Joint Sparsity Pattern Learning Based Channel Estimation for Massive MIMO-OTFS Systems

Kuo Meng, Shaoshi Yang, Xiao-Yang Wang, Yan Bu, Yurong Tang, Jianhua Zhang and Lajos Hanzo

Abstract—We propose a channel estimation scheme based on joint sparsity pattern learning (JSPL) for massive multi-input multi-output (MIMO) orthogonal time-frequency-space (OTFS) modulation aided systems. By exploiting the potential joint sparsity of the delay-Doppler-angle (DDA) domain channel, the channel estimation problem is transformed into a sparse recovery problem. To solve it, we first apply the *spike and slab* prior model to iteratively estimate the support set of the channel matrix, and a higher-accuracy parameter update rule relying on the identified support set is introduced into the iteration. Then the specific values of the channel elements corresponding to the support set are estimated by the orthogonal matching pursuit (OMP) method. Both our simulation results and analysis demonstrate that the proposed JSPL channel estimation scheme achieves an improved performance over the representative state-of-the-art baseline schemes, despite its reduced pilot overhead.

Index Terms—Bayesian learning, channel estimation, joint sparsity, massive MIMO, OTFS.

I. INTRODUCTION

Massive multi-input multi-output (MIMO) schemes are expected to find their way into next-generation mobile systems, due to their unrivalled advantages in spectral efficiency, energy efficiency and compact high-frequency form-factor [1]. Massive MIMO aided orthogonal frequency division multiplexing (OFDM) exhibits excellent robustness against the frequency-selective fading [2]. However, in high mobility scenarios (e.g., high speed railway, unmanned aerial vehicles, etc.), the Doppler spread may cause severe inter-carrier interference (ICI), hence degrading the performance [3].

To overcome this challenge, the orthogonal time-frequency-space (OTFS) modulation was developed in [4], which outperforms OFDM in doubly-selective channels. The OTFS modulation scheme maps its transmitted symbols to the delay-Doppler (DD) domain rather than to the time-frequency (TF)

K. Meng, S. Yang, and X.-Y. Wang are with the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: {meng_kuo, shaoshi.yang, wangxy_028}@bupt.edu.cn).

J. Zhang is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: jhzhang@bupt.edu.cn).

Y. Bu and Y. Tang are with the Department of Wireless and Terminal Technology, China Mobile Research Institute, Beijing 100053, China (e-mail: {buyan, tangyurong}@chinamobile.com).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, UK (e-mail: lh@ecs.soton.ac.uk).

domain of classic OFDM. For seamless integration into existing systems, OTFS can be implemented by adding a bespoke pre-processing module before the OFDM modulator, and a post-processing module at the receiver side. As a benefit, a time-varying channel can be converted into a near-time-invariant channel in the DD domain. Hence OTFS systems exhibit robust Doppler immunity in the face of high mobility and have attracted substantial research interests recently.

An important issue for OTFS is the accuracy of channel estimation. For single-antenna OTFS systems, sophisticated impulse-based channel estimation methods were studied in [5], [6]. The author of [5] suggested to use a pseudo-noise (PN) sequence for training, while Raviteja *et al.* [6] proposed an embedded pilot-aided channel estimation scheme requiring a guard region between pilot symbols. The performance of impulse-based methods are limited by the guard region and the pilot overhead, which makes it difficult to extend them to massive MIMO systems, where the pilot overhead is a major concern. To reduce the pilot overhead, the authors of [7] and [8] developed a joint pilot-and-data aided channel estimation algorithm. Srivastava *et al.* [9] proposed a guard-free pilot pattern and a sparse Bayesian learning (SBL) based channel estimation algorithm, and [10] explored the group sparsity for DD domains in MIMO-OTFS systems. Lei *et al.* [11] and [12] avoided guard region in SISO and small-scale MIMO systems with SBL methods using *Laplace prior*. The existing studies on SBL have not investigated the angular domain sparsity pattern, and the DDA domain joint sparsity in massive MIMO-OTFS systems also has to be researched. For massive MIMO-OTFS systems, Shen *et al.* [13] conceived a delay-Doppler-angle (DDA) domain sparse channel model and a 3D-structured orthogonal matching pursuit (SOMP) algorithm for downlink (DL) channel estimation. However, this scheme has a high pilot overhead, and assumes that the burst sparsity length of the angular domain is known in advance. Otherwise, the accuracy of the estimation is seriously affected. Liu *et al.* [14] exploited the reciprocity of time division duplex (TDD) and proposed an uplink-aided sparse DL channel estimation scheme, which reduces the pilot overhead. However, the application scenario of these contributions is limited, because they impose strong conditions, such as the *a priori* knowledge of the sparsity pattern and the TDD-based reciprocity, for allowing their schemes to exploit the channel sparsity.

In this paper, we eliminate the above-mentioned strong assumptions and conceive an improved channel estimation algorithm based on joint sparsity pattern learning (JSPL) under the powerful Bayesian framework, for fully exploiting the potential sparsity of the DDA domain channel in massive MIMO-OTFS systems. For more accurately learning the sparsity pattern, we firstly apply a flexible *spike and slab* prior

model [15], and derive its statistical channel model in the DDA domain under the Bayesian framework. Secondly, we propose a novel higher-accuracy parameter update rule to capture the joint sparsity pattern in DDA domain for obtaining the support set of the channel matrix. Finally, we transform the channel estimation problem into a simpler sparse signal recovery problem thanks to the finite support set of the channel, and solve it by the OMP algorithm. Our analysis and simulation results demonstrate that the proposed algorithm achieves better performance than the representative state-of-the-art baseline schemes, despite its reduced pilot overhead.

II. SYSTEM MODEL

A. Time-Varying Massive MIMO Channel Model

We consider a single-cell massive MIMO-OTFS system having N_T transmit antennas at the base station (BS) and U single-antenna users. For DL transmission operating in the frequency division duplex (FDD) mode[1], the users have to perform channel estimation and feed it back to the BS as a prerequisite for DL beamforming. Let us assume that the time-varying channel between a user and a BS antenna contains N_P dominant propagation paths. Then the multi-path DL channel at the time instant κ between the $(q+1)$th BS antenna ($q = 0, 1, 2, \cdots, N_T - 1$) and the uth user is given by

$$h_{u,q,\kappa,\ell} = \sum_{i=1}^{N_P} \alpha_{u,i} e^{j2\pi \nu_{u,i} \kappa T_s} \delta\left(\ell T_s - \tau_{u,i}\right) e^{j2\pi q \psi_{u,i}}, \quad (1)$$

where $\alpha_{u,i}$, $\nu_{u,i}$, $\tau_{u,i}$ and $\psi_{u,i}$ represent the propagation gain, Doppler shift, propagation delay and phase difference between two adjacent elements, corresponding to the ith path of the user u, respectively. Note that u, κ, ℓ and q are all integer values, and ℓ represents the index of the propagation delay $\tau_{u,i}$ in terms of the number of T_s. We denote the angle of departure (AoD) corresponding to the ith path of the user u as $\theta_{u,i}$. When a uniform linear array (ULA) of antennas is used, we have $\psi_{u,i} = \frac{d}{\lambda}\sin\theta_{u,i}$, where d is the antenna spacing and λ is the wavelength of the carrier frequency. Typically, $d = \lambda/2$ and $\theta_{u,i} \in [-\pi/2, \pi/2)$, thus we have $\psi_{u,i} \in [-1/2, 1/2)$. Finally, $\delta(\cdot)$ denotes the Dirac delta function, and T_s represents the sampling period or symbol duration.

B. OTFS Modulation and Demodulation

Consider a symbol sequence of length $N_\ell N_k$ that is rearranged into a two-dimensional data block $\mathbf{X}_q^{DD} \in \mathbb{C}^{N_\ell \times N_k}$ at the $(q+1)$th antenna, where N_ℓ and N_k represent the number of subcarriers and the number of symbols, respectively. The data block $\mathbf{X}_q^{DD}$ supposed to be processed in the delay-Doppler (DD) domain is firstly converted to the time-frequency (TF) domain data block $\underline{\mathbf{X}}_q \in \mathbb{C}^{N_\ell \times N_k}$ by the inverse symplectic finite Fourier transform (ISFFT) [4]. Then, $\underline{\mathbf{X}}_q$ is transformed into OFDM signal blocks by the inverse discrete Fourier transform (IDFT), followed by multiplying a transmitting window function $\mathbf{G}_{TX} \in \mathbb{C}^{N_\ell \times N_\ell}$. By inserting

---

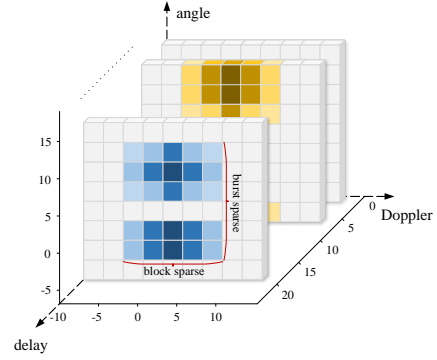[1]In fact, the proposed scheme does not depend on the specific duplex mode, hence it is also applicable to TDD.

Fig. 1. Sparsity pattern of the DDA domain channel model.

the cyclic prefix (CP) using the matrix $\mathbf{A}_{CP} \in \mathbb{C}^{(N_\ell + N_{CP}) \times N_\ell}$, the transmitted signal can be expressed as

$$\mathbf{s}_q = \text{vec}\{\mathbf{A}_{CP}\mathbf{G}_{TX}\mathbf{F}_{N_\ell}^H \underbrace{\mathbf{F}_{N_\ell}\mathbf{X}_q^{DD}\mathbf{F}_{N_k}^H}_{\text{ISFFT}}\}, \quad (2)$$

where N_{CP} is the length of CP, $\mathbf{F}_{N_\ell} \in \mathbb{C}^{N_\ell \times N_\ell}$ and $\mathbf{F}_{N_k} \in \mathbb{C}^{N_k \times N_k}$ are the DFT matrices, $(\cdot)^H$ denotes the conjugate transpose, and $\text{vec}\{\cdot\}$ denotes the parallel-to-serial conversion.

After the transmitted signal $\mathbf{s}_q$ passes through the channel, the signal $\mathbf{r}_u \in \mathbb{C}^{(N_\ell + N_{CP})N_k \times 1}$ received by the user u is rearranged as a matrix $\mathbf{R}_u \in \mathbb{C}^{(N_\ell + N_{CP}) \times N_k}$, i.e., $\mathbf{R}_u = \text{inverse\_vec}\{\mathbf{r}_u\}$. Then the signal removes the CP by using $\mathbf{A}_{RCP} \in \mathbb{C}^{N_\ell \times (N_\ell + N_{CP})}$ and passes through the receive window function $\mathbf{G}_{RX} \in \mathbb{C}^{N_\ell \times N_\ell}$. After applying the SFFT on the TF domain signal $\underline{\mathbf{Y}}_u \in \mathbb{C}^{N_\ell \times N_k}$, we obtain the DD domain data on the receiver side and it is expressed as

$$\mathbf{Y}_u^{DD} = \mathbf{F}_{N_\ell}^H \underbrace{\mathbf{F}_{N_\ell}\mathbf{G}_{RX}\mathbf{A}_{RCP}\mathbf{R}_u}_{\underline{\mathbf{Y}}_u} \mathbf{F}_{N_k}. \quad (3)$$

C. Channel Input-Output Analysis

We denote the $(\ell+1, k+1+N_k/2)$th entry of $\mathbf{Y}_u^{DD}$ as $y_{u,\ell,k}^{DD}$, which is expressed as [13]

$$y_{u,\ell,k}^{DD} \overset{N_k \to \infty}{=} \sum_{q=0}^{N_T-1} \sum_{\ell'=0}^{N_\ell-1} \sum_{k'=-N_k/2}^{N_k/2-1} x_{\ell',k',q}^{DD}$$
$$\times h_{u,\ell-\ell',k-k',q}^{DDS} e^{j2\pi \frac{\ell(k-k')}{N_k(N_\ell+N_{CP})}} + w_{u,\ell,k}^{DD}. \quad (4)$$

Eq. (4) represents the two-dimensional convolution of the DD domain, $h_{u,\ell-\ell',k-k',q}^{DDS}$ denotes the channel in the delay-Doppler-space (DDS) domain, and $w_{u,\ell,k}^{DD}$ is the noise satisfying a complex Gaussian distribution with zero mean and variance η. From (1) and (4) and considering a *spatial DFT* along the antenna index q [16], we can obtain the DDA domain channel as

$$h_{u,\ell,k,r}^{DDA} = \sum_{i=1}^{N_P} \alpha_{u,i} e^{j2\pi \nu_{u,i} T_s} \frac{\sin(\pi(\nu_{u,i}N_kT - k))}{\sin\left(\pi \frac{(\nu_{u,i}N_kT - k)}{N_k}\right)}$$
$$\times e^{j\pi \frac{(\nu_{u,i}N_kT - k)(N_k-1)}{N_k}} \delta\left(\ell T_s - \tau_{u,i}\right)$$
$$\times \frac{\sin(\pi(N_T\psi_{u,i} - r))}{\sin\left(\pi \frac{(N_T\psi_{u,i} - r)}{N_T}\right)} e^{j\pi \frac{(N_T\psi_{u,i} - r)(N_T-1)}{N_T}}, \quad (5)$$

where $T = (N_\ell + N_{\text{CP}})T_{\text{s}}$, and $r \in \{-N_{\text{T}}/2, -N_{\text{T}}/2 + 1, \cdots, N_{\text{T}}/2 - 1\}$ denotes the index of AoD corresponding to the antenna index q. From (5), we can verify that only few of the values of $h_{u,\ell,k,r}^{\text{DDA}}$ are nonzero elements, which are obtained when $r \to N_{\text{T}}\psi_{u,i}$, $\ell = \tau_{u,i}N_\ell\Delta f$ and $k \to \nu_{u,i}N_kT$, where Δf is subcarrier spacing. Thus the channel in the DDA domain is sparse. Moreover, the channel is block-sparse in the Doppler domain and burst-sparse in the angle domain, as shown in Fig. 1.

III. THE PROPOSED JSPL BASED CHANNEL ESTIMATION

A. Derivation of JSPL Based Channel Estimation Algorithm

Rather than one-dimensional sparsity, e.g., angle domain sparsity [17], massive MIMO-OTFS channel often exhibits a joint sparsity pattern in multiple domains. The performance of channel estimation can be improved if this multi-dimensional sparsity is properly exploited. For the sake of simplicity, we arrange $h_{u,\ell,k,r}^{\text{DDA}}$ into the column vector $\mathbf{h}_u \in \mathbb{C}^{N_\ell N_k N_{\text{T}} \times 1}$, where $\ell = 0, 1, \cdots, N_\ell - 1, k = -N_k/2, -N_k/2 + 1, \cdots, N_k/2 - 1$, and $r = -N_{\text{T}}/2, -N_{\text{T}}/2 + 1, \cdots, N_{\text{T}}/2 - 1$. We may omit the subscript u without confusion and convert Eq. (4) to

$$\mathbf{y} = (\mathbf{C} \star \mathbf{Z})\mathbf{h} + \mathbf{w} = \mathbf{\Phi}\mathbf{h} + \mathbf{w}, \qquad (6)$$

where $(\mathbf{C} \star \mathbf{Z}) = \left[\mathbf{C} \odot \mathbf{Z}_{-\frac{N_{\text{T}}}{2}}, \mathbf{C} \odot \mathbf{Z}_{-\frac{N_{\text{T}}}{2}+1}, \cdots, \mathbf{C} \odot \mathbf{Z}_r, \cdots, \mathbf{C} \odot \mathbf{Z}_{\frac{N_{\text{T}}}{2}-1} \right]$, with $\odot$ denoting the Hadamard product. $\mathbf{Z}_r \in \mathbb{C}^{N_\ell N_k \times N_\ell N_k}$ is a two-dimensional convolution matrix, whose $[(\ell + 1/2)N_k + k + 1, (\ell' + 1/2)N_k + k' + 1]$th element equals $\sum_{q=0}^{N_{\text{T}}-1} e^{j2\pi \frac{qr}{N_{\text{T}}}} x_{\ell-\ell',k-k',q}^{\text{DD}}$, and $\mathbf{C} \in \mathbb{C}^{N_\ell N_k \times N_\ell N_k}$ is a matrix with the $[(\ell + 1/2)N_k + k + 1, (\ell' + 1/2)N_k + k' + 1]$th element being $e^{j2\pi \frac{\ell(k-k')}{N_k(N_\ell+N_{\text{CP}})}}$. However, the sparsity pattern embedded in $\mathbf{h}$ is usually unknown *a priori*. To exploit the sparsity pattern of $\mathbf{h}$, we invoke the flexible *spike and slab* prior [15] from the Bayesian perspective, thus expressing the *a prior* distribution of $\mathbf{h}$ as

$$p(\mathbf{h}) = \prod_{n=1}^{N} \underbrace{[(1 - \lambda_n)\delta(h_n) + \lambda_n \mathcal{CN}(h_n \neq 0; 0, \mu)]}_{p(h_n)}, \quad (7)$$

where $\lambda_n \in (0, 1)$ represents the probability of h_n being nonzero[2], $N = N_\ell N_k N_{\text{T}} = M N_{\text{T}}$, h_n denotes the nth element of $\mathbf{h}$ and follows a complex Gaussian distribution $\mathcal{CN}(0, \mu)$ upon assuming Rayleigh fading except for $h_n = 0$. The likelihood function of the received signal is given by

$$p(\mathbf{y}|\mathbf{h}) = \mathcal{CN}(\mathbf{y}; \mathbf{\Phi}\mathbf{h}, \eta\mathbf{I}). \qquad (8)$$

Let $\mathcal{N}(x; \beta, \gamma)$ represent the expression $\frac{1}{\sqrt{2\pi\gamma}}e^{-\frac{(x-\beta)^2}{2\gamma}}$. Then the posterior distribution of $\text{Re}\{h_n\}$ can be written as[3]

$$p(\text{Re}\{h_n\}|\mathbf{y}; \beta_n, \gamma_n) = \frac{1}{\zeta_n} p(\text{Re}\{h_n\})\mathcal{N}(\text{Re}\{h_n\}; \beta_n, \gamma_n) \quad (9)$$

---

[2]λ_n essentially characterizes the sparsity degree of $\mathbf{h}$.

[3]Since our derivation aims to solve for the channel support set, namely the positions of the nonzero elements, rather than the value of the channel itself, it is sufficient to examine the real and imaginary parts of h_n separately. Thus we can simplify the derivation and the execution of the JSPL algorithm by invoking the one-dimensional real Gaussian distribution $\mathcal{N}$ twice: one for the real part and the other for the imaginary part of h_n.

upon using the approximate message passing (AMP) algorithm [18], which approximates the message passed from the node h_n to all the nodes y_m by a Gaussian distribution with a mean of β_n and variance of γ_n, according to the central limit theorem and factor graph. Here y_m, $m = 1, \cdots, M$, is the mth element of $\mathbf{y}$, while β_n and γ_n are iteratively updated (see Algorithm 1), and ζ_n is the normalization factor defined as

$$\begin{aligned} \zeta_n &\triangleq \int_{\text{Re}\{h_n\}} p(\text{Re}\{h_n\})\mathcal{N}(\text{Re}\{h_n\}; \beta_n, \gamma_n) \\ &= (1 - \lambda_n)\mathcal{N}(0; \beta_n, \gamma_n) + \lambda_n\mathcal{N}(0; \beta_n, \gamma_n + \mu). \end{aligned} \quad (10)$$

Based on (7) and (10), we rewrite (9) as

$$\begin{aligned} p(\text{Re}\{h_n\}|\mathbf{y}; \beta_n, \gamma_n) = &(1 - \varphi_n)\delta(h_n) \\ &+ \varphi_n\mathcal{N}\left(\text{Re}\{h_n\}; \frac{\mu\beta_n}{\gamma_n + \mu}, \frac{\mu\gamma_n}{\gamma_n + \mu}\right), \end{aligned} \quad (11)$$

where we have

$$\varphi_n = \frac{\lambda_n}{\zeta_n}\mathcal{N}(0; \beta_n, \gamma_n + \mu). \qquad (12)$$

From (11) we can obtain the first and second order statistics of $\text{Re}\{h_n\}$, i.e., the mean and variance, as follows:

$$\bar{h}_n = \varphi_n \frac{\mu\beta_n}{\gamma_n + \mu}, \qquad (13)$$

$$v_n = \varphi_n\left[\left(\frac{\mu\beta_n}{\gamma_n + \mu}\right)^2 + \frac{\mu\gamma_n}{\gamma_n + \mu}\right] - \bar{h}_n^2. \qquad (14)$$

To exploit the statistical characteristics of the channel, we define the hyperparameters as $\boldsymbol{\rho} \triangleq \{\boldsymbol{\lambda}, \mu, \eta\}$, where $\boldsymbol{\lambda} = \{\lambda_n\}$. Let us denote the estimate of a parameter at the tth iteration by $(\cdot)^{(t)}$, $t \geq 1$. Then $\boldsymbol{\rho}^{(t)}$ can be obtained by maximizing the expectation of the joint probability density function $p(\mathbf{h}, \mathbf{y})$ according to

$$\begin{aligned} \boldsymbol{\rho}^{(t+1)} &= \arg\max_{\boldsymbol{\rho}^{(t)}} \text{E}\left\{\ln p(\mathbf{h}, \mathbf{y})|\mathbf{y}; \boldsymbol{\rho}^{(t)}\right\} \\ &= \arg\max_{\boldsymbol{\rho}^{(t)}} \int_{\mathbf{h}} \ln[p(\mathbf{h})p(\mathbf{y}|\mathbf{h})]p(\text{Re}\{\mathbf{h}\}|\mathbf{y}; \boldsymbol{\rho}^{(t)}), \end{aligned} \quad (15)$$

where $p(\text{Re}\{\mathbf{h}\}|\mathbf{y}; \boldsymbol{\rho}^{(t)}) = \prod_n p(\text{Re}\{h_n\}|\mathbf{y}; \beta_n, \gamma_n)$. Since jointly optimizing all the variables in $\boldsymbol{\rho}$ is complex, we use the incremental expectation-maximization (EM) rule of [15] to update one variable at a time, while fixing the others. Firstly, to capture the channel's sparsity pattern, we update $\{\lambda_n\}$ for the above optimization problem as follows:

$$\begin{aligned} \lambda_n^{(t+1)} &= \arg\max_{\lambda_n \in (0,1)} \text{E}\left\{\ln[p(\mathbf{h})p(\mathbf{y}|\mathbf{h})]|\mathbf{y}; \boldsymbol{\rho}^{(t)}\right\} \\ &= \arg\max_{\lambda_n \in (0,1)} \int_{\mathbf{h}} [\ln p(\mathbf{h}) + C]p(\text{Re}\{\mathbf{h}\}|\mathbf{y}; \boldsymbol{\rho}^{(t)}) \\ &= \arg\max_{\lambda_n \in (0,1)} \int_{\mathbf{h}} \ln p(\mathbf{h})p(\text{Re}\{\mathbf{h}\}|\mathbf{y}; \boldsymbol{\rho}^{(t)}), \end{aligned} \quad (16)$$

where C denotes a constant, and (16) implies that for the update of λ_n, $p(\mathbf{y}|\mathbf{h})$ can be considered as a constant term to be ignored. To solve the problem (16), we obtain

$$\frac{d}{d\boldsymbol{\lambda}} \int_{\mathbf{h}} \ln p(\mathbf{h})p\left(\text{Re}\{\mathbf{h}\}|\mathbf{y}; \boldsymbol{\rho}^{(t)}\right) = 0. \qquad (17)$$

By means of Leibniz's formula and some approximate operations, we convert (17) to:

$$\sum_{n=1}^{N} \int_{h_n} p\left(\mathrm{Re}\left\{h_n\right\} \mid \mathbf{y}; \beta_n, \gamma_n\right) \frac{d}{d\lambda_n} \ln p(h_n) = 0, \quad (18)$$

where

$$\begin{aligned} \frac{d}{d\lambda_n} \ln p\left(h_n\right) &= \frac{\mathcal{CN}\left(h_n \neq 0; 0, \mu\right) - \delta\left(h_n\right)}{p\left(h_n\right)} \\ &= \begin{cases} \frac{1}{\lambda_n}, & h_n \neq 0. \\ -\frac{1}{1-\lambda_n}, & h_n = 0. \end{cases} \end{aligned} \quad (19)$$

To solve the integral term in (18), we define an interval $\mathcal{Z} \triangleq [-\varepsilon, \varepsilon]$, and represent the complement of $\mathcal{Z}$ by $\overline{\mathcal{Z}}$ in the real field. When $\varepsilon \to 0$ we have:

$$\lim_{\varepsilon \to 0} \int_{\mathrm{Re}\{h_n\} \in \mathcal{Z}} p\left(\mathrm{Re}\left\{h_n\right\} | \mathbf{y}; \beta_n, \gamma_n\right) = 1 - \varphi_n, \quad (20)$$

$$\lim_{\varepsilon \to 0} \int_{\mathrm{Re}\{h_n\} \in \overline{\mathcal{Z}}} p\left(\mathrm{Re}\left\{h_n\right\} | \mathbf{y}; \beta_n, \gamma_n\right) = \varphi_n. \quad (21)$$

By combining (19), (20) and (21), we can derive (18) as:

$$\sum_{n=1}^{N} \left[\frac{1}{\lambda_n} \varphi_n - \frac{1}{1-\lambda_n}\left(1 - \varphi_n\right)\right] = 0. \quad (22)$$

Hence, the update formula for λ_n is obtained as:

$$\lambda_n^{(t+1)} = \varphi_n^{(t)}. \quad (23)$$

From (23) we find that the update of $\{\lambda_n\}$ only depends on the nth element of the channel vector $\mathbf{h}$, which means it does not capture the joint sparsity pattern of the DDA channel. By considering the multi-burst sparsity of the angle domain, the supports of the channel exhibit two-dimensional multi-block sparsity in the Doppler-angle domain, which implies that the Doppler domain channel supports and the angular domain channel supports intersect across multiple blocks in the Doppler-angle plane. Therefore, we update $\{\lambda_n\}$ by using the adjacency-assisted method, as detailed below, in order to learn the sparsity pattern of the channel, and propose the following update rule:

$$\lambda_n^{(t+1)} = \frac{1}{\sum_a \xi_{n,a} |\Gamma(n,a)|} \sum_a \sum_{b=1}^{|\Gamma(n,a)|} \xi_{n,a} \varphi_{n,a,b}^{(t)}, \quad (24)$$

where $\Gamma(n,a)$ denotes the set of indices of the ath nearest elements in $\mathbf{h}$ for h_n. For example, the coordinate of h_n in the Doppler-angle domain is (k,r), then the set of its nearest elements $(a = 1)$ is denoted by $\Gamma(n,1) = \{(k-1,r), (k+1,r), (k, (r-1+N_T)_{N_T}), (k, (r+1+N_T)_{N_T})\}$ as shown in Fig. 2, and $(\cdot)_{N_T}$ represents modulo N_T. Case 1 represents the most conventional set of nearest neighbors containing the left and right elements of the current element in the Doppler domain and the up and down elements in the angle domain. Case 2 represents the set of nearest neighbors of a max Doppler boundary element containing the left element of the current element in the Doppler domain and the up and down elements in the angle domain. While Case 3 represents the set of nearest neighbors of an angle domain boundary element containing the left and right elements of the current element in the Doppler domain and the lower element and the cyclic mode-taking element due to the burst sparsity pattern in the angle domain. $\varphi_{n,a,b}^{(t)}$ represents the particular $\varphi_n^{(t)}$

Algorithm 1 JSPL-Based Channel Estimation Algorithm

Input: $\mathbf{y}$, $\boldsymbol{\Phi} = \{\Phi_{mn}\}$
Initialization: $t = 1$, $i = 1$, $\Omega = \emptyset$, $\Omega_{\mathrm{d}} = \emptyset$, $\Omega_{\mathrm{DA}} = \emptyset$, $S_m^{(0)} = 0$, $N_P = 0$, T_{MAX}, $\boldsymbol{\rho}^{(1)}$, ϵ_1, ϵ_2, $\bar{h}_n^{(1)}$, $v_n^{(1)}$
1: Update process parameters:
  $V_m^{(t)} = \sum_n |\Phi_{mn}|^2 v_n^{(t)}$
  $S_m^{(t)} = (y_m - \sum_n \Phi_{mn} \bar{h}_n^{(t)} + V_m^{(t)} S_m^{(t-1)})/(\eta^{(t)} + V_m^{(t)})$
2: Update variable parameters:
  $\gamma_n^{(t)} = \left[\sum_m \frac{|\Phi_{mn}|^2}{\eta^{(t)} + V_m^{(t)}}\right]^{-1}$
  $\beta_n^{(t)} = \bar{h}_n^{(t)} + \gamma_n^{(t)} \sum_m \Phi_{mn} S_m^{(t)}$
3: Update $\varphi_n^{(t)}$, $\bar{h}_n^{(t+1)}$ and $v_n^{(t+1)}$ with (12), (13) and (14)
4: Update $\boldsymbol{\rho}^{(t+1)}$ with (24), (25) and (26)
5: Let $t = t + 1$ and proceed to Step 1 until $\left\|\boldsymbol{\lambda}^{(t+1)} - \boldsymbol{\lambda}^{(t)}\right\|_2 < \epsilon_1 \left\|\boldsymbol{\lambda}^{(t)}\right\|_2$ or $t = T_{\mathrm{MAX}}$, s.t. $\forall n \in [1, \cdots, N]$
6: Reshape $\boldsymbol{\lambda}$ as a tensor $\boldsymbol{\Lambda}$
7: **if** $\|\boldsymbol{\Lambda}(\ell,:)\|_2 > \epsilon_2$ **then**
8: $\quad N_P = N_P + 1$
9: $\quad \Omega_{\mathrm{d}} = \Omega_{\mathrm{d}} \cup \ell$
10: **end if**
11: **for** $i \leq N_P$ **do**
12: $\quad$ **if** $\boldsymbol{\Lambda}(\ell_{\mathrm{d}}^{(i)}, k, r) > \epsilon_2$, s.t. $\ell_{\mathrm{d}}^{(i)} \in \Omega_{\mathrm{d}}$ **then**
13: $\quad\quad \Omega_{\mathrm{DA}}^{(i)} = \Omega_{\mathrm{DA}}^{(i)} \cup (k,r)$
14: $\quad$ **end if**
15: $\quad \Omega = \Omega \cup (\ell_{\mathrm{d}}^{(i)}, k, r)$, s.t. $\forall (k,r) \in \Omega_{\mathrm{DA}}^{(i)}$
16: $\quad \mathbf{h}^{(i)}|_{\underline{\Omega}} = \boldsymbol{\Phi}^{\dagger}|_{\underline{\Omega}} \times \mathbf{y}$, where $(\cdot)^{\dagger}$ denotes pseudo-inverse and $\underline{\Omega}$ indicates the positions of channel supports determined by Ω.
17: $\quad i = i + 1$
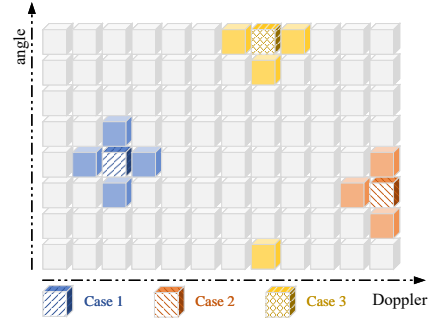18: **end for**
Output: $\mathbf{h}$

Fig. 2. Three cases of the set of the nearest elements.

---

corresponding to the bth element of $\Gamma(n,a)$, $\xi_{n,a}$ denotes the weight of $\varphi_{n,a,b}^{(t)}$, and $|\Gamma(n,a)|$ is the cardinality of $\Gamma(n,a)$. Secondly, the derivative of the objective function of (15) is calculated with respect to the individual parameters μ and η, which are then updated relying on the EM [15] as

$$\mu^{(t+1)} = \frac{1}{\sum_n \varphi_n^{(t)}} \sum_{n=1}^{N} \varphi_n^{(t)} \left[\left(\frac{\mu^{(t)} \beta_n^{(t)}}{\gamma_n^{(t)} + \mu^{(t)}}\right)^2 + \frac{\mu^{(t)} \gamma_n^{(t)}}{\gamma_n^{(t)} + \mu^{(t)}}\right], \quad (25)$$

$$\eta^{(t+1)} = \frac{1}{M} \sum_{m=1}^{M} \left[\left(\eta^{(t)} S_m^{(t)}\right)^2 + \frac{\eta^{(t)} V_m^{(t)}}{\eta^{(t)} + V_m^{(t)}}\right], \quad (26)$$

where $V_m^{(t)}$ and $S_m^{(t)}$ are intermediate parameters updated first in each iteration of our algorithm. Then, $\gamma_n^{(t)}$, $\beta_n^{(t)}$, $\varphi_n^{(t)}$, $\bar{h}_n^{(t+1)}$ and $v_n^{(t+1)}$ are updated in turn. Next, we update $\boldsymbol{\rho}^{(t+1)}$ to capture the sparsity pattern of the channel by setting an appropriate threshold ϵ_1 for $\boldsymbol{\lambda}$. Thus we can convert the channel estimation problem into a sparse signal recovery problem based on the DDA domain channel supports. Specifically, we

first obtain the channel supports in the delay domain, then by traversing each path we filter out the Doppler-angle domain two-dimensional multi-block channel supports corresponding to the path, and finally obtain the estimated channel. For clarity, the above procedures are summarized in Alg. 1.

B. Analysis of Computational Complexity and Pilot Overhead

The computational complexity of our JSPL algorithm is mainly determined by two parts. Firstly, the complexity of the hyperparameter learning part is determined by the update of parameters, which is characterized by $\mathcal{O}(T_{\text{MAX}}MN)$, where T_{MAX} denotes the maximum number of iterations and it is much smaller than MN. Secondly, in the sparse signal recovery part, the complexity is dominated by calculating the pseudo-inverse of $\mathbf{\Phi}$. According to compressed sensing theory [19], the actual dimension of the matrix involved in the inverse operation is $M \times Q$, where Q indicates the sparsity of the DDA domain channel and it is given by $Q = N_{\text{P}}D_{\text{MAX}}N_{\text{D}}$, with D_{MAX} being the index corresponding to the maximum Doppler shift and $N_{\text{D}} \approx N_{\text{T}}/10$ being the length of a single burst in the angle domain [17]. Thus when $M > Q$, the matrix has a left inverse, at which point the complexity is $\propto \max\{Q^3, Q^2M\} = \mathcal{O}(Q^2M)$. While when $M < Q$, the complexity is $\mathcal{O}(QM^2)$, so the complexity of the inverse calculation is $\propto \max\{\mathcal{O}(Q^2M), \mathcal{O}(QM^2)\}$. Considering the increased N_{T} in massive MIMO systems, Q represents the dominant contribution. Thus the complexity of JSPL is $\mathcal{O}(\frac{N_{\text{P}}D_{\text{MAX}}}{10}N_\ell^2 N_k^2 N_{\text{T}} + T_{\text{MAX}}N_\ell^2 N_k^2 N_{\text{T}})$, where N_{P} is typically much smaller than the other parameters. In massive MIMO systems, the complexity of JSPL is linearly increasing with N_T, thus it can be controlled within acceptable limit by right-sizing the OTFS resource blocks.

For the impulse-based channel estimation scheme, [6], the pilot overhead is $\propto N_\ell N_k N_{\text{T}}$, and the guard regions occupy substantial spectral resources. Additionally, upon increasing of the pilot overhead, the guard regions between different impulses may become insufficient, thus resulting in mutual interference. Therefore, it is unsuitable for massive MIMO systems. By contrast, the pilot overhead of channel estimation schemes based on sparse signal recovery is typically $\propto Q \log(N_\ell N_k N_{\text{T}})$, which is a lower bound. However, in reality the number of dominant path N_{P} and the multi-block sparsity of the Doppler-angle domain are usually unknown. As a result, the method of determining channel supports by correlation [13] often requires using more pilots than the lower-bound value to ensure high accuracy. Notably, our proposed JSPL scheme accurately estimates the channel supports by means of parameter learning, and the learning process can be offline nature, hence an ultra-low pilot overhead is incurred.

IV. SIMULATION RESULTS AND DISCUSSIONS

In this section, we compare the performance of our proposed JSPL algorithm and representative baseline schemes by numerical simulations. We use the 3GPP standardized channel model containing 6 dominant paths [20]. The carrier frequency is 4.9 GHz, the Duplex mode is FDD and the subcarrier spacing is $\Delta f = 15$kHz . The number of BS antennas is $N_{\text{T}} = 64$, and
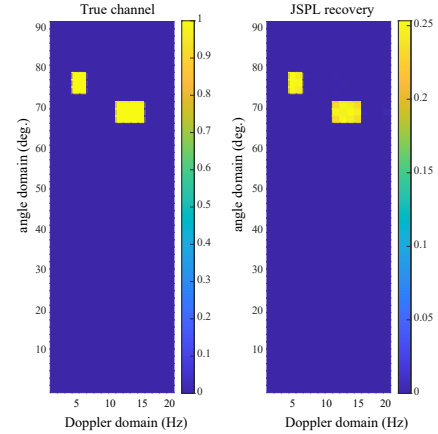

Fig. 3. The JSPL recovery (on the right) and the true channel (on the left) vs. the sparsity degree λ_n of the Doppler-angle domain channel matrix.

the UE has a single antenna. Then We set the size of the OTFS resource block to $(N_\ell, N_k) = (1024, 128)$, and CP = 16.6us. Other initializations concerning the learning can be found in Alg. 1. Finally, the receiver uses the popular message passing (MP) based algorithm relying on threshold detection [6].

In Fig. 3, we compare the JSPL recovery (on the right) with the true channel (on the left) as a function of the degree of sparsity λ_n – indicated by the accompanying color bar – of the Doppler-angle domain channel matrix. The channel supports are obtained by comparing the JSPL recovery to the threshold ϵ_2 (see Step 12 of Alg. 1). We observe that the accuracy of the two-dimensional channel support recovery in the Doppler-angle domain is high.

In Fig. 4, we evaluate the normalized mean square error (NMSE) performance of JSPL against the baseline schemes. Two vehicular speeds are considered: 120km/h and 360km/h. Since the impulse-based scheme is constrained by the guard region, while OMP and 3D-SOMP lack *a priori* knowledge about the number of dominant propagation paths and the channel supports, they are all inferior to our JSPL. Additionally, the performance improvement of JSPL at 120km/h is higher than at 360km/h. This is because the capability to estimate the Doppler spread for a given number of OTFS resource blocks tends to be degraded for a higher velocity.

Fig. 5 compares the NMSE performance of our JSPL and the baseline schemes under different pilot overheads and signal-to-noise ratio (SNR) values. We set the user's speed to 360km/h. It is observed that although our JSPL uses only 10% of the overhead of the baseline algorithms, it still outperforms the baselines thanks to its ability to learn the channel supports. In other words, our JSPL substantially reduces the pilot overhead, while ensuring a higher estimation accuracy than the baseline sparse signal recovery algorithms relying on correlation-aided channel support judgement. Moreover, the performance improvement of our JSPL becomes small with high pilot overhead, because the accuracy of determining channel supports by correlation is also significantly improved when there are sufficient pilots. However, such a high pilot overhead is unaffordable to massive MIMO systems.

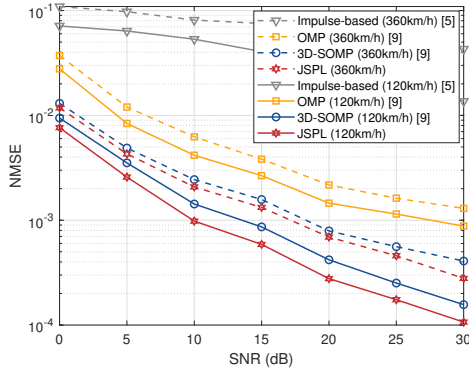Finally, in Fig. 6 we compare the bit error rate (BER)

Fig. 4. NMSE of our JSPL and the baseline schemes at different speeds.

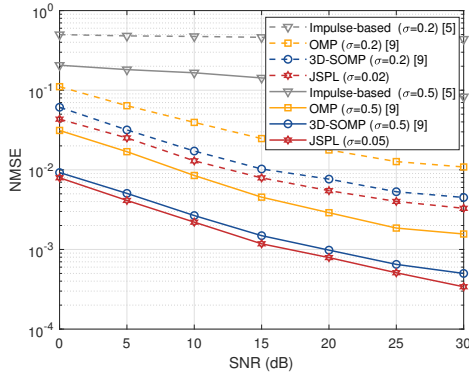Fig. 6. BER of our JSPL and the baseline schemes.

Fig. 5. NMSE of our JSPL and the baseline schemes with different pilot overhead $\sigma = \frac{\text{Number of resource blocks occupied by pilots}}{MN}$.

upon using different channel estimation algorithms and perfect channel knowledge. We set the user's speed to 360km/h and the pilot overhead to 50% for the baseline channel estimation algorithms (5% for JSPL and 0 for the perfect channel). It can be seen that our JSPL still outperforms the baseline channel estimation algorithms and exhibits a small gap w.r.t. the lower-bound provided by the perfect channel assumption.

V. CONCLUSION

We have proposed a JSPL based channel estimation algorithm for massive MIMO-OTFS systems under a Bayesian framework by exploiting the potential joint sparsity of the DDA domain channel. Due to the uncertainty of sparsity in realistic channels, we apply the *spike and slab* prior model to fit the channel and propose a new parameter update rule to estimate the channel support set by exploiting the channel's joint sparsity in multiple domains. Then we solve a sparse signal recovery problem based on the support set of the channel. Our analysis and simulation results demonstrate that the proposed algorithm exhibits competitive performance and greatly reduced pilot overhead at the cost of moderately increased computational complexity.

REFERENCES

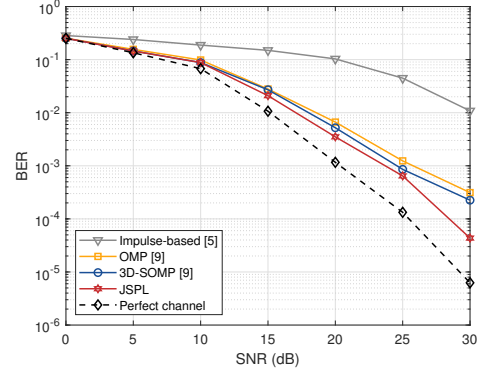[1] S. Yang, C. Zhou, T. Lv, and L. Hanzo, "Large-scale MIMO is capable of eliminating power-thirsty channel coding for wireless transmission of HEVC/H.265 video," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 57–63, Jun. 2016.

[2] A. Hu *et al.*, "An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Mar. 2014.

[3] T. Lv, S. Yang, and H. Gao, "Semi-blind channel estimation relying on optimum pilots designed for multi-cell large-scale MIMO systems," *IEEE Access*, vol. 4, pp. 1190–1204, Apr. 2016.

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

[5] K. R. Murali and A. Chockalingam, "On OTFS modulation for high-Doppler fading channels," in *Proc. IEEE Inf. Theory Appl. Workshop*, San Diego, USA, Feb. 2018, pp. 1–10.

[6] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay–Doppler channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906–4917, May 2019.

[7] H. B. Mishra, P. Singh, A. K. Prasad, and R. Budhiraja, "OTFS channel estimation and data detection designs with superimposed pilots," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2258–2274, Apr. 2022.

[8] W. Yuan *et al.*, "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.

[9] S. Srivastava *et al.*, "Bayesian learning aided sparse channel estimation for orthogonal time frequency space modulated systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8343–8348, Jul. 2021.

[10] ——, "Bayesian learning aided simultaneous row and group sparse channel estimation in orthogonal time frequency space modulated MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 635–648, 2021.

[11] L. Zhao, W.-J. Gao, and W. Guo, "Sparse Bayesian learning of delay-Doppler channel for OTFS system," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2766–2769, Dec. 2020.

[12] L. Zhao *et al.*, "Block sparse Bayesian learning-based channel estimation for MIMO-OTFS systems," *IEEE Commun. Lett.*, vol. 26, no. 4, pp. 892–896, 2022.

[13] W. Shen *et al.*, "Channel estimation for orthogonal time frequency space (OTFS) massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4204–4217, Aug. 2019.

[14] Y. Liu *et al.*, "Uplink-aided high mobility downlink channel estimation over massive MIMO-OTFS system," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1994–2009, Sep. 2020.

[15] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[16] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*, 2017.

[17] A. Liu, V. K. N. Lau, and W. Dai, "Exploiting burst-sparsity in massive MIMO with partial channel support information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7820–7830, Nov. 2016.

[18] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," in *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, Nov. 2009, pp. 18 914–18 919.

[19] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[20] *Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations*, 3GPP TR 25.996 V12.0.0, Sep. 2014.