Tensor-Based Low-Complexity Channel Estimation for mmWave Massive MIMO-OTFS Systems

Xianda Wu, Shaodan Ma, Xi Yang

Abstract—Orthogonal time frequency space (OTFS) modulation, collaborated with millimeter-wave (mmWave) massive multiple-input-multiple-output (MIMO), is a promising technology for next generation wireless communications in high mobility scenarios. However, one of the main challenges for mmWave massive MIMO-OTFS systems is the enormous computational complexity of channel estimation incurred by the huge OTFS symbol size and the large number of antennas. To address this issue, in this paper, a tensor-based orthogonal matching pursuit (OMP) channel estimation algorithm is proposed by exploiting the channel sparsity in the delay-Doppler-angle domain. In particular, we firstly propose a novel pilot design for the OTFS symbol structure in the frequency-time domain. Then, based on the proposed pilot structure, we formulate the channel estimation as a sparse signal recovery problem, and the tensor decomposition and parallel support detection are introduced into the tensor-based OMP algorithm to reduce the signal processing dimension significantly. Numerical simulations are performed to verify the superiority and the robustness of the proposed tensor-based OMP algorithm.

Keywords—OTFS, millimeter-wave, massive MIMO, channel estimation, compressed sensing

I. INTRODUCTION

Recently, orthogonal frequency division multiplexing (OFDM) modulation has been widely used in various wireless communication systems including the fifth generation cellular systems (5G), IEEE 802.11n, etc. However, the inter-carrier interference (ICI), which is caused by the time-varying characteristics of the wireless channels and consequently becomes one of the main challenges in OFDM systems, still remains unsolved. The ICI not only degrades the system performance, but also hampers the channel estimation and the signal detection[1]. The classical solution to overcome this nuisance from the fast time-varying channel characteristic is by designing a shorter OFDM symbol in the time domain. As such, the wireless channels can be regarded as time-invariant within several consecutive OFDM symbols. However, such classical method degrades the transmission efficiency to the coming beyond 5G (B5G) or the sixth generation (6G) cellular systems, since each symbol is added with cyclic prefix (CP) or guard interval (GI) in the transmission, where the size of CP or GI should be larger than the maximum channel length to avoid inter-symbol interference. In the high-speed scenarios of B5G and 6G, e.g., the high-speed train (HST) with the top speed up to 500 km/h[2,3], and the vehicular with top speed up to 300 km/h[4], the coherence time is strictly limited. Therefore, the symbol size in high-speed scenarios is generally much shorter than that of low-speed scenarios, while the length of CP or GI still remains the same. Hence, applying a short symbol for OFDM in high-speed scenarios to adapt to the high moving speed will also lead to a low system efficiency.

In order to meet the requirements of these high-mobility cases in future wireless communication systems, orthogonal time frequency spaces (OTFS) modulation has been proposed[5,6]. Unlike the OFDM technique, which only modulates the frequency domain signals into time domain by an inverse discrete Fourier transform (IDFT)[7], OTFS firstly converts the transmitted signals from the delay-Doppler domain to the frequency-time domain by an inverse symplectic finite Fourier transform (ISFFT), then conducts the multi-carrier transmission by utilizing the OFDM modulator. With the joint frequency-time-Doppler domain signal processing, OTFS can convert the time varying channels into invariant channels in the delay-Doppler domain and avoid the ICI. Therefore, OTFS demonstrates superior performance than
OFDM in high-mobility scenarios, and attracts tremendous attention in various research areas, including multiple-access investigation\cite{11}, modem structure design\cite{8}, low-complexity receiver design\cite{9,10}, iterative detection method\cite{11}, variational Bayes detector\cite{12}, etc.

Similar to other communication systems, channel estimation also plays an important role in OTFS communication systems. Most of existing studies for OTFS systems considered single-input-single-output (SISO) cases and proposed different channel estimation algorithms based on the threshold-based estimator\cite{13}, the pseudo-random noise (PN) sequences\cite{14}, the turbo compressed sensing (turbo-CS)\cite{15}, and so on.

Specially, the authors in Ref. [13] presented a pilot design, where only one pilot signal was located in the center of the pilot domain and was surrounded by guard intervals in the delay-Doppler domain of the OTFS symbol. Based on such a special pilot structure, the received pilots in the delay-Doppler domain can be used to estimate the channel responses once their power is beyond the given threshold. This channel estimation method is simple and fast, but it is difficult to select a proper threshold in practical applications. Moreover, the pilot transmission overhead is proportional to the Doppler frequency shift, which means that it is hard to extend this method into the cases of extremely high-mobility scenarios. Ref. [14] adopted PN sequences as the pilots in the delay-Doppler domain, so that the auto-correlation function was used to determine the significant path taps in the delay-Doppler domain. However, this PN-based method requires long pilots, and thus severely reduces the system efficiency. A turbo-CS based channel estimation algorithm was presented in Ref. [15], where the priors of the channel responses in the delay-Doppler domain were modeled by Markov random field (MRF). On the basis of the MRF, the turbo-CS algorithm exhibits a better performance than the other conventional message passing algorithms, such as the approximated message passing (AMP) algorithm. Nevertheless, enormous computation resources are needed to implement this turbo-CS algorithm, especially in the cases with large OTFS symbol size, since both the posterior and the prior of each channel responses in the delay-Doppler domain need to be updated in each iteration.

Note that the above studies\cite{13-15} mainly focus on SISO-OTFS systems. However, considering the rich spectrum resource in the millimeter-wave (mmWave) band and the high spectrum and energy efficiencies of massive massive multiple-input-multiple-output (MIMO) technique\cite{16-19}, mmWave massive MIMO combined with OTFS is also an attractive technique to boost the spectrum efficiency and is thus worth to be investigated. The authors in Refs. [20] and [21] have extended the threshold-based method to massive MIMO-OTFS systems, with the pilot length being proportional to the number of antenna pairs at the receiver and the transmitter. Whereas, this method greatly limits the spectral efficiency as the number of antennas goes to large. In order to overcome this problem, Ref. [4] proposed to estimate the massive MIMO-OTFS channels in the angular domain, in which the number of the parameters to be estimated is independent of the number of antennas. Based on the channel sparse properties in the delay-Doppler-angle domain, Ref. [4] proposed a three-dimension structured orthogonal matching pursuit (3D-SOMP) channel estimation algorithm. Unfortunately, this 3D-SOMP algorithm requires the weight calculation of each candidate in the delay-Doppler-angle domain per iteration. This consequently results in severe computation complexity and hinders the practical application of the 3D-SOMP algorithm when the candidate number in the delay-Doppler-angle domain becomes large.

To solve this problem, in this paper we propose a tensor-based orthogonal matching pursuit (OMP) channel estimation algorithm for the uplink transmission of mmWave massive MIMO-OTFS systems. The specific contributions are summarized as follows.

- We propose a novel OTFS symbol structure for the pilot design in the frequency-time domain. Compared to the existing delay-Doppler domain pilots, the proposed frequency-time domain pilots can be easily allocated in the multi-user scenarios without affecting the data allocation.
- Based on the proposed OTFS symbol structure, we propose a new rank-one pilot matrix design and formulate the channel estimation problem as a sparse signal recovery in the delay-Doppler-angle domain. Unlike the traditional greedy solutions to the sparse signal recovery, such as the OMP method\cite{22} and simultaneous OMP (S-OMP) algorithm\cite{23}, which only investigate the sparse signal in the delay-Doppler domain, our proposed pilot matrix design and tensor-based OMP algorithm can also bring the benefits from the angular domain and thus show better performance in many scenarios, which includes low signal-to-noise ratio (SNR) and varying velocity, etc.
- Tensor decomposition and new parallel support selection are introduced and proposed into the low complexity channel estimation algorithm, which leads to significant complexity reduction for mmWave massive MIMO-OTFS systems. The analysis shows that the computational complexity is reduced from the product of the numbers of sub-carriers and the OTFS symbols to their summation by the proposed tensor-based OMP algorithm.

The rest of the paper is organized as follows. In section II, we firstly introduce the OTFS modulation and then extend it to the case of mmWave massive MIMO systems. Section III presents the proposed tensor-based OMP channel estimation algorithm, which includes the pilot design, the sparse tensor reduction, and the enhanced OMP channel estimation. Nu-
merical results are provided in section IV, and section V concludes our work.

Notation: Underlined bold italic letter $\mathbf{X}$ denotes a tensor, bold italic letter $\mathbf{X}$ represents a matrix, and bold lowercase italic letter $x$ is a column vector. $[\mathbf{X}]_m$ and $[\mathbf{X}]_n$ indicate the $m$th row vector and the $n$th column vector of the matrix $\mathbf{X}$, respectively. The transpose, conjugate, conjugate transpose, and pseudo-inverse of a matrix are denoted by $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ and $(\cdot)^+$, respectively. The operator $\otimes$ is the Hadamard product, and $\otimes$ represents the Kronecker product. The operation $\text{vec}\{\cdot\}$ is a vectorization to a matrix, and $\text{diag}\{\cdot\}$ means a diagonal matrix with the input vector as the main diagonal. The notation $\times_n$ represents the mode-$n$ tensor product, and $\mathbf{I}_M$ is an $M \times M$ identity matrix.

II. SYSTEM MODEL

In this section, we first describe the detailed procedure of the OTFS modulation in SISO systems, and then extend it to mmWave massive MIMO systems.

A. OTFS Modulation

The OTFS modulation in SISO systems was firstly introduced in Ref. [5] and combined with mmWave communications in Ref. [6]. A generic description of frequency-time modulation is as follows.

- A lattice $\Lambda$ in the frequency-time domain is a sampling of the frequency and time axes at intervals $\Delta f$ and $T$ as
  \[
  \Lambda = \{(m\Delta f, nT), m = -\frac{M}{2}, \cdots, \frac{M}{2} - 1, n = 0, \cdots, N - 1\}. 
  \]  
  (1)

- An OTFS symbol has a total bandwidth of $M\Delta f$ Hz and a total duration of $NT$ seconds.

- A set of modulated symbols $\mathbf{X} \in \mathbb{C}^{M \times N}$ on the lattice $\Lambda$, is transmitted by $s(t)$ on the pulse waveform $g_{tx}(t)$ as
  \[
  s(t) = \sum_{m=-M/2}^{M/2-1} \sum_{n=0}^{N-1} \mathbf{X}[m + M/2, n]g_{tx}(t - nT)e^{j2\pi m\Delta f(t - nT)}, \quad (2)
  \]
  where $\mathbf{X}[k, l]$ represents the element in the $(k + 1)$th row and the $(l + 1)$th column of matrix $\mathbf{X}$.

- The inner product of the transmit pulse $g_{tx}(t)$ and its associated receive pulse $g_{rx}(t)$ is bi-orthogonal with respect to translation by frequency $\Delta f$ and time $T$ as
  \[
  A_{g_{rx},g_{tx}}(m,n) = \int g_{rx}^*(t)g_{tx}(t - nT)e^{j2\pi m\Delta f(t - nT)}dt = \delta(m)\delta(n), \quad (3)
  \]
  where $A_{g_{rx},g_{tx}}$ indicates the cross-ambiguity function, and $\delta(\cdot)$ is the Dirac delta function.

As Fig. 1 shows, the modulation in (2) is called the Heisenberg transform of frequency-time domain signals $\mathbf{X}$. The frequency-time domain signals $\mathbf{X}$ is modulated from the data sequence $\mathbf{X} \in \mathbb{C}^{M \times N}$ of the delay-Doppler domain by using an ISFFT and a transmit windowing function[5] as
\[
\mathbf{X} = \mathbf{W}_{tx} \otimes (\mathbf{F}_M \mathbf{X} \mathbf{F}_N^H), \quad (4)
\]
where $\mathbf{W}_{tx} \in \mathbb{C}^{M \times N}$ is a transmit window function. To simplify the analysis and mainly shed light on the channel estimation of OTFS systems, we assume $\mathbf{W}_{tx}$ is a matrix of all ones, and $\mathbf{F}_M$ and $\mathbf{F}_N$ are defined as
\[
\mathbf{F}_M = \begin{bmatrix} [\mathbf{F}_M]^T \frac{M}{2} + 1; \cdots; [\mathbf{F}_M]^T \frac{M}{2} - 1 \end{bmatrix}^T, \quad (5)
\]
\[
\mathbf{F}_N = \begin{bmatrix} [\mathbf{F}_N]^T \frac{N}{2} + 1; \cdots; [\mathbf{F}_N]^T \frac{N}{2} - 1 \end{bmatrix}^T, \quad (6)
\]
where $\mathbf{F}_M$ and $\mathbf{F}_N$ are $M$- and $N$-dimension discrete Fourier transform (DFT) matrices, respectively.
At the receiver, the received signal over the time-varying channel \( h(\tau, \nu) \) after removing the CP for each OTFS symbol can be given by

\[
    r(t) = \int_\tau \int_\nu h(\tau, \nu) s(t-\tau) e^{j 2 \pi v(t-\tau)} d\nu d\tau + v(t),
\]

where \( \tau \) indicates the time delay, \( \nu \) represents the Doppler frequency offset, and \( v(t) \) is the Gaussian noise with zero mean and variance \( \sigma^2 \). Then, the Wigner transform, which inverts the Heisenberg transform of (2) and is operated by the DFT matrix of the OFDM demodulator in Fig. 1, is applied to recover the transmitted signals\(^5\)

\[
    Y[m + M/2, n] = \int_s \int_\tau g_{rx}(t-\tau) r(t) e^{-j 2 \pi v(t-\tau)} dt|_{\tau=nT, \nu=\nu_{\max}} = 0,
\]

where \( m = -\frac{M}{2}, \cdots, \frac{M}{2} - 1 \), and \( n = 0, \cdots, N - 1 \). If \( h(\tau, \nu) \) has finite support bounded by \( (\tau_{\max}, \nu_{\max}) \), and \( A_{\nu_{\max}}(\tau, \nu) \) of (3) is equal to zero for \( \tau \in (nT - \tau_{\max}, nT + \tau_{\max}) \), \( \nu \in (m\Delta f - \nu_{\max}, m\Delta f + \nu_{\max}) \), then the received signals of (8) can be expressed as\(^5\)

\[
    Y = H \otimes X + V,
\]

where \( V \) is the addictive white Gaussian noise (AWGN), and \( H \) is the frequency-time domain channel and can be expressed by

\[
    H[m + M/2, n] = \int_\tau \int_\nu h(\tau, \nu) e^{j 2 \pi \nu nT} e^{-j 2 \pi (\nu + \nu_{\max})} \tau d\nu d\tau,
\]

where \( m = -\frac{M}{2}, \cdots, \frac{M}{2} - 1 \) and \( n = 0, \cdots, N - 1 \).

By considering the sparse feature of \( h(\tau, \nu) \) from the perspective of the time and Doppler domains, which is caused by the limited number of scatterers in the propagation environment, \( h(\tau, \nu) \) can be modeled in a sparse representation in terms of the time delay and the Doppler frequency shift\(^{20}\), i.e.,

\[
    h(\tau, \nu) = \sum_{i=1}^P h_i \delta(\tau - \tau_i) \delta(\nu - \nu_i),
\]

where \( P \) indicates the number of propagation paths, \( h_i \), then \( \tau_i \) and \( \nu_i \) represent the complex gain, time delays, and Doppler frequency shifts of the \( i \)-th path, respectively.

Let \( \kappa_i \) and \( \iota_i \) denote the delay and Doppler taps for the \( i \)-th path, respectively, \( \tau_i \) and \( \nu_i \) can be defined as

\[
    \tau_i = \kappa_i \frac{M\Delta f}{\nu_i}, \quad \nu_i = \frac{i\nu_i}{NT},
\]

where both \( \kappa_i \in [0, M - 1] \) and \( i\nu_i \in [-N/2, N/2 - 1] \) are integers. In this paper, we assume the resolutions \( 1/M\Delta f \) and \( 1/NT \) are sufficient to approximate the practical delays and Doppler frequency shifts.

**B. mmWave Massive MIMO-OTFS**

Previously, we present the detailed procedures of the OTFS modulation/demodulation for SISO communication systems.
In this subsection, we extend the OTFS to the MIMO scenarios and consider the uplink transmission of OTFS-based mmWave massive MIMO systems. Let $B$ represent the number of antenna elements of the uniform linear array (ULA) equipped at the base station (BS), and $U$ denote the number of single-antenna user terminals (UTs). The received signals after removing CP in the frequency-time domain of the $b$th antenna can be expressed as

$$Y_b = \sum_{u=1}^{U} H^{(u)}_b \odot X^{(u)} + V_b, \quad b = 1, \ldots, B,$$

where $H^{(u)}_b$ and $X^{(u)}$ represent the uplink frequency-time domain channel matrix and the uplink transmitted signals from the $u$th UT to the $b$th antenna of BS, respectively. $V_b$ indicates the AWGN matrix.

As illustrated in Fig. 2, the estimated $\hat{H}^{(u)}_b$ is essential to mmWave massive MIMO-OTFS systems. According to the property of the mmWave channels in the ULA, the channel responses in the same propagation path between two adjacent antennas have a phase shift $e^{-j2\pi \frac{d}{\lambda} \sin(\theta)}$, where $d$ is the spacing between two adjacent ULA elements, $\lambda$ represents the transmission carrier wavelength, and $\theta$ indicates the corresponding angle-of-arrival (AoA). Considering the sparse feature of OTFS channels (11), the channel responses in the same delay tap can be regarded within the same propagation path. Hence, by defining $\theta^m_u$ as the AoA of the $m$th delay tap between the $u$th UT and the BS, and considering the general OTFS channel model provided in (14), the OTFS channel matrix $H^{(u)}_b$ in (15) can be written as

$$H^{(u)}_b = \mathbf{F}_M D^{(u)}_b \mathbf{F}^H_N,$$

where $D^{(u)}_b \triangleq c^{(u)}_b D^{(u)}$ and

$$c^{(u)}_b = \text{diag}\{e^{-j2\pi \frac{d}{\lambda} \sin(\theta^1_u)}, \ldots, e^{-j2\pi \frac{d}{\lambda} \sin(\theta^L_u)}\},$$

where $c^{(u)}_b$ represents the effect caused by phase shifting in each delay tap. Note that $D^{(u)}_b$ has the same sparse property as $D^{(u)}$ since $c^{(u)}_b$ is a diagonal matrix with unitary exponential elements.

**III. Tensor-Based OMP Channel Estimation**

In this section, a novel frequency-time domain pilot design for mmWave massive MIMO-OTFS systems is firstly introduced. Then we propose a tensor-based low-complexity channel estimation algorithm on the basis of the OMP method.
link channel estimation, the pilots’ positions related different UTs in the lattice have no overlapping area.

Based on the proposed OTFS symbol’s structure, the received pilots related to the $i$th UT can be given by

$$\hat{Y}^{(u)} = [\hat{Y}_{1}^{(u)}, \cdots, \hat{Y}_{B}^{(u)}] = [\hat{H}_{1}^{(u)} \otimes X_{p}, \cdots, \hat{H}_{B}^{(u)} \otimes X_{p}] + \hat{V}^{(u)},$$  

where $\hat{Y}^{(u)} \in \mathbb{C}^{M_{p} \times B N_{p}}$, and $X_{p} \in \mathbb{C}^{M_{p} \times N_{p}}$ is the transmitted pilot matrix. Note that since the pilot contamination does not exist in the proposed OTFS symbol structures of Fig. 3, allocating the same pilots $X_{p}$ to each UT will not degrade the performance of channel estimation. Our proposed method is also applicable to the case in which each user is assigned with a distinct pilot matrix. The frequency-time domain channel matrix $\hat{H}_{b}^{(u)}$, $b = 1, \cdots, B$, can be expressed as

$$\hat{H}_{b}^{(u)} = \hat{F}_{M}^{(u)} D_{b}^{(u)} (\hat{F}_{N}^{(u)})^{H},$$  

where $\hat{F}_{M}^{(u)} = [\hat{F}_{M}]_{\Omega_{f,u}}$ and $\hat{F}_{N}^{(u)} = [\hat{F}_{N}]_{\Omega_{a,u}}$. The sets $\Omega_{f,u}$ and $\Omega_{a,u}$ represent the indices of pilots’ positions in the frequency and the time axes, respectively. For example, in the type-1 design shown in Fig. 3, the first UT is allocated by the following pilots’ position as $\Omega_{f,1} = [1, \cdots, 1 + (M_{p} - 1) | M / M_{p} |]$, and $\Omega_{a,1} = [1, 3, 5, \cdots, N - 1]$.

The uplink pilot matrix is designed as a rank-one matrix $X_{p} = x_{p,1} x_{p,2}^{T}$, where $x_{p,1} \in \mathbb{C}^{M_{p} \times 1}$ and $x_{p,2} \in \mathbb{C}^{N_{p} \times 1}$. Thus the received pilots from the $i$th UT to the $b$th antenna can be formulated as

$$\hat{Y}_{b}^{(u)} = \text{diag} \{ x_{p,1} \} \hat{F}_{M}^{(u)} D_{b}^{(u)} (\hat{F}_{N}^{(u)})^{H} \text{diag} \{ x_{p,2} \} + \hat{V}_{b}^{(u)},$$  

(20)

where both $\hat{X}_{p,1}$ and $\hat{X}_{p,2}$ are symmetric.

Based on (19) and (20), it is evident that the frequency-domain channel matrix $\hat{H}_{b}^{(u)}$, $b = 1, \cdots, B$, can be estimated individually in each BS’s antenna. However, such individual channel estimation at each antenna does not exploit the correlations in the angular domain among BS’s antennas. In the next subsection, we will derive the details of the joint channel estimation by making use of the received pilots of all BS’s antennas.

B. Low-Complexity Channel Estimation

In the proposed low-complexity channel estimation algorithm (namely the tensor-based OMP algorithm), three steps are involved, including the tensor decomposition, the parallel support detection, and the modified OMP channel estimation. These three steps are deliberately illustrated in sequence in the following.

From (19) and (20), for the purpose of well estimating the frequency-time domain channel matrices $\hat{H}_{b}^{(u)}$, $b = 1, \cdots, B$ for mmWave massive MIMO-OTFS systems with the aid of limited pilots, the sparsity of the delay-Doppler domain matrices $D_{b}^{(u)}$, $b = 1, \cdots, B$ is essential and exploited. These sparse matrices $D_{b}^{(u)}$ can be parameterized as variables of three dimensions, including time delay, Doppler frequency shift, and angles. Generally, these three dimensions are independent to each other. Thus, the tensor method, which is widely used to factorize the signals into more than two dimensions, can be utilized into channel estimation in this paper.

1) Tensor Decomposition: Let $\tilde{y}_{b}^{(u)} = \text{vec} \{ \hat{Y}_{b}^{(u)T} \}$, and $\tilde{Y}^{(u)} = [\tilde{y}_{1}^{(u)}, \cdots, \tilde{y}_{B}^{(u)}] \in \mathbb{C}^{M_{p} \times B N_{p}}$, then $\tilde{Y}^{(u)}$ can be given as

$$\tilde{Y}^{(u)} = \{ (\hat{X}_{p,1} \hat{F}_{M}^{(u)}) \otimes (\hat{X}_{p,2} (\hat{F}_{N}^{(u)})^{*}) \} \times$$

$$\{ \text{vec} \{ D_{1}^{(u)T} \}, \cdots, \text{vec} \{ D_{B}^{(u)T} \} \} + \hat{V}^{(u)},$$  

(21)

where

$$\{ \text{vec} \{ D_{1}^{(u)T} \}, \cdots, \text{vec} \{ D_{B}^{(u)T} \} \} |_{\Omega_{m}} = \{ D_{1}^{(u)T} |_{\Omega_{m}} a_{T}^{T} (\theta_{m}^{*}) \},$$  

(22)

where $\Omega_{m}$ represents the index set related to the $m$th delay and is defined as $\Omega_{m} = [(m - 1) N_{p} + 1 : m N_{p}]$, and $a(\theta) \in \mathbb{C}^{B \times 1}$ is the steering vector of ULA related to the angle $\theta$. If $G$ is sufficiently large, the off-grid deviation in angular domain can be ignored. Hence, the steering vector related to any $\theta_{m}$ can be represented by the linear combination of angle candidate matrix $A \in \mathbb{C}^{B \times G}$ and binary selection vector $q$, i.e., $a(\theta_{m}) = Aq$ where the $g$th element of $q$ is non-zero whereas others are zero. Note that the matrix $A$ can also be chosen as DFT matrices$^{[24]}$.

Therefore, the measurements $\tilde{Y}^{(u)}$ can be re-expressed as

$$\tilde{Y}^{(u)} = (\hat{X}_{p,1} \hat{F}_{M}^{(u)}) \otimes (\hat{X}_{p,2} (\hat{F}_{N}^{(u)})^{*}) \times$$

$$[(D^{(u)T})_{1} q_{1}^{T}, \cdots, (D^{(u)T})_{M} q_{M}^{T}] A^{T} + \hat{V}^{(u)},$$  

(23)

where $q_{m}$ is the binary selection vector of the $m$th delay tap for choosing the corresponding AoA $\theta_{m}$. The results after taking vectorization to the transpose of $\tilde{Y}^{(u)}$ can be given as

$$\tilde{y}^{(u)} = \text{vec} \{ (\tilde{Y}^{(u)})^{T} \} =$$

$$((\hat{X}_{p,1} \hat{F}_{M}^{(u)}) \otimes (\hat{X}_{p,2} (\hat{F}_{N}^{(u)})^{*}) \otimes A) d^{(u)} + \tilde{v}^{(u)},$$  

(24)

where $\tilde{v}^{(u)} \in \mathbb{C}^{M_{N} G}$ is a structured sparse vector as

$$\tilde{d}^{(u)} = \text{vec} \{ [q_{1} | D_{1}^{(u)T}]_{1}, \cdots, [q_{M} | D_{M}^{(u)T}]_{M} \}.$$  

(25)

Let $\tilde{q}^{(u)} \triangleq (\hat{X}_{p,1} \hat{F}_{M}^{(u)}) \otimes (\hat{X}_{p,2} (\hat{F}_{N}^{(u)})^{*}) \otimes A$, we can find that $\tilde{y}^{(u)}$ is the measurements and the computation of $\tilde{d}^{(u)}$ can
be regarded as a sparse signal recovery problem with dictionary matrix $\Phi^{(a)}$. Such problem can be solved by many greedy algorithms in the framework of compressed sensing, which includes OMP method\textsuperscript{[22]}. However, the computational complexity of the traditional OMP method is $\mathcal{O}(PM_pN_pBMN)$. As the size of the OTFS symbol and angular resolutions increase, the corresponding complexity of traditional OMP algorithm becomes unsustainable.

In order to reduce the computational complexity, one low-complexity solution is to focus on the channel estimation at each antenna based on $\tilde{y}_b^{(a)} = \text{vec}(\{\tilde{y}_b^{(a)}\})^T$, in which the OMP\textsuperscript{[22]} channel estimation can be performed by individual antenna with the computational complexity of $\mathcal{O}(PM_pN_pBMN)$. However, such traditional OMP method ignores the structural sparsity of the measurements received from different antennas. Another solution is to use S-OMP\textsuperscript{[23]} based on $\tilde{Y}^{(a)}$. Nevertheless, the computational complexities of these two methods are similar, except that the S-OMP exploits the common sparsity of different antennas in the delay-Doppler domain. As a consequence, both traditional OMP and S-OMP still have significant computational complexity and the structural sparsity is not well exploited in the angular domain. In order to overcome the challenge of the high computational complexity, we propose a tensor-based OMP algorithm by introducing the tensor decomposition to reduce the size of the sparse vector $\tilde{d}^{(u)}$.

Given the expression of $\tilde{y}^{(a)}$, the pilot measurements can be regarded as a tensor of three modes in Fig. 4 including delay, Doppler and angular domains, i.e.,

$$\tilde{Y}^{(a)} = D^{(a)} \times_1 A_1 \times_2 A_2 \times_3 A_3 + \tilde{V}^{(a)}, \quad (26)$$

where $A_n$ represents the matrix for tensor product $\times_n$ in mode-$n$, and can be expressed as

$$A_1 = \tilde{X}_{p,1} F^{(u)}_1, \quad (27)$$

$$A_2 = \tilde{X}_{p,2} F^{(u)}_2, \quad (28)$$

$$A_3 = A. \quad (29)$$

Furthermore, according to the unfolding properties of the tensor theory, we can respectively stack the received pilots from the $u$th UT in three modes as

$$\tilde{\bar{Y}}^{(a),1} = \tilde{Y}^{(a)} = A_1 \tilde{D}^{(a),1} + \tilde{V}^{(a),1}, \quad (30)$$

$$\tilde{\bar{Y}}^{(a),2} = [(\tilde{Y}_1^{(a)})^T, \cdots, (\tilde{Y}_B^{(a)})^T] = A_2 \tilde{D}^{(a),2} + \tilde{V}^{(a),2}, \quad (31)$$

$$\tilde{\bar{Y}}^{(a),3} = (\tilde{Y}^{(a)})^T = A_3 \tilde{D}^{(a),3} + \tilde{V}^{(a),3}, \quad (32)$$

where $\tilde{\bar{Y}}^{(a),n}$ is a mode-$n$ unfolding matrix of the tensor $\tilde{Y}^{(a)}$, and $\tilde{D}^{(a),n}$ can be given by

$$\tilde{D}^{(a),1} = [D_1^{(a)}, \cdots, D_B^{(a)}] (I_B \otimes (F_N^{(u)})^H \tilde{X}_{p,2}), \quad (33)$$

$$\tilde{D}^{(a),2} = [(D_1^{(a)})^T, \cdots, (D_B^{(a)})^T] (I_B \otimes (F_M^{(u)})^T \tilde{X}_{p,1}), \quad (34)$$

$$\tilde{D}^{(a),3} = [q_1(D_n^{(a)})_1, \cdots, q_M(D_n^{(a)})_M] \times (F_M^{(u)} \tilde{X}_{p,1} \otimes (F_N^{(u)})^H \tilde{X}_{p,2}). \quad (35)$$

2) Parallel Support Detection: It is evident that $D^{(a),n}$ is a row-sparse matrix and its sparsity is independent to the unfolding matrices in other modes. Hence, their supports can be parallely detected. In other words, since the accuracy of estimating supports of a mode does not affect that of the other modes, we can detect the supports of their row vector spaces simultaneously.

By considering the common row sparsity of each column vector in $D^{(a),n}$, a parallel support detection algorithm is proposed in Algorithm 1, where $\tilde{S}_n$ is the estimated support containing the sparse rows’ indices of $D^{(a),n}$, $n = 1, 2, 3$. The corresponding computational complexity of mode-1,2,3 are $\mathcal{O}(PM_pN_pBM)$, $\mathcal{O}(PM_pN_pBN)$, and $\mathcal{O}(PM_pN_BG)$, respectively.

The key idea of Algorithm 1 is borrowed from S-OMP\textsuperscript{[23]}. Since $D^k$ is sparse in row vector spaces, the first step of each iteration is to compute the contribution of each column vector of $A_n$ to the residual $R^k$ (from line 4 to 6). Once the support $g_{\max}$, which plays an important role in the current residual $R^k$, is determined, we combine it with the previous obtained supports and calculate their total contribution by orthogonal matching projection (from line 7 to 8). Finally, the residual is updated by removing the effects from the determined supports (line 9).

3) Modified OMP Channel Estimation: Based on the acquired supports $\tilde{S}_1$, $\tilde{S}_2$, and $\tilde{S}_3$, the sparse tensor $D^{(a)}$ with size $M \times N \times G$ can be greatly reduced as $\tilde{D}^{(a)} = [D^{(a)}]_{\tilde{S}_1,\tilde{S}_2,\tilde{S}_3}$, and the received measurements in (26) can be approximated by

$$\tilde{\bar{Y}}^{(a)} \approx \tilde{\bar{D}}^{(a)} \times_1 [A_1]_{\tilde{S}_1} \times_2 [A_2]_{\tilde{S}_2} \times_3 [A_3]_{\tilde{S}_3} + \tilde{V}^{(a)}, \quad (36)$$
and its vectored expression can be approximated by
\[ \hat{\mathbf{y}}^{(a)} \approx \left( \tilde{\mathbf{A}}_1 \otimes \tilde{\mathbf{A}}_2 \otimes \tilde{\mathbf{A}}_3 \right) \mathbf{d}^{(a)} + \hat{\mathbf{d}}^{(a)}, \]  
(37)
where \( \mathbf{d}^{(a)} \) is still sparse, and can be solved by the modified OMP method in Algorithm 2.

As shown in Algorithm 2, in each iteration, firstly, the support \( g_{\text{max}} \), whose corresponding vector provides the largest contribution to the residual \( r^k \), is determined by the matching operation (from line 4 to 10). Then, considering the vector spaces’ sparsity of \( \tilde{\mathbf{A}}_1 \otimes \tilde{\mathbf{A}}_2 \otimes \tilde{\mathbf{A}}_3 \), the pruning operation is performed (from line 11 to 13) to the vectors of the support \( g_{\text{max}} \). After that, the pursuit operation (from line 14 to 16) is applied by the orthogonal projection to the residual update.

Once the supports of the three domains, i.e., delay, Doppler and angle, are determined, the frequency-time domain channel matrix of each antenna can be calculated (from line 19 to 20). The computational complexity of the modified OMP channel estimation is only \( \mathcal{O}(P^4M_pN_pB) \).

C. Comparison and Discussion

The comparisons between the traditional OMP\(^{[22]}\), the traditional S-OMP\(^{[23]}\) and the proposed tensor-based OMP are listed in Tab. 1. As shown in the table, the computational complexity of the proposed tensor-based OMP algorithm is \( \mathcal{O}(PM_pN_pB(M + N + G)) + \mathcal{O}(P^4M_pN_pB) \), where the first term is related to the parallel support detection, and the second one corresponds to the modified OMP channel estimation. Since the path number \( P \) is much smaller than \( M, N \) and \( G \), the computational complexity of the proposed algorithms can be approximated as \( \mathcal{O}(PM_pN_pB(M + N + G)) \).

It is apparent that complexity of the tensor-based OMP method is much less than the traditional OMP and the S-OMP algorithms. Such benefit is brought by the parallel sparsity detection in each dimension. By exploiting the sparsity in multiple-dimensions via tensor decomposition, the size of candidates can be significantly reduced.

Hence, it is foreseeable that the tensor-based method can be also extended to the other classical greedy algorithms including generalized OMP\(^{[25]}\), subspace pursuit\(^{[26]}\), etc., for the purpose of improving the channel estimation performance, as the sparse signals can be also decomposed as a tensor with multiple independent dimensions.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Underlying principle</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional OMP</td>
<td>Exploit the sparsity in the delay-Doppler domain</td>
<td>( \mathcal{O}(PM_pN_pB) )</td>
</tr>
<tr>
<td>Traditional S-OMP</td>
<td>Exploit the sparsity in the delay-Doppler domain, and exploit the common sparsity in different antennas</td>
<td>( \mathcal{O}(PM_pN_pB) )</td>
</tr>
<tr>
<td>Proposed tensor-based OMP</td>
<td>Exploit the sparsity in the delay-Doppler-angle domain</td>
<td>( \mathcal{O}(PM_pN_pB(M + N + G)) )</td>
</tr>
</tbody>
</table>
**IV. SIMULATION RESULTS**

In this section, we demonstrate the performance of the proposed tensor-based OMP channel estimation algorithm from the perspective of the normalized mean square error (NMSE), i.e.,

\[ \text{NMSE} = \frac{1}{U} \sum_{u=1}^{U} \frac{\| \tilde{H}(u) - H(u) \|_2^2}{\| H(u) \|_2^2}. \]  

(38)

The adopted system parameters are as follows: The carrier frequency is 28 GHz and the sub-carrier spacing is 15 kHz. The number of sub-carriers is \( M = 512 \) with the interval number \( N = 128 \). The number of antennas of the ULA at BS is \( B = 64 \), and the number of path is \( P = 5 \) with maximum delay tap \( k_{\text{max}} = 16 \). As in Ref. [20], each delay tap has a single Doppler shift generated by the Jakes’ formula, i.e. \( v_p = v_{\text{max}} \cos(\phi_p) \), where \( v_{\text{max}} \) is the maximum Doppler shift determined by UT’s velocity, and \( \phi_p \) is uniformly distributed over \([-\pi, \pi]\). The number of angular grids we set is \( G = 64 \), and the AoAs are uniformly generated from \([-\pi, \pi]\). The pilot vectors, i.e. \( x_{p,1} \) and \( x_{p,2} \), are randomly generated with elements following independent normal distribution in the simulations.

Since the channel estimation can be regarded as a sparse signal recovery problem, the traditional greedy algorithms, e.g., OMP\(^{[22]}\) and S-OMP\(^{[23]}\), can be used to estimate the delay-Doppler domain channel matrix based on \( \tilde{y}_b^{(u)} = \text{vec}\{ (\tilde{Y}_b^{(u)})^T \} \) and \( \tilde{Y}^{(u)} = [\tilde{y}_1^{(u)}, \ldots, \tilde{y}_B^{(u)}] \), respectively. We assume that the sparsity, i.e., the path number, has been known for the traditional OMP, the traditional S-OMP, and our proposed tensor-based OMP.

In Fig. 5, the velocity of UT is set as 140 km/h, and the pilot design is type-1 presented in Fig. 3(a), where the numbers of pilots in the frequency and time axes are \( M_p = 20 \) and \( N_p = 64 \), respectively, such that the total pilot cost of one UT is around 2%. As shown in Fig. 5, the traditional S-OMP has better channel estimation accuracy than the traditional OMP in low SNR scenario since the traditional S-OMP exploits the structural sparsity of the delay-Doppler domain in each antenna. However, both of them do not consider the sparsity in the angular domain. Therefore, by full exploiting the angular sparsity, the proposed tensor-based OMP can significantly outperform the traditional OMP and the traditional S-OMP.

In Fig. 6 and Fig. 7, the type-1 pilot is adopted, and the velocity of UT and SNR are set as 140 km/h and 10 dB, respectively. In Fig. 7, we assume the maximal total pilot overhead for all UTs is 20%. As illustrated in Fig. 6 and Fig. 7, during the low pilot overhead region (i.e., pilot overhead < 1.5%), the traditional S-OMP performs better than the proposed tensor-based OMP. The reason is that the traditional S-OMP requires a lower pilot overhead and thus is more suitable for low pilot overhead scenarios. In the compressed theory, the pilot overhead requirement of S-OMP can be regarded as \( M_p N_p > O(\ln(MN/P)) \), while the proposed tensor-based OMP should satisfy both \( M_p > O(\ln(M/P)) \) and \( N_p > O(\ln(N/P)) \). However, Fig. 6 and Fig. 7 also demonstrate that, as the pilot overhead increases, the proposed...
The BER comparison against UT number in different methods (Pilot overhead is 4%, and the SNR is 10 dB.)

V. Conclusion

In this paper, we proposed a low complexity tensor-based OMP algorithm for the channel estimation of mmWave massive MIMO-OTFS systems. Specifically, we formulated the channel estimation as a sparse signal recovery problem, and provided a novel OTFS symbol structure for the pilots in the frequency-time domain. Based on these frequency-time domain pilots, the tensor-based support detection was proposed to parallelly reduce the dimensions of sparse channels in the delay, Doppler, and angular domains. As a consequence, the channels in mmWave massive MIMO-OTFS systems can be estimated in a low computational complexity. Numerical results verified the superior performance of the proposed tensor-based OMP channel estimation algorithm.

References


Xianda Wu received his B.S. and M.S. degrees from South China Normal University, Guangzhou, China, in 2013 and 2016, respectively. He is currently working towards his Ph.D. degree with University of Macau, Macao, China. From 2015 to 2016, he was with Southern University of Science and Technology, Shenzhen, China, as a Visiting Student. His current research interests include channel estimation, compressed sensing, statistical learning, and massive MIMO.

Shaodan Ma [corresponding author] received her double bachelor’s degree in science and economics, and her master’s degree in engineering, from Nankai University, Tianjin, China, in 1999 and 2002 respectively. She obtained her Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, China in 2006. From 2006 to 2011, she was a Postdoctoral Fellow in the University of Hong Kong. Since August 2011, she has been with University of Macau and is now an Associate Professor there. She was a Visiting Scholar in Princeton University in 2010. Her research interests are in the general areas of signal processing and communications, particularly, transceiver design, resource allocation and performance analysis.

Xi Yang received her B.S., M.S. and Ph.D. degrees from Southeast University, Nanjing, China, in 2013, 2016 and 2019, respectively. She is currently working as a Postdoctoral Fellow with the State Key Laboratory of Internet of Things for Smart City (University of Macau), Macao, China. Her current research interests include wireless communication system prototyping, massive MIMO, and millimeter wave communications.