistical characterization of what types of parameters are necessary and how they influence the channel, enabling us to gain valuable design insight before the transmission happens. Therefore, we are interested in an average sum-rate, which is defined as follows:

$$I_e(\mathbf{y}; \mathbf{x}_1, \cdots, \mathbf{x}_u, \cdots, \mathbf{x}_{K_u}) \triangleq \sum_{u=1}^{K_u} \mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}} \left[\log_2 \det \left(\mathbf{I}_L + \frac{E_u}{\widetilde{N}_0 + \sigma_u^2} \mathbf{G}_u^H \widehat{\mathbf{H}}_u^H \widehat{\mathbf{H}}_u \mathbf{G}_u \right) \right], \quad (45)$$

where $\mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}}[\cdot]$ denotes the expectation operation in terms of channel coefficients, delay, and Doppler indices of all users, respectively. Note that the expectation is due to the fact that the sum-rate in (43) highly depends on the channel matrix $\mathbf{H}_u$, which contains the channel coefficients, delay, and Doppler indices of all users. While the average rate expression in (45) is a statistic, it is usually used for offering insight through numerical analysis, where the exact channel information is still required when calculating the sum-rate in (45). To shed light on the obtained average rate, let us consider an upper bound of (45). According to Jensen's inequality and the property of determinant operation, we have

$$I_e(\mathbf{y}; \mathbf{x}_1, \cdots, \mathbf{x}_u, \cdots, \mathbf{x}_{K_u}) \leq \sum_{u=1}^{K_u} \log_2 \det \left(\mathbf{I}_L + \mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}} \left[\frac{E_u}{\widetilde{N}_0 + \sigma_u^2} \mathbf{G}_u^H \widehat{\mathbf{H}}_u^H \widehat{\mathbf{H}}_u \mathbf{G}_u \right] \right), \quad (46)$$

where the equality is obtained if and only if the non-negative eigenvalues of the matrix $\mathbf{G}_u^H \mathbf{H}_u^H \mathbf{H}_u \mathbf{G}_u$ of all users share the same value [27]. However, the statistical characterizations remain unobservable in the upper bound (46) due to the presence of MUI in the denominator, resulting in the expectation of a ratio instead of a linear component. To solve this, we resort to first-order Taylor approximation in [46] and the closed-form expression of the approximated sum-rate with statistical characterizations is given in the following lemma.

Lemma 4: Considering an arbitrary multi-user transmission scheme, e.g., DDMA, the mapping matrix for the u-th user,

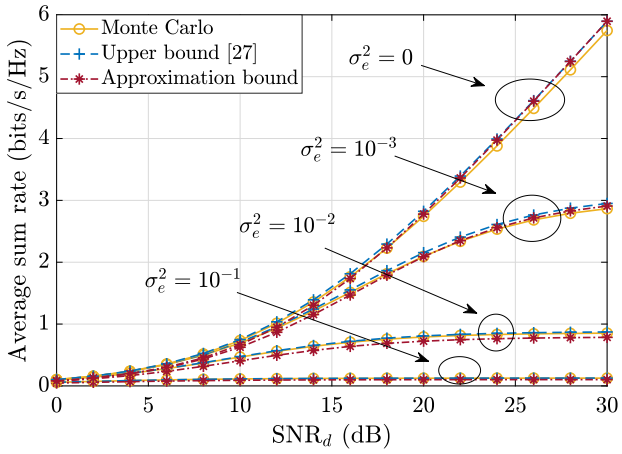

Fig. 5. The achievable sum-rate for four users following the power profile in (12). The Monte Carlo sum-rate is calculated based on (45), the upper bound is calculated based on [27], and the approximation bound is calculated based on (47). The range of delay and Doppler indices for four users are given as $l_{\min} \in [0, 0, 6, 6]$, $l_{\max} \in [4, 4, 10, 10]$, and $k_{\max} \in [3, 7, 3, 7]$, respectively.

$\mathbf{G}_u$, the corresponding estimated channel matrix, $\widehat{\mathbf{H}}_u$, and the variance of estimation error, σ_e^2, the closed-form expression of the approximation on the achievable sum-rate is given by

$$I_e \approx I_{\text{approx}} = \sum_{u=1}^{K_u} \log_2 \left(1 + \frac{E_u \Theta_u}{\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}} \right), \tag{47}$$

where we define

$$\Theta_u = \frac{P(e^{-l_{\min,u}} - e^{-l_{\max,u}})}{2k_{\max,u}(l_{\max,u} - l_{\min,u})} \tag{48}$$

as the *a priori* channel statistical information.

Proof: The proof is given in Appendix D.

According to (47), we define the *effective* SINR for the u-th user by

$$\text{SINR}_u = \frac{E_u \Theta_u}{\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}}, \tag{49}$$

where Θ_u is given in (48). To verify the correctness of our approximation, we present the comparison of the average achievable sum-rate among various approaches in Fig. 5. Specifically, we consider the OTFS systems with four users, $M = 16$, $N = 8$, $P = 4, L = 8$, practical power-delay and power-Doppler profiles in (12), and DDMA. The Monte Carlo method is based on (45), the upper bound based on [27], and the approximation bound based on (47) with various maximum delay and Doppler indices are considered and compared. Here, we define $\text{SNR}_d = E_{\text{tot}}/\widetilde{N}_0$. As shown in the figure, the derived approximated sum-rate shows a close match with the Monte Carlo calculation and the corresponding upper bound across various SNR levels and estimation error conditions. It is noteworthy that a larger estimation error significantly degrades the sum-rate performance under the same SNR conditions. Additionally, as the SNR increases, rate floors become evident, which aligns well with our previous analysis presented after Lemma 3.

B. Power Allocation for Multi-User Transmissions Based on Approximated Average Sum-Rate

It is interesting to notice that the expressions in (47) and (49) are highly related to the symbol energy and statistical CSI of the intended user and interference users, providing the possibility of power allocation at the transmitter side before transmissions. Therefore, we in the following present the power allocation that maximizes the approximation of the sum-rate based on the statistical CSI. Note that the rate optimal power allocation tends to assign a larger power to users with a better channel condition according to the water-filling principle. Consequently, the weak users are often power-hungry if the user powers are allocated purely to maximize the achievable rate. Therefore, we in the following also consider the user fairness to ensure each user can transmit some information. Mathematically, this is translated into a weighted sum of each user's rates and we therefore obtain the following optimization problem.

$$\max_{E_u, u=1,\ldots,K_u} \sum_{u=1}^{K_u} w_u \log_2 \left(1 + \frac{E_u \Theta_u}{\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}} \right) \tag{50a}$$

$$\text{subject to}: \sum_{u=1}^{K_u} E_u \le E_{\text{tot}}, \tag{50b}$$

$$E_u \le E_{\max}, \forall u = 1, \cdots, K_u, \tag{50c}$$

where w_u denotes the weighting factor of the u-th user and E_{tot} denotes the total symbol power budget of the system. $E_{\max}$ in (50c) puts an upper limit on the transmission power of each user, whose value is a constant that depends on the hardware limitations of the power amplifier. Additionally, (50c) can prevent excessive power allocation to any single user, thereby enhancing user fairness. Moreover, it is important to emphasize that although the specific expressions of the weighting factor may vary based on different user fairness principles, our derivations remain consistently valid, and we will give one example for demonstration purposes in the numerical results.

C. Optimal Power Control Based on Fractional Programming

It can be observed from (50a) that the main components of the objective function, i.e., the SINR terms, are in the ratio form, and a logarithm function is also involved in the equation that is nondecreasing and concave. Consequently, the optimization problem is non-convex. To solve this problem, we can apply the quadratic transform [47] to each SINR term to obtain the optimal solution. Specifically, the problem (50) can be reformulated as

$$\max_{E_u, u=1,\ldots,K_u} f(\mathcal{E}, \boldsymbol{\beta}) \tag{51a}$$

$$\text{subject to}: \sum_{u=1}^{K_u} E_u \le E_{\text{tot}}, \tag{51b}$$

$$E_u \le E_{\max}, \forall u = 1, \cdots, K_u, \tag{51c}$$

Here, $\mathcal{E}$ denotes the collection comprising the symbol power for each user, which is defined as $\mathcal{E} = \{E_1, \ldots, E_u, \ldots, E_{K_u}\}$. Similarly, $\boldsymbol{\beta}$ represents the set of all auxiliary variables, with $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_u, \ldots, \beta_{K_u}\}$. The new objective function $f(\mathcal{E}, \boldsymbol{\beta})$ is given as

$$\begin{aligned} &f(\mathcal{E}, \boldsymbol{\beta}) \\ &= \sum_{u=1}^{K_u} w_u \log_2 \\ &\quad \times \Bigg(1 + \underbrace{2\beta_u \sqrt{E_u \Theta_u} - \beta_u^2 \left(\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}\right)}_{g(\mathcal{E}, \beta_u)}\Bigg). \end{aligned} \tag{52}$$

According to [47], the optimal β_u for fixed $\mathcal{E}$ can be found in closed-form by taking $\frac{\partial f}{\partial \beta_u} = 0$ as

$$\beta_u^\star = \frac{\sqrt{E_u \Theta_u}}{\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}}, \forall u = 1, \ldots, K_u. \tag{53}$$

On the other hand, the function $g(\mathcal{E}, \beta_u)$ in (52) is concave in $\mathcal{E}$ for fixed β_u [47]. Therefore, the function $f(\mathcal{E}, \boldsymbol{\beta})$ is concave in $\mathcal{E}$ for fixed β_u, due to the concave and nondecreasing of the logarithm function. With a fixed β_u, the optimal solution for $\mathcal{E}$ can be obtained by a numerical convex optimization tool. The solution details are summarized in Algorithm 1.

Algorithm 1 Optimal Power Allocation With User Fairness

Initialization: Initialize $\mathcal{E}$ by uniform allocating the power, i,e, $E_u = \frac{E_{\text{tot}}}{K_u}, \forall u = 1, \cdots, K_u$.

repeat

Step 1: Update $\boldsymbol{\beta}$ by (53).

Step 2: Update $\mathcal{E}$ by solving the convex optimization problem (51) with fixed $\boldsymbol{\beta}$.

until The value in function (52) converges.

The main complexity within each iteration of Algorithm 1 arises from Step 2, as Step 1 involves only closed-form initialization. Step 2 requires solving problem (51) using a numerical convex optimization tool such as CVX, where the complexity depends on the number of variables and constraints. Specifically, problem (51) involves K_u variables and $K_u + 1$ affine constraints, leading to a computational complexity of $O(K_u^3(K_u + 1))$ [48]. Considering $\mathcal{I}$ as the number of iterations, the total computational complexity of Algorithm 1 is $O(\mathcal{I} \cdot K_u^3 \cdot (K_u + 1))$. As will be shown later in the numerical results section, the number of iterations is quite small, where Algorithm 1 can typically converge within 10 iterations and can converge within 5 iterations with only a slight performance reduction. Therefore, our proposed design can be effectively solved within the period when the statistical CSI remains unchanged, which is typically on the order of seconds [31].

D. Closed-Form Optimal Power Control

It is worth noting that while we achieved the optimal power allocation for each user in the previous subsection, the solution was derived through numerical optimization offering limited engineering insight. More importantly, there is in general no analytical formula for the global optimal solution of problem (50), despite various sub-optimal solutions with either closed-form or low complexity can be found in the literature [47], [49]. To shed light on the power allocation based on statistical CSI in OTFS systems, we here slightly reformulate the optimization problem and derive a closed-form solution for optimal power allocation. According to the chain rule in [42], we can rewrite the approximated achievable sum-rate in (47) as follows:

$$\begin{aligned} I_{\text{approx}} &= \sum_{u=1}^{K_u} \log_2 \left(1 + \frac{E_u \Theta_u}{\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}}\right) \\ &= \log_2 \left(1 + \frac{\sum_{u=1}^{K_u} E_u \Theta_u}{\widetilde{N}_0}\right). \end{aligned} \tag{54}$$

To ensure user fairness, we introduce a constraint requiring each user to achieve a minimum rate, rather than capturing the notion of user fairness via weighting factors, as in problem (50). Consequently, the optimization problem to maximize the sum-rate can be formulated as follows:

$$\max_{E_u, u=1,\ldots,K_u} \log_2 \left(1 + \frac{\sum_{u=1}^{K_u} E_u \Theta_u}{\widetilde{N}_0}\right) \tag{55a}$$

$$\text{subject to}: \ (50b), \ (50c),$$

$$I_u \geq I_{\min,u}, \forall u = 1, \cdots, K_u, \tag{55b}$$

where $I_u = \log_2 \left(1 + \frac{E_u \Theta_u}{\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}}\right)$ denotes the achievable rate for the u-th user and $I_{\min,u}$ represents the corresponding minimum constant rate requirement. Note that the problem (55) is convex except for constraint (55b). To resolve this, (55b) can be equivalently rewritten as an affine constraint as follows:

$$E_u \Theta_u \geq (2^{I_{\min,u}} - 1) \left(\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}\right). \tag{56}$$

Building upon this, we can derive the closed-form solution for optimal power allocation by applying the KKT conditions, whose global optimality is guaranteed due to the convexity of the problem [50]. The corresponding Lagrangian function is expressed as follows:

$$\begin{aligned} &\mathcal{L}(\mathcal{E}, \lambda, \boldsymbol{\mu}) \\ &= \log_2 \left(1 + \frac{\sum_{u=1}^{K_u} E_u \Theta_u}{\widetilde{N}_0}\right) - \lambda \left(\sum_{u=1}^{K_u} E_u - E_{\text{tot}}\right) \\ &\quad - \sum_{u=1}^{K_u} \mu_u \left((2^{I_{\min,u}} - 1) \left(\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'}\right) - E_u \Theta_u\right), \end{aligned} \tag{57}$$

where λ and the set $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_u, \ldots, \mu_{K_u}\}$ represent the nonnegative Lagrangian multipliers. Then, the dual problem of (55) is given by

$$\min_{\lambda, \boldsymbol{\mu} \geq 0} \max_{\mathcal{E}} \ \mathcal{L}(\mathcal{E}, \lambda, \boldsymbol{\mu}). \tag{58}$$

The KKT conditions pertinent to this optimization are listed as follows:

$$\lambda^\star \geq 0, \tag{59a}$$

$$\mu_u^\star \geq 0, \forall u = 1, \ldots, K_u, \tag{59b}$$

$$\lambda \left(\sum_{u=1}^{K_u} E_u - E_{\text{tot}} \right) = 0, \tag{59c}$$

$$\mu_u \left((2^{I_{\min,u}} - 1) \left(\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'} \right) - E_u \Theta_u \right) = 0, \quad \forall u = 1, \ldots, K_u, \tag{59d}$$

$$\frac{\partial \mathcal{L}}{\partial E_u} = \frac{\Theta_u}{(\widetilde{N}_0 + \sum_{u=1}^{K_u} E_u \Theta_u) \ln(2)} - \lambda - \Theta_u (2^{I_{\min,u}} - 1) \sum_{k=1}^{u-1} \mu_k + \Theta_u \mu_u = 0. \tag{59e}$$

By substituting $u = 1$ into (59e), we can isolate the negative components involving the summation of μ_k in (59e), thereby simplifying the analysis and directly examining λ as follows:

$$\lambda = \frac{\Theta_1}{(\widetilde{N}_0 + \sum_{u=1}^{K_u} E_u \Theta_u) \ln(2)} + \Theta_1 \mu_1 > 0, \tag{60}$$

which subsequently implies

$$\sum_{u=1}^{K_u} E_u = E_{\text{tot}}. \tag{61}$$

Furthermore, by setting $\frac{\partial \mathcal{L}}{\partial E_1} = \frac{\partial \mathcal{L}}{\partial E_2} = 0$ and assuming that the users are sorted in descending order based on the quality of the *a priori* channel statistical information, i.e., $\Theta_1 \geq \cdots \geq \Theta_{K_u}$, we have

$$\Theta_2 \mu_2 = \left(\Theta_1 + \Theta_2 (2^{I_{\min,u}} - 1) \right) \mu_1 + \frac{\Theta_1 - \Theta_2}{(\widetilde{N}_0 + \sum_{u=1}^{K_u} E_u \Theta_u) \ln(2)} > 0. \tag{62}$$

Similarly, by setting $\frac{\partial \mathcal{L}}{\partial E_1} = \frac{\partial \mathcal{L}}{\partial E_u} = 0, \forall u = 2, \ldots, K_u$, we have

$$\mu_u > 0, \forall u = 2, \ldots, K_u, \tag{63}$$

which subsequently implies

$$E_u \Theta_u = (2^{I_{\min,u}} - 1) \left(\widetilde{N}_0 + \sum_{u'=u+1}^{K_u} E_{u'} \Theta_{u'} \right), \quad \forall u = 2, \ldots, K_u, \tag{64}$$

leading to

$$I_u = I_{\min,u}, \forall u = 2, \ldots, K_u. \tag{65}$$

Based on (64), the optimal power allocation for the K_u-th user $E_{K_u}^\star$ can be obtained as

$$E_{K_u}^\star = \frac{(2^{I_{\min,K_u}} - 1) \widetilde{N}_0}{\Theta_{K_u}}, \tag{66}$$

and the optimal power allocation for the $(K_u - 1)$-th user $E_{K_u - 1}^\star$ can be obtained as

$$E_{K_u - 1}^\star = \frac{(2^{I_{\min,K_u - 1}} - 1) \widetilde{N}_0 2^{I_{\min,K_u}}}{\Theta_{K_u - 1}}. \tag{67}$$

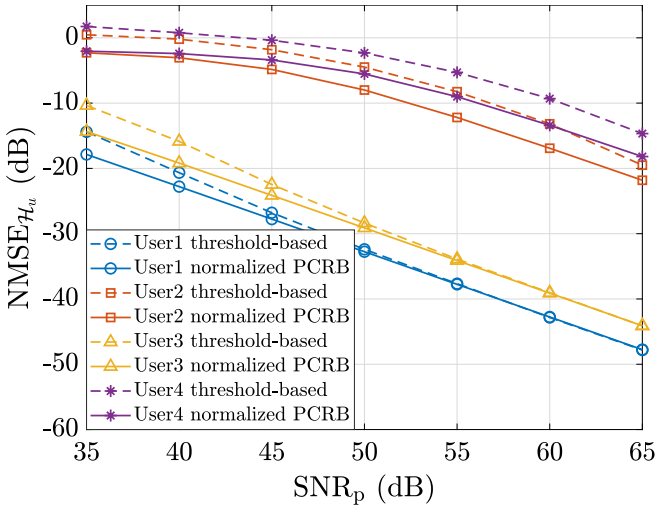

Fig. 6. NMSE and NPCRB comparison for four users.

Following the similar steps above and considering constraint (50c), the closed-form solution of optimal power allocation can be given as

$$E_u^\star = \frac{N_0}{\Theta_u} (2^{I_{\min,u}} - 1) 2^{\sum_{u'=u+1}^{K_u} I_{\min,u}}, \forall u = 2, \ldots, K_u, \tag{68}$$

$$E_1^\star = \min \left\{ E_{\max}, \; E_{\text{tot}} - \sum_{u=2}^{K_u} E_u^\star \right\}. \tag{69}$$

From (64)-(69), it can be observed that the optimal power allocation strategy at the transmitter side is to allocate the power according to the statistical CSI's quality of all users, except for the user with the best statistical channel condition, to ensure the minimum achievable rate can be satisfied. Subsequently, any excess power available, after satisfying all rate constraints, is allocated to the user with the best statistical CSI condition to maximize the achievable sum rate. Furthermore, since $E_u^\star$ in (68) represents the power required to satisfy the minimum achievable rate constraints, it is necessary that $E_u^\star \leq E_{\max}$ holds in (68) to ensure the feasibility of this allocation strategy.

V. Numerical Results

In this section, we present the simulation results for multi-user OTFS transmission to verify our proposed designs. If not specified otherwise, we set $M = 64$, $N = 16$, $L = 16$, $P = 4$, and $K_u = 4$. Without loss of generality, we fix the noise PSD as $N_0 = 0$ dBm, such that the pilot SNR $\text{SNR}_p = \frac{\mathcal{P}_u}{N_0}$ shares the same value as the pilot power $\mathcal{P}_u$. The delay and Doppler indices are uniformly generated from $[l_{\min}, l_{\max}]$ and $[-k_{\max}, k_{\max}]$, respectively. The channel coefficients are generated randomly from the exponential power-delay and uniform power-Doppler profiles based on (12). Furthermore, we assume that all users' power profiles are jointly normalized from the view of the receiver side. The range of delay and Doppler indices for four users are given as $l_{\min} \in [0, 0, 6, 6]$, $l_{\max} \in [4, 4, 10, 10]$, and $k_{\max} \in [3, 7, 3, 7]$, respectively.

We first demonstrate the PCRB and NMSE performance for different users individually in terms of pilot SNR SNR_p

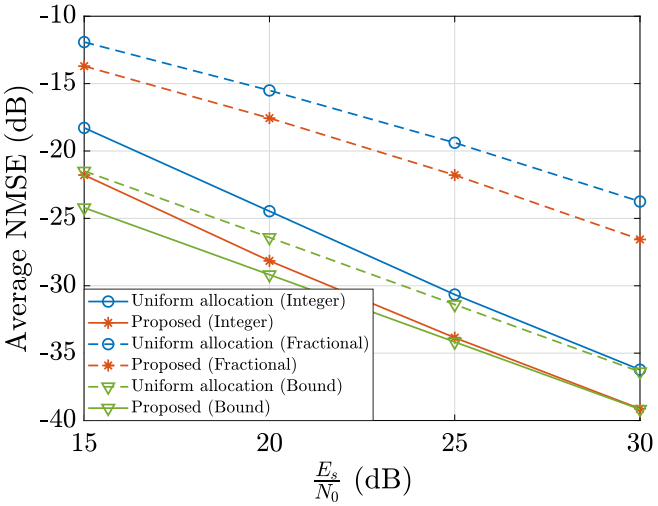

Fig. 7. NMSE comparison between uniform and proposed allocations with both integer and fractional delay and Doppler.

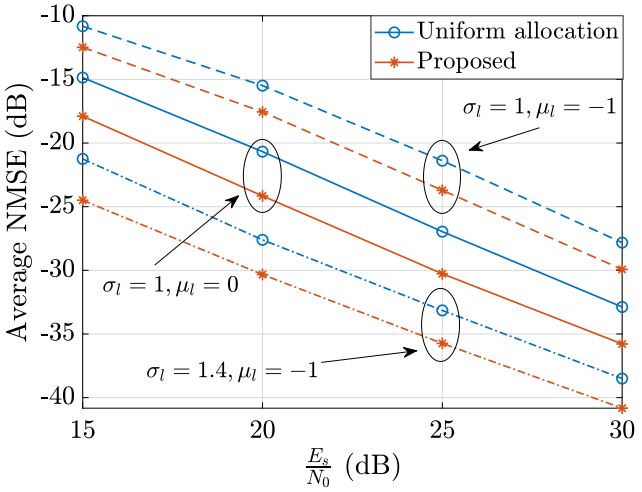

Fig. 8. NMSE comparison between uniform and proposed allocations with different power profiles.

in Fig. 6. The solid lines denote the normalized PCRB, and the dashed lines denote the corresponding NMSE. We define the channel matrix $\mathcal{H}_u$ by discretizing the channel $h_u(\tau, \nu)$ in (5). Consequently, the NMSE for u-th user is given by $\text{NMSE}_{\mathcal{H}_u} = \frac{||\hat{\mathcal{H}}_u - \mathcal{H}_u||^2}{||\mathcal{H}_u||^2}$, where $\hat{\mathcal{H}}_u$ denotes the estimated channel. It can be observed from Fig. 6 that the $\text{NMSE}_{\mathcal{H}_u}$ of all users coincide with the corresponding normalized PCRB at high pilot SNRs, where the user who obtains better statistical channel condition will coincide with corresponding PCRB at lower pilot SNR as expected. This is because the estimation accuracy depends on the pilot and channel power, owing to the convolution property of DD domain channels, where a better statistic channel condition indicates a larger average channel power, and therefore reducing the required pilot power for sufficiently accurate estimation performance. This observation is obvious for user 1 and user 3 but is less obvious for user 2 and user 4 as their statistical channel conditions are not as good. The users with different minimum and maximum delay and maximum Doppler indices have different estimation performances, which shows that pilot power allocation among multiple users to improve the overall estimation performances is possible. Additionally, it is shown in Fig. 6 that the estimation performance for both user 2 and user 3 is worse than that of user 1. Specifically, user 2 experiences such estimation degradation due to a larger minimum and maximum delay, while the performance of user 3 suffers due to a larger maximum Doppler shift compared to user 1, respectively. Nonetheless, the performance degradation of user 2 is significantly more severe than that of user 3. This indicates that the user's estimation performance is more sensitive to the delay range rather than the Doppler range by considering the exponential delay and uniform Doppler power profiles. Note that user 4 has the largest delay and Doppler indices and obtains the worst estimation performance as expected.

Next, Fig. 7 compares the average NMSE performances with different pilot power allocations, i.e., $\text{NMSE} = \frac{1}{K_u}\sum_u^{K_u} \text{NMSE}_{\mathcal{H}_u}$, in terms of transmission SNR $\frac{E_s}{N_0}$. Note that E_{tot} denotes the total symbol power budget and E_s denotes the normalized average data symbol energy, i.e., $E_s = E_{\text{tot}}/K_u$. The total pilot power $\mathcal{P} = 55$ dB is considered. We can observe from Fig. 7 that the proposed scheme obtains around 3 dB average NMSE performance improvement than the conventional uniform allocation scheme under the same total pilot power and transmission SNRs. The theoretical bound closely aligns with the actual NMSE at high SNR levels and demonstrates consistent performance improvement across all SNR conditions. Furthermore, the proposed scheme can provide NMSE performance improvement even in the presence of fractional delay and Doppler shifts. As shown in Fig. 7, the proposed scheme can provide around 2 dB improvement compared to the uniform allocation scheme, while a 3 dB improvement can be observed in a higher SNR regime. Additionally, it is important to emphasize that the pilot overhead is dominated by the number of guard symbols, which only depends on the maximum delay and Doppler of the channel, and is independent from pilot and data SNR [8]. Therefore, our approach does not require additional pilot overhead compared to the scheme in [8]. Furthermore, the NMSE performance of each user is not influenced by the number of served users, as each user's pilots are isolated from those of others by sufficient guard space.

Moreover, we consider different power profiles to demonstrate the robustness of our proposed pilot power allocation scheme. Specifically, we assume the average channel power follows a Laplace distribution [51] in delay and uniform distribution in Doppler, which is given by

$$\gamma_{p,u} = \frac{1}{2\sigma_l} e^{-\frac{|l_{p,u} - \mu_l|}{\sigma_l}} \times \frac{1}{2k_{\max,u}}, \tag{70}$$

where $\mu_l \in \mathbb{R}$ is a location parameter and $\sigma_l > 0$ is a scale parameter of the Laplace distribution. Note that the Laplace distribution is also within the family of exponential distributions, which includes two adjustable parameters that can be useful for studying the robustness of our proposed schemes. As shown in Fig. 8, our proposed allocation achieves improvements in NMSE performance across various power profile shapes, thereby confirming the robustness of our scheme.

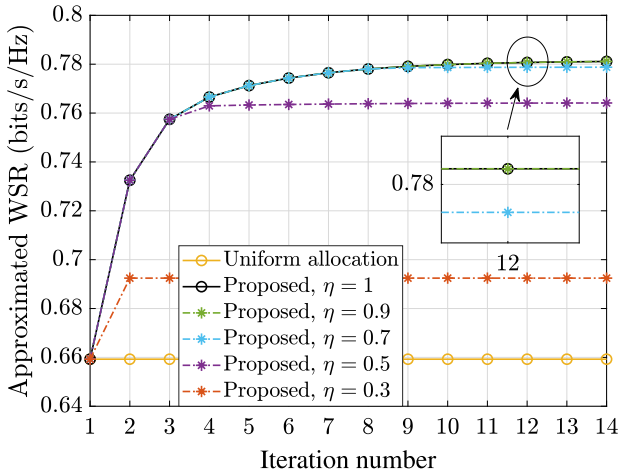

Fig. 9. Convergence of Algorithm 1.

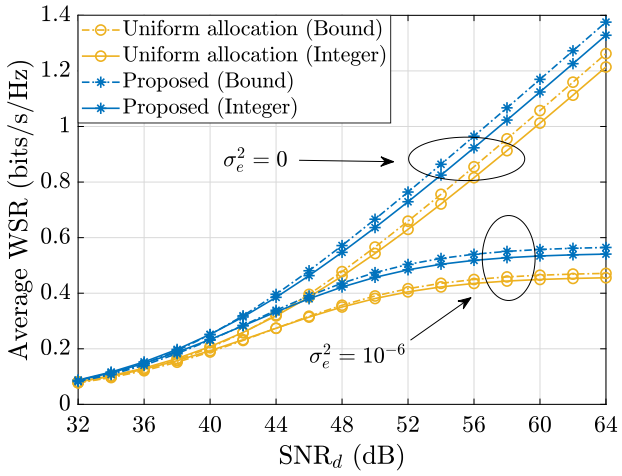

Fig. 10. Achievable average WSR comparison between uniform and proposed allocations.

Next, we validate and provide insights into the proposed average sum-rate optimization algorithms. The SNR$_d$ is defined by SNR$_d = E_{\text{tot}}/\widetilde{N}_0$ and the weighting factor for the users are $[0.022, 0.033, 0.322, 0.623]$, which is selected based on corresponding channel conditions. Specifically, Fig. 9 shows the convergence of the fractional programming algorithm adopted in Section IV-C, where SNR$_d$ = 52 dB and $\eta = \frac{E_{\max}}{E_{\text{tot}}}$ denotes the proportion of the maximum transmission power for each user relative to the total power budget. Note that $E_{\max}$ may exceed $\frac{E_{\text{tot}}}{K_u}$, i.e., $\eta \geq \frac{1}{K_u}$. This is because users with poor channel quality are often allocated more power to enhance fairness. However, it is not always necessary for all users to transmit at their maximum power $E_{\max}$ due to MUI management and energy efficiency considerations. It is observed that Algorithm 1 typically converges within 10 iterations and achieves an approximated weighted sum-rate (WSR) improvement over the uniform allocation scheme. Additionally, while setting the maximum transmission power of users equal to the total power budget, i.e., $\eta = 1$, can achieve the largest performance improvement, it tends to allocate most of the power to a single user, leading to a less fair power allocation, as previously analyzed after (50). Therefore, reducing the maximum transmission power is necessary to enhance user fairness. The influence of the maximum transmission power is minimal when η values are 0.9 and 0.7. However, the WSR slightly decreases when $\eta = 0.5$ and continues to decrease as η is reduced to 0.3. This reduction occurs because the maximum transmission power constraint may prevent the users from achieving the optimal power level. To strike an effective balance among convergence speed, WSR performance, and user fairness, we will adopt $\eta = 0.5$ in the following analysis.

The comparison of the average WSR between different symbol power allocations across various SNR and estimation error conditions is shown in Fig. 10. From Fig. 10, it is evident that the approximation on the sum-rate derived in (47) is in close agreement with the actual sum-rate. Additionally, it is noteworthy that the sum-rate achieved by using our proposed allocation design is superior to that of the traditional uniform allocation per symbol under the same total power budget.

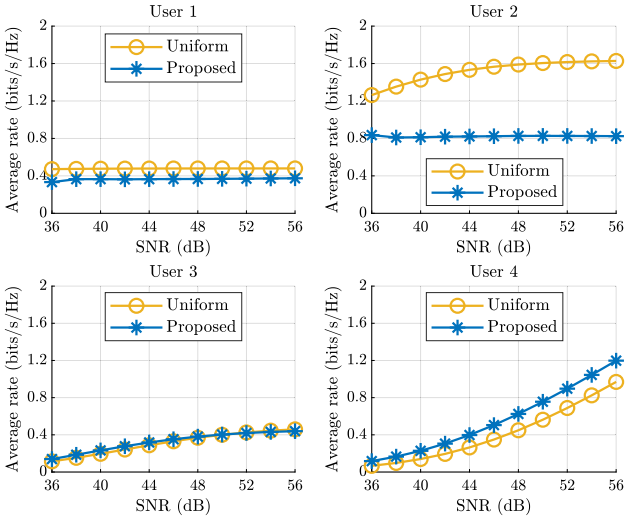

Fig. 11. Average achievable rate comparison between uniform and proposed allocations for each user.

This improvement in WSR can still be maintained even in the presence of estimation errors. The average achievable rates for each user under different allocation schemes are presented in Fig. 11. As shown, the achievable rate for user 4 improves with the adoption of our proposed allocation, while the rates for user 1 and user 3 decrease slightly, and the rate for user 2 decreases more significantly. This is not unexpected, as user 4 is assigned the highest weight to enhance its rate and ensure user fairness. Although the rate for user 2 shows a significant decrease, it still maintains a relatively good performance due to its favorable channel quality. However, user 1, despite having the best statistical channel quality, achieves a lower rate under both the uniform and proposed allocation schemes. This occurs because the strong users usually face a large amount of MUI while the weaker users are less influenced owing to the SIC decoding [52]. As a result, increasing the achievable rate of weaker users may degrade the achievable rate of stronger users as the MUI is enhanced. Additionally, it can be noticed that the performance of user 1 to user 3 show rate floors due to nontrivial MUI. Consequently,

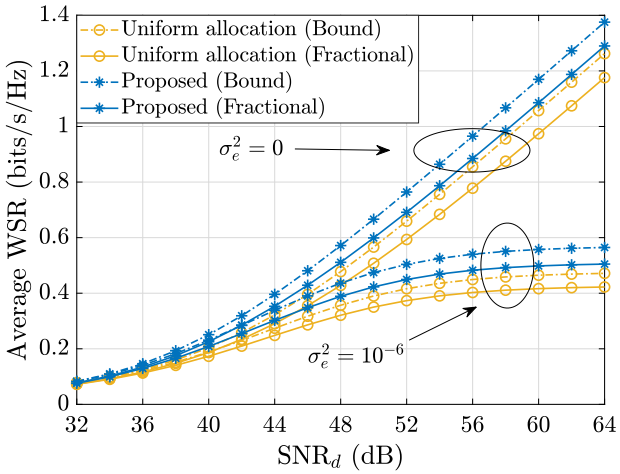

Fig. 12. Achievable average WSR comparison between uniform and proposed allocations with fractional delay and Doppler shifts.

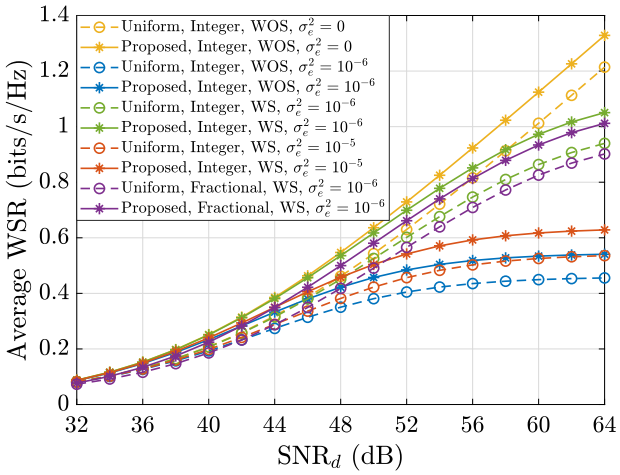

Fig. 13. Achievable average WSR comparison between uniform and proposed allocations, accounting for channel sparsity and estimation error.

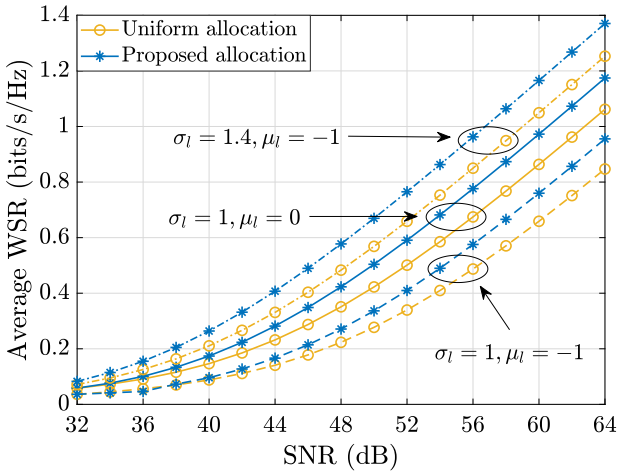

Fig. 14. Achievable average WSR comparison between uniform and proposed allocations with different power profiles.

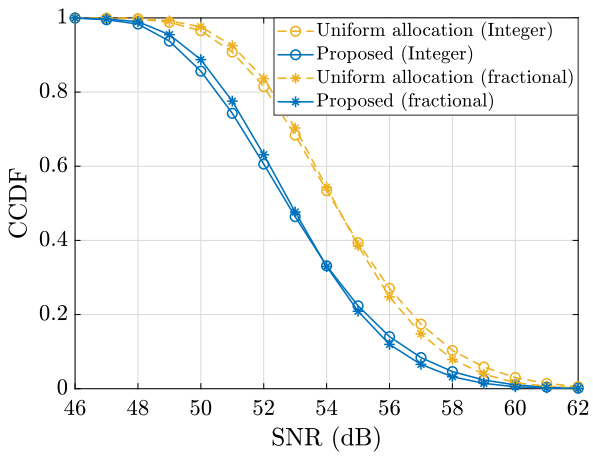

Fig. 15. CCDF comparison between uniform and proposed allocations.

serving more users would intensify MUI, causing the rate floor to become more severe and appear under lower SNR conditions. A potential solution for this is to consider user grouping, which will be studied in our future work.

Moreover, we demonstrate in Fig. 12 that the proposed scheme can provide a similar average WSR performance improvement compared to the integer case even when considering fractional delay and Doppler shifts. Note that the widening gap between the actual achievable sum-rate and the derived approximation in the fractional case compared to the integer case arises from the symbol spreading due to the presence of fractional delay and Doppler shifts. However, it is important to emphasize that the performance improvement on average WSR remains consistent, demonstrating the effectiveness of our proposed design in dealing with fractional delay and Doppler shifts. Furthermore, Fig. 13 demonstrates the impact of estimation error on WSR while accounting for channel sparsity. Specifically, "with sparsity" (WS) represents that errors only occur in non-zero elements in $\mathbf{H}_u$ in (41), whereas "without sparsity" (WOS) represents errors occur in each element in $\mathbf{H}_u$. As shown in Fig. 13, the WSR of the WS scenario outperforms that of the WOS scenario under the same error variance and SNR conditions, which aligns well with our analysis. Additionally, by replacing the power profile in (12) by (70), it can be observed from Fig. 14 that our proposed scheme robustly provides average WSR improvements across various power profiles. Alternatively, we demonstrate the achievable sum-rate performance via the complementary cumulative distribution function (CCDF) as shown in Fig. 15. We define the SNR as SNR $= E/\widetilde{N}_0$, where $E = (NE_{\text{tot}} + \mathcal{P})/(MN)$ denotes the transmission power. The sum-rate threshold is set to the average sum-rate of uniform allocation at SNR $= 54$ dB and the CCDF computes the probability that the sum-rate in each channel realization, i.e., OTFS frame, falls below this threshold. From Fig. 15, it can be observed that the proposed allocation exhibits a lower probability of falling below the threshold compared to the uniform allocation under various SNR conditions and even in the presence of fractional delay and Doppler shifts. This indicates that the proposed scheme achieves better sum-rate performance than the uniform allocation scheme.

Furthermore, the cumulative distribution function (CDF) of user rate is presented in Fig. 16, where the proposed

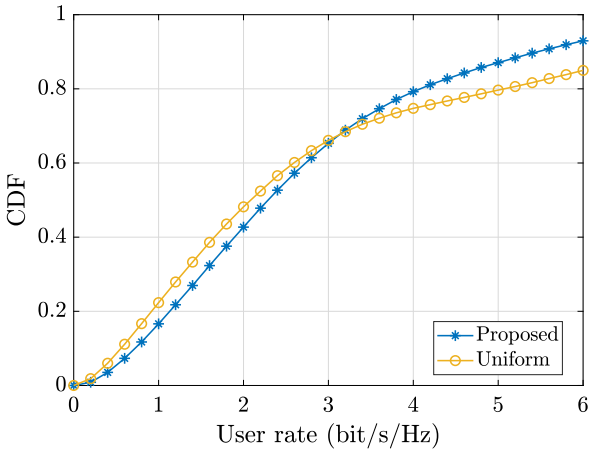

Fig. 16. The CDF of the uniform allocation scheme and the proposed closed-form allocation scheme.

allocation is based on the closed-form solution in Section IV-D with $I_{\min,u} = 2$ bit/s/Hz, $\forall u = 2, \ldots, K_u$ and $\text{SNR}_d = 48$ dB. We note that under uniform allocation, some users' rate may fall below the minimum rate requirement because, without prior channel information, the transmitter is unable to allocate power effectively to satisfy these constraints. This intuitively leads to less fairness compared to our proposed allocation scheme, as evidenced in Fig. 16. Specifically, it can be observed that for over 60% users, which includes weaker users of particular interest, there is an increase in rate performance compared to that of a uniform allocation scheme. This demonstrates that our proposed allocation scheme can improve the performance of low-rate users and therefore elevate the quality of user experience and enhance fairness.

VI. Conclusion

In this paper, we conducted the design of uplink multi-user OTFS transmissions based on statistical CSI. Specifically, we proposed a pilot power allocation for multi-user OTFS transmissions based on the derived PCRBs, which was derived in the closed-form by adopting the *a priori* channel statistical information. Based on the PCRB, an optimization problem for pilot power allocation was formulated and the optimal allocation was obtained. Next, we provided an in-depth analysis of the achievable sum-rate for uplink multi-user OTFS transmissions based on the statistical CSI. Specifically, we derived a closed-form expression on the approximation of the achievable sum-rate, which only relates to the *a priori* channel statistical information. Then, an optimization problem for symbol power allocation was formulated by considering user fairness and statistical CSI, whose optimal solution was obtained by the fractional programming approach. Numerical results validated our analysis and presented an approximate 3 dB NMSE improvement in terms of channel estimation accuracy, compared to the conventional embedded pilot scheme. Our results have also shown a performance improvement in achievable sum-rate by adopting the proposed power allocation, compared to the uniform allocation scheme. Additionally, we believe that our work could be extended to more complex scenarios, such as joint power allocation between pilots and data signals in scenarios where transmit power is limited. This extension could be formulated as an optimization problem aimed at maximizing the achievable sum-rate, while being subject to constraints on channel estimation accuracy and total transmission power. It is also noteworthy that our study focuses on the uncoded case, and incorporating channel coding would significantly reduce the required SINR. Therefore, exploring code design could be a promising approach to further enhance power efficiency [53].

Appendix A
Proof of Lemma 1

Recall that $\ln p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{z}) = C - \frac{1}{N_0}\sum_{k,l}|y[k,l] - z[k,l]|^2$, where constant C is independent from $\boldsymbol{\theta}$. For ease of derivation, we omit $[k,l]$ and take 1st-order partial derivative as

$$\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{z})}{\partial \theta_i} = -\frac{1}{N_0}\sum_{k,l}\left(-\frac{\partial z}{\partial \theta_i}(y^* - z^*) - \frac{\partial z^*}{\partial \theta_i}(y - z)\right). \tag{71}$$

Note that $\mathbb{E}_y[y - z] = \mathbb{E}_y[y^* - z^*] = 0$ because z is the mean of y. We can prove the regularity condition of PCRB is satisfied as $\mathbb{E}_{\boldsymbol{\theta}\mathbf{y}}\left[\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{z})}{\partial \theta_i}\right] = 0$. Next, we take the second-order partial derivatives as

$$\begin{aligned}&\frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{z})}{\partial \theta_i \partial \theta_j}\\ &= -\frac{1}{N_0}\sum_{k,l}\Big(-\frac{\partial^2 z}{\partial \theta_i \partial \theta_j}(y^* - z^*) + \frac{\partial z}{\partial \theta_i}\frac{\partial z^*}{\partial \theta_j}\\ &\quad - \frac{\partial^2 z^*}{\partial \theta_i \partial \theta_j}(y - z) + \frac{\partial z^*}{\partial \theta_i}\frac{\partial z}{\partial \theta_j}\Big),\end{aligned} \tag{72}$$

where we apply $\mathbb{E}_y[y - z] = \mathbb{E}_y[y^* - z^*] = 0$ again to simplify the result as

$$\begin{aligned}&-\mathbb{E}_{\boldsymbol{\theta}\mathbf{y}}\left[\frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{z})}{\partial \theta_i \partial \theta_j}\right]\\ &= \frac{1}{N_0}\sum_{k,l}\left(\frac{\partial z}{\partial \theta_i}\frac{\partial z^*}{\partial \theta_j} + \frac{\partial z^*}{\partial \theta_i}\frac{\partial z}{\partial \theta_j}\right)\\ &= \frac{2}{N_0}\mathbb{E}_{\boldsymbol{\theta}}\left[\Re\left\{\sum_{k,l}\left(\frac{\partial z}{\partial \theta_i}\right)^*\left(\frac{\partial z}{\partial \theta_j}\right)\right\}\right].\end{aligned} \tag{73}$$

Now, we verify that $\mathbf{J}_r$ is a diagonal matrix. Recall that $z[k,l]$ contains $x_u[[k - k_{p,u}]_N, [l - l_{p,u}]_M]$, which is all zeros except for the pilot. First, let θ_i and θ_j be two arbitrary parameters associated to two different paths. We can observe that equation (73) contains the component $x_u[[k - k_{p,u}]_N, [l - l_{p,u}]_M]^* x_u[[k - k_{p,u}]_N, [l - l_{p,u}]_M]$, which equals to zero because the $(k_{p,u}, l_{p,u})$ for different paths is different. Next, we consider the case where θ_i and θ_j are two parameters from the same path. If we fix $\theta_i = \Re(h_{p,u})$, then for $\theta_j = \Im(h_{p,u})$, we have $\mathbf{J}_p^{(i,j)} = \Re\{j\} = 0$; for $\theta_j = l_{p,u}$ or $\theta_j = k_{p,u}$, we have $\mathbf{J}_p^{(i,j)} = \mathbb{E}_{\boldsymbol{\theta}}[h_{p,u}] = 0$. For $\theta_i = l_{p,u}$ and $\theta_j = k_{p,u}$, we have $\mathbf{J}_p^{(i,j)} = \mathbb{E}_{\boldsymbol{\theta}}[k_{p,u}] = 0$. The remaining off-diagonal elements can be proved to be zero in a similar approach. ■

$$
\begin{aligned}
&I(\mathbf{y};\mathbf{x}_1,\cdots,\mathbf{x}_{K_u}|\widehat{\mathbf{H}}_1,\cdots,\widehat{\mathbf{H}}_{K_u})\\
&=\sum_{u=1}^{K_u}\left(\log_2(\pi e)^L\det\left(\sum_{i=u}^{K_u}E_i\widehat{\mathbf{H}}_i\mathbf{G}_i\mathbf{G}_i^H\widehat{\mathbf{H}}_i^H+\widetilde{N}_0\mathbf{I}\right)-\log_2(\pi e)^L\det\left(\sum_{j=u+1}^{K_u}E_j\widehat{\mathbf{H}}_j\mathbf{G}_j\mathbf{G}_j^H\widehat{\mathbf{H}}_j^H+\widetilde{N}_0\mathbf{I}\right)\right)\\
&=\sum_{u=1}^{K_u}\log_2\left(\det\left[\mathbf{I}+\left(E_u\widehat{\mathbf{H}}_u\mathbf{G}_u\mathbf{G}_u^H\widehat{\mathbf{H}}_u^H\right)\left(\sum_{j=u+1}^{K_u}E_j\widehat{\mathbf{H}}_j\mathbf{G}_j\mathbf{G}_j^H\widehat{\mathbf{H}}_j^H+\widetilde{N}_0\mathbf{I}\right)^{-1}\right]\right)\\
&=\sum_{u=1}^{K_u}\log_2\left(\det\left[\mathbf{I}_L+\frac{E_u}{\widetilde{N}_0+\sigma_u^2}\mathbf{G}_u^H\widehat{\mathbf{H}}_u^H\widehat{\mathbf{H}}_u\mathbf{G}_u\right]\right).
\end{aligned}
\tag{79}
$$

Appendix B
Proof of Lemma 2

Recall that $\ln p(\mathbf{h})=\sum_u\sum_p-\frac{\Re(h_{p,u})^2+\Im(h_{p,u})^2}{\gamma_{p,u}}-\ln(\pi\gamma_{p,u})$, whose first-order partial derivatives are given by

$$
\frac{\partial\ln p(\mathbf{h})}{\partial\theta_i}=\begin{cases}-\dfrac{2\Re(h_{p,u})}{\gamma_{p,u}}, & 1\le i\le PK_u,\\ -\dfrac{2\Im(h_{p,u})}{\gamma_{p,u}}, & PK_u<i\le 2PK_u,\end{cases}
\tag{74}
$$

and for $2PK_u<i\le 4PK_u$, we have

$$
\frac{\partial\ln p(\mathbf{h})}{\partial\theta_i}=-\frac{1}{\gamma_{p,u}}\frac{\partial\gamma_{p,u}}{\partial\theta_i}+\frac{|h_{p,u}|^2}{\gamma_{p,u}^2}\frac{\partial\gamma_{p,u}}{\partial\theta_i}.
\tag{75}
$$

Note that we can prove the regularity condition by substituting $\mathbb{E}[h_{p,u}]=0$ and $\mathbb{E}[|h_{p,u}|^2]=\gamma_{p,u}$ into (74) and (75), respectively. After taking the second-order partial derivatives and expectations, we can obtain

$$
\mathbf{J}_p^{(i,i)}=\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{2}{\gamma_{p,u}}\right],1\le i\le 2PK_u.
\tag{76}
$$

Note that we obtain the results for $2PK_u<i,j\le 4PK_u$ in a similar approach by using $\mathbb{E}[h_{p,u}]=0$, $\mathbb{E}[|h_{p,u}|^2]=\gamma_{p,u}$, and $\mathbb{E}[k_{p,u}]=0$ suitably. Finally, we give hints on proving the diagonal property of $\mathbf{J}_p$. If $\theta_i=\Re(h_{p,u})$, then $\mathbf{J}_r^{(i,j)}=\mathbb{E}_{\boldsymbol{\theta}}[\Re\{h_{p,u}\}]=0,\forall i\ne j$. If we fix $\theta_i=k_{p,u}$, then $\mathbf{J}_r^{(i,j)}=0,\forall i\ne j$ because $p(\mathbf{h})$ is independent of $k_{p,u}$. ■

Appendix C
Proof of Lemma 3

Based on (8) and (41), the DD domain input-output relationship can be rewritten as

$$
\mathbf{y}=\sum_{u=1}^{K_u}\widehat{\mathbf{H}}_u\mathbf{x}_u+\underbrace{\sum_{u=1}^{K_u}\Delta\mathbf{H}_u\mathbf{x}_u+\mathbf{w}}_{\mathbf{n}},
\tag{77}
$$

where $\mathbf{n}$ represents the equivalent noise, accounting for the effects of imperfect CSI. The covariance of $\mathbf{n}$ can be calculated under the Gaussian approximation, i.e.,

$$
\begin{aligned}
\mathbb{E}[\mathbf{n}\mathbf{n}^H]&=N_0\mathbf{I}_{MN}+\sum_{u=1}^{K_u}E_u\mathbb{E}[\Delta\mathbf{H}_u\mathbf{G}_u\mathbf{G}_u^H\Delta\mathbf{H}_u^H],\\
&=\underbrace{N_0\mathbf{I}_{MN}+E_{\text{tot}}L\sigma_e^2\mathbf{I}_{MN}}_{\widetilde{N}_0\mathbf{I}_{MN}}.
\end{aligned}
\tag{78}
$$

According to the chain rule [44] and the property of the matrix determinant, we have (79), as shown at the top of this page. ■

Appendix D
Proof of Lemma 4

For the sake of presentation, we recall the average sum-rate I_e is upper bounded in (46) as follows

$$
\begin{aligned}
&I_e(\mathbf{y};\mathbf{x}_1,\cdots,\mathbf{x}_u,\cdots,\mathbf{x}_{K_u})\\
&\le\sum_{u=1}^{K_u}\log_2\det\left(\mathbf{I}_L+\mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}}\left[\frac{E_u}{\widetilde{N}_0+\sigma_u^2}\mathbf{G}_u^H\widehat{\mathbf{H}}_u^H\widehat{\mathbf{H}}_u\mathbf{G}_u\right]\right).
\end{aligned}
\tag{80}
$$

In order to obtain the closed-form expression of the approximated sum-rate with statistical characterization, we apply first-order Taylor approximation [46] on $\mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}}\left[\frac{E_u}{\widetilde{N}_0+\sigma_u^2}\mathbf{G}_u^H\widehat{\mathbf{H}}_u^H\widehat{\mathbf{H}}_u\mathbf{G}_u\right]$ in (80) as follows

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}}\left[\frac{E_u}{\widetilde{N}_0+\sigma_u^2}\mathbf{G}_u^H\widehat{\mathbf{H}}_u^H\widehat{\mathbf{H}}_u\mathbf{G}_u\right]\\
&\overset{(a)}{=}\frac{E_u\mathbf{G}_u^H\mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}}\left[\widehat{\mathbf{H}}_u^H\widehat{\mathbf{H}}_u\right]\mathbf{G}_u}{\widetilde{N}_0+\frac{1}{L}\text{Tr}\left(\sum_{u'=u+1}^{K_u}E_{u'}\mathbf{G}_{u'}^H\mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}}\left[\widehat{\mathbf{H}}_{u'}^H\widehat{\mathbf{H}}_{u'}\right]\mathbf{G}_{u'}\right)}\\
&\overset{(b)}{=}\frac{E_u\mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}}\left[\sum_{p=1}^{P}|h_{p,u}|^2\right]}{\widetilde{N}_0+\sum_{u'=u+1}^{K_u}E_{u'}\mathbb{E}_{\mathbf{h},\mathbf{l},\mathbf{k}}\left[\sum_{p=1}^{P}|h_{p,u'}|^2\right]}\mathbf{I}_L\\
&\overset{(c)}{=}\frac{E_u\frac{P(e^{-l_{\min,u}}-e^{-l_{\max,u}})}{2k_{\max,u}(l_{\max,u}-l_{\min,u})}}{\widetilde{N}_0+\sum_{u'=u+1}^{K_u}E_{u'}\frac{P(e^{-l_{\min,u'}}-e^{-l_{\max,u'}})}{2k_{\max,u'}(l_{\max,u'}-l_{\min,u'})}}\mathbf{I}_L,
\end{aligned}
\tag{81}
$$

where (a) denotes the application of first-order Taylor approximation, (b) is due to $\mathbf{G}_u^H\mathbf{G}_u=\mathbf{I}_L$ for arbitrary orthogonal multiple access scheme, and (c) denotes substituting the practical power-delay and power-Doppler profile in (12). After substituting (81) into the right hand side of (80) and according to the property of determinant operation, I_e is approximated as follows

$$
I_e\approx\sum_{u=1}^{K_u}\log_2
$$

$$\times \left(1+\frac{E_u \frac{P(e^{-l_{\min,u}}-e^{-l_{\max,u}})}{2k_{\max,u}(l_{\max,u}-l_{\min,u})}}{\widetilde{N}_0+\sum_{u'=u+1}^{K_u} E_{u'} \frac{P(e^{-l_{\min,u'}}-e^{-l_{\max,u'}})}{2k_{\max,u'}(l_{\max,u'}-l_{\min,u'})}}\right). \tag{82}$$

■

REFERENCES

[1] M. Nie, S. Li, and D. Mishra, "Improving channel estimation performance for uplink OTFS transmissions: Pilot design based on a posteriori Cramér–Rao bound," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2023, pp. 301–306.

[2] R. Hadani et al., "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[3] Z. Wei et al., "Orthogonal time-frequency space modulation: A promising next-generation waveform," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 136–144, Aug. 2021.

[4] Z. Wei et al., "Off-grid channel estimation with sparse Bayesian learning for OTFS systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7407–7426, Sep. 2022.

[5] Y. Shan, F. Wang, Y. Hao, J. Yuan, J. Hua, and Y. Xin, "Off-grid channel estimation using grid evolution for OTFS systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 9549–9565, Aug. 2024.

[6] S. Li, W. Yuan, Z. Wei, and J. Yuan, "Cross domain iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2227–2242, Apr. 2022.

[7] S. Li, "Hybrid MAP and PIC detection for OTFS modulation," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 7193–7198, Jul. 2021.

[8] P. Raviteja et al., "Embedded pilot-aided channel estimation for OTFS in delay-Doppler channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906–4917, May 2019.

[9] S. Li, W. Yuan, Z. Wei, J. Yuan, B. Bai, and G. Caire, "On the pulse shaping for delay-Doppler communications," in *Proc. IEEE Global Commun. Conf.*, Dec. 2023, pp. 4909–4914.

[10] W. Yuan, S. Li, Z. Wei, J. Yuan, and D. W. K. Ng, "Data-aided channel estimation for OTFS systems with a superimposed pilot and data transmission scheme," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1954–1958, Sep. 2021.

[11] L. Gaudio, G. Colavolpe, and G. Caire, "OTFS vs. OFDM in the presence of sparsity: A fair comparison," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4410–4423, Jun. 2022.

[12] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1822–1835, Mar. 2018.

[13] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, "On the fundamental tradeoff of integrated sensing and communications under Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5723–5751, Sep. 2023.

[14] Z. Ren et al., "Fundamental CRB-rate tradeoff in multi-antenna ISAC systems with information multicasting and multi-target sensing," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3870–3885, Apr. 2024.

[15] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "On the effectiveness of OTFS for joint radar parameter estimation and communication," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5951–5965, Sep. 2020.

[16] F. Liu, Z. Yuan, Q. Guo, Z. Wang, and P. Sun, "Message passing-based structured sparse signal recovery for estimation of OTFS channels with fractional Doppler shifts," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7773–7785, Dec. 2021.

[17] J. Pan, "Cramer-rao low bound of channel estimation for orthogonal time frequency space modulation system," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 9646–9658, Oct. 2021.

[18] 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, Jun. 2023.

[19] D. Mishra and E. G. Larsson, "Optimal channel estimation for reciprocity-based backscattering with a full-duplex MIMO reader," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1662–1677, Mar. 2019.

[20] A. Mehrotra, S. Srivastava, S. Asifa, A. K. Jagannatham, and L. Hanzo, "Online Bayesian learning-aided sparse CSI estimation in OTFS modulated MIMO systems for ultra-high-Doppler scenarios," *IEEE Trans. Commun.*, vol. 72, no. 4, pp. 2182–2200, Apr. 2024.

[21] B. C. Pandey, S. K. Mohammed, P. Raviteja, Y. Hong, and E. Viterbo, "Low complexity precoding and detection in multi-user massive MIMO OTFS downlink," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4389–4405, May 2021.

[22] J. Sun, Z. Wang, and Q. Huang, "A closed-form minimum BER precoder for orthogonal time frequency space systems," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1898–1902, Aug. 2022.

[23] B. Cao, Z. Xiang, and P. Ren, "Low complexity transmitter precoding for MU MIMO-OTFS," *Digit. Signal Process.*, vol. 115, Aug. 2021, Art. no. 103083.

[24] S. Li, J. Yuan, P. Fitzpatrick, T. Sakurai, and G. Caire, "Delay-Doppler domain tomlinson-harashima precoding for OTFS-based downlink MU-MIMO transmissions: Linear complexity implementation and scaling law analysis," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 2153–2169, Apr. 2023.

[25] C. Liu, S. Li, W. Yuan, X. Liu, and D. W. K. Ng, "Predictive precoder design for OTFS-enabled URLLC: A deep learning approach," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 7, pp. 2245–2260, Jul. 2023.

[26] C. Shen, J. Yuan, and H. Lin, "Error performance of rectangular pulse-shaped OTFS with practical receivers," *IEEE Wireless Commun. Lett.*, vol. 11, no. 12, pp. 2690–2694, Dec. 2022.

[27] R. Chong, S. Li, J. Yuan, and D. W. K. Ng, "Achievable rate upper-bounds of uplink multiuser OTFS transmissions," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 791–795, Apr. 2022.

[28] 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.

[29] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 957–961, Jan. 2018.

[30] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1994.

[31] F. Hlawatsch and G. Matz, *Wireless Communications Over Rapidly Time-Varying Channels*. New York, NY, USA: Academic, 2011.

[32] S. Li, J. Yuan, W. Yuan, Z. Wei, B. Bai, and D. W. K. Ng, "Performance analysis of coded OTFS systems over high-mobility channels," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6033–6048, Sep. 2021.

[33] Y. Ma et al., "Characteristics of channel spreading function and performance of OTFS in high-speed railway," *IEEE Trans. Wireless Commun.*, vol. 22, no. 10, pp. 7038–7054, Oct. 2023.

[34] Y. Ma, G. Ma, B. Ai, N. Wang, and Z. Zhong, "Impacts of time-varying channel spreading function on the OTFS modulation system," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 10744–10749, Jul. 2024.

[35] A. Lozano and N. Jindal, "Are yesterday-s information-theoretic fading models and performance metrics adequate for the analysis of today's wireless systems?" *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 210–217, Nov. 2012.

[36] A. F. Molisch, *Wireless Communications*, vol. 34. Hoboken, NJ, USA: Wiley, 2012.

[37] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and Statistics for Engineers and Scientists*, vol. 5. New York, NY, USA: MacMillan, 1993.

[38] S. K. Mohammed, "Derivation of OTFS modulation from first principles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 7619–7636, Aug. 2021.

[39] S. Li et al., "Fundamentals of delay-Doppler communications: Practical implementation and extensions to OTFS," 2024, *arXiv:2403.14192*.

[40] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[41] J. Dauwels, "Computing Bayesian Cramér–Rao bounds," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Sep. 2005, pp. 425–429.

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

[43] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[44] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1999.

[45] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[46] H. Seltman. (2012). *Approximations for Mean and Variance of a Ratio*. [Online]. Available: http://www.stat.cmu.edu/∼hseltman/files/ratio.pdf

[47] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.
[48] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink MISO OFDMA-URLLC systems," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7184–7200, Nov. 2020.
[49] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.
[50] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[51] T. Eltoft, T. Kim, and T.-W. Lee, "On the multivariate Laplace distribution," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 300–303, May 2006.
[52] Z. Wei, J. Guo, D. W. K. Ng, and J. Yuan, "Fairness comparison of uplink NOMA and OMA," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–6.
[53] Z. Wei, D. W. K. Ng, and J. Yuan, "Joint pilot and payload power control for uplink MIMO-NOMA with MRC-SIC receivers," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 692–695, Apr. 2018.

Mingcheng Nie (Graduate Student Member, IEEE) received the bachelor's and the Master of Philosophy degrees in electrical engineering from the University of New South Wales (UNSW), Australia, in 2021 and 2024, respectively. His research interests include delay-Doppler domain signal processing, resource allocation, integrated sensing and communication (ISAC), and physical layer security.

Shuangyang Li (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Xidian University, China, in 2013, 2016, and 2021, respectively, and the Ph.D. degree from the University of New South Wales (UNSW), Australia, in 2022. He is currently a Research Assistant at the Technical University of Berlin (TU-Berlin). Prior to that, he was a Research Associate at the University of Western Australia (UWA). His research interests include signal processing, channel coding, applied information theory, and their applications to communication systems, with a specific focus on waveform designs. He was a recipient of the Marie Skłodowska-Curie Actions (MSCA) Fellowship 2022. He received the Best Paper Award from IEEE ICC 2023 and the Best Workshop Paper Award from IEEE WCNC 2023. He was listed in the World's Top 2% Scientists by Stanford University for citation impact 2024 and was a recipient of the Best Young Researcher Award 2024 from the IEEE ComSoc EMEA region. He frequently serves as the Organizer/Chair for workshops and tutorials on related topics of orthogonal time frequency space (OTFS) in IEEE flagship conferences and is a Founding Member and currently the Co-Chair of the special interest group (SIG) on OTFS. He is currently an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS.

Deepak Mishra (Senior Member, IEEE) received the B.Tech. degree in electronics and communication engineering from GGSIPU, New Delhi, India, in 2012, and the Ph.D. degree in electrical engineering from Indian Institutes of Technology (IIT) at Delhi, India, in 2017. He is a Senior Lecturer at the School of Electrical Engineering and Telecommunications, UNSW Sydney, where he joined in 2019. Previously, he was a Post-Doctoral Researcher at Linköping University, Sweden. He has been a Visiting Researcher at Northeastern University, USA; the University of Rochester, USA; Huawei Technologies France, Southwest Jiaotong University, China; and the Queen's University of Belfast, U.K. His research interests include AI-enabled wireless sensing, energy harvesting and MIMO backscattering, underwater communications, physical layer security, signal processing, and green cooperative networks. He has been a recipient of the IBM Ph.D. Fellowship, the Raman Charpak Fellowship, the Endeavour Research Fellowship, and Australian Discovery Early Career Researcher Award. He received five Exemplary Reviewer awards for IEEE journals. He is an Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON INTELLIGENT VEHICLES, IEEE WIRELESS COMMUNICATIONS LETTERS, and IEEE ACCESS. He has been a Guest Editor for multiple special issues and a Topical Advisory Panel Member of the MDPI *Energies* journal. He was named a Best Editor of IEEE WIRELESS COMMUNICATIONS LETTERS in 2023.

Jinhong Yuan (Fellow, IEEE) received the B.E. and Ph.D. degrees in electronics engineering in 1991 and 1997, respectively. From 1997 to 1999, he was a Research Fellow with the School of Electrical Engineering, University of Sydney, Sydney, Australia. In 2000, he joined the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, where he is currently the Head of the School. He has published two books, five book chapters, more than 300 papers in telecommunications journals and conference proceedings, and 50 industrial reports. His current research interests include error control coding and information theory, communication theory, wireless communications, and delay-Doppler domain signal processing and communications. He is a co-inventor of one patent on MIMO systems and four patents on low-density-parity-check codes. He has co-authored four Best Paper Awards and one Best Poster Award, including the Best Paper Award from IEEE International Conference on Communications, Kansas City, USA, in 2018; the Best Paper Award from IEEE Wireless Communications and Networking Conference, Cancun, Mexico, in 2011; and the Best Paper Award from the IEEE International Symposium on Wireless Communications Systems, Trondheim, Norway, in 2007. He served as the IEEE NSW Chapter Chair for Joint Communications/Signal Processions/Ocean Engineering Chapter from 2011 to 2014. He served as an Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS from 2012 to 2017 and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2019 to 2024. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS.

Derrick Wing Kwan Ng (Fellow, IEEE) received the bachelor's (Hons.) and the Master of Philosophy degrees in electronic engineering from The Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2006 and 2008, respectively, and the Ph.D. degree from The University of British Columbia, Vancouver, BC, Canada, in November 2012. Following his Ph.D., he was a Senior Post-Doctoral Fellow at the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany. He is currently a Scientia Associate Professor with the University of New South Wales, Sydney, NSW, Australia. His research interests include global optimization, integrated sensing and communication (ISAC), physical layer security, IRS-assisted communication, UAV-assisted communication, wireless information and power transfer, and green (energy-efficient) wireless communications.

Dr. Ng has been recognized as a Highly Cited Researcher by Clarivate Analytics (Web of Science) since 2018. He was a recipient of Australian Research Council (ARC) Discovery Early Career Researcher Award 2017; the IEEE Communications Society Leonard G. Abraham Prize 2023; the IEEE Communications Society Stephen O. Rice Prize 2022; the Best Paper Awards at the WCSP 2020 and 2021; the IEEE TCGCC Best Journal Paper Award 2018; INISCOM 2018; IEEE International Conference on Communications (ICC) 2018, 2021, 2023, and 2024; IEEE International Conference on Computing, Networking and Communications (ICNC) 2016; IEEE Wireless Communications and Networking Conference (WCNC) 2012; IEEE Global Telecommunication Conference (Globecom) 2011, 2021, and 2023; and IEEE Third International Conference on Communications and Networking in China 2008. From January 2012 to December 2019, he served as an Editorial Assistant to the Editor-in-Chief for IEEE TRANSACTIONS ON COMMUNICATIONS. He is an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and an Associate Editor-in-Chief of IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.