



**HAL**  
open science

## Iterative channel estimation and data detection algorithm for MIMO-OTFS systems

Rabah Ouchikh, Thierry Chonavel, Abdeldjalil Aissa El Bey, Mustapha  
Djeddou

► **To cite this version:**

Rabah Ouchikh, Thierry Chonavel, Abdeldjalil Aissa El Bey, Mustapha Djeddou. Iterative channel estimation and data detection algorithm for MIMO-OTFS systems. *Digital Signal Processing*, 2023, 143, pp.104234. 10.1016/j.dsp.2023.104234 . hal-04229160

**HAL Id: hal-04229160**

**<https://imt.hal.science/hal-04229160v1>**

Submitted on 6 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Iterative channel estimation and data detection algorithm for MIMO-OTFS systems

Rabah Ouchikh<sup>a</sup>, Thierry Chonavel<sup>b</sup>, Abdeldjalil Aïssa-El-Bey<sup>b</sup>, Mustapha Djeddou<sup>a</sup>

<sup>a</sup>*Laboratoire Télécommunications, Ecole Militaire Polytechnique, Bordj El-Bahri, Algeria.*

<sup>b</sup>*IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France.*

---

## Abstract

Channel estimation in high-mobility environments is a challenging problem for advanced mobile communication systems (5G and beyond). In this manuscript, we first propose an iterative algorithm for channel estimation and data detection in the delay-Doppler domain for multiple-input multiple-output orthogonal time frequency space system. Then, in order to increase the spectral efficiency of the system, we use a superimposed pilot pattern. The proposed algorithm takes advantage from the sparse nature of the channel in the delay-Doppler domain and iterates between message passing-aided data detection and data-aided channel estimation. For channel estimation, we propose two algorithms. The first one consists in estimating all channel parameters, including the number of path gains, delay taps, Doppler taps, and channel gains by using a mean-field approximation and the variational Bayesian expectation maximization algorithm. The second one, based on the fact that delay and Doppler taps remain unchanged for a rather long period of time, uses an MMSE approach combined with Cholesky decomposition to only estimate channel gains in each transmitted frame. For data detection, we adapt the message-passing algorithm proposed in the literature. We also derive a lower bound on the signal-to-interference-plus-noise ratio of the proposed scheme, and maximize it by optimally allocating power between pilots and data symbols. Finally, we compare the complexity and the performance in terms of normalized mean square error, bit error rate, and spectral efficiency against existing methods. Simulation results, conducted in high-mobility scenarios show that the proposed algorithm achieves a good compromise between complexity and performance.

*Keywords:* OTFS, MIMO, channel estimation, data detection, superimposed pilot pattern.

---

## 1. Introduction

Future advanced mobile-communications systems require high reliability, low latency and high spectral efficiency (SE) communications [1, 2]. Orthogonal frequency division multiplexing (OFDM), which is the

---

*Email addresses:* ouchikh16rabah@gmail.com (Rabah Ouchikh), abdeljalil.aissaelbey@imt-atlantique.fr (Thierry Chonavel), thierry.chonavel@imt-atlantique.fr (Abdeldjalil Aïssa-El-Bey), djeddou.mustapha@gmail.com (Mustapha Djeddou)

4 most popular modulation technique used in today's mobile-communication systems, offers high throughput  
5 and high SE but its performance deteriorates in high mobility scenarios, such as high-speed vehicular, high-  
6 speed trains and even in aircraft flying scenarios [3]. These performance degradations are caused by the  
7 doubly-dispersive nature of the wireless channel in high mobility environments. This nature causes severe  
8 inter-carrier-interference in OFDM systems.

9 Orthogonal time frequency space (OTFS) modulation, which has been recently proposed in [4, 5], is  
10 a promising solution. Its robustness and better performance, compared to OFDM, are attractive for high-  
11 speed vehicular communication systems. The secret of OTFS is that it transforms a doubly-selective wireless  
12 channel to an almost flat one in the delay-Doppler (DD) domain thanks to the inverse symplectic finite Fourier  
13 transform (ISFFT). The fact that the doubly-dispersive channel is transformed into a flat fading channel  
14 can be exploited to reduce the bit error rate (BER). The sparsity of the channel in the DD domain can also  
15 be exploited to reduce the pilot overhead required to estimate a rapidly time-varying channel.

16 As with OFDM, multiple-input multiple-output (MIMO) can also be combined with OTFS and benefit  
17 from the diversity to further increase the transmission rate [6–10]. To ensure robust data transmission in a  
18 MIMO-OTFS system, efficient channel estimation and data detection algorithms are required at the receiver  
19 side. Several channel estimation schemes for MIMO-OTFS systems have been proposed in the literature  
20 [11–21]. Let us briefly review the most recent and popular of them.

21 In [11], a channel estimation scheme in the DD domain designed for MIMO-OTFS system is suggested.  
22 The designed channel estimation method uses impulses in the DD grid as pilots for estimation. Thanks to the  
23 pilot pattern used in [11] for pilots, data symbols and guard intervals, there is no interference between pilots  
24 and data symbols. Simulation results show that the proposed channel estimation scheme for MIMO-OTFS  
25 achieves good performance and outperforms the MIMO-OFDM system under high-Doppler scenarios.

26 In [12], a 3D orthogonal matching pursuit (3D-OMP) algorithm is suggested to solve the challenging  
27 downlink channel estimation problem for massive MIMO-OTFS system. First, the authors of [12] show that  
28 the MIMO-OTFS channel exhibits a 3D-structured sparsity. The channel is block sparse along the Doppler  
29 dimension because the system bandwidth is much greater than the Doppler shift of a path, i.e., the only one  
30 non-zero block is concentrated around zero, but its length is unknown. It is normal sparse along the delay  
31 dimension because the number of dominant propagation paths is limited. It is also burst sparse along the  
32 angle dimension, this is due to the fact that the angle-of-departure spread of a path at the base station is  
33 usually small. Although the non-zero burst lengths can be depicted as constant, it is unknown where each  
34 burst starts. Then, the downlink channel estimation is formulated as a sparse recovery problem. Simulations  
35 show that the designed algorithm can acquire a good channel state information (CSI) with reduced pilot  
36 overhead.

37 In [13], a downlink massive MIMO-OTFS channel estimation is regarded from a Bayesian perspective.  
38 First, the massive MIMO-OTFS signal model in the time domain is derived. Then, the variational Bayesian  
39 framework is adopted to recover the uplink channel parameters (the angle, the delay, the Doppler shift, and  
40 the channel gain) for each physical scattering path. Next, the reciprocity of the channel between the uplink  
41 and the downlink is exploited to reconstruct the parameters for the downlink massive channels at the base  
42 station. Simulations results confirm the validity and robustness of the proposed scheme.

43 In [14], authors use a tensor-based OMP channel estimation algorithm that exploits the channel sparsity  
44 in the DD-angle domain. A novel pilot design for OTFS in the time-frequency domain is firstly suggested.  
45 Then, based on this pilot pattern, the channel estimation is formulated as a sparse recovery problem and the  
46 tensor decomposition, and parallel support detection are introduced into the tensor-based OMP algorithm  
47 to reduce the signal processing dimension significantly. Numerical results show the superiority and the  
48 robustness of the proposed algorithm.

49 In [15], a scheme to acquire the state of massive MIMO-OTFS channels has been designed. A pilot  
50 pattern is then proposed to reduce pilot overhead and save memory consumption. Then, a channel estimation  
51 algorithm based on a modified sensing matrix is designed to acquire the downlink CSI. Numerical results  
52 show that this scheme has significant advantages over previous algorithms.

53 In [16], a low pilot overhead channel estimation scheme for cyclic prefix (CP)-OFDM-based massive  
54 MIMO-OTFS system is suggested. First, the CP-OFDM-based massive MIMO-OTFS system channel with  
55 antenna directivity pattern is analysed, and the burst sparsity in the angle domain is transformed into block  
56 sparsity by using non-uniform Fourier transform. Then, to solve the problem that the pilot overhead grows  
57 linearly with the number of antennas, a three-dimensional dynamic support detect algorithm is suggested.  
58 Simulation results show that the proposed algorithm has lower pilot overhead and higher channel estimation  
59 accuracy compared to the conventional OMP and the 3D-structured OMP algorithms.

60 A new receiver architecture based on the basic expansion model (BEM) OTFS is designed in [19] for high  
61 mobility communications with Doppler spread channel. First, the analytical BEM-OTFS system model is  
62 derived. Then, a low-order generalized complex exponential aided rough channel estimation with low pilot  
63 overhead is suggested. Finally, a high-resolution generalized complex exponential BEM model with a large  
64 BEM order is adopted for the channel estimation refinement and equalization. The data symbols are exploited  
65 as pseudo-pilots, leading to higher estimation accuracy. Simulation results show that the suggested method  
66 outperforms existing solutions in terms of BER and mean squared error (MSE) for channel estimation, while  
67 featuring low pilot overhead.

68 In another work [20], a new channel estimation scheme for MIMO-OTFS is suggested. The 2D structure  
69 in Doppler-angle domain for channel is considered and characterized via a local Beta process. The uplink

70 channel estimation problem is formulated as a sparse recovery problem, and it is solved via the estimation of  
 71 the delay, the Doppler shift and the angle of the channel using sparse Bayesian learning. Therefore, downlink  
 72 channel estimation is performed with the help of these estimated parameters. Simulation results show that  
 73 the proposed scheme achieves good performance of channel estimation and also shows robust adaptation to  
 74 variable Doppler.

75 In [21], to improve signal reconstruction in high-Doppler scenarios, a compressive sampling matching  
 76 pursuit (CoSaMP) algorithm is suggested for the downlink channel estimation in massive MIMO-OTFS  
 77 system. Simulation results show the gain brought by the CoSaMP algorithm with interleaving compared to  
 78 the state-of-the-art channel estimation schemes with a lower computational complexity.

79 In [17], in order to reduce the pilot overhead, training duration and the pre-processing complexity, an  
 80 end-to-end input-output MIMO-OTFS model is derived and a new model for sparse channel estimation, in  
 81 which pilots are placed in the time-frequency domain, is suggested. The two main contributions of this work  
 82 is that the suggested channel estimation scheme can estimate the fractional Doppler shifts efficiently and the  
 83 SE of the system is increased compared to the previous state-of-the-art methods.

84 The authors in [18] proposed an efficient channel estimation method for MIMO-OTFS systems. The  
 85 channel estimation problem is formulated as a block sparse recovery problem and it is solved via the designed  
 86 block sparse Bayesian learning with block reorganization method. Simulation results show that the suggested  
 87 method outperforms previous state-of-the-art methods in terms of performance and noise robustness.

88 To the best of our knowledge, except [22], [23], and [24], all channel estimation algorithms for SISO-OTFS  
 89 modulation in the DD domain use two groups of channel estimation methods. The first group concern the  
 90 conventional pilot aided (CPA) design with a super-frame architecture, where two frames are sent for each  
 91 transmission: one frame for channel estimation and another for symbol detection [11]. The second one is  
 92 the embedded pilot (EP)-aided channel estimation, where a pilot pattern regrouping pilots, data symbols,  
 93 and guard intervals in the same frame is used [25]. Both schemes suffer from the degradation of the SE of  
 94 the system. This problem becomes even more difficult in a MIMO-OTFS system due to the large number of  
 95 guard intervals required in the DD domain.

96 The second group of channel estimation methods, which is the most widely used in the literature, can  
 97 be further divided, according to the type of pilot pattern, into three schemes. We will define these three  
 98 schemes through three methods: a reference method [25] and two recent methods [17, 18], one method for  
 99 each scheme. In the first scheme, which is called pilot type-1 [25], pilots are designed with higher power than  
 100 data and guard intervals are inserted around pilot to alleviate the interference of data symbols. Using pilot  
 101 type-1, the channel estimation step is simple and less complex as it is performed with a simple threshold  
 102 method. However, the performance is poor with high pilot overhead (low SE) and high peak-to-average

103 power ratio (PAPR) due to guard intervals and different power design for data and pilots. To remedy these  
 104 shortcomings, the second and third schemes were proposed. In the second scheme, called pilot type-2 [18],  
 105 there is no guard intervals between the pilots and data symbols. The pilot type-2 tolerates some interference  
 106 between the data symbols and part of the pilots, but on the other hand it improves the pilot overhead by  
 107 removing the guard intervals and therefore improves the SE. In the third scheme, called pilot type-3 [17],  
 108 another strategy has been adopted, which consists in directly transmitting pilots over the time-frequency  
 109 domain grid for estimating the DD domain CSI. This leads to a reduction of the pilot overhead, training  
 110 duration and pre-processing complexity.

111 Although the pilot type-2 and pilot type-3 schemes offer better SE than the pilot type-1 scheme and the  
 112 CPA design, the pilot overhead of both schemes increases with increasing delay and Doppler spread of the  
 113 channel. For example, in a high Doppler spread channel, more pilots are needed for good channel estimation.  
 114 This increases the pilot overhead and decreases the SE. Hence, the need for a superimposed pilot pattern  
 115 and data symbols, where pilots and data symbols are superimposed in the same locations in the DD grid.

116 Superimposed schemes allow for the cancellation of pilot overhead due to the non-use of guard intervals  
 117 and the non-allocation of a particular space for pilots, but their use for MIMO systems requires the develop-  
 118 ment of a powerful channel estimation and data detection scheme to manage the interference between pilots  
 119 and data symbols. Furthermore, to improve SE and to minimize the BER, the transmitted power must be  
 120 optimally distributed between the pilots and data symbols. Considering the challenges mentioned earlier,  
 121 the objective of this work is to develop an iterative scheme for channel estimation and data detection in the  
 122 DD domain for MIMO-OTFS systems, utilizing superimposed pilot pattern to increase the SE. The main  
 123 contributions of this work can be summarized as follows:

- 124 1. We propose an algorithm for MIMO-OTFS channel estimation and data detection that uses super-  
 125 imposed pilot pattern in the DD domain and benefices from the sparse nature of the channel in this  
 126 domain. The proposed design mitigates the interference between the pilots and data symbols by iter-  
 127 ating between message passing-aided data detection and data-aided channel estimation in the DD  
 128 domain, and has better BER and SE. For channel estimation, we design two formulations and two  
 129 solutions to solve this problem. The first formulation consists in estimating the channel parameters,  
 130 including delay taps, Doppler taps and path gains, as a sparse recovery problem. This problem has  
 131 been solved by adapting the suggested algorithm for SISO-OTFS system in our previous works [26, 27]  
 132 to MIMO-OTFS system. This algorithm uses a mean-field approximation and VB-EM algorithm. The  
 133 second formulation consists in estimating a compact vector containing only the path gains, based on  
 134 the fact that the delay taps and Doppler taps remain unchanged for a given period. This problem has

135 been solved via a low complexity MMSE and the Cholesky decomposition.

- 136
- 137 2. For data detection step, we propose to this context an adapted version of the message-passing algo-  
138 rithm (AMPA). This algorithm is computationally-efficient as it exploits the channel sparsity in the  
139 DD domain. In its detection process, it takes into account the interference of the pilots on data symbols.  
140
- 141 3. We derive a lower bound on the signal-to-interference-plus-noise-ratio (SINR) of the proposed scheme,  
142 which is subsequently optimized to derive the optimal pilot power expression. It is demonstrated that  
143 the optimal power level reduces the BER and increases further the SE of the proposed scheme.  
144
- 145 4. We numerically validate the narrowness of the optimal distribution of transmitted power between pilots  
146 and data symbols and show its effect on the NMSE, BER and SE of the proposed scheme. Then, we  
147 compare the complexity and the performance of the proposed algorithm against four state-of-the-art  
148 methods, named: embedded pilots (EP) [25], block sparse Bayesian learning with block reorganization  
149 (BSBL-BR) [18], row-group OMP (RG-OMP), and row-group bayesian learning (RG-BL) [17], in terms  
150 of SE, normalize mean squared error (NMSE) and BER. Finally, we show the good compromise achieved  
151 by the proposed scheme between NMSE and BER performance, SE, computational complexity and  
152 PAPR.

153 The rest of the paper is structured as follows. Section 2 is reserved for the description of the MIMO-  
154 OTFS system model using superimposed pilot pattern. The suggested algorithm for channel estimation  
155 and data detection is detailed in Section 3. Section 4 is dedicated to the derivation of the optimal power  
156 distribution between pilots and data symbols. The complexity analysis of the proposed solution as well  
157 as the comparison of this complexity with the state-of-the-art methods are presented in Section 5. The  
158 performance of the designed scheme is evaluated in Section 6 by various experiments. Finally, conclusions  
159 are given in Section 7.

160 **Notations:** We denote by  $a$ ,  $\mathbf{a}$  and  $\mathbf{A}$ , a scalar, a column vector and a matrix, respectively. Symbols  
161  $\otimes$  and  $\odot$  denote the Kronecker and the Hadamard products, respectively. Operators  $\text{vec}(\cdot)$  and  $\text{vec}_{m,n}^{-1}(\cdot)$   
162 designate the column vectorization of an  $m \times n$  matrix into an  $mn \times 1$  vector and the invectorization of an  
163  $mn \times 1$  vector to an  $m \times n$  matrix, respectively.  $\delta(\cdot)$  is the Dirac-delta function and  $\mathbb{E}\{\cdot\}$  is the expectation  
164 operation. The notation  $\text{diag}\{d_1, d_2, \dots, d_N\}$  denotes an  $N \times N$  diagonal matrix with entry  $(i, i)$  equal to  $d_i$ ,  
165 and the notation  $\text{blkdiag}\{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_N\}$  denotes a block diagonal matrix with  $\mathbf{B}_i$  the  $i^{\text{th}}$  block diagonal

166 entry. The modulo- $M$  operator is given by  $[\cdot]_M$ . Hermitian transposition is denoted by the superscript  $(\cdot)^H$ .  
 167 Matrices  $\mathbf{I}_N$ ,  $\mathbf{F}_n$ , and  $\mathbf{F}_n^H$  represent the  $N \times N$  identity matrix, the  $n$ -point DFT matrix, and the  $n$ -point  
 168 IDFT matrix.

## 169 2. MIMO-OTFS system model

170 We consider a MIMO-OTFS system whose parameters are given in Table 1. Its block diagram is shown  
 171 in Fig. 1.

Table 1: Parameters of a MIMO-OTFS system.

Parameter	Physical signification
$M, N$	delay bins, Doppler bins
$N_t, N_r$	number of Tx and Rx antennas
$\mathbf{W}_t, \mathbf{W}_r$	transmit pulse matrix, receive pulse matrix
$f_c$	carrier frequency
$\Delta t(s), \Delta f(Hz)$	slot duration, subcarrier spacing
$k_\nu, l_\tau$	maximum Doppler tap, maximum delay tap
$N\Delta t, M\Delta f$	frame duration, frame bandwidth
$t_c$	coherence time

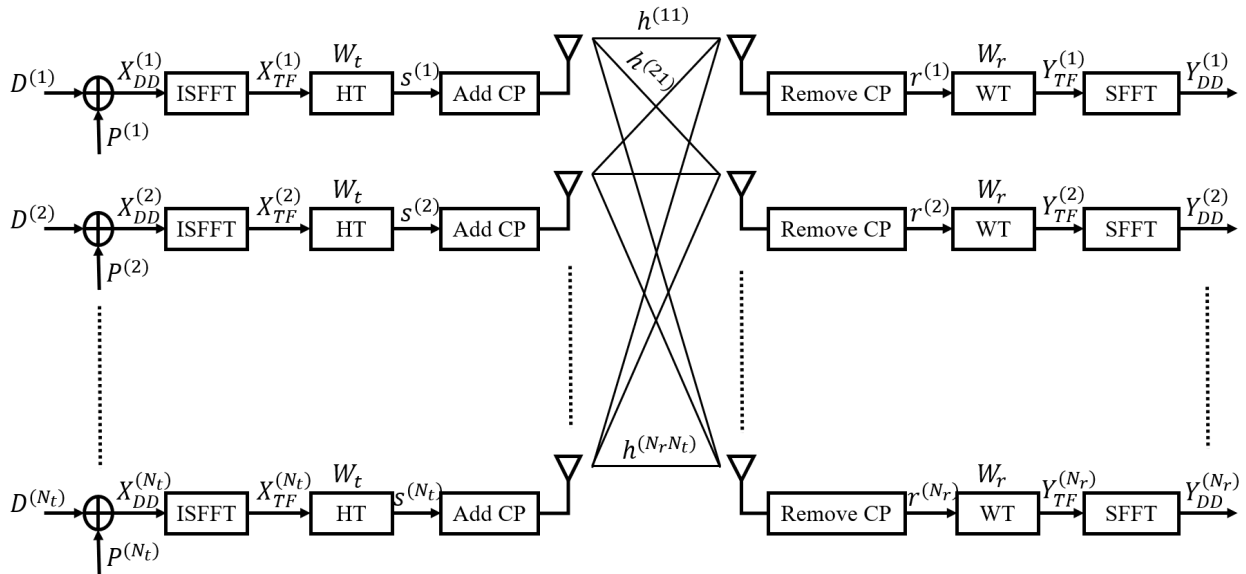


Figure 1: Block diagram of a MIMO-OTFS system: (I)SFFFT: (Inverse) symplectic finite Fourier transform, HT: Heisenberg transform, WT: Wigner transform, CP: Cyclic prefix, TF: Time-frequency.

172 At the transmitter side, the 2D grid of symbols in the DD domain  $\mathbf{X}_{DD}$  is formed, for each antenna, by  
 173 a superposition of data symbols and pilots as follows:



$$\begin{cases} \mathbf{X}_{DD}^{(1)} = \mathbf{D}^{(1)} + \mathbf{P}^{(1)}, \\ \vdots \\ \mathbf{X}_{DD}^{(N_t)} = \mathbf{D}^{(N_t)} + \mathbf{P}^{(N_t)}, \end{cases} \quad (1)$$

174 where  $\mathbf{D}^{(i)} \in \mathbb{C}^{M \times N}$  is the matrix formed by data symbols and  $\mathbf{P}^{(i)} \in \mathbb{C}^{M \times N}$  is the matrix formed by the  
175 pilots in the  $i$ -th branch.

176 The DD domain signal  $\mathbf{X}_{DD}$ , in each branch, is then transformed into a time-frequency domain signal  
177  $\mathbf{X}_{TF}$  via an ISFFT as follows:

$$\begin{cases} \mathbf{X}_{TF}^{(1)}(n, m) = \frac{1}{\sqrt{MN}} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} \mathbf{X}_{DD}^{(1)}(k, l) e^{j2\pi(\frac{nk}{N} - \frac{ml}{M})}, \\ \vdots \\ \mathbf{X}_{TF}^{(N_t)}(n, m) = \frac{1}{\sqrt{MN}} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} \mathbf{X}_{DD}^{(N_t)}(k, l) e^{j2\pi(\frac{nk}{N} - \frac{ml}{M})}. \end{cases} \quad (2)$$

178 Next,  $\mathbf{X}_{TF}$  is converted to waveform  $s(t)$  using the Heisenberg transform as

$$\begin{cases} \mathbf{s}^{(1)} = (\mathbf{F}_N^H \otimes \mathbf{W}_t) \mathbf{x}_{DD}^{(1)}, \\ \vdots \\ \mathbf{s}^{(N_t)} = (\mathbf{F}_N^H \otimes \mathbf{W}_t) \mathbf{x}_{DD}^{(N_t)}, \end{cases} \quad (3)$$

179 where  $\mathbf{s}^{(i)} = \text{vec}(\mathbf{S}^{(i)})$  with  $\mathbf{S}^{(i)} = \mathbf{W}_t \mathbf{F}_M^H (\mathbf{F}_M \mathbf{X}_{DD}^{(i)} \mathbf{F}_N^H) = \mathbf{W}_t \mathbf{X}_{DD}^{(i)} \mathbf{F}_N^H$ , and  $\mathbf{x}_{DD}^{(i)} = \text{vec}(\mathbf{X}_{DD}^{(i)})$ .  $\mathbf{W}_t$  is a  
180 pulse shaping matrix (for rectangular pulse  $\mathbf{W}_t = \mathbf{I}_M$ ). One CP is added to the signal  $s(t)$  in each branch  
181 before its transmission.

182 The MIMO wireless channel in the DD domain is sparse with few parameters. The channel between the  
183  $t$ -th transmitting antenna and the  $r$ -th receiving antenna has  $P$  taps. Thus, the baseband channel impulse  
184 response can be represented as

$$h_{rt}(\tau, \nu) = \sum_{i=1}^P h_i^{(rt)} \delta(\tau - \tau_i) \delta(\nu - \nu_i), \quad (4)$$

185 where  $h_i^{(rt)}$ ,  $\nu_i$ , and  $\tau_i$  are the complex channel gain, the Doppler shift, and the delay of the  $i$ -th path,  
186 respectively. The  $i$ -th delay and Doppler taps  $(l_i, k_i)$  can be written as  $l_i = \tau_i M \Delta f$ ,  $k_i = \nu_i N \Delta t$ .

187 We derive the linear system describing the input/output relations of the MIMO channel as

$$\begin{cases} \mathbf{r}^{(1)} = \mathbf{H}_{11}\mathbf{s}^{(1)} + \dots + \mathbf{H}_{1N_t}\mathbf{s}^{(N_t)} + \mathbf{n}^{(1)}, \\ \mathbf{r}^{(2)} = \mathbf{H}_{21}\mathbf{s}^{(1)} + \dots + \mathbf{H}_{2N_t}\mathbf{s}^{(N_t)} + \mathbf{n}^{(2)}, \\ \vdots \\ \mathbf{r}^{(N_r)} = \mathbf{H}_{N_r,1}\mathbf{s}^{(1)} + \dots + \mathbf{H}_{N_r,N_t}\mathbf{s}^{(N_t)} + \mathbf{n}^{(N_r)}, \end{cases} \quad (5)$$

188 where  $\mathbf{H}_{rt} = \sum_{i=1}^P h_i^{(rt)} \mathbf{\Pi}^{l_i} \mathbf{\Delta}^{k_i} \in \mathbb{C}^{MN \times MN}$  is the corresponding channel matrix between the  $t$ -th transmit-  
189 ting antenna and the  $r$ -th receiving antenna with  $\mathbf{\Pi} \in \mathbb{C}^{MN \times MN}$ , is the permutation matrix (forward cyclic  
190 shift):

$$\mathbf{\Pi} = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}. \quad (6)$$

191  $\mathbf{\Delta} = \text{diag}[\alpha^0, \alpha^1, \dots, \alpha^{MN-1}] \in \mathbb{C}^{MN \times MN}$ , where  $\alpha = e^{\frac{j2\pi}{MN}}$ .  $\mathbf{n}^{(i)} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2)$  is an additive complex  
192 Gaussian noise variable at the  $i$ -th receiving antenna.

193 Let  $\mathbf{r}_m = [(\mathbf{r}^{(1)})^T, (\mathbf{r}^{(2)})^T, \dots, (\mathbf{r}^{(N_r)})^T]^T$ ,  $\mathbf{s}_m = [(\mathbf{s}^{(1)})^T, (\mathbf{s}^{(2)})^T, \dots, (\mathbf{s}^{(N_t)})^T]^T$ ,  $\mathbf{n}_m = [(\mathbf{n}^{(1)})^T, (\mathbf{n}^{(2)})^T, \dots, (\mathbf{n}^{(N_r)})^T]^T$ ,  
194 and

$$\mathbf{H}_M = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \cdots & \mathbf{H}_{1N_t} \\ \mathbf{H}_{21} & \mathbf{H}_{22} & \cdots & \mathbf{H}_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N_r,1} & \mathbf{H}_{N_r,2} & \cdots & \mathbf{H}_{N_r,N_t} \end{pmatrix}, \quad (7)$$

195 Thus,

$$\mathbf{r}_m = \mathbf{H}_M \mathbf{s}_m + \mathbf{n}_m, \quad (8)$$

196 where  $\mathbf{r}_m, \mathbf{n}_m \in \mathbb{C}^{MNN_r \times 1}$ ,  $\mathbf{s}_m \in \mathbb{C}^{MNN_t \times 1}$ , and  $\mathbf{H}_M \in \mathbb{C}^{MNN_r \times MNN_t}$ .

197 The received signal in the DD domain, for the  $i$ -th antenna, is written in a vector form as  $\mathbf{y}_{DD}^{(i)} =$   
198  $(\mathbf{F}_N \otimes \mathbf{W}_r) \mathbf{r}^{(i)}$ , where  $\mathbf{y}_{DD}^{(i)} = \text{vec}(\mathbf{Y}_{DD}^{(i)})$ . Based on the fact that  $\mathbf{r}^{(i)} = \sum_{j=1}^{N_t} \mathbf{H}_{ij} \mathbf{s}^{(j)} + \mathbf{n}^{(i)}$  and  $\mathbf{s}^{(j)} =$   
199  $(\mathbf{F}_N^H \otimes \mathbf{W}_t) \mathbf{x}_{DD}^{(j)}$ , we get

$$\mathbf{y}_m = \mathcal{H} \mathbf{x}_m + \tilde{\mathbf{n}}, \quad (9)$$

200 where  $\mathbf{y}_m = [(\mathbf{y}_{DD}^{(1)})^T, (\mathbf{y}_{DD}^{(2)})^T, \dots, (\mathbf{y}_{DD}^{(N_r)})^T]^T$ ,  $\mathbf{x}_m = [(\mathbf{x}_{DD}^{(1)})^T, (\mathbf{x}_{DD}^{(2)})^T, \dots, (\mathbf{x}_{DD}^{(N_t)})^T]^T$ ,  $\tilde{\mathbf{n}} = [((\mathbf{F}_N \otimes \mathbf{W}_r)\mathbf{n}^{(1)})^T, ((\mathbf{F}_N \otimes$   
 201  $\mathbf{W}_r)\mathbf{n}^{(2)})^T, \dots, ((\mathbf{F}_N \otimes \mathbf{W}_r)\mathbf{n}^{(N_r)})^T]^T$ , and

$$\mathbf{H} = \begin{pmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} & \cdots & \mathcal{H}_{1N_t} \\ \mathcal{H}_{21} & \mathcal{H}_{22} & \cdots & \mathcal{H}_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{H}_{N_r,1} & \mathcal{H}_{N_r,2} & \cdots & \mathcal{H}_{N_r,N_t} \end{pmatrix}, \quad (10)$$

202 with

$$\begin{aligned} \mathcal{H}_{ij} &= (\mathbf{F}_N \otimes \mathbf{W}_r) \mathbf{H}_{ij} (\mathbf{F}_N^H \otimes \mathbf{W}_t), \\ &= (\mathbf{F}_N \otimes \mathbf{W}_r) \left( \sum_{p=1}^P h_p^{(ij)} \mathbf{\Pi}^{l_p} \mathbf{\Delta}^{k_p} \right) (\mathbf{F}_N^H \otimes \mathbf{W}_t), \\ &= \sum_{p=1}^P h_p^{(ij)} \mathbf{\Lambda}_p, \end{aligned} \quad (11)$$

203 where  $\mathbf{\Lambda}_p = (\mathbf{F}_N \otimes \mathbf{W}_r) \mathbf{\vartheta}_p (\mathbf{F}_N^H \otimes \mathbf{W}_t)$ , and  $\mathbf{\vartheta}_p = \mathbf{\Pi}^{l_p} \mathbf{\Delta}^{k_p}$ . Since  $(\mathbf{F}_N \otimes \mathbf{W}_r)$  is a unitary matrix and  
 204  $\mathbf{n}^{(i)} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{MN})$ ,  $\tilde{\mathbf{n}}$  and  $\mathbf{n}$  share the same distribution. The additive noise vector  $\tilde{\mathbf{n}}$  has a mean  
 205  $\mathbf{m}_{\tilde{\mathbf{n}}} = \mathbb{E}\{\tilde{\mathbf{n}}\} = \mathbf{0}_{MNN_r}$  and covariance matrix

$$\mathbf{C}_{\tilde{\mathbf{n}}} = \mathbb{E}\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\} = \sigma_n^2 \mathbf{I}_{MNN_r}. \quad (12)$$

206 The transmitted vector  $\mathbf{x}_m$  contains pilots and data symbols. It is written in the form  $\mathbf{x}_m = \mathbf{p} + \mathbf{d}$ ,  
 207 where  $\mathbf{p} = [(\mathbf{p}^{(1)})^T, (\mathbf{p}^{(2)})^T, \dots, (\mathbf{p}^{(N_t)})^T]^T$ , and  $\mathbf{d} = [(\mathbf{d}^{(1)})^T, (\mathbf{d}^{(2)})^T, \dots, (\mathbf{d}^{(N_t)})^T]^T$  with  $\mathbf{p}^{(i)} = \text{vec}(\mathbf{P}^{(i)})$  and  
 208  $\mathbf{d}^{(i)} = \text{vec}(\mathbf{D}^{(i)})$ . Thus, equation (9) can be written as follows:

$$\mathbf{y}_m = \mathcal{H}(\mathbf{p} + \mathbf{d}) + \tilde{\mathbf{n}} = \mathbf{y}_p + \mathbf{y}_d + \tilde{\mathbf{n}}, \quad (13)$$

209 where  $\mathbf{y}_p = \mathcal{H}\mathbf{p}$  and  $\mathbf{y}_d = \mathcal{H}\mathbf{d}$ . The term  $\mathbf{y}_p = \mathcal{H}\mathbf{p}$  can also be written as

$$\mathbf{y}_p = \mathbf{\Phi}_p \mathbf{h}_m, \quad (14)$$

210 where  $\mathbf{\Phi}_p = \text{blkdiag}(\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_{N_r}) \in \mathbb{C}^{MNN_r \times PN_tN_r}$  is a block diagonal matrix with  $\mathbf{B}_1 = \mathbf{B}_2 = \dots =$   
 211  $\mathbf{B}_{N_r} = [\mathbf{\Lambda}_1 \mathbf{p}^{(1)}, \dots, \mathbf{\Lambda}_P \mathbf{p}^{(1)}, \dots, \mathbf{\Lambda}_1 \mathbf{p}^{(N_t)}, \dots, \mathbf{\Lambda}_P \mathbf{p}^{(N_t)}]$ , and  $\mathbf{h}_m = [h_1^{(11)}, \dots, h_P^{(11)}, \dots, h_1^{(N_r N_t)}, \dots, h_P^{(N_r N_t)}]^T$  is an  
 212  $(N_t N_r P \times 1)$  vector formed by the channel coefficients of all paths.  $\mathbf{h}_m$  has zero mean and the following  
 213 covariance matrix:

$$\begin{aligned}
\mathbf{C}_{\mathbf{h}_m} &= \mathbb{E}\{\mathbf{h}_m \mathbf{h}_m^H\}, \\
&= \text{diag}\{\sigma_{h_1}^2, \dots, \sigma_{h_P}^2, \dots, \sigma_{h_{N_r N_t}}^2\}.
\end{aligned} \tag{15}$$

214 Assuming pilot symbols are chosen independently with uniform distribution, the matrix  $\Phi_p$  has a zero  
215 mean and a covariance matrix given as follows:

$$\mathbf{C}_{\Phi_p} = \mathbb{E}\{\Phi_p \Phi_p^H\} = N_t P \sigma_p^2 \mathbf{I}_{M N N_r}. \tag{16}$$

*Proof.* : Since  $\Phi_p = \text{blkdiag}(\mathbf{B}_1, \dots, \mathbf{B}_{N_r})$ , its covariance matrix can be expressed as follows:

$$\begin{aligned}
\mathbf{C}_{\Phi_p} &= \mathbb{E}\{\Phi_p \Phi_p^H\}, \\
&= \text{blkdiag}(\mathbb{E}\{\mathbf{B}_1 \mathbf{B}_1^H\}, \dots, \mathbb{E}\{\mathbf{B}_{N_r} \mathbf{B}_{N_r}^H\}).
\end{aligned} \tag{17}$$

216 Letting  $\mathbb{E}\{\mathbf{p}^{(n)} (\mathbf{p}^{(n)})^H\} = \sigma_p^2 \mathbf{I}_{MN}$ , the expression  $\mathbb{E}\{\mathbf{B}_i \mathbf{B}_i^H\}$  is expressed as follows:

$$\begin{aligned}
\mathbb{E}\{\mathbf{B}_i \mathbf{B}_i^H\} &= \sum_{n=1}^{N_t} \sum_{i=1}^P \mathbf{\Lambda}_i \mathbb{E}\{\mathbf{p}^{(n)} (\mathbf{p}^{(n)})^H\} \mathbf{\Lambda}_i^H, \\
&= \sigma_p^2 \sum_{n=1}^{N_t} \sum_{i=1}^P \mathbf{\Lambda}_i \mathbf{\Lambda}_i^H.
\end{aligned} \tag{18}$$

217 Since  $\mathbf{\Lambda}_i = (\mathbf{F}_N \otimes \mathbf{W}_r) \boldsymbol{\vartheta}_i (\mathbf{F}_N^H \otimes \mathbf{W}_t)$ ,  $\boldsymbol{\vartheta}_i = \mathbf{\Pi}^{l_i} \mathbf{\Delta}^{k_i}$ , and by taking  $\mathbf{G}_r = \mathbf{F}_N \otimes \mathbf{W}_r$  and  $\mathbf{G}_t = \mathbf{F}_N^H \otimes \mathbf{W}_t$ ,  
218 we have

$$\begin{aligned}
\mathbf{\Lambda}_i \mathbf{\Lambda}_i^H &= \mathbf{G}_r \boldsymbol{\vartheta}_i \mathbf{G}_t \mathbf{G}_t^H \boldsymbol{\vartheta}_i^H \mathbf{G}_r^H, \\
&= \mathbf{G}_r \boldsymbol{\vartheta}_i \boldsymbol{\vartheta}_i^H \mathbf{G}_r^H, \\
&= \mathbf{G}_r \mathbf{\Pi}^{l_i} \mathbf{\Delta}^{k_i} (\mathbf{\Delta}^{k_i})^H (\mathbf{\Pi}^{l_i})^H \mathbf{G}_r^H, \\
&= \mathbf{G}_r \mathbf{\Pi}^{l_i} (\mathbf{\Pi}^{l_i})^H \mathbf{G}_r^H,
\end{aligned} \tag{19}$$

219 Note that  $\mathbf{\Pi} \mathbf{\Pi}^H = \mathbf{I}_{MN}$ , that is  $\mathbf{\Pi}^H = \mathbf{\Pi}^{-1}$ . Then,

$$\mathbf{\Pi}^{l_i} (\mathbf{\Pi}^{l_i})^H = \mathbf{I}_{MN}, \tag{20}$$

220 and inserting (20) into (19) we get  $\mathbf{\Lambda}_i \mathbf{\Lambda}_i^H = \mathbf{G}_r \mathbf{G}_r^H = \mathbf{I}_{MN}$ . Then,

$$\mathbb{E}\{\mathbf{B}_i \mathbf{B}_i^H\} = N_t P \sigma_p^2 \mathbf{I}_{MN}, \tag{21}$$

221 leading thus to Eq. (16). □

222 Let  $\mathbf{h} = [(\mathbf{h}^{(11)})^T, \dots, (\mathbf{h}^{(21)})^T, \dots, (\mathbf{h}^{(N_r N_t)})^T]^T$  being the  $LN_t N_r$  sparse vector of channel, where  $\mathbf{h}^{(ij)} \in$   
223  $\mathbb{C}^L$  is a sparse vector of channel between the  $i$ -th Tx antenna and the  $j$ -th Rx antenna containing only  $P$

224 non-zero elements. Equation (14) can also take the following form:

$$\mathbf{y}_p = (\mathcal{K} \odot \Psi) \mathbf{h} = \mathbf{A} \mathbf{h}, \quad (22)$$

225 where  $\mathcal{K} \in \mathbb{C}^{MN N_r \times LN_t N_r}$  is the pilots matrix given as follows:

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}_{11} & \mathcal{K}_{12} & \cdots & \mathcal{K}_{1N_t} \\ \mathcal{K}_{21} & \mathcal{K}_{22} & \cdots & \mathcal{K}_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{K}_{N_r 1} & \mathcal{K}_{N_r 2} & \cdots & \mathcal{K}_{N_r N_t} \end{pmatrix}, \quad (23)$$

226 with  $L = (2k_\nu + 1)(l_\tau + 1)$ , and the elements of the sub-matrices  $\mathcal{K}_{ij}$  of the matrix  $\mathcal{K}$  are calculated as follows:

227  $\mathcal{K}_{ij}[i', l'(2k_\nu + 1) + k' + k_\nu] = \mathbf{P}^{(i)}[(k - k')_N, (l - l')_M]$ , for  $i' = 0 : MN$ ,  $l' \in [0, l_\tau]$  and  $k' \in [-k_\nu, k_\nu]$ .

228  $\Psi \in \mathbb{C}^{MN N_r \times LN_t N_r}$  is an additional phase shift matrix given by

$$\Psi = \begin{pmatrix} \Psi_{11} & \Psi_{12} & \cdots & \Psi_{1N_t} \\ \Psi_{21} & \Psi_{22} & \cdots & \Psi_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{N_r 1} & \Psi_{N_r 2} & \cdots & \Psi_{N_r N_t} \end{pmatrix}, \quad (24)$$

229 where the elements of the sub-matrices  $\Psi_{ij}$  of the matrix  $\Psi$  are calculated as follows:  $\Psi_{ij}[i', l'(2k_\nu + 1) +$

230  $k' + k_\nu] = \exp(k'(l - l')/MN)$ .

231 Therefore, equation (13) can take the two following forms:

$$\mathbf{y}_m = \mathcal{H} \mathbf{d} + \Phi_p \mathbf{h}_m + \tilde{\mathbf{n}}. \quad (25)$$

$$\mathbf{y}_m = \mathcal{H} \mathbf{d} + \mathbf{A} \mathbf{h} + \tilde{\mathbf{n}}. \quad (26)$$

### 232 3. Proposed algorithms

233 In this section, we detail the proposed algorithm for channel estimation and symbol detection. The

234 first step in the proposed algorithm is the estimation of the number of channel paths  $P$  and the delay and

235 Doppler taps  $\{l_i, k_i\}_{i=1:P}$ . Since  $\{l_i, k_i\}_{i=1:P}$  remain unchanged for a period  $T_s$ , for a doubly-underspread

236 (DU) channel [28], the estimation of these parameters is done once every  $T_s = N_T T$  seconds, where  $N_T$  is

237 the number of frames in the period  $T_s$  and  $T$  is the frame duration. Only the channel gains  $\{h_i\}_{i=1:N_t N_r P}$

238 will be estimated in each OTFS frame. For this purpose, we propose the frame architecture shown in Fig 2.  
 239 In each branch of the MIMO-OTFS system, the first frame ( $F_{T_1}$ ) is used for channel estimation including the  
 240 number of channel paths  $P$ , the delay and Doppler taps  $\{l_i, k_i\}_{i=1:P}$ , and the channel gains  $\{h_i\}_{i=1:N_t N_r P}$ .  
 241 In each of the other frames ( $F_{T_2}, \dots, F_{T_N}$ ), only the estimation of the channel gains is considered.

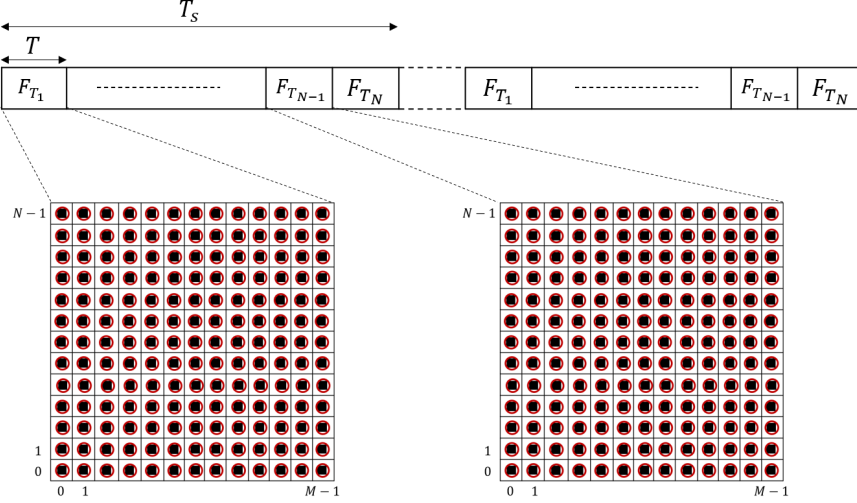


Figure 2: Frame architecture (■: data, ○: pilot).

### 242 3.1. Iterative SoBaP-AMPA algorithm for channel estimation and data detection

243 This algorithm operates in the  $F_{T_1}$  transmitted frame and iterates between data-aided channel estimation  
 244 and AMPA-assisted data detection.

#### 245 3.1.1. Data-aided channel estimation (SoBaP algorithm)

246 The model (26) can be rewritten as follows:

$$\mathbf{y}_m = \sum_{i=1}^{LN_t N_r} s_i c_i \mathbf{a}_i + \tilde{\mathbf{v}}_d, \quad (27)$$

where  $\mathbf{a}_i$  is the  $i$ -th column of  $\mathbf{A}$ ,  $\mathbf{h} = \mathbf{s} \odot \mathbf{c} = [s_1 c_1, s_2 c_2, \dots, s_{LN_t N_r} c_{LN_t N_r}]^T$ , where  $\mathbf{s} \in \{0, 1\}^{LN_t N_r}$  is the support vector of channel and  $\mathbf{c}$  denotes the vector of channel gains.  $\tilde{\mathbf{v}}_d = \mathbf{H}\mathbf{d} + \tilde{\mathbf{n}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\tilde{\mathbf{v}}_d})$  (see proof) is a mixture involving the additive noise, the channel contributions and the data symbols.  $\tilde{\mathbf{v}}_d$  can also take the form  $\tilde{\mathbf{v}}_d = \mathbf{\Phi}_d \mathbf{h}_m + \tilde{\mathbf{n}}$ , where  $\mathbf{\Phi}_d = \text{blkdiag}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_{N_r}) \in \mathbb{C}^{MN N_r \times PN_t N_r}$  with  $\mathbf{C}_1 = \mathbf{C}_2 = \dots = \mathbf{C}_{N_r} = [\mathbf{\Lambda}_1 \mathbf{d}^{(1)}, \dots, \mathbf{\Lambda}_P \mathbf{d}^{(1)}, \dots, \mathbf{\Lambda}_1 \mathbf{d}^{(N_t)}, \dots, \mathbf{\Lambda}_P \mathbf{d}^{(N_t)}]$ . In the same way as  $\mathbf{\Phi}_p$ , one can easily verify that  $\mathbf{\Phi}_d$  has a zero mean and a covariance matrix given as follows:

$$\mathbf{C}_{\mathbf{\Phi}_d} = \mathbb{E}\{\mathbf{\Phi}_d \mathbf{\Phi}_d^H\} = N_t P \sigma_d^2 \mathbf{I}_{MN N_r}. \quad (28)$$

*Proof.* : Since  $\mathbf{m}_{\mathbf{h}_m} = \mathbb{E}\{\mathbf{h}_m\} = \mathbf{0}_{PN_tN_r}$  and  $\mathbf{m}_{\tilde{\mathbf{n}}} = \mathbb{E}\{\tilde{\mathbf{n}}\} = \mathbf{0}_{MNN_r}$ , the mean of  $\tilde{\mathbf{v}}_d$  is

$$\mathbf{m}_{\tilde{\mathbf{v}}_d} = \mathbb{E}\{\tilde{\mathbf{v}}_d\} = \mathbf{0}_{MNN_r}, \quad (29)$$

and its covariance matrix is calculated as follows:

$$\begin{aligned} \mathbf{C}_{\tilde{\mathbf{v}}_d} &= \mathbb{E}\{\tilde{\mathbf{v}}_d\tilde{\mathbf{v}}_d^H\}, \\ &= \mathbb{E}\{\Phi_d\mathbf{h}_m\mathbf{h}_m^H\Phi_d^H\} + \mathbb{E}\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\}. \end{aligned} \quad (30)$$

247 The expression  $\mathbb{E}\{\Phi_d\mathbf{h}_m\mathbf{h}_m^H\Phi_d^H\}$  can be evaluated based on the following property [29]: if  $\mathbf{X}$  is a  $K \times L$   
 248 random matrix that satisfies  $\mathbb{E}\{\mathbf{X}\mathbf{X}^H\} = \sigma_x^2\mathbf{I}_K$ , then for any  $L \times L$  hermitian matrix  $\mathbf{Y}$ ,  $\mathbb{E}\{\mathbf{X}\mathbf{Y}\mathbf{X}^H\} =$   
 249  $\frac{\text{Tr}(\mathbf{Y})}{L}\mathbb{E}\{\mathbf{X}\mathbf{X}^H\}$ . Therefore,

$$\begin{aligned} \mathbb{E}\{\Phi_d\mathbf{h}_m\mathbf{h}_m^H\Phi_d^H\} &= \frac{\text{Tr}(\mathbf{h}_m\mathbf{h}_m^H)}{PN_tN_r}\mathbb{E}\{\Phi_d\Phi_d^H\}, \\ &= \left(\frac{\sum_{i=1}^{PN_tN_r}\sigma_{h_i}^2}{PN_tN_r}\right)\mathbb{E}\{\Phi_d\Phi_d^H\}. \end{aligned} \quad (31)$$

250 By replacing the expressions (12), (28) and (31) in (30) we get

$$\mathbf{C}_{\tilde{\mathbf{v}}_d} = \left(\sigma_n^2 + \frac{\sigma_d^2}{N_r} \sum_{i=1}^{PN_tN_r} \sigma_{h_i}^2\right)\mathbf{I}_{MNN_r}. \quad (32)$$

251

□

252 Letting  $\mathbf{c}_s$  the  $(PN_tN_r \times 1)$  vector and  $\mathbf{A}_s$  the  $(MNN_r \times PN_tN_r)$  matrix constructed, respectively, from  
 253 the elements of  $\mathbf{c}$  and  $\mathbf{A}$  considering the indices  $i$  where  $s_i \neq 0$ . Thus,

$$p(\mathbf{y}_m|\mathbf{c}, \mathbf{s}) = \mathcal{CN}(\mathbf{A}_s\mathbf{c}_s, \mathbf{C}_{\tilde{\mathbf{v}}_d}). \quad (33)$$

254 The inputs of the vector  $\mathbf{h}$  are modeled by a Bernoulli-Gaussian model. This model allows to take into  
 255 account the sparsity of the vector  $\mathbf{h}$ . We suppose that  $\mathbf{c}$  follows the probabilistic model given by

$$p(\mathbf{c}|\mathbf{s}) = \prod_{i=1}^{LN_tN_r} p(c_i|s_i), \quad (34)$$

256 where  $p(c_i|s_i) = \mathcal{CN}(0, \sigma_{s_i}^2)$  and  $\sigma_1^2 \gg \sigma_0^2$ .

257 With the model (33), (34), the vector  $\mathbf{y}_m$  can be seen as a noisy mixture of atoms specified by the  
 258 support  $\mathbf{s}$ . It is worth noting that the variances  $\sigma_{c_i}^2$  of Gaussian distributions are independent of the support  
 259  $\mathbf{s}$ . It is also to be noted that, to detect the location of spikes, the variables  $s_i$  corresponding to the inputs

260 of  $\mathbf{h}$  are assumed to be independent Bernoulli random variables ( $s_i = 1$  if a spike is present at  $h_i$  and  
 261  $s_i = 0$  otherwise). The unstructured sparsity treated here can be modelled using a standard product of the  
 262 Bernoulli distributions:

$$p(\mathbf{s}) = \prod_{i=1}^{LN_t N_r} p(s_i), \quad (35)$$

263 where  $p(s_i) = \text{Ber}(p_i)$  and  $p_i = p(b_i = 1) = 1 - p(b_i = 0)$ .

264 Furthermore, in modern approaches, instead of trying to approximate the means of the posterior distri-  
 265 butions from MCMC simulations, one tries to iteratively compute an exact variational approximation of the  
 266 target posterior distribution.

267 For the sparse channel estimation problem that we would like to solve in this work, we propose a MAP-  
 268 estimator, which correspond to the optimal Bayesian estimator using Bayesian cost. Therefore, the first  
 269 approach to solve the channel estimation problem consists in solving the following joint MAP problem:

$$(\hat{\mathbf{s}}, \hat{\mathbf{c}}) = \arg \max_{\mathbf{s}, \mathbf{c}} \log p(\mathbf{s}, \mathbf{c} | \mathbf{y}_m). \quad (36)$$

270 It has been shown in [30] that the solution set of the joint MAP problem (36) is the same as the standard  
 271 sparse recovery problem using the Bernoulli-Gaussian model (34), (35). This close connection motivates the  
 272 use of the Bernoulli-Gaussian model in the channel estimation problem treated in this work.

273 More specifically, we are interested here in the determination of channel support  $\mathbf{s}$  using the MAP  
 274 criterion. The decision minimizing the decision error probability on support  $\mathbf{s}$  is given by:

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \mathcal{A}} \log p(\mathbf{s} | \mathbf{y}_m), \quad (37)$$

275 where  $\mathcal{A} = \{0, 1\}^{LN_t N_r}$ .

276 The evaluation of  $\log p(\mathbf{s} | \mathbf{y}_m)$  for all  $2^{\text{card}(\mathcal{A})}$  sequences of  $\mathbf{s} \in \mathcal{A}$  is required for solving problem (37).  
 277 This makes problem (37) complex.

278 In order to reduce this complexity, individual decision on each input of the support vector  $\mathbf{s}$  is classically  
 279 considered. This decision is taken from a marginalized MAP estimation problem, conducting to

$$\hat{s}_i = \arg \max_{s_i \in \{0, 1\}} \log p(s_i | \mathbf{y}_m). \quad (38)$$

280 Even if the problem (38) seems easy to solve because the search space contains only two elements ( $s_i \in$   
 281  $\{0, 1\}$ ), the evaluation of  $p(s_i | \mathbf{y}_m)$  remains intractable due to the costly marginalization of  $p(\mathbf{s} | \mathbf{y}_m)$  over



282 the  $s_j$ 's, for  $j \neq i$ . To solve this problem, we compute a tractable surrogate  $q(s_i)$  of  $p(s_i|\mathbf{y}_m)$ . The  
 283 procedure adopted here allows to compute an approximation  $q(s_i)$  of  $p(s_i|\mathbf{y}_m)$  and is named the mean-field  
 284 approximation. It is summarized in Appendix A. In this case, problem (38) becomes

$$\hat{s}_i = \arg \max_{s_i \in \{0,1\}} \log q(s_i). \quad (39)$$

285 Based on the mean-field approximation, problem (39) can be solved by a threshold method as follows:

$$\hat{s}_i = \begin{cases} 1, & \text{if } q(s_i = 1) > \rho, \\ 0, & \text{otherwise,} \end{cases} \quad (40)$$

286 where  $\rho = 0.5$ , which is a value that minimizes the Bayes risk when uniform and equal costs are selected  
 287 [31, Section II.B]. Once the support vector  $\mathbf{s}$  is estimated, the vector of channel gains  $\mathbf{c}$  can be estimated  
 288 by MAP estimate  $\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} \log p(\mathbf{c}|\hat{\mathbf{s}}, \mathbf{y}_m)$ :

$$\begin{aligned} \hat{\mathbf{c}}_{\hat{\mathbf{s}}} &= (\mathbf{A}_{\hat{\mathbf{s}}}^T \mathbf{A}_{\hat{\mathbf{s}}} + \mathbf{\Delta})^{-1} \mathbf{A}_{\hat{\mathbf{s}}}^T \mathbf{y}_m, \\ \text{and } \hat{c}_i &= 0 \quad \text{if } s_i = 0, \end{aligned} \quad (41)$$

289 where  $\mathbf{\Delta} = \text{diag}[\sigma^2/\sigma_{c_1}^2, \sigma^2/\sigma_{c_2}^2, \dots, \sigma^2/\sigma_{c_{LN_t N_r}}^2]$ .  $\mathbf{A}_{\hat{\mathbf{s}}}$  and  $\hat{\mathbf{c}}_{\hat{\mathbf{s}}}$  denote the corresponding columns of  $\mathbf{A}$  and the  
 290 entries of  $\mathbf{c}$  limited to  $\hat{\mathbf{s}}$ , respectively. This solution reduces to the least-square solution when  $\sigma^2 \ll \sigma_{c_i}^2$   
 291 and to matched filtering when  $\sigma^2$  is large. The proposed algorithm for channel estimation is summarized in  
 292 Algorithm 1.

---

**Algorithm 1** Channel estimation algorithm

---

**Require:**  $\mathbf{y}_m \in \mathbb{C}^{MNN_r}$ ,  $\mathbf{A} \in \mathbb{C}^{MNN_r \times LN_t N_r}$

**Ensure:**  $\hat{\mathbf{s}} \in \{0, 1\}^{LN_t N_r}$ ,  $\hat{\mathbf{c}} \in \mathbb{C}^{LN_t N_r \times 1}$

- 1:  $p(\mathbf{s}) = \prod_k p(s_k)$
  - 2: prior mean for  $\mathbf{c}$ :  $\mathbf{m} = \mathbf{0}_{LN_t N_r}$
  - 3: probability  $\mathbf{q}$ :  $\mathbf{q}^{(0)} \sim (\mathcal{U}_{[0,1]})_{1:LN_t N_r}$
  - 4:  $\mathbf{r}$ :  $\mathbf{r}^{(0)} = \mathbf{y}_m - \mathbf{A}(\mathbf{s} \odot \mathbf{m})$
  - 5: **while**  $k \leq K$  and  $|q(s_i^{(k)}) - q(s_i^{(k-1)})| < \epsilon$  **do**
  - 6:   **for**  $l = 1 : LN_t N_r$  **do**
  - 7:      $\Sigma(s_k|\mathbf{y}_m) = \frac{\sigma_{c_k}^2 \sigma^2}{\sigma^2 + s_k \sigma_{c_k}^2 \mathbf{A}_k^T \mathbf{A}_k}$
  - 8:      $m(s_l|\mathbf{y}_m)^{(k)} = s_l \frac{\sigma_{c_l}^2}{\sigma^2 + s_l \sigma_{c_l}^2 \mathbf{A}_l^T \mathbf{A}_l} \mathbf{r}_l^T \mathbf{A}_l$
  - 9:      $q_l^{(k)} = q(s_l|\mathbf{y}_m)^{(k)} \propto \sqrt{\Sigma(s_l|\mathbf{y}_m)} e^{\left(\frac{1}{2} \frac{m(s_l|\mathbf{y}_m)^2}{\Sigma(s_l|\mathbf{y}_m)}\right)} p(s_l)$
  - 10:      $\mathbf{r}^{(k)}: \mathbf{r}^{(k)} = \mathbf{r}^{(k)} - \mathbf{A}_l(s_l^{(k)}) m(s_l|\mathbf{y}_m)^{(k)}$
  - 11:   **end for**
  - 12: **end while**
  - 13: estimate support  $\mathbf{s}$ :  $\hat{\mathbf{s}} = (\mathbf{q} > 0.5)$
  - 14: estimate  $\mathbf{c}$  conditional to  $\hat{\mathbf{s}}$ :  $\hat{\mathbf{c}}_{\hat{\mathbf{s}}} = (\mathbf{A}_{\hat{\mathbf{s}}}^T \mathbf{A}_{\hat{\mathbf{s}}} + \mathbf{\Delta})^{-1} \mathbf{A}_{\hat{\mathbf{s}}}^T \mathbf{y}_m$  and  $\hat{c}_k = 0$  if  $s_k = 0$
-

293 3.1.2. AMPA-aided data detection

294 Once the channel estimation is done, the obtained results, namely the channel coefficients as well as the  
 295 delay and Doppler taps will be used to form the following model allowing the detection of data symbols:

$$\mathbf{y}_d = \mathbf{y}_m - \mathbf{A}\mathbf{h} = \hat{\mathcal{H}}\mathbf{d} + \tilde{\mathbf{w}}_e, \quad (42)$$

296 where  $\tilde{\mathbf{w}}_e = \mathbf{A}(\mathbf{h} - \hat{\mathbf{h}}) + \tilde{\mathbf{n}}$  is formed by the additive noise and the channel estimation error  $\mathbf{h} - \hat{\mathbf{h}}$ . The mean  
 297  $m_{\tilde{w}_e(k)}$  and variance  $var_{\tilde{w}_e(k)}$  of the  $k$ -th element of  $\tilde{\mathbf{w}}_e$  are given as follows:

$$m_{\tilde{w}_e(k)} = \mathbb{E}\{\tilde{w}_e(k)\} = 0, \quad (43)$$

$$var_{\tilde{w}_e(k)} = \sigma_n^2 + \sigma_p^2 M_h, \quad (44)$$

298 where  $M_h = \mathbb{E}\{\|\mathbf{h} - \hat{\mathbf{h}}\|^2\}$  is the MSE of channel estimation.

299 *Proof.* : Letting  $\mathbf{e}_h = \mathbf{A}(\mathbf{h} - \hat{\mathbf{h}})$ , then  $\tilde{\mathbf{w}}_e = \mathbf{e}_h + \tilde{\mathbf{n}}$ . One can easily verify that the  $(l + kM)$ -th element of  
 300 the vector  $\mathbf{e}_h$  can be written as follows:

$$e_h(l + kM) = \sum_{j=1}^{LN_t N_r} A[l + kM, j](h(j) - \hat{h}(j)), \quad (45)$$

301 where  $A[l + kM, j] = \beta(l + kM, j)\mathbf{P}^{(i)}[(k - k_j)_N, (l - l_j)_M]$  and  $\beta_j(l + kM)$  is an additional shift phase given  
 302 as follows:

$$\beta(p, q) = \begin{cases} e^{-j2\pi \frac{p}{N}} e^{j2\pi \frac{k_i((m-l_q)_M)}{MN}}, & \text{if } p = \lambda_q, \text{ and } m < l_q \\ e^{j2\pi \frac{k_q((m-l_q)_M)}{MN}}, & \text{if } p = \lambda_q, \text{ and } m \geq l_q \\ 0, & \text{otherwise,} \end{cases} \quad (46)$$

303 where  $\lambda_q = [M - l_q]_M + M[n - k_q]_N$ .

304 Because all pilot symbols have a zero mean and an equal power of  $\sigma_p^2$ , the mean and the variance of  
 305  $e_h(l + kM)$  are given as follows:

$$m_{e_h(l+kM)} = \mathbb{E}\{e_h(l + kM)\} = 0, \quad (47)$$

and

$$\begin{aligned} var\{e_h(l + kM)\} &= \sigma_p^2 \sum_{i=1}^{LN_t N_r} \mathbb{E}\{|h(i) - \hat{h}(i)|^2\}, \\ &= \sigma_p^2 \mathbb{E}\{\|\mathbf{h} - \hat{\mathbf{h}}\|^2\} = \sigma_p^2 M_h. \end{aligned} \quad (48)$$

306 Finally, since  $e_h$  and  $\tilde{\mathbf{n}}$  are statistically independent and based on (12), (47) and (48), the mean and the  
 307 variance of  $\tilde{w}_e(k)$  are calculated as follows:

$$m_{\tilde{w}_e(k)} = \mathbb{E}\{e_h(k)\} + \mathbb{E}\{\tilde{\mathbf{n}}(k)\} = 0, \quad (49)$$

and

$$\text{var}\{\tilde{w}_e(k)\} = \text{var}\{e_h(k)\} + \text{var}\{\tilde{\mathbf{n}}(k)\} = \sigma_n^2 + \sigma_p^2 M_h, \quad (50)$$

308 where  $M_h = \mathbb{E}\{\|\mathbf{h} - \hat{\mathbf{h}}\|^2\}$ .

309 □

310 The aim here is to estimate the data symbol vector  $\mathbf{d}$  from  $\mathbf{y}_d$ ,  $\hat{\mathcal{H}}$ ,  $m_{\tilde{w}_e(k)}$  and  $\text{var}_{\tilde{w}_e(k)}$ . For this purpose,  
 311 we adapt the low-complexity MP algorithm proposed in [32], which is suitable for uncoded OTFS, taking  
 312 advantage of the channel sparsity in the DD domain.

313 Based on (42), and by observing that  $\hat{\mathcal{H}}$  is sparse: each row of  $\hat{\mathcal{H}}$  contains only  $PN_t$  non-zero elements  
 314 and each column of  $\hat{\mathcal{H}}$  contains only  $PN_r$  non-zero elements, the system is modelled as a sparsely connected  
 315 factor graph with  $MNN_t$  variable nodes and  $MNN_r$  observation nodes corresponding, respectively, to  $\mathbf{d}$   
 316 and  $\mathbf{y}_d$ . The factor graph of the MP algorithm, which consists of observation nodes and variable nodes, is  
 317 shown in Fig 3.

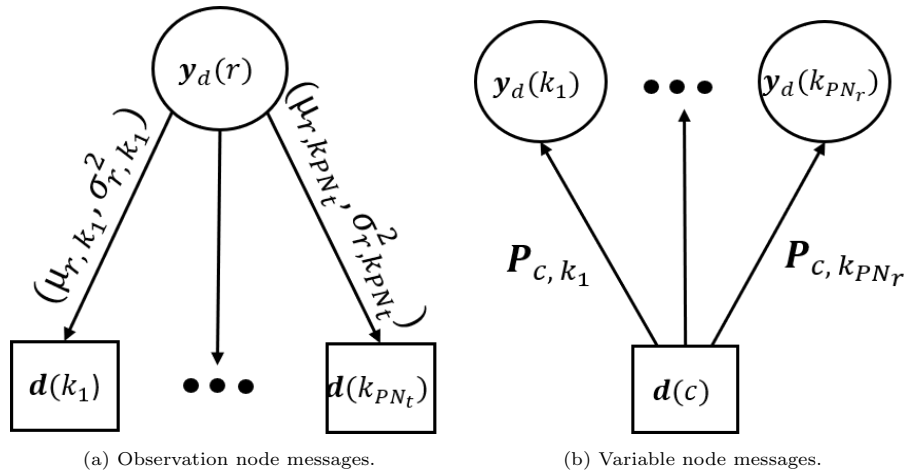


Figure 3: Messages in factor graph.

318 Thanks to the channel sparsity in the DD domain, each variable node  $\mathbf{d}(c)$  is connected to only  $PN_r$   
 319 observation nodes  $\{\mathbf{y}_d(k_i), k_i \in \mathcal{I}_c\}$ , and each observation node  $\mathbf{y}_d(r)$  is connected to only  $PN_t$  variable  
 320 nodes  $\{\mathbf{d}(k_i), k_i \in \mathcal{I}_r\}$ , where  $\mathcal{I}_c$  and  $\mathcal{I}_r$  denote, respectively, the sets of non-zero indices in the  $c$ -th column  
 321 and the  $r$ -th row of  $\hat{\mathcal{H}}$ .

322 The application of the MAP criterion on (42) gives:

$$\hat{\mathbf{d}} = \arg \max_{\mathbf{d} \in \mathbb{A}^{NMN_r}} p(\mathbf{d} | \mathbf{y}_d, \hat{\mathcal{H}}), \quad (51)$$

323 where  $\mathbb{A}$  denote the employed finite modulation alphabet (e.g., QAM).

324 Problem (51) is demanding in terms of complexity. Therefore, the following symbol by symbol MAP  
325 criterion is used:

$$\begin{aligned} \hat{\mathbf{d}}(r) &= \arg \max_{a_j \in \mathbb{A}} p(\mathbf{d}(r) = a_j | \mathbf{y}_d, \hat{\mathcal{H}}), \\ &= \arg \max_{a_j \in \mathbb{A}} \frac{1}{|\mathbb{A}|} p(\mathbf{y}_d | \mathbf{d}(r) = a_j, \hat{\mathcal{H}}), \\ &\approx \arg \max_{a_j \in \mathbb{A}} \prod_{k \in \mathcal{I}_r} p(\mathbf{y}_d(k) | \mathbf{d}(r) = a_j, \hat{\mathcal{H}}), \end{aligned} \quad (52)$$

326 for  $r = 0 : MN_t - 1$ .

327 It is worth noting that, thanks to the sparsity of the matrix  $\hat{\mathcal{H}}$ , the elements of vector  $\mathbf{y}_m$  are almost  
328 independent for a given  $\mathbf{d}(k)$ . It should also be noted that the entries of the vector of data symbols  $\mathbf{d}$   
329 are assumed to be equally likely. Problem (52) can be solved using the AMPA, which will be detailed  
330 below. The message that passes to the observation node  $\mathbf{y}_d(c)$ , for  $c \in \mathcal{I}_r$ , from the variable node  $\mathbf{d}(r)$ , for  
331  $r = 0 : MN_t - 1$  is the probability mass function (pmf) given as follows

$$\mathbf{p}_{rc} = \{p_{rc}(a_i) | a_i \in \mathbb{A}\}. \quad (53)$$

332 The relationship between the  $r$ -th observation node  $\mathbf{y}_d(r)$  and the  $c$ -th variable node  $\mathbf{d}(c)$  is given as  
333 follows:

$$\mathbf{y}_d(r) = \mathbf{d}(c) \hat{\mathcal{H}}(r, c) + \phi_{rc}, \quad (54)$$

334 where  $\hat{\mathcal{H}}(r, c)$  denotes the element in the  $r$ -th row and  $c$ -th column of  $\hat{\mathcal{H}}$ , and  $\phi_{rc}$  is the interference-plus-noise  
335 term expressed as follows:

$$\phi_{rc} = \sum_{k \in \mathcal{I}_r, k \neq c} \mathbf{d}(k) \hat{\mathcal{H}}(r, k) + \tilde{w}_e(r), \quad (55)$$

336 Using the central-limit theorem, the first term of equation (55) is approximated by a Gaussian distri-  
337 bution.  $\tilde{w}_e(r)$  follows also a Gaussian distribution. Therefore,  $\phi_{rc}$  is approximated to a Gaussian random  
338 variable.

339 In the  $t$ -th iteration of the AMPA, messages containing the mean  $\mu_{r,c}^{(t)}$  and the variance  $(\sigma_{r,c}^{(t)})^2$  of  $\phi_{rc}$   
 340 passes from the observation node  $\mathbf{y}_d(r)$  to the variable node  $\mathbf{d}(c)$ , where  $c \in \mathcal{I}_r$ . The major difference  
 341 between the versions of MP algorithm proposed in the literature [11, 32] and the AMPA proposed here lies  
 342 in the calculation of  $\mu_{r,c}^{(t)}$  and  $(\sigma_{r,c}^{(t)})^2$ . This difference is mainly due to the two following points:

- 343 • The previous versions of the MP algorithm assume perfect knowledge of CSI, i.e., the MIMO matrix  
 344  $\mathcal{H}$  is perfectly known. In our case, we do not know  $\mathcal{H}$  but we have an estimate  $\hat{\mathcal{H}}$ .
- 345
- 346 • Unlike previous versions where the interference term contains only additive noise, in our case, this  
 347 term is a function of additive noise  $\tilde{\mathbf{n}}$  and the term  $\mathbf{e}_h$  containing the symbols of pilots and the error  
 348 of channel estimation (42)-(44). This mixture is not present in previous versions of the MP algorithm  
 349 for OTFS due to the use of guard intervals between pilot symbols and data symbols, which decreases  
 350 the SE of the system.

351 Using the mean and variance of  $\tilde{w}_e(k)$  that we derived in (49) and (50), the mean  $\mu_{r,c}^{(t)}$  and the variance  
 352  $(\sigma_{r,c}^{(t)})^2$  of the interference term are given as follows:

$$\mu_{r,c}^{(t)} = \sum_{k \in \mathcal{I}_r, k \neq c} \sum_{i=1}^{|\mathbb{A}|} a_i p_{k,r}^{(t-1)}(a_i) \hat{\mathcal{H}}(r, k) + m_{\tilde{w}_e(k)}, \quad (56)$$

353 and

$$(\sigma_{r,c}^{(t)})^2 = \sum_{k \in \mathcal{I}_r, k \neq c} \sum_{i=1}^{|\mathbb{A}|} |a_i|^2 p_{k,r}^{(t-1)}(a_i) |\hat{\mathcal{H}}(r, k)|^2 - |\mu_{r,c}^{(t)}|^2 + \text{var}\{\tilde{w}_e(k)\}. \quad (57)$$

354 The next step in the AMPA is to calculate the entries of the pmf vector  $\mathbf{P}_{r,c}^{(t)}$  that passes as a message  
 355 from the variable node  $\mathbf{d}(c)$  to the observation node  $\mathbf{y}_d(r)$ . The elements of the pmf vector are calculated  
 356 using a damping method as in [11]:

$$p_{c,r}^{(t)}(a_i) = \Delta p_{c,r}^{(t-1)}(a_i) + (1 - \Delta) p_{c,r}^{(t-2)}(a_i), \quad (58)$$

357 where  $\Delta \in (0, 1]$  is the damping factor allowing the improvement of the convergence of the AMPA, and

$$p_{c,r}^{(t)} \propto \prod_{k \in \mathcal{I}_c, k \neq r} \exp\left(-\frac{|\mathbf{y}_d(k) + \mu_{k,c}^{(t)} + \hat{\mathcal{H}}(k, c) a_i|}{(\sigma_{k,c}^{(t)})^2}\right). \quad (59)$$

358 The AMPA stops if the maximum number of iteration  $n_{iter}$  is reached or if  $|p_{c,r}^{(t)}(a_i) - p_{c,r}^{(t-1)}(a_i)| < \epsilon$ ,  
 359 where  $\epsilon$  is a small value.

360 Finally, decision on the detected data symbols is given as follows:

$$\hat{\mathbf{d}}(c) = \arg \max_{a_i \in \mathbb{A}} p_c^{(t)}(a_i), \quad (60)$$

361 for  $c = 0 : NMN_t - 1$ . The AMPA is summarized in Algorithm 2.

---

**Algorithm 2** AMPA for MIMO-OTFS data detection

---

**Require:**  $\mathbf{y}_d, \hat{\mathcal{H}}, \tilde{w}_\epsilon(k): m_{\tilde{w}_\epsilon(k)}, var_{\tilde{w}_\epsilon(k)}, n_{iter}$

**Ensure:**  $\hat{\mathbf{d}}$

- 1: Iteration index  $t = 0$
  - 2: The probability mass function (pmf):  $\mathbf{p}_{rc}^{(0)} = \frac{1}{|\mathbb{A}|}$  for  $r = 0 : NMN_t - 1$  and  $c \in \mathcal{I}_r$
  - 3: **while**  $t \leq n_{iter}$  and  $|p_{c,r}^{(t)}(a_i) - p_{c,r}^{(t-1)}(a_i)| < \epsilon$  **do**
  - 4:   The mean  $\mu_{r,c}^{(t)}$  and the variance  $(\sigma_{r,c}^{(t)})^2$  of the interference term  $\phi_{rc}$  pass from the observation node  $\mathbf{y}_d(r)$  to the variable node  $\mathbf{d}(c)$
  - 5:   The pmf vector  $\mathbf{p}_{r,c}^{(t)}$  is updated and passes as a message from the variable node  $\mathbf{d}(c)$  to the observation node  $\mathbf{y}_d(r)$
  - 6:   The distribution of the transmitted symbols  $\mathbf{d}$  is updated
  - 7:    $t = t + 1$
  - 8: **end while**
- 

362 *3.2. Iterative LMMSE-AMPA algorithm for channel estimation and data detection*

363 As previously stated, since the number of paths  $P$ , delay and Doppler taps  $\{l_i, k_i\}_{i=1:P}$  remain un-  
 364 changed for a period of time  $T_s$ , only the channel gains vector  $\{h_i\}_{i=1:P}$  will be estimated in the frames  
 365  $F_{T_2}, F_{T_3}, \dots, F_{T_N}$ . In the same way as above, we propose an iterative algorithm for channel estimation and  
 366 data detection. For channel estimation, we propose a low complexity MMSE. For data symbol detection,  
 367 the AMAP is used.

368 *3.2.1. Data-aided channel estimation (LMMSE algorithm)*

369 Based on the fact that  $\mathbf{y}_d = \Phi_d \mathbf{h}_m$ , equation (25) can be written as follows:

$$\mathbf{y}_m = \Phi_d \mathbf{h}_m + \Phi_p \mathbf{h}_m + \tilde{\mathbf{n}}. \quad (61)$$

370 Letting  $\hat{\Phi}_d^{(0)}$  the first estimate of  $\Phi_d$ , the equation (61) is re-written as follows:

$$\begin{aligned} \mathbf{y}_m &= (\Phi_p + \hat{\Phi}_d^{(0)}) \mathbf{h}_m + (\Phi_d - \hat{\Phi}_d^{(0)}) \mathbf{h}_m + \tilde{\mathbf{n}}, \\ &= \Phi_{pd}^{(0)} \mathbf{h}_m + \boldsymbol{\mu}_d^{(0)}, \end{aligned} \quad (62)$$

371 where  $\boldsymbol{\mu}_d^{(i)} = (\Phi_d - \hat{\Phi}_d^{(i)}) \mathbf{h}_m + \tilde{\mathbf{n}}$  is the noise-plus-interference vector whose mean is given as follows:

$$\mathbf{m}_{\boldsymbol{\mu}_d^{(i)}}^{(i)} = \mathbb{E}\{\boldsymbol{\mu}_d^{(i)}\} = \mathbf{0}_{MN N_r}, \quad (63)$$

and in the same way as (32), its covariance matrix is derived as follows:

$$\begin{aligned} \mathbf{C}_{\boldsymbol{\mu}_d^{(i)}}^{(i)} &= \mathbb{E}\{\boldsymbol{\mu}_d^{(i)}(\boldsymbol{\mu}_d^{(i)})^H\}, \\ &= \left( \sigma_n^2 + \frac{2\sigma_d^2}{N_r} \sum_{i=1}^{PN_t N_r} \sigma_{h_i}^2 \right) \mathbf{I}_{MN N_r}. \end{aligned} \quad (64)$$

372 From (62) and based on (63), (64), the MMSE estimate of the channel vector  $\mathbf{h}_m$  in the  $i$ -th iteration is  
373 given as follows:

$$\hat{\mathbf{h}}_m^{(i)} = ((\boldsymbol{\Phi}_{pd}^{(i-1)})^H (\mathbf{C}_{\boldsymbol{\mu}_d^{(i-1)}}^{(i-1)})^{-1} \boldsymbol{\Phi}_{pd}^{(i-1)} + \mathbf{C}_{\mathbf{h}_m}^{-1})^{-1} (\boldsymbol{\Phi}_{pd}^{(i-1)})^H (\mathbf{C}_{\boldsymbol{\mu}_d^{(i-1)}}^{(i-1)})^{-1} \mathbf{y}_m. \quad (65)$$

374 It is worth noting that this estimator benefits from the channel sparsity in the DD domain by calculating  
375 the inverse of a  $(PN_t N_r \times PN_t N_r)$  matrix, where  $P \ll MN$ .

376 Equation (65) requires the inversion of the matrix  $\mathbf{M} = ((\boldsymbol{\Phi}_{pd}^{(i-1)})^H (\mathbf{C}_{\boldsymbol{\mu}_d^{(i-1)}}^{(i-1)})^{-1} \boldsymbol{\Phi}_{pd}^{(i-1)} + \mathbf{C}_{\mathbf{h}_m}^{-1})$ . As  $\mathbf{M}$   
377 is a bounded positive definite Hermitian matrix (see Appendix B) with a lower and upper bound  $PN_t$ , we  
378 propose a low complexity MMSE (LMMSE) by computing  $\mathbf{M}^{-1}$  via the Cholesky decomposition. Algorithm  
379 3 summarize the estimate of the channel vector  $\mathbf{h}_m$  in the  $i$ -th iteration.

---

**Algorithm 3** LMMSE algorithm for channel estimation

---

**Require:**  $\boldsymbol{\Phi}_{pd}^{(i-1)}$ ,  $(\mathbf{C}_{\boldsymbol{\mu}_d^{(i-1)}}^{(i-1)})^{-1}$ ,  $\mathbf{C}_{\mathbf{h}_m}^{-1}$ ,  $\mathbf{y}_m$

**Ensure:**  $\hat{\mathbf{h}}_m^{(i)}$

- 1: Compute the Hermitian bounded matrix  $\mathbf{M} = (\boldsymbol{\Phi}_{pd}^{(i-1)})^H (\mathbf{C}_{\boldsymbol{\mu}_d^{(i-1)}}^{(i-1)})^{-1} \boldsymbol{\Phi}_{pd}^{(i-1)} + \mathbf{C}_{\mathbf{h}_m}^{-1}$
  - 2: Compute the Cholesky decomposition of  $\mathbf{M} = \mathbf{L}\mathbf{D}\mathbf{L}^H$
  - 3: Solve the system  $\mathbf{L}\mathbf{s}_1 = (\boldsymbol{\Phi}_{pd}^{(i-1)})^H (\mathbf{C}_{\boldsymbol{\mu}_d^{(i-1)}}^{(i-1)})^{-1} \mathbf{y}_m$
  - 4: Solve the system  $\mathbf{D}\mathbf{s}_2 = \mathbf{s}_1$
  - 5: Solve the system  $\mathbf{L}^H \mathbf{s}_3 = \mathbf{s}_2$
  - 6:  $\hat{\mathbf{h}}_m^{(i)} = \mathbf{s}_3$
- 

### 3.2.2. AMPA-aided data detection

380 Once the channel is estimated, the AMPA is used for data detection. The observation vector  $\tilde{\mathbf{y}}_d^{(i)}$  at the  
381  $i$ -th iteration used for the symbol detection can be expressed as follows:

$$\tilde{\mathbf{y}}_d^{(i)} = \mathbf{y}_m - \hat{\mathcal{H}}^{(i)} \mathbf{p} = \hat{\mathcal{H}} \mathbf{d} + \tilde{\mathbf{w}}_{\tilde{\epsilon}}^{(i)}, \quad (66)$$

383 where  $\tilde{\mathbf{w}}_e^{(i)} = (\mathcal{H} - \hat{\mathcal{H}}^{(i)})\mathbf{p} + \tilde{\mathbf{n}}$ . In the same way as (43) and (44) we can easily show that the mean of the  
 384  $k$ -th element of  $\tilde{\mathbf{w}}_e^{(i)}$  is given by  $m_{\tilde{w}_e^{(i)}}(k) = 0$  and its variance is expressed as  $var_{\tilde{w}_e^{(i)}}(k) = \sigma_n^2 + \sigma_p^2 M_h^{(i)}$ .

385 The channel estimation and data detection algorithm designed in this work is outlined in Algorithm 4

---

**Algorithm 4** Channel estimation and data detection algorithm

---

**Require:**  $\mathbf{y}_m, \mathbf{A}, N_{iter}, \Phi_{pd}, (\mathbf{C}_{\mu_d})^{-1}, \mathbf{C}_{\mathbf{h}_m}^{-1}$ , frame index  $f_P$

**Ensure:**  $\hat{\mathbf{d}}, \hat{P}, \{\hat{P}_i, \hat{k}_i\}_{i=1:P}, \{\hat{h}_i\}_{i=1:\hat{P}N_tN_r}$

- 1: **while**  $n \leq N_{iter}$  &  $|\mathbf{h}_m^{(n)} - \mathbf{h}_m^{(n-1)}| < \epsilon$  **do**
  - 2:   **if**  $f_P == 1$  **then**
  - 3:     Use the SoBaP algorithm (Algorithm 1):  $\{\hat{P}, \{\hat{l}_i, \hat{k}_i\}_{i=1:\hat{P}}, \{\hat{h}_i\}_{i=1:\hat{P}N_tN_r}\} = \text{SoBaP}(\mathbf{y}_m, \mathbf{A})$
  - 4:   **else**
  - 5:     Use the LMMSE algorithm (Algorithm 3):  $\{\{\hat{h}_i\}_{i=1:\hat{P}N_tN_r}\} = \text{LMMSE}(\mathbf{y}_m, \Phi_{pd}, (\mathbf{C}_{\mu_d})^{-1}, \mathbf{C}_{\mathbf{h}_m}^{-1})$
  - 6:   **end if**
  - 7:   Compute:  $\mathbf{y}_d, \hat{\mathcal{H}}, \tilde{w}_e(k) : m_{\tilde{w}_e(k)}, var_{\tilde{w}_e(k)}$
  - 8:   Use the AMPA (Algorithm 2):  $\hat{\mathbf{d}} = \text{MAP}(\mathbf{y}_d, \hat{\mathcal{H}}, \tilde{w}_e(k) : m_{\tilde{w}_e(k)}, var_{\tilde{w}_e(k)})$
  - 9: **end while**
- 

386 **4. Optimal power distribution between pilots and data symbols**

387 In the proposed scheme, the pilots and data symbols are superimposed in the same DD locations. There-  
 388 fore, the power must be optimally distributed between the pilots and data symbols to maximise the SINR.  
 389 Consequently, the maximization of SINR leads to the maximization of SE and the minimization of BER  
 390 [33]. In what follows, we will derive the SINR lower bound which will be maximised to obtain the optimal  
 391 pilot power  $\sigma_{p,opt}^2$  and the optimal data symbol power  $\sigma_{d,opt}^2 = 1 - \sigma_{p,opt}^2$ . For the sake of simplicity, we will  
 392 consider one iteration of the proposed algorithm. Eq. (66) can be rewritten as follows:

$$\begin{aligned}
 \tilde{\mathbf{y}}_d &= \mathbf{y}_m - \Phi_p \hat{\mathbf{h}}_m, \\
 &= \Phi_d \mathbf{h}_m + \Phi_p (\mathbf{h}_m - \hat{\mathbf{h}}_m) + \tilde{\mathbf{n}}, \\
 &= \Phi_d \hat{\mathbf{h}}_m + (\Phi_d + \Phi_p) (\mathbf{h}_m - \hat{\mathbf{h}}_m) + \tilde{\mathbf{n}}, \\
 &= \Phi_d \hat{\mathbf{h}}_m + \tilde{\mathbf{w}}_e,
 \end{aligned} \tag{67}$$

393 where  $\tilde{\mathbf{w}}_e = (\Phi_d + \Phi_p) \tilde{\mathbf{h}}_m + \tilde{\mathbf{n}}$  and  $\tilde{\mathbf{h}}_m = (\mathbf{h}_m - \hat{\mathbf{h}}_m)$ .

394 Using (67), the  $[k, l]$ -th element of the received DD signal at the Rx  $j$ -th antenna can be expressed as  
 395 [22]



$$\begin{aligned}
Y_d^{(j)}[k, l] &= \sum_{i=1}^{N_t} \sum_{p=1}^P \hat{h}_{\beta,p}^{(ji)} D^{(i)}[[k - k_p]_N, [l - l_p]_M] + \sum_{i=1}^{N_t} \sum_{p=1}^P \tilde{h}_{\beta,p}^{(ji)} (D^{(i)}[[k - k_p]_N, [l - l_p]_M] \\
&\quad + P^{(i)}[[k - k_p]_N, [l - l_p]_M]) + \tilde{n}[k, l],
\end{aligned} \tag{68}$$

396 where  $\hat{h}_{\beta,p}^{(ji)} = \hat{h}_{m,p}^{(ji)} \beta[k, l]$  and  $\tilde{h}_{\beta,p}^{(ji)} = (\hat{h}_{m,p}^{(ji)} - \hat{h}_{m,p}^{(ji)}) \beta[k, l]$ .

397 Equation (68) can be rewritten as

$$\begin{aligned}
Y_d^{(j)}[k, l] &= \tilde{\mathbf{d}}^T \hat{\mathbf{h}}_\beta + (\tilde{\mathbf{d}} + \tilde{\mathbf{p}})^T \tilde{\mathbf{h}}_\beta + \tilde{n}[k, l], \\
&= \tilde{\mathbf{d}}^T \hat{\mathbf{h}}_\beta + v[k, l],
\end{aligned} \tag{69}$$

398 where  $v[k, l] = (\tilde{\mathbf{d}} + \tilde{\mathbf{p}})^T \tilde{\mathbf{h}}_\beta + \tilde{n}[k, l]$  is the noise-plus-interference term,  $\tilde{\mathbf{d}} \in \mathbb{C}^{PN_t \times 1}$ ,  $\tilde{\mathbf{p}} \in \mathbb{C}^{PN_t \times 1}$  denote  
399 data and pilot vectors. Their  $p$ -th element is given by  $D^{(i)}[[k - k_p]_N, [l - l_p]_M]$  and  $P^{(i)}[[k - k_p]_N, [l - l_p]_M]$ ,  
400 respectively. The  $p$ -th element of  $\tilde{\mathbf{h}}_\beta$  and  $\hat{\mathbf{h}}_\beta$  are the scalars  $\tilde{h}_{\beta,p}^{(ji)}$  and  $\hat{h}_{\beta,p}^{(ji)}$ , respectively.

401 Using (69), the SINR of the  $[k, l]$ -th symbol can be expressed as follows:

$$\text{SINR}[k, l] = \frac{\mathbb{E}\{|\tilde{\mathbf{d}}^T \hat{\mathbf{h}}_\beta|^2\}}{\mathbb{E}\{|v[k, l]|^2\}}. \tag{70}$$

402 The numerator of (70) depends on the channel estimate and its denominator depends on the estimation  
403 error. These two entities are independent because, for a linear MMSE estimator, the estimation error is  
404 orthogonal to the observations [34] and the noise is independent of data symbols. Since  $\mathbb{E}\{\tilde{\mathbf{d}}^* \tilde{\mathbf{d}}^T\} = \sigma_d^2 \mathbf{I}_{PN_t}$   
405 and  $\beta[k, l]$  is a phase factor, the numerator of (70) can be expressed as

$$\begin{aligned}
\mathbb{E}\{|\tilde{\mathbf{d}}^T \hat{\mathbf{h}}_\beta|^2\} &= \mathbb{E}\{(\hat{\mathbf{h}}_\beta)^H \mathbb{E}\{\tilde{\mathbf{d}}^* \tilde{\mathbf{d}}^T\} \hat{\mathbf{h}}_\beta\}, \\
&= \sigma_d^2 \mathbb{E}\{\|\hat{\mathbf{h}}_m\|^2\},
\end{aligned} \tag{71}$$

406 The term  $\mathbb{E}\{\|\hat{\mathbf{h}}_m\|^2\} = \text{Tr}(\mathbb{E}\{\hat{\mathbf{h}}_m (\hat{\mathbf{h}}_m)^H\})$  can be calculated using (67) as follows:

$$\text{Tr}(\mathbb{E}\{\hat{\mathbf{h}}_m (\hat{\mathbf{h}}_m)^H\}) = \sigma_h^2 - M_h, \tag{72}$$

407 where  $\sigma_h^2 = \text{Tr}(\mathbf{C}_{\mathbf{h}_m}) = \sum_{i=1}^{PN_t N_r} \sigma_{h_i}^2$  and  $M_h$  is the MSE of channel estimation defined in (50). From (71)  
408 and (72), the expression of the SINR numerator can be expressed as follows:

$$\mathbb{E}\{|\tilde{\mathbf{d}}^T \hat{\mathbf{h}}_\beta|^2\} = \sigma_d^2 (\sigma_h^2 - M_h). \tag{73}$$

409 The denominator of (70) can be simplified as follows:

$$\begin{aligned}
\mathbb{E}\{|v[k, l]|^2\} &= \mathbb{E}\{(\tilde{\mathbf{h}}_\beta)^H (\mathbb{E}\{\tilde{\mathbf{d}}^* \tilde{\mathbf{d}}^T\} + \tilde{\mathbf{p}}^* \tilde{\mathbf{p}}^T) \tilde{\mathbf{h}}_\beta\} + \mathbb{E}\{|\tilde{n}[k, l]|^2\}, \\
&= \sigma_d^2 \mathbb{E}\{\|\tilde{\mathbf{h}}_m\|^2\} + \text{Tr}(\mathbb{E}\{(\tilde{\mathbf{h}}_\beta \tilde{\mathbf{h}}_\beta)^H\}) + \tilde{\mathbf{p}}^* \tilde{\mathbf{p}}^T + \sigma_n^2.
\end{aligned} \tag{74}$$

410 The second term of (74) can be simplified using the following property: for two positive semi-definite  
411 matrices  $\mathbf{X} \in \mathbb{C}^{K \times K}$  and  $\mathbf{Y} \in \mathbb{C}^{K \times K}$ ,  $\text{Tr}(\mathbf{X}\mathbf{Y}) \leq \text{Tr}(\mathbf{X})\text{Tr}(\mathbf{Y})$  [35]:

$$\text{Tr}(\mathbb{E}\{(\tilde{\mathbf{h}}_\beta \tilde{\mathbf{h}}_\beta)^H\} \tilde{\mathbf{p}}^* \tilde{\mathbf{p}}^T) \leq \text{Tr}(\mathbb{E}\{(\tilde{\mathbf{h}}_\beta \tilde{\mathbf{h}}_\beta)^H\}) \text{Tr}(\tilde{\mathbf{p}}^* \tilde{\mathbf{p}}^T) = PN_t \sigma_p^2 M_h. \tag{75}$$

412 Using (75), (74) can be simplified as follows:

$$\mathbb{E}\{|v[k, l]|^2\} \leq \sigma_d^2 M_h + PN_t \sigma_p^2 M_h + \sigma_n^2. \tag{76}$$

413 By replacing (73) and (76) in (70), the SINR expression becomes

$$\text{SINR} \geq \frac{\sigma_d^2 (\sigma_h^2 - M_h)}{\sigma_d^2 M_h + PN_t \sigma_p^2 M_h + \sigma_n^2}. \tag{77}$$

414 To evaluate the expression of the SINR, it is necessary to calculate the MSE  $M_h$  for the superimposed-  
415 aided channel estimation. We show that  $M_h$  is lower bounded as

$$M_h \geq \frac{(PN_t N_r)^2 (\sigma_n^2 + 2\sigma_h^2 \sigma_d^2 / N_r)}{PMN N_t N_r \sigma_p^2 + \tilde{\tau}_h^2 (\sigma_n^2 + 2\sigma_h^2 \sigma_d^2 / N_r)}. \tag{78}$$

416 where  $\tilde{\tau}_h^2 = \sum_{i=1}^{PN_t N_r} \frac{1}{\sigma_{h_i}^2}$ .

*Proof.* : For any positive definite matrix  $\mathbf{X} \in \mathbb{C}^{K \times K}$ ,  $\text{Tr}(\mathbf{X}^{-1}) \geq \frac{K^2}{\text{Tr}(\mathbf{X})}$  [35]. Using this result,  $M_h = \text{Tr}((\mathbf{\Phi}_p)^H (\mathbf{C}_{\mu_d})^{-1} \mathbf{\Phi}_p + \mathbf{C}_{\mathbf{h}_m}^{-1})^{-1}$  can be expressed as follows:

$$M_h \geq \frac{(PN_t N_r)^2}{\text{Tr}((\mathbf{\Phi}_p)^H (\mathbf{C}_{\mu_d})^{-1} \mathbf{\Phi}_p + \mathbf{C}_{\mathbf{h}_m}^{-1})}. \tag{79}$$

Using (32) and the property used in (31) and letting  $\tilde{\tau}_h^2 = \sum_{i=1}^{PN_t N_r} \frac{1}{\sigma_{h_i}^2}$ , the denominator of (79) can be simplified as follows:

$$\text{Tr}((\mathbf{\Phi}_p)^H (\mathbf{C}_{\mu_d})^{-1} \mathbf{\Phi}_p + \mathbf{C}_{\mathbf{h}_m}^{-1}) = \frac{\text{Tr}((\mathbf{\Phi}_p)^H \mathbf{\Phi}_p)}{\sigma_n^2 + 2\sigma_h^2 \sigma_d^2 / N_r} + \tilde{\tau}_h^2. \tag{80}$$

We have, by using (17) and (21)

$$\text{Tr}((\mathbf{\Phi}_p)^H \mathbf{\Phi}_p) = PMN N_t N_r \sigma_p^2. \tag{81}$$

417 We get the desired result in (78) by substituting (81) in (80) then (80) in (79). □

418 By accounting for (78) in (77), and by substituting  $\sigma_d^2 = 1 - \sigma_p^2$ , the SINR becomes

$$\text{SINR} \geq \frac{a_0 + a_1\sigma_p^2 + a_2\sigma_p^4}{b_0 + b_1\sigma_p^2 + b_2\sigma_p^4}, \quad (82)$$

419 where  $a_0 = -N_r^3 N_t^2 P^2 \sigma_n^2 - 2N_r^2 N_t^2 P^2 \sigma_h^2 + N_r \sigma_h^2 \tilde{\tau}_h^2 \sigma_n^2 + 2\sigma_h^4 \tilde{\tau}_h^2$ ,  $a_1 = -N_r^3 N_t^2 P^2 \sigma_n^2 - MN N_r^2 N_t P \sigma_h^2 -$   
 420  $4N_r^2 N_t^2 P^2 \sigma_h^2 + N_r \sigma_h^2 \tilde{\tau}_h^2 \sigma_n^2 + 4\sigma_h^4 \tilde{\tau}_h^2$ ,  $a_2 = -2\sigma_h^4 \tilde{\tau}_h^2 + 2N_r^2 N_t^2 P^2 \sigma_h^2 + MN N_r^2 N_t P \sigma_h^2$ ,  $b_0 = N_r^3 N_t^2 P^2 \sigma_n^2 +$   
 421  $2N_r^2 N_t^2 P^2 \sigma_h^2 + N_r \tilde{\tau}_h^2 \sigma_n^4 + 2\sigma_h^2 \tilde{\tau}_h^2 \sigma_n^2$ ,  $b_1 = -2\sigma_h^2 \tilde{\tau}_h^2 \sigma_n^2 - 4N_r^2 N_t^2 P^2 \sigma_h^2 + MN N_r^2 N_t P \sigma_n^2 - N_r^3 N_t^2 P^2 \sigma_n^2 + 2N_r^2 N_t^3 P^3 \sigma_h^2 +$   
 422  $N_r^3 N_t^3 P^3 \sigma_n^2$  and  $b_2 = -2N_r^2 N_t^3 P^3 \sigma_h^2 + 2N_r^2 N_t^2 P^2 \sigma_h^2$ .

423 In order to determine the optimal pilot power, the first step is to differentiate the lower bound of SINR,  
 424 and set the resulting equation equal to zero. Once this equation has been solved, the optimal pilot power  
 425 can be obtained as

$$\sigma_{p,opt}^2 = \left| \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right|, \quad (83)$$

426 where  $a = b_1 a_2 - b_2 a_1$ ,  $b = 2b_0 a_2 - 2b_2 a_0$  and  $c = b_0 a_1 - b_1 a_0$ .

## 427 5. Complexity analysis

428 We will focus on the most complex term for each operation. For the sake of simplicity, we assume that  
 429  $N_t = N_r = N_a$ .

430 The complexity of the proposed technique contains two terms. The first one concerns the channel esti-  
 431 mation, named  $C_{ce}$ , while the second one concerns the data detection, named  $C_{dd}$ . Therefore, the overall  
 432 complexity of the proposed algorithm is  $C = C_{ce} + C_{dd}$ .

433 The complexity of one iteration of the detection algorithm requires the computation of (56), (57), (58),  
 434 and (60). Each of (56), (57), and (58) has a complexity of  $\mathcal{O}(MNN_a P|\mathbb{A}|)$ . In addition, the complexity of  
 435 (58) is  $\mathcal{O}(MNN_a |\mathbb{A}|)$ . So, the overall complexity of the data detection algorithm is dominated by  $C_{dd} =$   
 436  $\mathcal{O}(n_{iter}^{(dd)} MNN_a P|\mathbb{A}|)$ , where  $n_{iter}^{(dd)}$  is the number of iteration required to the convergence of the data detection  
 437 algorithm.

438 For the channel estimation step, we have two proposed algorithms. For the SoBaP algorithm, the most  
 439 complex operation per iteration is the update (A.9), which is  $\mathcal{O}(MNLN_a^3)$ . Therefore, the overall complexity  
 440 for the SoBaP algorithm is dominated by  $C_{ce}^{(sobap)} = \mathcal{O}(n_{iter}^{(ce)} MNLN_a^3)$ , where  $n_{iter}^{(ce)}$  is the number of iteration  
 441 required to the convergence of the SoBaP algorithm.

442 The LMMSE algorithm, is divided into five steps. The first one is the computation of the Hermitian  
 443 bounded matrix  $\mathbf{M} = (\mathbf{\Phi}_{pd}^{(i-1)})^H (\mathbf{C}_{\mu_d}^{(i-1)})^{-1} \mathbf{\Phi}_{pd}^{(i-1)} + \mathbf{C}_{\mathbf{h}_m}^{-1}$ , its computational cost is  $\frac{1}{2}(Q^2 + 3Q + 2)R$ , where  
 444  $Q = N_a P$  and  $R = P N_a^2$ . The second step is the Cholesky decomposition, which requires  $\frac{1}{2}(Q^2 + 3Q)R$   
 445 operation. The third and the fifth steps can be solved by band forward and backward substitutions and  
 446 each of them require  $QR$  operation. The fourth step, which consists in solving a diagonal system requires  
 447  $R$  operation. Therefore, the overall complexity of the the proposed LMMSE algorithm is dominated by  
 448  $C_{ce}^{(lmmse)} = \mathcal{O}(N_a^4 P^3)$ . Finally, the complexity of the channel estimation step of the proposed algorithm  
 449 is an average between the complexity of SoBaP and LMMSE, it is given as  $C_{ce} = \frac{C_{ce}^{sobap} + (N_f - 1)C_{ce}^{lmmse}}{N_f} \approx$   
 450  $\mathcal{O}\left(\frac{(N_f - 1)N_a P^3 + n_{iter}^{(ce)} MNL}{N_f} N_a^3\right)$ .

451 It should be noted that  $(N_f - 1)N_a P^3 \ll n_{iter}^{(ce)} MNL$  in practice, this means that the complexity of the  
 452 channel estimation step reduced to  $C_{ce} = \mathcal{O}\left(\frac{n_{iter}^{(ce)} MNL}{N_f} N_a^3\right)$ .

453 Table 2 summarizes the complexity of the proposed channel estimation algorithm and that of each of the  
 454 state-of-the-art methods listed in this paper.

Table 2: Computational complexities of the proposed and existing channel estimation methods.

Scheme	Computational complexity
EP [25]	$\mathcal{O}(N l_\tau N_a)$
BSBL-BR [18]	$\mathcal{O}(N_{iter} N_a^3 L K^2)$
RG-OMP [17]	$\mathcal{O}(M^3 N_p^3)$
RG-BL [17]	$\mathcal{O}(G^3 N_a^3)$
SoBaP-LMMSE-AMPA	$\mathcal{O}\left(\frac{n_{iter}^{(ce)} MNL}{N_f} N_a^3\right)$

455 Note that  $K = (2k_\nu + 2Q + 1)(l_\tau + 1)$  and  $G = (M_\tau + 1)(G_\nu + 1)$ . We note also  $C_m$  the complexity of  
 456 the method  $m$  and we compare the complexities of all the methods.

457 It is clear that the EP is the least complex of all the methods considered here because its complexity  
 458 varies linearly with  $N$ . It is also clear that RG-OMP is the most complex due to the  $M^3$  factor in the  
 459 expression of its complexity.

460 Since  $k_\nu < N_\nu < N$ ,  $l_\tau < M_\tau < M$ ,  $N_\nu \ll G_\nu$ , and  $N_{iter} \approx 10$  in practice, we have  $N_{iter} L K^2 < G^3$ ,  
 461 which leads to  $C_{BSBL-BR} < C_{RG-BL}$ .

462 Since  $n_{iter}^{(ce)} \approx 10$ , we can also easily check that  $n_{iter}^{(ce)} MNL / N_f < N_{iter} L K^2$ , so, we have  $C_{SoBaP-LMMSE-AMPA} <$   
 463  $C_{BSBL-BR}$ . We can conclude that  $C_{EP} < C_{SoBaP-LMMSE-AMPA} < C_{BSBL-BR} < C_{RG-BL} < C_{RG-OMP}$ .

## 464 6. Simulation results

465 In this section, we first evaluate the performance of the proposed algorithm in terms of NMSE, BER and  
 466 SE. Then, we compare obtained results against four state-of-the-art methods: EP (pilot type-1), BSBL-BR

467 (pilot type-2), RG-OMP and RG-BL (pilot type-3).

468 *6.1. Simulation setting*

469 The simulation parameters are given in Table 3.

Table 3: Simulation parameters.

Parameter	Value	Parameter	Value
$f_c$	4 GHz	$\Delta f$	15 kHz
$(N, M)$	(16, 16)	$(N_t, N_r)$	(2, 2)
$t_c$	541 $\mu s$	$T$	1 $ms$
Modulation	'BPSK'	$(W_t, W_r)$	'rectangular'

470 For the channel delay model, we use a 5-tap DD model whose parameters are given in Table 4. Each  
 471 delay tap owns a single Doppler shift in the form  $\nu_k = \nu_{max} \cos(\theta_k)$ , where  $\nu_{max}$  is the maximum Doppler  
 472 shift of the channel and  $\theta_k \sim \mathcal{U}_{[0,\pi]}$ . The maximum delay tap  $l_\tau = 4$  and the maximum Doppler tap  $k_\nu = 2$   
 473 correspond to a high mobility scenario with a maximum relative speed  $\nu = 500$  km/h.

Table 4: Power delay profile [22].

Path no.	1	2	3	4	5
delay ( $\mu s$ )	2.08	5.20	8.33	11.46	20.8
Path power (dB)	1	-1.80	-3.57	-5.38	-8.86

474 The NMSE expression used in this work is given as follows:

$$\text{NMSE} = 1 - \left( \frac{|\mathbf{h}^H \hat{\mathbf{h}}|}{\|\mathbf{h}\|_2 \|\hat{\mathbf{h}}\|_2} \right)^2. \quad (84)$$

475 We assume that each location in the DD grid has a power of  $\sigma_x^2$ . The total power of each OTFS frame  
 476 is  $MN\sigma_x^2$ . Thus, by taking  $\sigma_x^2 = \sigma_d^2 + \sigma_p^2 = 1$ , the total power for each OTFS frame is fixed to  $MN$ . For a  
 477 fair comparison with the state-of-the-art schemes, this total power must remain the same for each scheme.  
 478 For the EP scheme [25], due to the insertion of guard intervals between the pilots and data symbols, all the  
 479 power that is assigned to the guard interval positions must be assigned to the pilots and data symbols, i.e.,  
 480  $\sigma_p^2 = \sigma_d^2 = \frac{MN}{MN+N_t-L_p}$ . Therefore, we have a total power per frame of  $N_t\sigma_p^2 + (MN - L_p)\sigma_d^2 = MN$ . For  
 481 BSBL-BR method [18], there is no guard intervals between the  $L$  pilots and the  $MN - L$  data symbols.  
 482 Thus, by taking  $\sigma_p^2 = \sigma_d^2 = 1$ , we have a total power per frame of  $\sigma_p^2 L + \sigma_d^2 (MN - L) = MN$ . For RG-OMP  
 483 and RG-BL methods [17], in each frame,  $MN_p$  symbols are pilots in a block of  $M(N + N_p)$  symbols. By  
 484 taking  $\sigma_p^2 = \sigma_d^2 = \frac{N}{N+N_p}$ , we have a total power per frame of  $MN\sigma_d^2 + MN_p\sigma_p^2 = MN$ .

485 Table 5 presents a summary of the optimal pilot power  $\sigma_{p,opt}^2$  for various SNR values, as calculated using  
 486 the expression derived in equation (83). It is evident from the table that, based on the given simulation

487 parameters, the average optimal pilot power is approximately 0.3. To maximize the SE and minimize the  
 488 BER of the proposed superimposed-aided design, it is recommended to allocate 30% of the total power to  
 489 pilots and 70% to data, which is also confirmed by numerical verification.

Table 5: Optimal pilot power with  $P = 5$  and  $M = N = 16$ .

SNR (dB)	0	5	10	15	20
$\sigma_{p,opt}^2$	0.2520	0.2995	0.3135	0.3178	0.3191

### 490 6.2. Stationarity of the channel

491 The channel used here is a doubly-underspread (DU) channel. For this type of channels, the delay and  
 492 Doppler taps  $\{l_i, k_i\}_{i=1:P}$  remain practically constant over a period of time. Here, we first check that the  
 493 channel used is indeed DU. Then, we look for the number of frames  $N_f$  for which the delay and Doppler  
 494 taps  $\{l_i, k_i\}_{i=1:P}$  remain almost constant.

495 It is shown in [28] that a DU channel satisfies the following condition:  $\Delta\nu_{max}\Delta\tau_{max} \ll \nu_{max}\tau_{max} \ll 1$ .  
 496  $\Delta\nu_{max} = 2\nu_{max}\sin(\delta/2)$  is the maximum Doppler correlation lag where  $\delta$  denotes the maximum angular  
 497 spread of the scatters.  $\Delta\tau_{max} = \omega/c$  is the maximum delay correlation lag where  $\omega$  denotes the maximum  
 498 spatial extension. In our case,  $\nu_{max} = 1850$  Hz and  $\tau_{max} = 20.8 \mu s$ , thus  $\nu_{max}\tau_{max} = 0.03848$ . Letting  
 499  $\delta = 3^\circ$  and  $\omega = 30 m$  [28],  $\Delta\nu_{max} = 96.9$  Hz and  $\Delta\tau_{max} = 0.1 \mu s$ , thus,  $\Delta\tau_{max}\Delta\nu_{max} = 9.69 \cdot 10^{-7}$ . Finally,  
 500 we see that  $\Delta\nu_{max}\Delta\tau_{max} \ll \nu_{max}\tau_{max} \ll 1$ . Therefore, the channel used here is a DU channel.

501 We have  $N_f = T_s/T$ , where  $T$  is the frame duration given as  $T = N/\Delta f$  and  $T_s$  denotes the duration for  
 502 which  $\{l_i, k_i\}_{i=1:P}$  remain practically constant. This duration is computed as  $T_s = 1/\Delta\nu_{max}$ . In our case,  
 503  $T_s = 10.3 ms$ , and  $T = 1 ms$ , so,  $N_f \approx 10$ . It is worth noting that, for constant  $N$ ,  $\Delta f$  and  $\delta$ , the number  
 504 of frames  $N_f$  is proportional to  $\nu_{max}$ . Thus, for speeds below  $500 km/h$ ,  $N_f$  increases.

### 505 6.3. Influence of the power of pilots on BER performance

506 The power allocated to the pilots influences the performance of the system. We investigate the variation  
 507 of the BER as a function of the power of the pilots  $\sigma_p^2$ . Fig 4 shows the variation of the BER of the proposed  
 508 algorithm by varying the power of the pilots in the interval  $[0.1; 0.9]$  for SNR = 5, 10, 15 dB.

509 We observe that, for the three SNR values, the minimum BER is reached at  $\sigma_p^2 = \sigma_{p,opt}^2 \approx 30\%$ . The  
 510 degradation of the BER performance away from  $\sigma_{p,opt}^2$  is due to the following two factors: If  $\sigma_p^2 < \sigma_{p,opt}^2$ , the  
 511 degradation is due to a poor channel estimation and if  $\sigma_{p,opt}^2 < \sigma_p^2$ , this degradation is due to the reduced  
 512 data power since we have set  $\sigma_d^2 = 1 - \sigma_p^2$ .

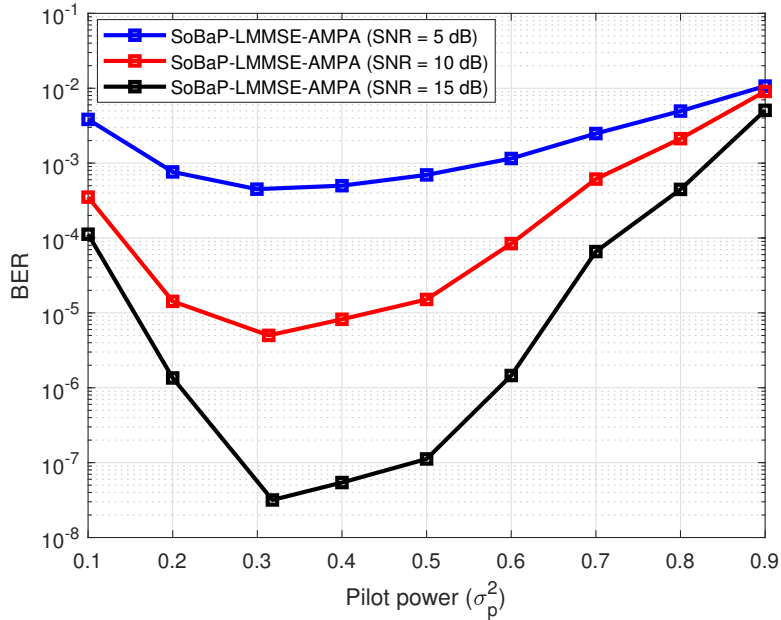


Figure 4: BER versus pilot power performance for the proposed algorithm with  $N_t = N_r = 2$ ,  $N = M = 16$  and SNR = 5, 10, 15 dB.

513 We now investigate in Fig 5 the BER of the proposed scheme, for different pilot powers. We observe from  
 514 this figure that the proposed scheme exhibits the lowest BER when the optimal power allocation is used.  
 515 This confirms the optimal pilot power which is calculated analytically in section 4 and shown in Table 5.

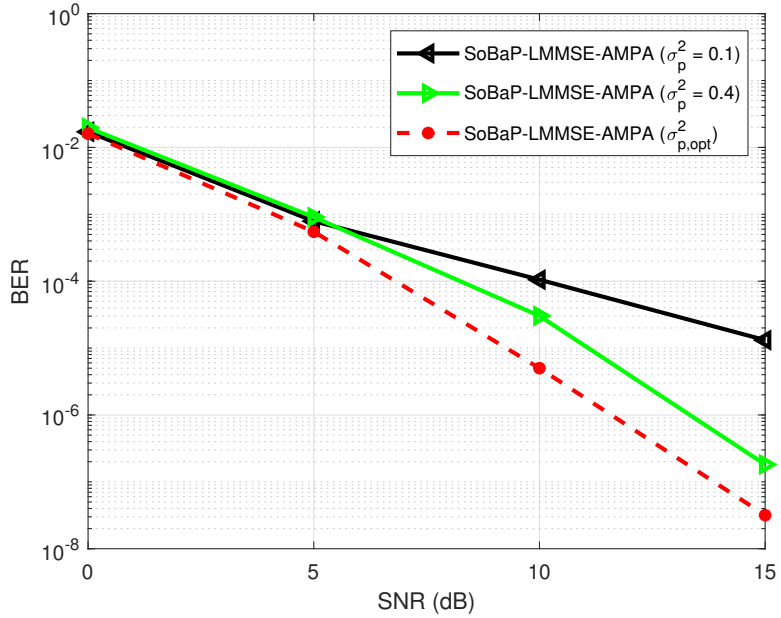


Figure 5: Effect of pilot power on the BER of the proposed scheme with  $N_t = N_r = 2$ ,  $N = M = 16$ .

516 *6.4. Convergence of the SoBaP-LMMSE-AMPA algorithm*

517 Fig 6 shows the BER performance of the proposed SoBaP-LMMSE-AMPA algorithm versus the number  
 518 of iterations with  $SNR = 5, 15$  dB,  $\sigma_p^2 = \sigma_{p,opt}^2$  dB, and BPSK modulation. From Fig 6, it can be seen that,  
 519 for both SNR values, the BER decreases with increasing number of iterations, and it saturates after about 5  
 520 or 6 iterations. The convergence and the effectiveness of the proposed SoBaP-LMMSE-AMPA algorithm are  
 521 confirmed by the steady states reached by the two curves of the BER after a few iterations. It can be also  
 522 observed that, as expected, the scenario with higher SNR exhibits superior convergence speed and detection  
 523 accuracy.

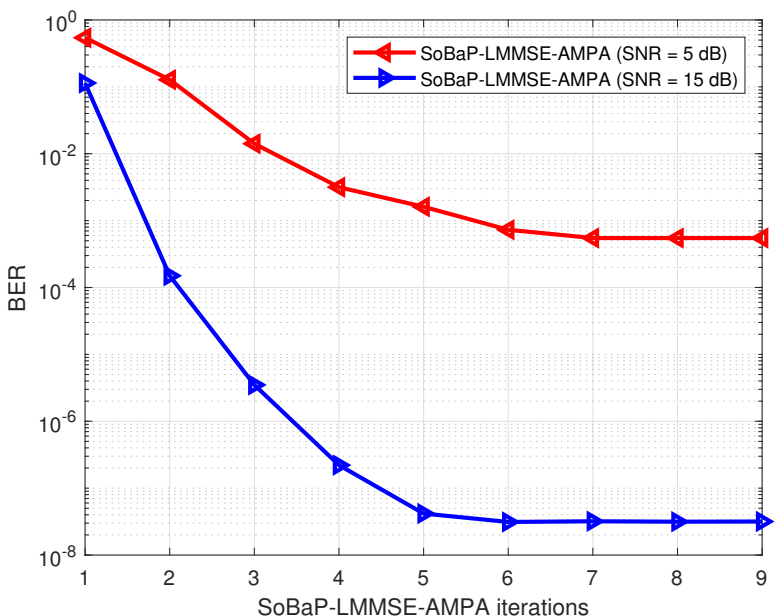


Figure 6: BER performance of the SoBaP-LMMSE-AMPA algorithm versus the number of iterations with  $N_t = N_r = 2$ ,  $N = M = 16$ ,  $\sigma_p^2 = \sigma_{p,opt}^2$  and SNR = 5, 15 dB.

524 *6.5. Channel NMSE performance*

525 We now investigate the variation of NMSE as a function of SNR. Fig 7 shows a comparison in terms  
 526 of NMSE between the proposed algorithm and EP [25], BSBSL-BR [18], and RG-OMP and RG-BL [17]  
 527 methods.

528 From Fig 7, we observe that the proposed algorithm has the lowest NMSE value compared to the other  
 529 methods, it exceeds the RG-BL, which is the best performing state-of-the-art method, by about 5 dB at  
 530  $SNR = 10$  dB. The EP method is most affected by noise owing to the threshold method.



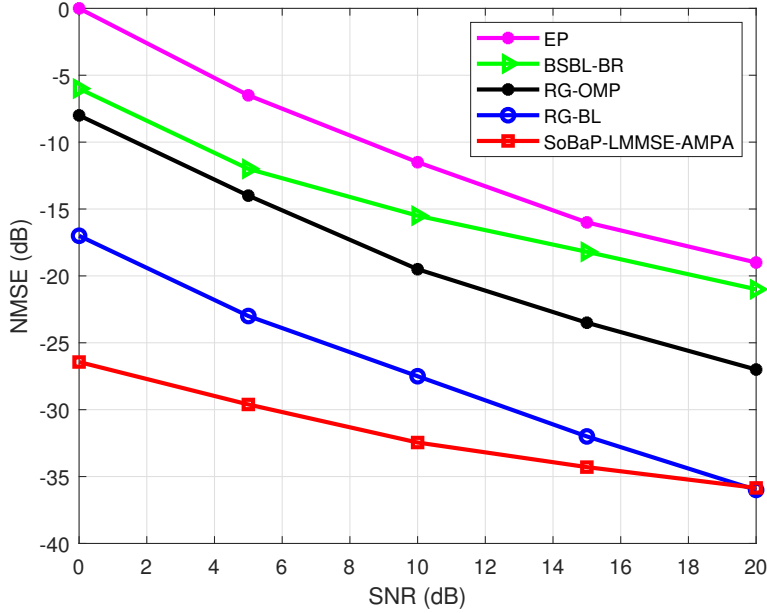


Figure 7: NMSE versus SNR performance for the MIMO-OTFS system with  $N_t = N_r = 2$ ,  $N = M = 16$ , and  $\sigma_p^2 = \sigma_{p,opt}^2$ .

### 531 6.6. BER performance evaluation

532 We now investigate the variation of BER as a function of SNR. Fig 8 shows the BER over SNR perfor-  
 533 mance of the proposed algorithm and EP [25], BSBSL-BR [18], and RG-OMP and RG-BL [17] methods.

534 From Fig 8, we observe that the EP and the BSBL-BR methods perform almost similarly. We also  
 535 observe that the proposed algorithm outperforms all the other methods. It exceeds the RG-BL method by  
 536 about 3 dB at  $BER = 10^{-4}$ .

### 537 6.7. SE performance evaluation

538 We now investigate the average SE as a function of SNR. The SE expression of a scheme  $s$  is given as  
 539 follows [22]:

$$\mathcal{R}_s = (1 - \eta_s) \log_2(1 + \text{SINR}_s), \quad (85)$$

540 where  $s = \text{EP, BSBL-BR, RG-OMP, RG-BL, SoBaP-LMMSE-AMPA}$ ,  $\eta_s$  denotes the pilot overhead related  
 541 to the scheme  $s$ .

542 The pilot overhead  $\eta_s$  is calculated using the frame-structure for the scheme  $s$ . The SINR expressions  
 543 for all the schemes are derived in the same way as in [22]. Table 6 shows the expressions of  $\text{SINR}_s$  and  $\eta_s$   
 544 for each scheme  $s$ .

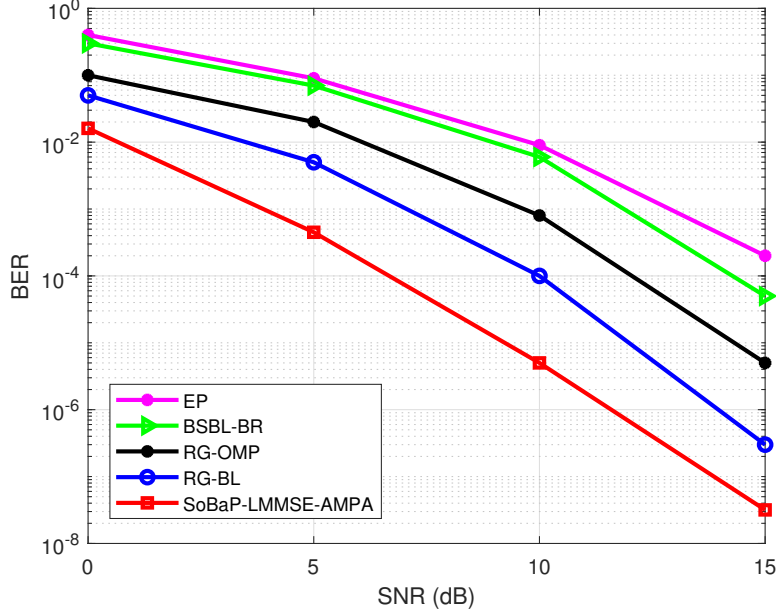


Figure 8: BER versus SNR performance for the MIMO-OTFS system with  $N_t = N_r = 2$ ,  $N = M = 16$ , and  $\sigma_p^2 = \sigma_{p,opt}^2$ .

Table 6: Expressions of  $\text{SINR}_s$  and  $\eta_s$ .

Scheme ( $s$ )	$\eta_s$	$\text{SINR}_s$
EP [25]	$\frac{L_p}{MN}$	$\frac{(\sigma_h^2 - M_{h,EP})\sigma_d^2}{\sigma_n^2 + \sigma_d^2 M_{h,EP}}$
BSBL-BR [18]	$\frac{L}{MN}$	$\frac{(\sigma_h^2 - M_{h,BSBL-BR})\sigma_d^2}{\sigma_n^2 + \sigma_d^2 M_{h,BSBL-BR}}$
RG-OMP [17]	$\frac{N_p}{N_p + N}$	$\frac{(\sigma_h^2 - M_{h,RG-OMP})\sigma_d^2}{\sigma_n^2 + \sigma_d^2 M_{h,RG-OMP}}$
RG-BL [17]	$\frac{N_p}{N_p + N}$	$\frac{(\sigma_h^2 - M_{h,RG-BL})\sigma_d^2}{\sigma_n^2 + \sigma_d^2 M_{h,RG-BL}}$
SoBaP-LMMSE-AMPA	0	$\frac{(\sigma_h^2 - M_{h,PS})\sigma_{d,opt}^2}{\sigma_n^2 + \sigma_{d,opt}^2 M_{h,PS} + \sigma_{p,opt}^2 P N_t M_{h,PS}}$

545 where  $L_p = (4k_\nu + 1)(N_t l_\tau + l_\tau + N_t)$ ,  $N_p$  is the number of pilots along time-axis for the RG-OMP and  
 546 RG-BL schemes and  $M_{h,s}$  is the channel MSE of scheme  $s$ .

547 Fig 9 shows the average SE comparison between the proposed scheme and the EP, BSBL-BR, RG-OMP,  
 548 and RG-BL schemes for  $N = M = 16$  and for different values of  $l_\tau$  and  $k_\nu$ . It is not surprising to see from  
 549 this figure that the SE of all the methods increases as the SNR increases. It is important to note that the  
 550 SE of the proposed scheme exhibits a significant increase compared to that of the EP design regardless of  
 551 the values of  $l_\tau$  and  $k_\nu$ . We observe from this figure that as  $l_\tau$  and/or  $k_\nu$  increase, the SE of all state-of-  
 552 the-art schemes degrades. This is because the pilot overhead of these schemes increases with  $l_\tau$  and  $k_\nu$ . The  
 553 proposed designs, in contrast, avoid this pilot overhead. We also see that for  $l_\tau = 4$  and  $k_\nu = 1$ , the SE of the  
 554 proposed scheme is close to those of BSBL-BR, RG-OMP, and RG-BL schemes, because the pilot overhead  
 555 decrease with  $l_\tau$  and/or  $k_\nu$ .

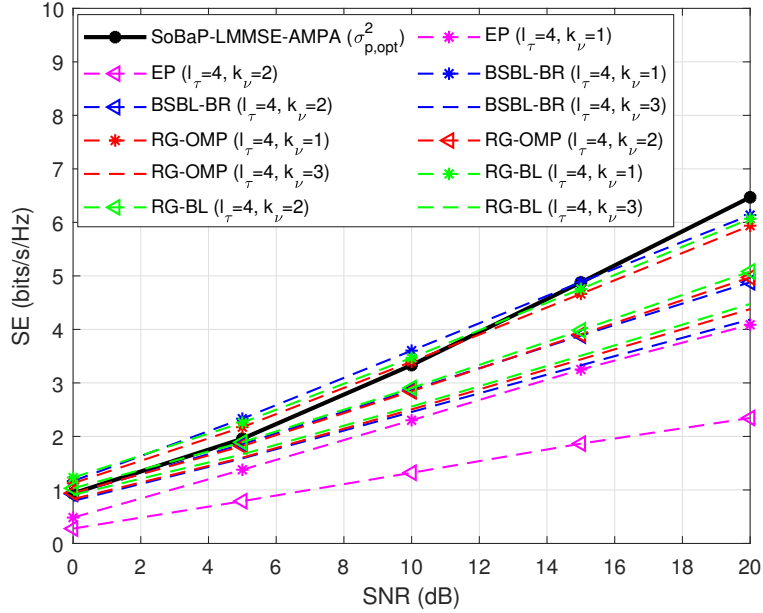


Figure 9: SE versus SNR performance for the MIMO-OTFS system with  $N_t = N_r = 2$  and  $N = M = 16$ .

## 556 7. Conclusion

557 In this manuscript, we have proposed an iterative algorithm for channel estimation and data detection  
558 in the delay-Doppler domain for MIMO-OTFS systems that we have named SoBaP-LMMSE-AMPA. To  
559 improve the spectral efficiency of the system, we adopted a superimposed pilot pattern. The proposed  
560 algorithm iterates between message passing-aided data detection and data-aided channel estimation. For  
561 the channel estimation step, two algorithms have been proposed. The first one consists in estimating all the  
562 parameters of the channel, including the number of channel paths, delay taps, Doppler taps, and channel  
563 gains via mean-field approximation and the so-called VB-EM algorithm. The second one is based on LMMSE  
564 combined with Cholesky decomposition. This second solution is only used to estimate the channel gains when  
565 the delay and Doppler taps remain unchanged. For data detection, we adapted the MP algorithm to our  
566 context. We have also derived a lower bound on the signal-to-interference-plus-noise ratio of the proposed  
567 scheme, and maximized it by optimally allocating power between pilots and data symbols. The proposed  
568 algorithm has been compared to four state-of-the-art methods, including a recognized MP method, and  
569 three recent solutions (BSBL-BR, RG-OMP and RG-BL). Complexity analysis and simulation results have  
570 shown that the proposed algorithm achieves a good compromise with the state-of-the-art methods in terms  
571 of computation complexity and performs significantly better in terms of NMSE, BER, and SE.

572 **Appendix A. Calculation of the mean field approximation**

573 The methodology called Mean Field (MF) approximation is adopted here to compute an approximation  
 574  $q(s_i)$  of the posterior probability  $p(s_i|\mathbf{y}_m)$ . The MF approximation [36, 37] of  $p(\boldsymbol{\theta}|\mathbf{y}_m)$ , where  $\boldsymbol{\theta} = (\mathbf{s}, \mathbf{c})$  and  
 575  $p(\boldsymbol{\theta}|\mathbf{y}_m)$  its posterior distribution, is the surrogate distribution  $q^*(\boldsymbol{\theta})$  which satisfies

$$q^*(\boldsymbol{\theta}) = \arg \min_{q(\boldsymbol{\theta})} \left\{ \int_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) \log \left( \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{y}_m)} \right) d\boldsymbol{\theta} \right\}, \quad (\text{A.1})$$

subject to

$$q(\boldsymbol{\theta}) = \prod_{k=1}^K q(\boldsymbol{\theta}_k), \quad \int_{\boldsymbol{\theta}_k} q(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k = 1 \quad \forall k \in [1, K]. \quad (\text{A.2})$$

576 Successive minimizations of the Kullback-Leibler divergence [38] with respect to the parameters of factors  
 577  $q(\boldsymbol{\theta}_i)$  can solve the problem (A.1), (A.2) [39]. The procedure given in [36] named VB-EM algorithm [40–42],  
 578 is ensured to converge to a saddle point or a (local or global) maximum of problem (A.1), (A.2) under mild  
 579 conditions [36].

To approximate the marginals  $p(\boldsymbol{\theta}_i|\mathbf{y}_m)$ , the MF approximations offer a good framework. Indeed,

$$\begin{aligned} p(\boldsymbol{\theta}_i|\mathbf{y}_m) &= \int_{\boldsymbol{\theta}_{-i}} p(\boldsymbol{\theta}|\mathbf{y}_m) d\boldsymbol{\theta}_{-i}, \\ &\simeq \int_{\boldsymbol{\theta}_{-i}} q(\boldsymbol{\theta}|\mathbf{y}_m) d\boldsymbol{\theta}_{-i}, \\ &\simeq q(\boldsymbol{\theta}_i|\mathbf{y}_m), \end{aligned} \quad (\text{A.3})$$

580 where the last equality stems from (A.2).

581 Here, we consider the particular case where the MF approximation  $q(\mathbf{c}, \mathbf{s})$  of  $p(\mathbf{c}, \mathbf{s}|\mathbf{y}_m)$  simply writes  
 582  $q(\mathbf{c}, \mathbf{s}) = \prod_i q(c_i, s_i)$ . Together with models (34), (35), the corresponding VB-EM update is given as follows:

$$q(c_i, s_i|\mathbf{y}_m) = q(c_i|s_i, \mathbf{y}_m)q(s_i|\mathbf{y}_m), \quad (\text{A.4})$$

583 where

$$q(c_i|s_i, \mathbf{y}_m) = \mathcal{N}(m(s_i), \Sigma(s_i)) \quad (\text{A.5})$$

$$q(s_i|\mathbf{y}_m) \propto \sqrt{\Sigma(s_i)} \exp \left( \frac{1}{2} \frac{m(s_i)^2}{\Sigma(s_i)} \right) p(s_i) \quad (\text{A.6})$$

584 and

$$\Sigma(s_i|\mathbf{y}_m) = \frac{\sigma_{c_i}^2 \sigma^2}{\sigma^2 + s_i \sigma_{c_i}^2 \mathbf{A}_i^T \mathbf{A}_i}, \quad (\text{A.7})$$

$$m(b_i|\mathbf{y}_m) = s_i \frac{\sigma_{c_i}^2}{\sigma^2 + s_i \sigma_{c_i}^2 \mathbf{A}_i^T \mathbf{A}_i} \mathbf{r}_i^T \mathbf{A}_i, \quad (\text{A.8})$$

$$\mathbf{r}_i = \mathbf{y}_m - \sum_{l \neq i} q(s_l = 1) m(s_l = 1) \mathbf{A}_l, \quad (\text{A.9})$$

From (A.3), an approximation of  $p(s_k|\mathbf{y}_m)$  derive easily from the following relationship:

$$p(s_i|\mathbf{y}_m) \simeq \int q(c_i, s_i|\mathbf{y}_m) dc_i = q(s_i). \quad (\text{A.10})$$

## 585 Appendix B. Analysis for positive definiteness of matrix $\mathbf{M}$

586 Let the matrix  $\mathbf{X} \in \mathbb{C}^{K \times K}$ . It is said to be positive definite if it satisfies the following property: For any  
587 non-zero vector  $\mathbf{z} \in \mathbb{C}^K$ ,  $\mathbf{z}^H \mathbf{X} \mathbf{z}$  is a strictly positive real number [43].

588 The matrix  $\mathbf{M}$  is given by  $\mathbf{M} = ((\Phi_{pd})^H (\mathbf{C}_{\mu_d})^{-1} \Phi_{pd} + \mathbf{C}_{h_m}^{-1})$ , where  $\mathbf{C}_{\mu_d} = \lambda \mathbf{I}_{MN_t N_r}$ , with  $\lambda$  is  
589 a positive number given by  $\lambda = \left( \sigma_n^2 + \frac{2\sigma_d^2}{N_r} \sum_{i=1}^{PN_t N_r} \sigma_{h_i}^2 \right)$  and  $\mathbf{C}_{h_m} = \text{diag}\{\sigma_{h_1}^2, \sigma_{h_2}^2, \dots, \sigma_{h_{PN_t N_r}}^2\}$ . For all  
590 non-zero  $\mathbf{z} \in \mathbb{C}^{PN_t N_r \times 1}$ , we have

$$\begin{aligned} \mathbf{z}^H \mathbf{M} \mathbf{z} &= \mathbf{z}^H ((\Phi_{pd})^H (\mathbf{C}_{\mu_d})^{-1} \Phi_{pd}^{(i-1)} + \mathbf{C}_{h_m}^{-1}) \mathbf{z}, \\ &= \mathbf{z}^H (\Phi_{pd})^H (\lambda \mathbf{I}_{MN_t N_r})^{-1} \Phi_{pd} \mathbf{z} + \mathbf{z}^H \left( \text{diag}\{\sigma_{h_1}^2, \sigma_{h_2}^2, \dots, \sigma_{h_{PN_t N_r}}^2\} \right)^{-1} \mathbf{z}, \\ &= \frac{1}{\lambda} \mathbf{z}^H (\Phi_{pd})^H \Phi_{pd} \mathbf{z} + \sum_{i=1}^{PN_t N_r} \frac{|z_i|^2}{\sigma_{h_i}^2}, \\ &= \frac{1}{\lambda} (\Phi_{pd} \mathbf{z})^H \Phi_{pd} \mathbf{z} + \sum_{i=1}^{PN_t N_r} \frac{|z_i|^2}{\sigma_{h_i}^2}, \\ &= \frac{1}{\lambda} \|\Phi_{pd} \mathbf{z}\|^2 + \sum_{i=1}^{PN_t N_r} \frac{|z_i|^2}{\sigma_{h_i}^2}. \end{aligned} \quad (\text{B.1})$$

591 We have  $\frac{1}{\lambda} \|\Phi_{pd} \mathbf{z}\|^2 + \sum_{i=1}^{PN_t N_r} \frac{|z_i|^2}{\sigma_{h_i}^2} > 0$ , which means that  $\mathbf{M}$  is a positive definite matrix.

## 592 References

- 593 [1] P. Singh, H. B. Mishra, R. Budhiraja, Low-complexity linear MIMO-OTFS receivers, in: 2021 IEEE  
594 International Conference on Communications Workshops (ICC Workshops), IEEE, 2021, pp. 1–6.
- 595 [2] K. Xu, Z. Shen, Y. Wang, X. Xia, Location-aided mMIMO channel tracking and hybrid beamforming for  
596 high-speed railway communications: An angle-domain approach, IEEE Systems Journal 14 (1) (2019)  
597 93–104.

- 598 [3] P. Singh, A. Gupta, H. B. Mishra, R. Budhiraja, Low-complexity ZF/MMSE MIMO-OTFS receivers  
599 for high-speed vehicular communication, *IEEE Open Journal of the Communications Society* 3 (2022)  
600 209–227.
- 601 [4] A. Monk, R. Hadani, M. Tsatsanis, S. Rakib, OTFS-orthogonal time frequency space, arXiv preprint  
602 arXiv:1608.02993.
- 603 [5] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, R. Calderbank, Orthogonal  
604 time frequency space modulation, in: *2017 IEEE Wireless Communications and Networking Conference*  
605 (WCNC), IEEE, 2017, pp. 1–6.
- 606 [6] G. Surabhi, R. M. Augustine, A. Chockalingam, On the diversity of uncoded OTFS modulation in  
607 doubly-dispersive channels, *IEEE transactions on wireless communications* 18 (6) (2019) 3049–3063.
- 608 [7] Z. Ding, R. Schober, P. Fan, H. V. Poor, OTFS-NOMA: An efficient approach for exploiting heteroge-  
609 nous user mobility profiles, *IEEE Transactions on Communications* 67 (11) (2019) 7950–7965.
- 610 [8] R. M. Augustine, G. Surabhi, A. Chockalingam, Space-time coded OTFS modulation in high-doppler  
611 channels, in: *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, IEEE, 2019, pp.  
612 1–6.
- 613 [9] G. Surabhi, A. Chockalingam, Low-complexity linear equalization for  $2 \times 2$  MIMO-OTFS signals, in:  
614 *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications*  
615 (SPAWC), IEEE, 2020, pp. 1–5.
- 616 [10] M. Li, S. Zhang, F. Gao, P. Fan, O. A. Dobre, A new path division multiple access for the massive  
617 MIMO-OTFS networks, *IEEE journal on selected areas in communications* 39 (4) (2020) 903–918.
- 618 [11] M. K. Ramachandran, A. Chockalingam, MIMO-OTFS in high-doppler fading channels: Signal detection  
619 and channel estimation, in: *2018 IEEE Global Communications Conference (GLOBECOM)*, IEEE,  
620 2018, pp. 206–212.
- 621 [12] W. Shen, L. Dai, J. An, P. Fan, R. W. Heath, Channel estimation for orthogonal time frequency space  
622 (OTFS) massive MIMO, *IEEE Transactions on Signal Processing* 67 (16) (2019) 4204–4217.
- 623 [13] Y. Liu, S. Zhang, F. Gao, J. Ma, X. Wang, Uplink-aided high mobility downlink channel estimation  
624 over massive MIMO-OTFS system, *IEEE Journal on Selected Areas in Communications* 38 (9) (2020)  
625 1994–2009.

- 626 [14] X. Wu, S. Ma, X. Yang, Tensor-based low-complexity channel estimation for mmwave massive MIMO-  
627 OTFS systems, *Journal of Communications and Information Networks* 5 (3) (2020) 324–334.
- 628 [15] D. Shi, W. Wang, L. You, X. Song, Y. Hong, X. Gao, G. Fettweis, Deterministic pilot design and  
629 channel estimation for downlink massive MIMO-OTFS systems in presence of the fractional doppler,  
630 *IEEE Transactions on Wireless Communications* 20 (11) (2021) 7151–7165.
- 631 [16] T. Li, C. Han, R. Yao, Y. Fan, X. Zuo, Low pilot overhead channel estimation for CP-OFDM-based  
632 massive MIMO OTFS system, *IET Communications* 16 (10) (2022) 1071–1082.
- 633 [17] S. Srivastava, R. K. Singh, A. K. Jagannatham, L. Hanzo, Bayesian learning aided simultaneous row  
634 and group sparse channel estimation in orthogonal time frequency space modulated MIMO systems,  
635 *IEEE Transactions on Communications* 70 (1) (2021) 635–648.
- 636 [18] L. Zhao, J. Yang, Y. Liu, W. Guo, Block sparse bayesian learning-based channel estimation for MIMO-  
637 OTFS systems, *IEEE Communications Letters* 26 (4) (2022) 892–896.
- 638 [19] Y. Liu, Y. L. Guan, D. González, Near-optimal BEM OTFS receiver with low pilot overhead for high-  
639 mobility communications, *IEEE Transactions on Communications* 70 (5) (2022) 3392–3406.
- 640 [20] F. Zhang, W. Ji, L. Qiu, Channel estimation for massive MIMO-OTFS systems via sparse bayesian  
641 learning with 2-D local beta process, in: *2022 IEEE Wireless Communications and Networking Confer-*  
642 *ence (WCNC)*, IEEE, 2022, pp. 1383–1388.
- 643 [21] R. Muzavazi, O. O. Oyerinde, Channel estimation and data detection schemes for orthogonal time  
644 frequency space massive MIMO systems, *Computers and Electrical Engineering* 102 (2022) 108215.
- 645 [22] H. B. Mishra, P. Singh, A. K. Prasad, R. Budhiraja, OTFS channel estimation and data detection  
646 designs with superimposed pilots, *IEEE transactions on wireless communications* 21 (4) (2021) 2258–  
647 2274.
- 648 [23] W. Yuan, S. Li, Z. Wei, J. Yuan, D. W. K. Ng, Data-aided channel estimation for OTFS systems with a  
649 superimposed pilot and data transmission scheme, *IEEE wireless communications letters* 10 (9) (2021)  
650 1954–1958.
- 651 [24] Y. Liu, Y. L. Guan, D. González, BEM OTFS receiver with superimposed pilots over channels with  
652 doppler and delay spread, in: *ICC 2022-IEEE International Conference on Communications*, IEEE,  
653 2022, pp. 2411–2416.

- 654 [25] P. Raviteja, K. T. Phan, Y. Hong, Embedded pilot-aided channel estimation for OTFS in delay–doppler  
655 channels, *IEEE transactions on vehicular technology* 68 (5) (2019) 4906–4917.
- 656 [26] R. Ouchikh, A. Aïssa-El-Bey, T. Chonavel, M. Djeddou, Sparse channel estimation algorithms for OTFS  
657 system, *IET Communications*.
- 658 [27] R. Ouchikh, A. Aïssa-El-Bey, T. Chonavel, M. Djeddou, Iterative channel estimation and data detection  
659 algorithm for OTFS modulation, in: *ICASSP 2022-2022 IEEE International Conference on Acoustics,  
660 Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 5263–5267.
- 661 [28] G. Matz, On non-WSSUS wireless fading channels, *IEEE transactions on wireless communications* 4 (5)  
662 (2005) 2465–2478.
- 663 [29] P. Singh, H. B. Mishra, A. K. Jagannatham, K. Vasudevan, Semi-blind, training, and data-aided channel  
664 estimation schemes for MIMO-FBMC-OQAM systems, *IEEE Transactions on Signal Processing* 67 (18)  
665 (2019) 4668–4682.
- 666 [30] C. Herzet, A. Drémeau, Bayesian pursuit algorithms, in: *2010 18th European Signal Processing Con-  
667 ference*, IEEE, 2010, pp. 1474–1478.
- 668 [31] O. Rabaste, T. Chonavel, Estimation of multipath channels with long impulse response at low SNR via  
669 an MCMC method, *IEEE Transactions on Signal Processing* 55 (4) (2007) 1312–1325.
- 670 [32] P. Raviteja, K. T. Phan, Y. Hong, E. Viterbo, Interference cancellation and iterative detection for  
671 orthogonal time frequency space modulation, *IEEE transactions on wireless communications* 17 (10)  
672 (2018) 6501–6515.
- 673 [33] P. Singh, H. B. Mishra, A. K. Jagannatham, K. Vasudevan, L. Hanzo, Uplink sum-rate and power  
674 scaling laws for multi-user massive MIMO-FBMC systems, *IEEE Transactions on Communications*  
675 68 (1) (2019) 161–176.
- 676 [34] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall, Inc., 1993.
- 677 [35] Z. Ulukök, R. Türkmen, On some matrix trace inequalities, *Journal of Inequalities and Applications*  
678 2010 (2010) 1–8.
- 679 [36] A. Drémeau, C. Herzet, L. Daudet, Boltzmann machine and mean-field approximation for structured  
680 sparse decompositions, *IEEE Transactions on Signal Processing* 60 (7) (2012) 3425–3438.
- 681 [37] M. J. Beal, *Variational algorithms for approximate Bayesian inference*, University of London, University  
682 College London (United Kingdom), 2003.



- 683 [38] T. Van Erven, P. Harremos, Rényi divergence and kullback-leibler divergence, *IEEE Transactions on*  
684 *Information Theory* 60 (7) (2014) 3797–3820.
- 685 [39] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM  
686 algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1) (1977) 1–22.
- 687 [40] T. P. Minka, A family of algorithms for approximate bayesian inference, Ph.D. thesis, Massachusetts  
688 Institute of Technology (2001).
- 689 [41] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West, et al., The variational  
690 bayesian em algorithm for incomplete data: with application to scoring graphical model structures,  
691 *Bayesian statistics* 7 (453-464) (2003) 210.
- 692 [42] C. M. Bishop, N. M. Nasrabadi, *Pattern recognition and machine learning*, Vol. 4, Springer, 2006.
- 693 [43] R. A. Horn, C. R. Johnson, *Matrix analysis*, Cambridge university press, 2012.