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

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

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

Rabah Ouchikh[a], Thierry Chonavel[b], Abdeldjalil Aïssa-El-Bey[b], Mustapha Djeddou[a]

[a] *Laboratoire Télécommunications, Ecole Militaire Polytechnique, Bordj El-Bahri, Algeria.*
[b] *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), abdeldjalil.aissaelbey@imt-atlantique.fr (Thierry Chonavel), thierry.chonavel@imt-atlantique.fr (Abdeldjalil Aïssa-El-Bey), djeddou.mustapha@gmail.com (Mustapha Djeddou)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2. For data detection step, we propose to this context an adapted version of the message-passing algorithm (AMPA). This algorithm is computationally-efficient as it exploits the channel sparsity in the DD domain. In its detection process, it takes into account the interference of the pilots on data symbols.

3. We derive a lower bound on the signal-to-interference-plus-noise-ratio (SINR) of the proposed scheme, which is subsequently optimized to derive the optimal pilot power expression. It is demonstrated that the optimal power level reduces the BER and increases further the SE of the proposed scheme.

4. We numerically validate the narrowness of the optimal distribution of transmitted power between pilots and data symbols and show its effect on the NMSE, BER and SE of the proposed scheme. Then, we compare the complexity and the performance of the proposed algorithm against four state-of-the-art methods, named: embedded pilots (EP) [25], block sparse Bayesian learning with block reorganization (BSBL-BR) [18], row-group OMP (RG-OMP), and row-group bayesian learning (RG-BL) [17], in terms of SE, normalize mean squared error (NMSE) and BER. Finally, we show the good compromise achieved by the proposed scheme between NMSE and BER performance, SE, computational complexity and PAPR.

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

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

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

2. MIMO-OTFS system model

We consider a MIMO-OTFS system whose parameters are given in Table 1. Its block diagram is shown 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
$\boldsymbol{W}_t$, $\boldsymbol{W}_r$	transmit pulse matrix, receive pulse matrix
f_c	carrier frequency
$\Delta t(s)$, $\Delta f(Hz)$	slot duration, subcarrier spacing
k_ν, l_τ	maximum Doppler tap, maximum delay tap
$N\Delta t$, $M\Delta f$	frame duration, frame bandwidth
t_c	coherence time

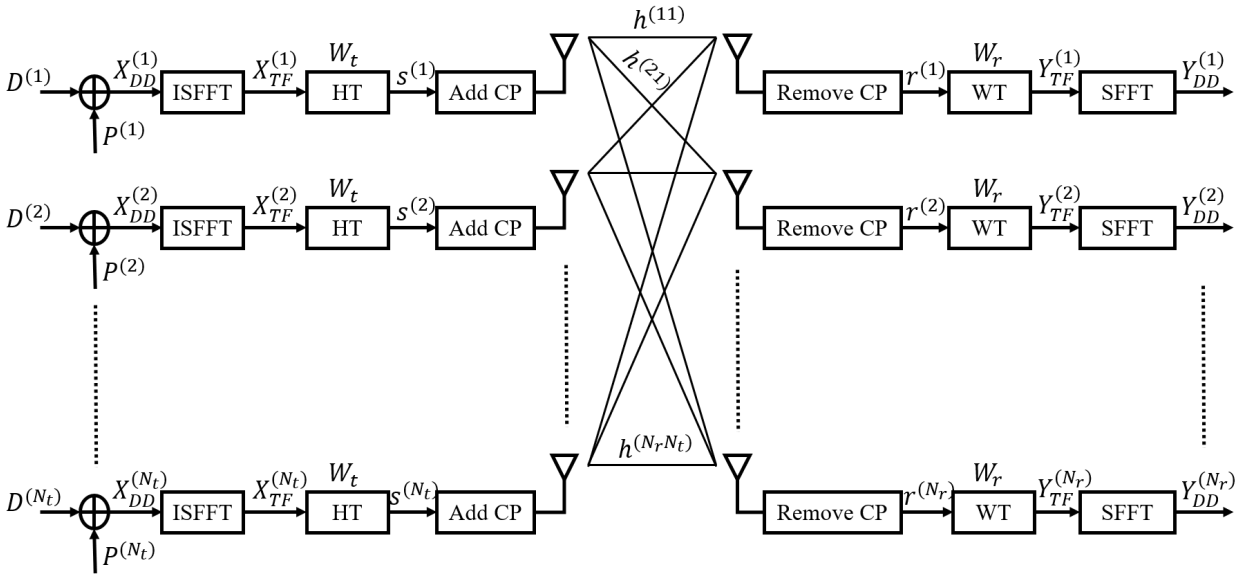


Figure 1: Block diagram of a MIMO-OTFS system: (I)SFFT: (Inverse) symplectic finite Fourier transform, HT: Heisenberg transform, WT: Wigner transform, CP: Cyclic prefix, TF: Time-frequency.

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{cases} \boldsymbol{r}^{(1)} = \boldsymbol{H}_{11}\boldsymbol{s}^{(1)} + ... + \boldsymbol{H}_{1N_t}\boldsymbol{s}^{(N_t)} + \boldsymbol{n}^{(1)}, \\ \boldsymbol{r}^{(2)} = \boldsymbol{H}_{21}\boldsymbol{s}^{(1)} + ... + \boldsymbol{H}_{2N_t}\boldsymbol{s}^{(N_t)} + \boldsymbol{n}^{(2)}, \\ \quad \vdots \\ \boldsymbol{r}^{(N_r)} = \boldsymbol{H}_{N_r1}\boldsymbol{s}^{(1)} + ... + \boldsymbol{H}_{N_rN_t}\boldsymbol{s}^{(N_t)} + \boldsymbol{n}^{(N_r)}, \end{cases} \tag{5}$$

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

$$\boldsymbol{\Pi} = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{6}$$

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

Let $\boldsymbol{r}_m = [(\boldsymbol{r}^{(1)})^T, (\boldsymbol{r}^{(2)})^T, ..., (\boldsymbol{r}^{(N_r)})^T]^T$, $\boldsymbol{s}_m = [(\boldsymbol{s}^{(1)})^T, (\boldsymbol{s}^{(2)})^T, ..., (\boldsymbol{s}^{(N_t)})^T]^T$, $\boldsymbol{n}_m = [(\boldsymbol{n}^{(1)})^T, (\boldsymbol{n}^{(2)})^T, ..., (\boldsymbol{n}^{(N_r)})^T]^T$, and

$$\boldsymbol{H}_M = \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} & \cdots & \boldsymbol{H}_{1N_t} \\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} & \cdots & \boldsymbol{H}_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{H}_{N_r1} & \boldsymbol{H}_{N_r2} & \cdots & \boldsymbol{H}_{N_rN_t} \end{pmatrix}, \tag{7}$$

Thus,

$$\boldsymbol{r}_m = \boldsymbol{H}_M \boldsymbol{s}_m + \boldsymbol{n}_m, \tag{8}$$

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

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

$$\boldsymbol{y}_m = \boldsymbol{\mathcal{H}}\boldsymbol{x}_m + \widetilde{\boldsymbol{n}}, \tag{9}$$

where $\boldsymbol{y}_m = [(\boldsymbol{y}_{DD}^{(1)})^T, (\boldsymbol{y}_{DD}^{(2)})^T, ..., (\boldsymbol{y}_{DD}^{(N_r)})^T]^T$, $\boldsymbol{x}_m = [(\boldsymbol{x}_{DD}^{(1)})^T, (\boldsymbol{x}_{DD}^{(2)})^T, ..., (\boldsymbol{x}_{DD}^{(N_t)})^T]^T$, $\widetilde{\boldsymbol{n}} = [((\boldsymbol{F}_N \otimes \boldsymbol{W}_r)\boldsymbol{n}^{(1)})^T, ((\boldsymbol{F}_N \otimes \boldsymbol{W}_r)\boldsymbol{n}^{(2)})^T, ..., ((\boldsymbol{F}_N \otimes \boldsymbol{W}_r)\boldsymbol{n}^{(N_r)})^T]^T$, and

$$\boldsymbol{\mathcal{H}} = \begin{pmatrix} \boldsymbol{\mathcal{H}}_{11} & \boldsymbol{\mathcal{H}}_{12} & \cdots & \boldsymbol{\mathcal{H}}_{1N_t} \\ \boldsymbol{\mathcal{H}}_{21} & \boldsymbol{\mathcal{H}}_{22} & \cdots & \boldsymbol{\mathcal{H}}_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{H}}_{N_r1} & \boldsymbol{\mathcal{H}}_{N_r2} & \cdots & \boldsymbol{\mathcal{H}}_{N_rN_t} \end{pmatrix}, \tag{10}$$

with

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

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

$$\boldsymbol{C}_{\widetilde{\boldsymbol{n}}} = \mathbb{E}\{\widetilde{\boldsymbol{n}}\widetilde{\boldsymbol{n}}^H\} = \sigma_n^2 \boldsymbol{I}_{MNN_r}. \tag{12}$$

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

$$\boldsymbol{y}_m = \boldsymbol{\mathcal{H}}(\boldsymbol{p} + \boldsymbol{d}) + \widetilde{\boldsymbol{n}} = \boldsymbol{y}_p + \boldsymbol{y}_d + \widetilde{\boldsymbol{n}}, \tag{13}$$

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

$$\boldsymbol{y}_p = \boldsymbol{\Phi}_p \boldsymbol{h}_m, \tag{14}$$

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

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{h}_m} &= \mathbb{E}\{\boldsymbol{h}_m \boldsymbol{h}_m^H\}, \\ &= \text{diag}\{\sigma^2_{h_1^{(11)}}, ..., \sigma^2_{h_P^{(11)}}, ..., \sigma^2_{h_P^{(N_r N_t)}}\}. \end{aligned} \tag{15}$$

Assuming pilot symbols are chosen independently with uniform distribution, the matrix $\boldsymbol{\Phi}_p$ has a zero mean and a covariance matrix given as follows:

$$\boldsymbol{C}_{\boldsymbol{\Phi}_p} = \mathbb{E}\{\boldsymbol{\Phi}_p \boldsymbol{\Phi}_p^H\} = N_t P \sigma_p^2 \boldsymbol{I}_{MNN_r}. \tag{16}$$

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

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

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

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

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

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

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

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

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

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

leading thus to Eq. (16). □

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

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

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

where $\boldsymbol{\mathcal{K}} \in \mathbb{C}^{MNN_r \times LN_tN_r}$ is the pilots matrix given as follows:

$$\boldsymbol{\mathcal{K}} = \begin{pmatrix} \boldsymbol{\mathcal{K}}_{11} & \boldsymbol{\mathcal{K}}_{12} & \cdots & \boldsymbol{\mathcal{K}}_{1N_t} \\ \boldsymbol{\mathcal{K}}_{21} & \boldsymbol{\mathcal{K}}_{22} & \cdots & \boldsymbol{\mathcal{K}}_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{K}}_{N_r1} & \boldsymbol{\mathcal{K}}_{N_r2} & \cdots & \boldsymbol{\mathcal{K}}_{N_rN_t} \end{pmatrix}, \tag{23}$$

with $L = (2k_\nu+1)(l_\tau+1)$, and the elements of the sub-matrices $\boldsymbol{\mathcal{K}}_{ij}$ of the matrix $\boldsymbol{\mathcal{K}}$ are calculated as follows: $\boldsymbol{\mathcal{K}}_{ij}[i', l'(2k_\nu+1)+k'+k_\nu] = \boldsymbol{P}^{(i)}[(k-k')_N, (l-l')_M]$, for $i' = 0 : MN$, $l' \in [0, l_\tau]$ and $k' \in [-k_\nu, k_\nu]$. $\boldsymbol{\Psi} \in \mathbb{C}^{MNN_r \times LN_tN_r}$ is an additional phase shift matrix given by

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} & \cdots & \boldsymbol{\Psi}_{1N_t} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} & \cdots & \boldsymbol{\Psi}_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{N_r1} & \boldsymbol{\Psi}_{N_r2} & \cdots & \boldsymbol{\Psi}_{N_rN_t} \end{pmatrix}, \tag{24}$$

where the elements of the sub-matrices $\boldsymbol{\Psi}_{ij}$ of the matrix $\boldsymbol{\Psi}$ are calculated as follows: $\boldsymbol{\Psi}_{ij}[i', l'(2k_\nu+1)+k'+k_\nu] = \exp(k'(l-l')/MN)$.

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

$$\boldsymbol{y}_m = \boldsymbol{\mathcal{H}}\boldsymbol{d} + \boldsymbol{\Phi}_p\boldsymbol{h}_m + \widetilde{\boldsymbol{n}}. \tag{25}$$

$$\boldsymbol{y}_m = \boldsymbol{\mathcal{H}}\boldsymbol{d} + \boldsymbol{A}\boldsymbol{h} + \widetilde{\boldsymbol{n}}. \tag{26}$$

3. Proposed algorithms

In this section, we detail the proposed algorithm for channel estimation and symbol detection. The first step in the proposed algorithm is the estimation of the number of channel paths P and the delay and 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 (DU) channel [28], the estimation of these parameters is done once every $T_s = N_T T$ seconds, where N_T is the number of frames in the period T_s and T is the frame duration. Only the channel gains $\{h_i\}_{i=1:N_tN_rP}$

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

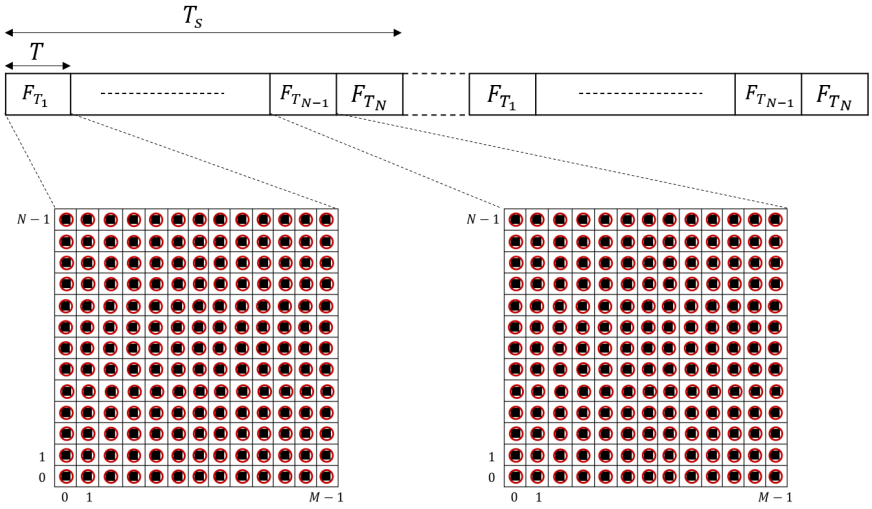


Figure 2: Frame architecture (■: data, ○: pilot).

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

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

3.1.1. Data-aided channel estimation (SoBaP algorithm)

The model (26) can be rewritten as follows:

$$\boldsymbol{y}_m = \sum_{i=1}^{LN_tN_r} s_i c_i \boldsymbol{a}_i + \widetilde{\boldsymbol{v}}_d, \tag{27}$$

where $\boldsymbol{a}_i$ is the i-th column of $\boldsymbol{A}$, $\boldsymbol{h} = \boldsymbol{s} \odot \boldsymbol{c} = [s_1c_1, s_2c_2, ..., s_{LN_tN_r}c_{LN_tN_r}]^T$, where $\boldsymbol{s} \in \{0,1\}^{LN_tN_r}$ is the support vector of channel and $\boldsymbol{c}$ denotes the vector of channel gains. $\widetilde{\boldsymbol{v}}_d = \boldsymbol{\mathcal{H}}\boldsymbol{d} + \widetilde{\boldsymbol{n}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C}_{\widetilde{\boldsymbol{v}}_d})$ (see proof) is a mixture involving the additive noise, the channel contributions and the data symbols. $\widetilde{\boldsymbol{v}}_d$ can also take the form $\widetilde{\boldsymbol{v}}_d = \boldsymbol{\Phi}_d \boldsymbol{h}_m + \widetilde{\boldsymbol{n}}$, where $\boldsymbol{\Phi}_d = \text{blkdiag}(\boldsymbol{C}_1, \boldsymbol{C}_2, ..., \boldsymbol{C}_{N_r}) \in \mathbb{C}^{MNN_r \times PN_tN_r}$ with $\boldsymbol{C}_1 = \boldsymbol{C}_2 = ... = \boldsymbol{C}_{N_r} = [\boldsymbol{\Lambda}_1\boldsymbol{d}^{(1)}, ..., \boldsymbol{\Lambda}_P\boldsymbol{d}^{(1)}, ..., \boldsymbol{\Lambda}_1\boldsymbol{d}^{(N_t)}, ..., \boldsymbol{\Lambda}_P\boldsymbol{d}^{(N_t)}]$. In the same way as $\boldsymbol{\Phi}_p$, one can easily verify that $\boldsymbol{\Phi}_d$ has a zero mean and a covariance matrix given as follows:

$$\boldsymbol{C}_{\boldsymbol{\Phi}_d} = \mathbb{E}\{\boldsymbol{\Phi}_d \boldsymbol{\Phi}_d^H\} = N_t P \sigma_d^2 \boldsymbol{I}_{MNN_r}. \tag{28}$$

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

$$\boldsymbol{m}_{\widetilde{\boldsymbol{v}}_d} = \mathbb{E}\{\widetilde{\boldsymbol{v}}_d\} = \boldsymbol{0}_{MNN_r}, \tag{29}$$

and its covariance matrix is calculated as follows:

$$\begin{aligned} \boldsymbol{C}_{\widetilde{\boldsymbol{v}}_d} &= \mathbb{E}\{\widetilde{\boldsymbol{v}}_d \widetilde{\boldsymbol{v}}_d^H\}, \\ &= \mathbb{E}\{\boldsymbol{\Phi}_d \boldsymbol{h}_m \boldsymbol{h}_m^H \boldsymbol{\Phi}_d^H\} + \mathbb{E}\{\widetilde{\boldsymbol{n}} \widetilde{\boldsymbol{n}}^H\}. \end{aligned} \tag{30}$$

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

$$\begin{aligned} \mathbb{E}\{\boldsymbol{\Phi}_d \boldsymbol{h}_m \boldsymbol{h}_m^H \boldsymbol{\Phi}_d^H\} &= \frac{\mathrm{Tr}(\boldsymbol{h}_m \boldsymbol{h}_m^H)}{PN_tN_r} \mathbb{E}\{\boldsymbol{\Phi}_d \boldsymbol{\Phi}_d^H\}, \\ &= \left(\frac{\sum_{i=1}^{PN_tN_r} \sigma_{h_i}^2}{PN_tN_r} \right) \mathbb{E}\{\boldsymbol{\Phi}_d \boldsymbol{\Phi}_d^H\}. \end{aligned} \tag{31}$$

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

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

□

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

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

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

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

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

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

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

$$p(\boldsymbol{s}) = \prod_{i=1}^{LN_tN_r} p(s_i), \tag{35}$$

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

Furthermore, in modern approaches, instead of trying to approximate the means of the posterior distributions from MCMC simulations, one tries to iteratively compute an exact variational approximation of the target posterior distribution.

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

$$(\hat{\boldsymbol{s}}, \hat{\boldsymbol{c}}) = \underset{\boldsymbol{s},\boldsymbol{c}}{\arg\max} \log p(\boldsymbol{s}, \boldsymbol{c}|\boldsymbol{y}_m). \tag{36}$$

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

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

$$\hat{\boldsymbol{s}} = \underset{\boldsymbol{s}\in\mathcal{A}}{\arg\max} \log p(\boldsymbol{s}|\boldsymbol{y}_m), \tag{37}$$

where $\mathcal{A} = \{0, 1\}^{LN_tN_r}$.

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

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

$$\hat{s}_i = \underset{s_i\in\{0,1\}}{\arg\max} \log p(s_i|\boldsymbol{y}_m). \tag{38}$$

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

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

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

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

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

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

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

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

Algorithm 1 Channel estimation algorithm

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

3.1.2. AMPA-aided data detection

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

$$\boldsymbol{y}_d = \boldsymbol{y}_m - \boldsymbol{A}\boldsymbol{h} = \hat{\boldsymbol{\mathcal{H}}}\boldsymbol{d} + \widetilde{\boldsymbol{w}}_e, \tag{42}$$

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

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

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

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

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

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

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

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

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

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

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

and

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

Finally, since $\boldsymbol{e}_h$ *and* $\widetilde{\boldsymbol{n}}$ *are statistically independent and based on (12), (47) and (48), the mean and the variance of* $\widetilde{w}_e(k)$ *are calculated as follows:*

$$m_{\widetilde{w}_e(k)} = \mathbb{E}\{e_h(k)\} + \mathbb{E}\{\widetilde{n}(k)\} = 0, \tag{49}$$

and

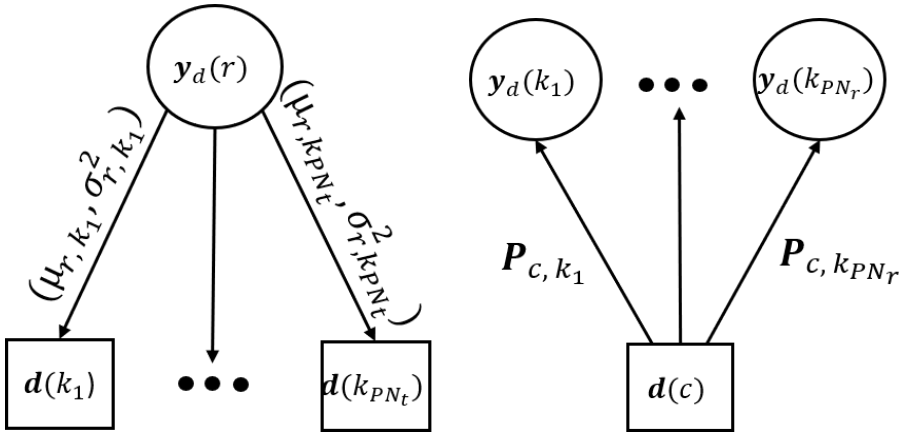
$$var\{\widetilde{w}_e(k)\} = var\{e_h(k)\} + var\{\widetilde{n}(k)\} = \sigma_n^2 + \sigma_p^2 M_h, \tag{50}$$

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

□

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

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



(a) Observation node messages. (b) Variable node messages.

Figure 3: Messages in factor graph.

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

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

$$\hat{\boldsymbol{d}} = \underset{\boldsymbol{d}\in\mathbb{A}^{NMN_r}}{\arg\max}\, p(\boldsymbol{d}|\boldsymbol{y}_d, \hat{\boldsymbol{\mathcal{H}}}), \tag{51}$$

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

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

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

for $r = 0 : MNN_t - 1$.

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

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

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

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

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

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

Using the central-limit theorem, the first term of equation (55) is approximated by a Gaussian distribution. $\widetilde{w}_e(r)$ follows also a Gaussian distribution. Therefore, ϕ_{rc} is approximated to a Gaussian random variable.

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 ϕ_{rc} passes from the observation node $\boldsymbol{y}_d(r)$ to the variable node $\boldsymbol{d}(c)$, where $c \in \mathcal{I}_r$. The major difference between the versions of MP algorithm proposed in the literature [11, 32] and the AMPA proposed here lies in the calculation of $\mu_{r,c}^{(t)}$ and $(\sigma_{r,c}^{(t)})^2$. This difference is mainly due to the two following points:

- The previous versions of the MP algorithm assume perfect knowledge of CSI, i.e., the MIMO matrix $\boldsymbol{\mathcal{H}}$ is perfectly known. In our case, we do not know $\boldsymbol{\mathcal{H}}$ but we have an estimate $\hat{\boldsymbol{\mathcal{H}}}$.

- Unlike previous versions where the interference term contains only additive noise, in our case, this term is a function of additive noise $\widetilde{\boldsymbol{n}}$ and the term $\boldsymbol{e}_h$ containing the symbols of pilots and the error of channel estimation (42)-(44). This mixture is not present in previous versions of the MP algorithm for OTFS due to the use of guard intervals between pilot symbols and data symbols, which decreases the SE of the system.

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

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

and

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

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

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

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

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

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

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

$$\hat{\boldsymbol{d}}(c) = \underset{a_i \in \mathbb{A}}{\arg\max}\ p_c^{(t)}(a_i), \tag{60}$$

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

Algorithm 2 AMPA for MIMO-OTFS data detection

Require: $\boldsymbol{y}_d$, $\hat{\boldsymbol{\mathcal{H}}}$, $\widetilde{w}_e(k)$: $m_{\widetilde{w}_e(k)}$, $var_{\widetilde{w}_e(k)}$, n_{iter}

Ensure: $\hat{\boldsymbol{d}}$

1: Iteration index $t = 0$
2: The probability mass function (pmf): $\boldsymbol{p}_{rc}^{(0)} = \frac{1}{|\mathbb{A}|}$ 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 ϕ_{rc} pass from the observation node $\boldsymbol{y}_d(r)$ to the variable node $\boldsymbol{d}(c)$
5: The pmf vector $\boldsymbol{p}_{r,c}^{(t)}$ is updated and passes as a message from the variable node $\boldsymbol{d}(c)$ to the observation node $\boldsymbol{y}_d(r)$
6: The distribution of the transmitted symbols $\boldsymbol{d}$ is updated
7: $t = t + 1$
8: **end while**

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

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

3.2.1. Data-aided channel estimation (LMMSE algorithm)

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

$$\boldsymbol{y}_m = \boldsymbol{\Phi}_d \boldsymbol{h}_m + \boldsymbol{\Phi}_p \boldsymbol{h}_m + \widetilde{\boldsymbol{n}}. \tag{61}$$

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

$$\begin{aligned} \boldsymbol{y}_m &= (\boldsymbol{\Phi}_p + \hat{\boldsymbol{\Phi}}_d^{(0)})\boldsymbol{h}_m + (\boldsymbol{\Phi}_d - \hat{\boldsymbol{\Phi}}_d^{(0)})\boldsymbol{h}_m + \widetilde{\boldsymbol{n}}, \\ &= \boldsymbol{\Phi}_{pd}^{(0)} \boldsymbol{h}_m + \boldsymbol{\mu}_d^{(0)}, \end{aligned} \tag{62}$$

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

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

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

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

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

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

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

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

Algorithm 3 LMMSE algorithm for channel estimation

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

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

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

3.2.2. AMPA-aided data detection

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

$$\widetilde{\boldsymbol{y}}_d^{(i)} = \boldsymbol{y}_m - \hat{\boldsymbol{\mathcal{H}}}^{(i)}\boldsymbol{p} = \hat{\boldsymbol{\mathcal{H}}}\boldsymbol{d} + \widetilde{\boldsymbol{w}}_{\tilde{e}}^{(i)}, \tag{66}$$

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

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: $\boldsymbol{y}_m$, $\boldsymbol{A}$, N_{iter}, $\boldsymbol{\Phi}_{pd}$, $(\boldsymbol{C}_{\boldsymbol{\mu}_d})^{-1}$, $\boldsymbol{C}_{\boldsymbol{h}_m}^{-1}$, frame index f_P

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

```
while n ≤ N_iter & |h_m^(n) − h_m^(n−1)| < ε do
    if f_P == 1 then
        Use the SoBaP algorithm (Algorithm 1): {P̂, {l̂_i, k̂_i}_{i=1:P̂}, {ĥ_i}_{i=1:P̂N_tN_r}} = SoBaP (y_m, A)
    else
        Use the LMMSE algorithm (Algorithm 3): {{ĥ_i}_{i=1:P̂N_tN_r}} = LMMSE (y_m, Φ_pd, (C_μd)^{-1}, C_hm^{-1})
    end if
    Compute: y_d, Ĥ, w̃_e(k) : m_{w̃_e(k)}, var_{w̃_e(k)}
    Use the AMPA (Algorithm 2): d̂ = MAP (y_d, Ĥ, w̃_e(k) : m_{w̃_e(k)}, var_{w̃_e(k)})
end while
```

4. Optimal power distribution between pilots and data symbols

In the proposed scheme, the pilots and data symbols are superimposed in the same DD locations. Therefore, the power must be optimally distributed between the pilots and data symbols to maximise the SINR. Consequently, the maximization of SINR leads to the maximization of SE and the minimization of BER [33]. In what follows, we will derive the SINR lower bound which will be maximised to obtain the optimal 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 consider one iteration of the proposed algorithm. Eq. (66) can be rewritten as follows:

$$\begin{aligned}
\widetilde{\boldsymbol{y}}_d &= \boldsymbol{y}_m - \boldsymbol{\Phi}_p \hat{\boldsymbol{h}}_m, \\
&= \boldsymbol{\Phi}_d \boldsymbol{h}_m + \boldsymbol{\Phi}_p \left(\boldsymbol{h}_m - \hat{\boldsymbol{h}}_m\right) + \widetilde{\boldsymbol{n}}, \\
&= \boldsymbol{\Phi}_d \hat{\boldsymbol{h}}_m + (\boldsymbol{\Phi}_d + \boldsymbol{\Phi}_p) \left(\boldsymbol{h}_m - \hat{\boldsymbol{h}}_m\right) + \widetilde{\boldsymbol{n}}, \\
&= \boldsymbol{\Phi}_d \hat{\boldsymbol{h}}_m + \widetilde{\boldsymbol{w}}_e,
\end{aligned} \tag{67}$$

where $\widetilde{\boldsymbol{w}}_e = (\boldsymbol{\Phi}_d + \boldsymbol{\Phi}_p)\,\tilde{\boldsymbol{h}}_m + \widetilde{\boldsymbol{n}}$ and $\tilde{\boldsymbol{h}}_m = \left(\boldsymbol{h}_m - \hat{\boldsymbol{h}}_m\right)$.

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

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

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

Equation (68) can be rewritten as

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$M_h\geq\frac{(PN_tN_r)^2}{Tr((\boldsymbol{\Phi}_p)^H(\boldsymbol{C}_{\boldsymbol{\mu}_d})^{-1}\boldsymbol{\Phi}_p+\boldsymbol{C}_{\boldsymbol{h}_m}^{-1})}. \tag{79}$$

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

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

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

$$Tr((\boldsymbol{\Phi}_p)^H\boldsymbol{\Phi}_p)=PMNN_tN_r\sigma_p^2. \tag{81}$$

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

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

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

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

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

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

where $a = b_1a_2 - b_2a_1$, $b = 2b_0a_2 - 2b_2a_0$ and $c = b_0a_1 - b_1a_0$.

5. Complexity analysis

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

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

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

For the channel estimation step, we have two proposed algorithms. For the SoBaP algorithm, the most complex operation per iteration is the update (A.9), which is $\mathcal{O}(MNLN_a^3)$. Therefore, the overall complexity 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 required to the convergence of the SoBaP algorithm.

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

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 channel estimation step reduced to $C_{ce} = \mathcal{O}(\frac{n_{iter}^{(ce)} MNL}{N_f} N_a^3)$.

Table 2 summarizes the complexity of the proposed channel estimation algorithm and that of each of the 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	$\mathcal{O}(\frac{n_{iter}^{(ce)} MNL}{N_f} N_a^3)$

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 the method m and we compare the complexities of all the methods.

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

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 $N_{iter} L K^2 < G^3$, which leads to $C_{\text{BSBL-BR}} < C_{\text{RG-BL}}$.

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 $C_{\text{SoBaP-LMMSE-AMPA}} < C_{\text{BSBL-BR}}$. We can conclude that $C_{EP} < C_{\text{SoBaP-LMMSE-AMPA}} < C_{\text{BSBL-BR}} < C_{\text{RG-BL}} < C_{\text{RG-OMP}}$.

6. Simulation results

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

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

6.1. Simulation setting

The simulation parameters are given in Table 3.

Table 3: Simulation parameters.

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

For the channel delay model, we use a 5-tap DD model whose parameters are given in Table 4. Each delay tap owns a single Doppler shift in the form $\nu_k = \nu_{max}\cos(\theta_k)$, where ν_{max} is the maximum Doppler shift of the channel and $\theta_k \sim \mathcal{U}_{[0,\pi]}$. The maximum delay tap $l_\tau = 4$ and the maximum Doppler tap $k_\nu = 2$ 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 (μs)	2.08	5.20	8.33	11.46	20.8
Path power (dB)	1	-1.80	-3.57	-5.38	-8.86

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

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

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

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

parameters, the average optimal pilot power is approximately 0.3. To maximize the SE and minimize the BER of the proposed superimposed-aided design, it is recommended to allocate 30% of the total power to 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
$\sigma^2_{p,opt}$	0.2520	0.2995	0.3135	0.3178	0.3191

6.2. Stationarity of the channel

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

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

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 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, $T_s = 10.3\ ms$, and $T = 1\ ms$, so, $N_f \approx 10$. It is worth noting that, for constant N, Δf and δ, the number of frames N_f is proportional to v_{max}. Thus, for speeds below 500 km/h, N_f increases.

6.3. Influence of the power of pilots on BER performance

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

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

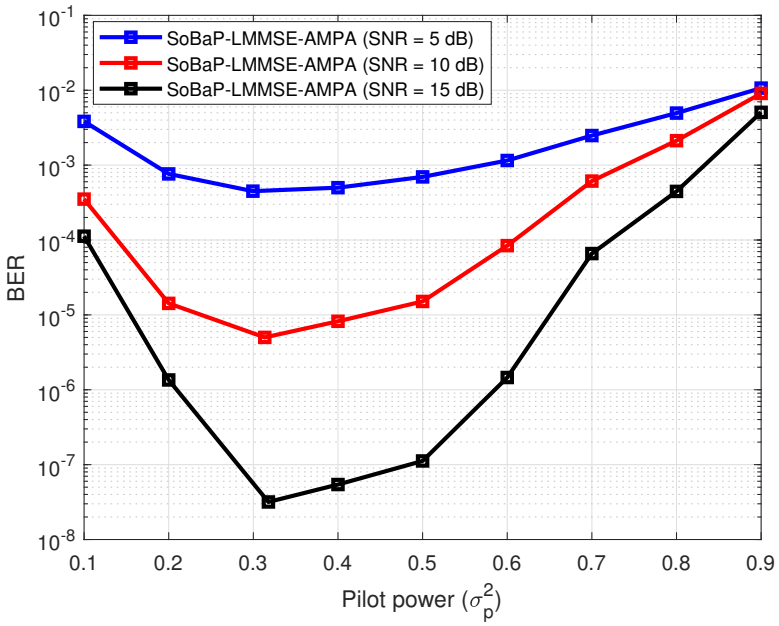


Figure 4: BER versus pilot power performance for the proposed algorithm with $N_t = N_r = 2$, $N = M = 16$ and SNR = 5, 10, 15 dB.

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

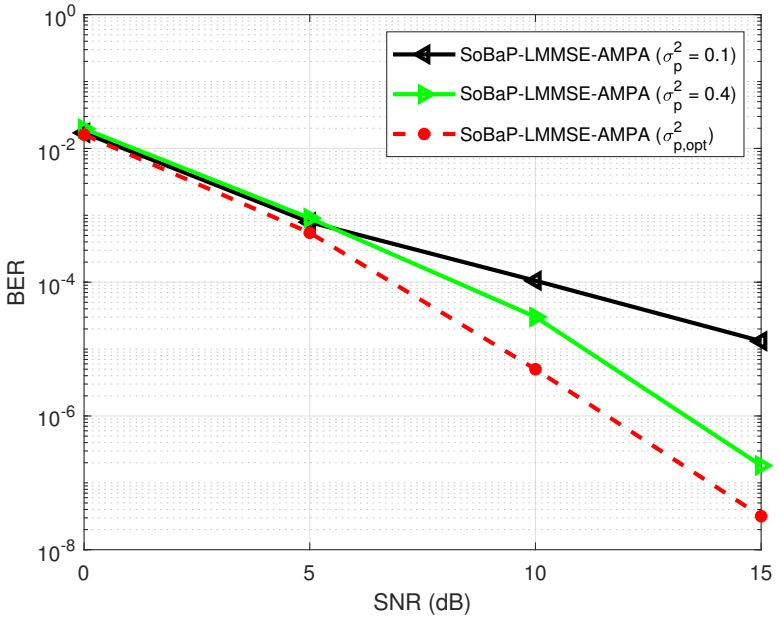


Figure 5: Effect of pilot power on the BER of the proposed scheme with $N_t = N_r = 2$, $N = M = 16$.

6.4. Convergence of the SoBaP-LMMSE-AMPA algorithm

Fig 6 shows the BER performance of the proposed SoBaP-LMMSE-AMPA algorithm versus the number 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, for both SNR values, the BER decreases with increasing number of iterations, and it saturates after about 5 or 6 iterations. The convergence and the effectiveness of the proposed SoBaP-LMMSE-AMPA algorithm are confirmed by the steady states reached by the two curves of the BER after a few iterations. It can be also observed that, as expected, the scenario with higher SNR exhibits superior convergence speed and detection accuracy.

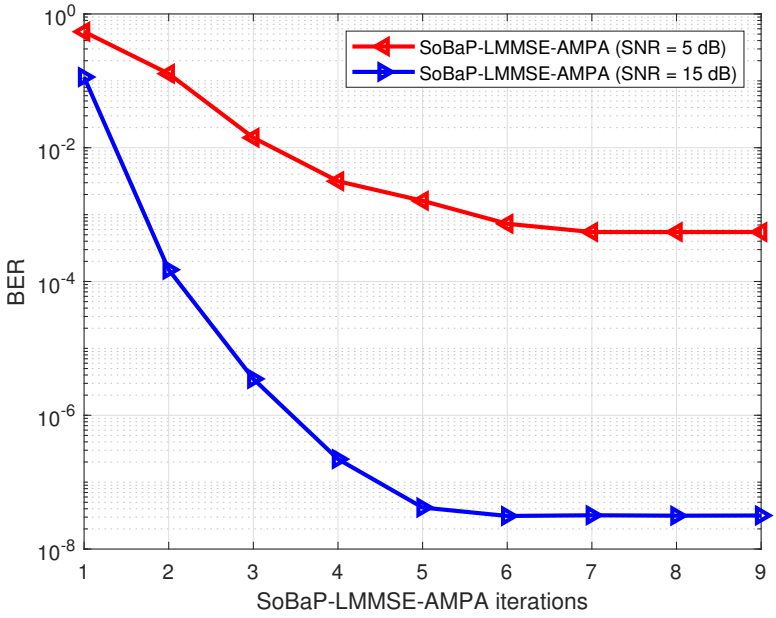


Figure 6: BER performance of the SoBaP-LMMSE-AMPA algorithm versus the number of iterations with $N_t = N_r = 2$, $N = M = 16$, $\sigma_p^2 = \sigma_{p,opt}^2$ and SNR = 5, 15 dB.

6.5. Channel NMSE performance

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

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

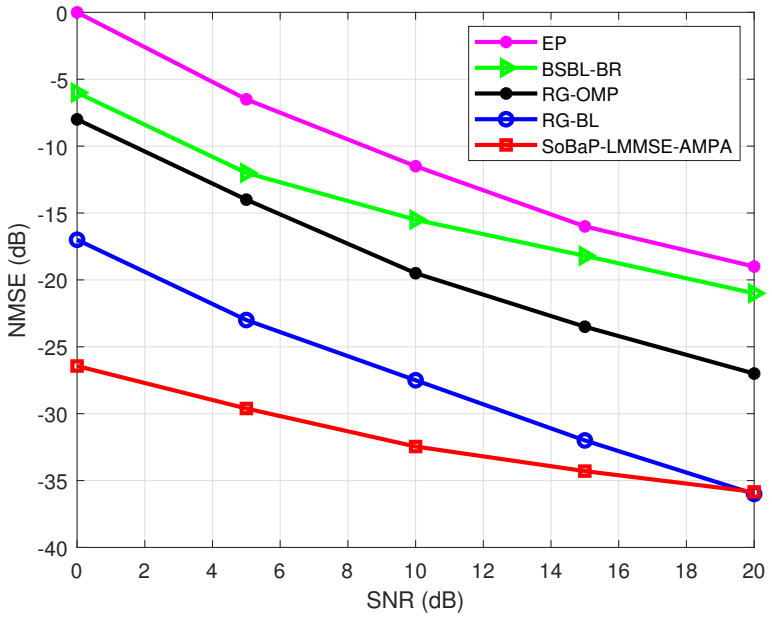


Figure 7: NMSE versus SNR performance for the MIMO-OTFS system with $N_t = N_r = 2$, $N = M = 16$, and $\sigma_p^2 = \sigma_{p,opt}^2$.

6.6. BER performance evaluation

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

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

6.7. SE performance evaluation

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

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

where $s =$ EP, BSBL-BR, RG-OMP, RG-BL, SoBaP-LMMSE-AMPA, η_s denotes the pilot overhead related to the scheme s.

The pilot overhead η_s is calculated using the frame-structure for the scheme s. The SINR expressions for all the schemes are derived in the same way as in [22]. Table 6 shows the expressions of SINR_s and η_s for each scheme s.

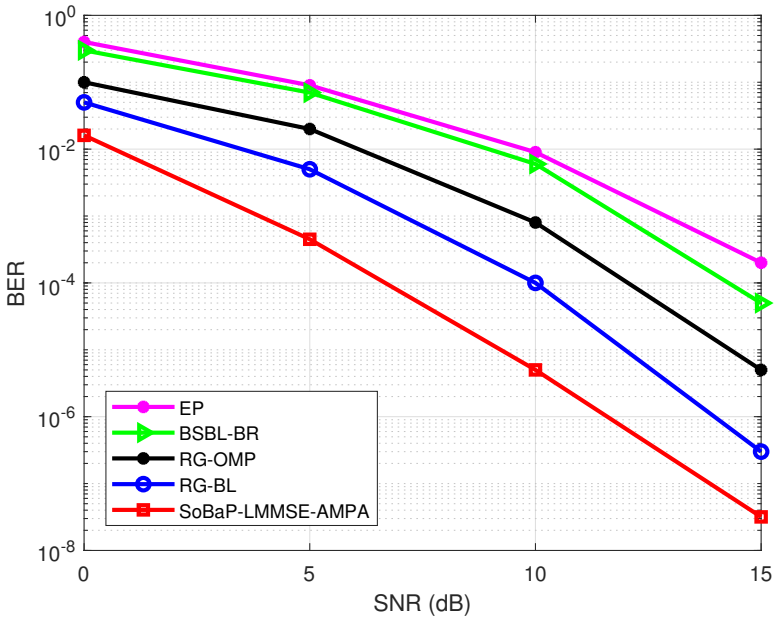


Figure 8: BER versus SNR performance for the MIMO-OTFS system with $N_t = N_r = 2$, $N = M = 16$, and $\sigma_p^2 = \sigma_{p,opt}^2$.

Table 6: Expressions of SINR_s and η_s.

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

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 RG-BL schemes and $M_{h,s}$ is the channel MSE of scheme s.

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

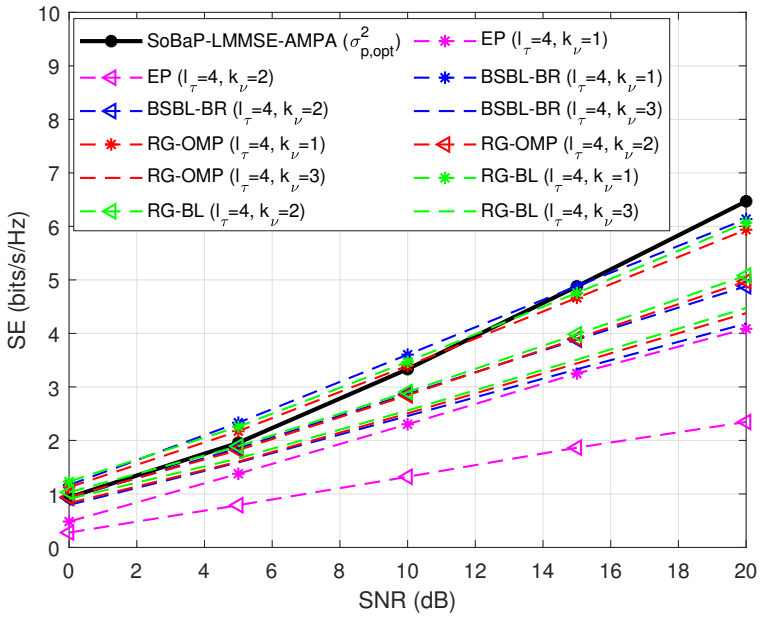


Figure 9: SE versus SNR performance for the MIMO-OTFS system with $N_t = N_r = 2$ and $N = M = 16$.

7. Conclusion

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

Appendix A. Calculation of the mean field approximation

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

$$q^\star(\boldsymbol{\theta}) = \arg\min_{q(\boldsymbol{\theta})} \left\{ \int_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) \log \left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{y}_m)} \right) \mathrm{d}\boldsymbol{\theta} \right\}, \tag{A.1}$$

subject to

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

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

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

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

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

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

$$q(c_i, s_i|\boldsymbol{y}_m) = q(c_i|s_i, \boldsymbol{y}_m) q(s_i|\boldsymbol{y}_m), \tag{A.4}$$

where

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

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

and

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

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

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

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

$$p(s_i|\boldsymbol{y}_m) \simeq \int q(c_i, s_i|\boldsymbol{y}_m)\mathrm{d}c_i = q(s_i). \tag{A.10}$$

Appendix B. Analysis for positive definiteness of matrix M

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

The matrix $\boldsymbol{M}$ is given by $\boldsymbol{M} = ((\boldsymbol{\Phi}_{pd})^H(\boldsymbol{C}_{\boldsymbol{\mu}_d})^{-1}\boldsymbol{\Phi}_{pd} + \boldsymbol{C}_{\boldsymbol{h}_m}^{-1})$, where $\boldsymbol{C}_{\boldsymbol{\mu}_d} = \lambda\boldsymbol{I}_{MNN_r}$, with λ is a positive number given by $\lambda = \left(\sigma_n^2 + \frac{2\sigma_d^2}{N_r}\sum_{i=1}^{PN_tN_r}\sigma_{h_i}^2\right)$ and $\boldsymbol{C}_{\boldsymbol{h}_m} = \mathrm{diag}\{\sigma_{h_1}^2, \sigma_{h_2}^2, ..., \sigma_{h_{PN_tN_r}}^2\}$. For all non-zero $\boldsymbol{z} \in \mathbb{C}^{PN_tN_r\times 1}$, we have

$$\begin{aligned}
\boldsymbol{z}^H\boldsymbol{M}\boldsymbol{z} &= \boldsymbol{z}^H((\boldsymbol{\Phi}_{pd})^H(\boldsymbol{C}_{\boldsymbol{\mu}_d})^{-1}\boldsymbol{\Phi}_{pd}^{(i-1)} + \boldsymbol{C}_{\boldsymbol{h}_m}^{-1})\boldsymbol{z},\\
&= \boldsymbol{z}^H(\boldsymbol{\Phi}_{pd})^H(\lambda\boldsymbol{I}_{MNN_r})^{-1}\boldsymbol{\Phi}_{pd}\boldsymbol{z} + \boldsymbol{z}^H\left(\mathrm{diag}\{\sigma_{h_1}^2, \sigma_{h_2}^2, ..., \sigma_{h_{PN_tN_r}}^2\}\right)^{-1}\boldsymbol{z},\\
&= \frac{1}{\lambda}\boldsymbol{z}^H(\boldsymbol{\Phi}_{pd})^H\boldsymbol{\Phi}_{pd}\boldsymbol{z} + \sum_{i=1}^{PN_tN_r}\frac{|z_i|^2}{\sigma_{h_i}^2},\\
&= \frac{1}{\lambda}(\boldsymbol{\Phi}_{pd}\boldsymbol{z})^H\boldsymbol{\Phi}_{pd}\boldsymbol{z} + \sum_{i=1}^{PN_tN_r}\frac{|z_i|^2}{\sigma_{h_i}^2},\\
&= \frac{1}{\lambda}\|\boldsymbol{\Phi}_{pd}\boldsymbol{z}\|^2 + \sum_{i=1}^{PN_tN_r}\frac{|z_i|^2}{\sigma_{h_i}^2}.
\end{aligned} \tag{B.1}$$

We have $\frac{1}{\lambda}\|\boldsymbol{\Phi}_{pd}\boldsymbol{z}\|^2 + \sum_{i=1}^{PN_tN_r}\frac{|z_i|^2}{\sigma_{h_i}^2} > 0$, which means that $\boldsymbol{M}$ is a positive definite matrix.

References

[1] P. Singh, H. B. Mishra, R. Budhiraja, Low-complexity linear MIMO-OTFS receivers, in: 2021 IEEE International Conference on Communications Workshops (ICC Workshops), IEEE, 2021, pp. 1–6.

[2] K. Xu, Z. Shen, Y. Wang, X. Xia, Location-aided mMIMO channel tracking and hybrid beamforming for high-speed railway communications: An angle-domain approach, IEEE Systems Journal 14 (1) (2019) 93–104.

[3] P. Singh, A. Gupta, H. B. Mishra, R. Budhiraja, Low-complexity ZF/MMSE MIMO-OTFS receivers for high-speed vehicular communication, IEEE Open Journal of the Communications Society 3 (2022) 209–227.

[4] A. Monk, R. Hadani, M. Tsatsanis, S. Rakib, OTFS-orthogonal time frequency space, arXiv preprint arXiv:1608.02993.

[5] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, R. Calderbank, Orthogonal time frequency space modulation, in: 2017 IEEE Wireless Communications and Networking Conference (WCNC), IEEE, 2017, pp. 1–6.

[6] G. Surabhi, R. M. Augustine, A. Chockalingam, On the diversity of uncoded OTFS modulation in doubly-dispersive channels, IEEE transactions on wireless communications 18 (6) (2019) 3049–3063.

[7] Z. Ding, R. Schober, P. Fan, H. V. Poor, OTFS-NOMA: An efficient approach for exploiting heterogenous user mobility profiles, IEEE Transactions on Communications 67 (11) (2019) 7950–7965.

[8] R. M. Augustine, G. Surabhi, A. Chockalingam, Space-time coded OTFS modulation in high-doppler channels, in: 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), IEEE, 2019, pp. 1–6.

[9] G. Surabhi, A. Chockalingam, Low-complexity linear equalization for $2\times$ 2 MIMO-OTFS signals, in: 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), IEEE, 2020, pp. 1–5.

[10] M. Li, S. Zhang, F. Gao, P. Fan, O. A. Dobre, A new path division multiple access for the massive MIMO-OTFS networks, IEEE journal on selected areas in communications 39 (4) (2020) 903–918.

[11] M. K. Ramachandran, A. Chockalingam, MIMO-OTFS in high-doppler fading channels: Signal detection and channel estimation, in: 2018 IEEE Global Communications Conference (GLOBECOM), IEEE, 2018, pp. 206–212.

[12] W. Shen, L. Dai, J. An, P. Fan, R. W. Heath, Channel estimation for orthogonal time frequency space (OTFS) massive MIMO, IEEE Transactions on Signal Processing 67 (16) (2019) 4204–4217.

[13] Y. Liu, S. Zhang, F. Gao, J. Ma, X. Wang, Uplink-aided high mobility downlink channel estimation over massive MIMO-OTFS system, IEEE Journal on Selected Areas in Communications 38 (9) (2020) 1994–2009.

[14] X. Wu, S. Ma, X. Yang, Tensor-based low-complexity channel estimation for mmwave massive MIMO-OTFS systems, Journal of Communications and Information Networks 5 (3) (2020) 324–334.

[15] D. Shi, W. Wang, L. You, X. Song, Y. Hong, X. Gao, G. Fettweis, Deterministic pilot design and channel estimation for downlink massive MIMO-OTFS systems in presence of the fractional doppler, IEEE Transactions on Wireless Communications 20 (11) (2021) 7151–7165.

[16] T. Li, C. Han, R. Yao, Y. Fan, X. Zuo, Low pilot overhead channel estimation for CP-OFDM-based massive MIMO OTFS system, IET Communications 16 (10) (2022) 1071–1082.

[17] S. Srivastava, R. K. Singh, A. K. Jagannatham, L. Hanzo, Bayesian learning aided simultaneous row and group sparse channel estimation in orthogonal time frequency space modulated MIMO systems, IEEE Transactions on Communications 70 (1) (2021) 635–648.

[18] L. Zhao, J. Yang, Y. Liu, W. Guo, Block sparse bayesian learning-based channel estimation for MIMO-OTFS systems, IEEE Communications Letters 26 (4) (2022) 892–896.

[19] Y. Liu, Y. L. Guan, D. González, Near-optimal BEM OTFS receiver with low pilot overhead for high-mobility communications, IEEE Transactions on Communications 70 (5) (2022) 3392–3406.

[20] F. Zhang, W. Ji, L. Qiu, Channel estimation for massive MIMO-OTFS systems via sparse bayesian learning with 2-D local beta process, in: 2022 IEEE Wireless Communications and Networking Conference (WCNC), IEEE, 2022, pp. 1383–1388.

[21] R. Muzavazi, O. O. Oyerinde, Channel estimation and data detection schemes for orthogonal time frequency space massive MIMO systems, Computers and Electrical Engineering 102 (2022) 108215.

[22] H. B. Mishra, P. Singh, A. K. Prasad, R. Budhiraja, OTFS channel estimation and data detection designs with superimposed pilots, IEEE transactions on wireless communications 21 (4) (2021) 2258–2274.

[23] W. Yuan, S. Li, Z. Wei, J. Yuan, D. W. K. Ng, Data-aided channel estimation for OTFS systems with a superimposed pilot and data transmission scheme, IEEE wireless communications letters 10 (9) (2021) 1954–1958.

[24] Y. Liu, Y. L. Guan, D. González, BEM OTFS receiver with superimposed pilots over channels with doppler and delay spread, in: ICC 2022-IEEE International Conference on Communications, IEEE, 2022, pp. 2411–2416.

[25] P. Raviteja, K. T. Phan, Y. Hong, Embedded pilot-aided channel estimation for OTFS in delay–doppler channels, IEEE transactions on vehicular technology 68 (5) (2019) 4906–4917.

[26] R. Ouchikh, A. Aïssa-El-Bey, T. Chonavel, M. Djeddou, Sparse channel estimation algorithms for OTFS system, IET Communications.

[27] R. Ouchikh, A. Aïssa-El-Bey, T. Chonavel, M. Djeddou, Iterative channel estimation and data detection algorithm for OTFS modulation, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, pp. 5263–5267.

[28] G. Matz, On non-WSSUS wireless fading channels, IEEE transactions on wireless communications 4 (5) (2005) 2465–2478.

[29] P. Singh, H. B. Mishra, A. K. Jagannatham, K. Vasudevan, Semi-blind, training, and data-aided channel estimation schemes for MIMO-FBMC-OQAM systems, IEEE Transactions on Signal Processing 67 (18) (2019) 4668–4682.

[30] C. Herzet, A. Drémeau, Bayesian pursuit algorithms, in: 2010 18th European Signal Processing Conference, IEEE, 2010, pp. 1474–1478.

[31] O. Rabaste, T. Chonavel, Estimation of multipath channels with long impulse response at low SNR via an MCMC method, IEEE Transactions on Signal Processing 55 (4) (2007) 1312–1325.

[32] P. Raviteja, K. T. Phan, Y. Hong, E. Viterbo, Interference cancellation and iterative detection for orthogonal time frequency space modulation, IEEE transactions on wireless communications 17 (10) (2018) 6501–6515.

[33] P. Singh, H. B. Mishra, A. K. Jagannatham, K. Vasudevan, L. Hanzo, Uplink sum-rate and power scaling laws for multi-user massive MIMO-FBMC systems, IEEE Transactions on Communications 68 (1) (2019) 161–176.

[34] S. M. Kay, Fundamentals of statistical signal processing: estimation theory, Prentice-Hall, Inc., 1993.

[35] Z. Ulukök, R. Türkmen, On some matrix trace inequalities, Journal of Inequalities and Applications 2010 (2010) 1–8.

[36] A. Drémeau, C. Herzet, L. Daudet, Boltzmann machine and mean-field approximation for structured sparse decompositions, IEEE Transactions on Signal Processing 60 (7) (2012) 3425–3438.

[37] M. J. Beal, Variational algorithms for approximate Bayesian inference, University of London, University College London (United Kingdom), 2003.

[38] T. Van Erven, P. Harremos, Rényi divergence and kullback-leibler divergence, IEEE Transactions on Information Theory 60 (7) (2014) 3797–3820.

[39] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society: Series B (Methodological) 39 (1) (1977) 1–22.

[40] T. P. Minka, A family of algorithms for approximate bayesian inference, Ph.D. thesis, Massachusetts Institute of Technology (2001).

[41] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West, et al., The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures, Bayesian statistics 7 (453-464) (2003) 210.

[42] C. M. Bishop, N. M. Nasrabadi, Pattern recognition and machine learning, Vol. 4, Springer, 2006.

[43] R. A. Horn, C. R. Johnson, Matrix analysis, Cambridge university press, 2012.