# Joint channel estimation and data detection for high rate OTFS systems

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

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

# Joint channel estimation and data detection for high rate OTFS systems†

Rabah Ouchikh*[1] | Thierry Chonavel[2] | Abdeldjalil Aïssa-El-Bey[2] | Mustapha Djeddou[3]

[1]Laboratoire Télécommunications, Ecole Militaire Polytechnique, Bordj El-Bahri, Algeria
[2]Lab-STICC, UMR CNRS 6285, F-29238 Brest, IMT Atlantique, Brest, France
[3]Department of Electronics, National Polytechnic School, Algiers, Algeria

**Correspondence**
*Rabah Ouchikh, Email: ouchikh16rabah@gmail.com

**Summary**

This paper proposes a new pilot pattern in the Delay-Doppler (DD) domain for the Orthogonal Time Frequency Space (OTFS) system. In contrast to the embedded-pilot schemes, guard intervals are not used so as increase the spectral efficiency. Also, compared to the superimposed design where data symbols and pilots are arranged on the entire DD grid, in the proposed rearrangement, the number of pilots used is only spread over a sub-grid of the DD grid. Hence, the interference of pilots with data symbols is reduced. Afterwards, an algorithm for channel estimation (CE) and symbol detection in the DD domain benefiting from the sparsity of the DD channel is designed. The sparse CE step is formulated as a specific marginalization of the maximum a posteriori (MAP) criterion by providing a Bayesian approach via the MF approximation that involves the VB-EM algorithm. Detection of data symbols is done using a low complexity MP algorithm. We also propose an interference cancellation (IC) scheme to mitigate contamination of data by pilots that is run after each CE step. To achieve a high CE accuracy, based on the mean mutual incoherence property (MIP), a pilot optimization problem for OTFS is formulated, and develop a Simulated Annealing-based algorithm to solve it. Finally, simulation results show that the proposed scheme achieves a good compromise between spectral efficiency, complexity, and performance in terms of Bit Error Rate (BER) and Normalized Mean Square Error (NMSE) when compared to literature benchmarks.

**KEYWORDS:**
OTFS, channel estimation, Bayesian approach, pilot sequence optimization, SA, IC scheme

## 1 | INTRODUCTION

Future mobile communication systems generation (5G and beyond) are envisioned to support reliable and high throughput communication even in high mobility scenarios such as in high-speed trains and Unmanned Aerial Vehicle (UAV) communication systems [1,2,3]. However, the most popular modulation technology deployed in 5G mobile communication systems is based on Orthogonal Frequency Division Multiplexing (OFDM) which suffers from performance deterioration in high mobility environments [2]. Hence, we need other modulations techniques to deal with this problem. The OTFS modulation which has been recently proposed [4,5,6] is a promising solution thanks to its robustness against channel-induced Doppler shift compared to OFDM. OTFS

†Joint channel estimation and data detection for high rate OTFS systems.
[0]**Abbreviations:** OTFS, orthogonal time frequency and space. VB-EM, variational Bayesian expectation-maximization. SA, simulated annealing. MF, mean-field

multiplexes data symbols onto 2D orthogonal basis functions in the Delay-Doppler (DD) domain to deal with time-varying multipath channel dynamics at high speeds. OTFS modulation can be realized by adding preprocessing and post-processing blocks to OFDM. In practice, efficient CE and data detection algorithms are crucial for the successful realization of OTFS systems[7]. Several CE schemes in the DD domain have been recently developed in the OTFS literature. We present DD-CE algorithms reported in the literature under three groups.

The first group of algorithms uses schemes that employ an entire frame for transmitting pilots. This type of scheme uses the first OTFS frame for estimating the channel and subsequent frames for transmitting data[1,8,9,10,11]. In Zhang et al.,[8] a 2D Turbo compressed sensing (CS) algorithm for CE is proposed where the support matrix of the DD channel is modeled by using a Markov random field (MRF) and the Bernoulli Gaussian (BG) distribution. In Rasheed et al.,[9] an uplink DD-CE algorithm is developed for OTFS multiple access (OTFS-MA) systems. In this last paper, the CE problem is formulated as a sparse recovery problem that is solved using orthogonal matching pursuit (OMP) and modified subspace pursuit (MSP) algorithms. In Khan and Mohammed,[1] two low complexity algorithms for CE in OTFS with fractional delay and Doppler are proposed. The first one is based on a Modified Maximum Likelihood CE (M-MLE) while the second one is termed the two step method (TSE), where the joint 2D estimation of the DD shift of a path in the M-MLE method is further decoupled into two separate 1D estimation steps. In Gómez-Cuba,[10] an OMP with Binary-division Refinement (OMPBR) algorithm is suggested for CS-based OTFS-CE with continuous parameter estimation in the DD domain. Whereas, in Murali and Chockalingam,[11] a pseudo-random noise (PN) pilot-based CE scheme in the DD domain is developed, benefiting from the sparse DD impulse response which reflects the actual physical geometry of the wireless channel.

A second group of algorithms involves embedding both pilots and data symbols in the same OTFS frame by providing guard intervals around pilots to avoid data interference in the CE process[12,13,14,15,16,17,18,19,20]. In Raviteja et al.,[12] data symbols, pilots, and guard intervals are arranged adequately in the DD grid to avoid interference at the receiver side. Also, the used CE scheme is based on a thresholding method. In Ramachandran and Chockalingam,[13] a CE scheme using pulses in the DD domain as pilots adapted for MIMO-OTFS is proposed. The equivalent MIMO-OTFS channel matrix is obtained using a single MIMO-OTFS frame thanks to sufficient spacing between pilots and data in the DD domain. In Shen et al.,[14] a 3D-structured OMP-based algorithm is used for downlink CE that exploits the 3D-structured channel sparsity in the Delay-Doppler-Angle domain in Massive-MIMO-OTFS systems. In Liu et al.,[15] an uplink-aided high mobility downlink CE scheme for the Massive-MIMO-OTFS networks is suggested. MIMO-OTFS signal model along the uplink and reciprocity between the uplink and the downlink are formulated to recover the parameters of the channel for each physical scattering path. In Zhao et al.,[16] a new pilot pattern and a sparse Bayesian learning (SBL)-based CE algorithm is developed to minimize pilot overhead and pilot power. In this last paper, a sparse signal prior model is constructed as a hierarchical Laplace prior and the expectation-maximisation (EM) algorithm is used to update the parameters in this prior model. In Qu et al.,[17] a low-dimensional subspace estimation of continuous Doppler-spread channel in OTFS systems is proposed, while a set of transform domain basis functions is designed to span a low-dimensional subspace for modeling the OTFS channel. Then, the CE is performed by estimating a few projection coefficients of ECRs in the developed subspace, with training pilots. In Zhao et al.,[18] the CE problem for MIMO-OTFS is formulated as a block sparse signal recovery problem, which is solved by the proposed block SBL with block reorganization (BSBL-BR) method. In Guo et al.,[19] a joint iterative CE and data detection algorithm for OTFS and a novel MMSE-DD domain channel estimator is constructed. In Liu et al.,[20] a prior channel statistics-based scheme is suggested to maximize the system ergodic capacity by optimizing the CE overhead while ensuring the high-quality performance of the OTFS over DD channels.

The last group of CE strategies adopts a superimposed pilot scheme, where pilots and data symbols are spread in the DD domain[21,22,23]. In Yuan et al.[21], a data-aided CE algorithm for a superimposed pilot and data transmission scheme is suggested to improve the SE. To accurately estimate the channel and detect the data symbols, the channel is coarsely estimated based on the pilot symbol, followed by an iterative process that detects the data symbols and refines the channel estimate. In Yuan et al.[23], the pilot pattern proposed in Yuan et al.[21] is adapted to uplink channel estimation for the ISAC-assisted OTFS modulation. In Mishra et al.,[22] pilots and data symbols are superimposed in the DD grid. This configuration disperses pilots energy over the entire DD domain and achieves a higher SE compared to previous systems using GIs.

This paper is an extended version of the paper by Ouchikh R et al.[24] The main difference is the adopted pilot pattern. The pattern suggested in this manuscript belongs to the third group cited above where pilots and data symbols are arranged in the same DD frame but share only a sub-grid of the entire DD grid. Whereas, in Ouchikh R et al.[24] an algorithm for channel estimation and symbol detection is suggested for an arrangement of pilots and data in each location of the DD grid. The proposed configuration in this paper allows some of the data to be unaffected by pilots, thus increasing the performance of CE and data detection, especially at high SNRs. We then investigate an iterative CE and data detection algorithm in the DD domain. The proposed

algorithm, which benefits from channel sparsity in the DD domain, iterates between message passing-aided data detection and data-aided CE. The CE step is based on a Bayesian approach via the VB-EM algorithm. Pilot contamination in the proposed pilot pattern is addressed via an interference cancellation step. Finally, we address pilot design as an optimization problem and develop a simulated annealing-based algorithm to solve it where the optimization criterion is the average MIP criteria. The main contributions can be summarized as follows:

- We propose a novel pilot pattern in the DD domain for the OTFS system, without guard intervals between pilots and data symbols to improve spectral efficiency (SE). In this scheme the pilots are superimposed on the data symbols in a delay-Doppler sub-grid. The idea here is to let the pilots and data symbols interfere with each other to eliminate pilot overhead and thus increase SE.
- To achieve high CE accuracy, we formulate the pilot design as an optimization problem based on the average MIP criterion and develop an SA-based algorithm to solve it. This optimization problem is non-convex, which means that conventional convex optimization algorithms are not guaranteed to reach a global minimum. This led us to take advantage of the SA algorithm to solve this problem.
- An iterative CE and data detection algorithm in the DD domain is designed where the CE step is formulated as the marginalization of the MAP problem by providing a Bayesian approach. The data detection step is based on a low complexity MP algorithm. The detection step of this algorithm is the only common point between the current paper and the one in Ouchikh et al.[24]. For the CE step, it is based on the VB-EM algorithm in both works but in the present paper is done using only a sub-grid of the entire DD grid, contrary to Ouchikh R et al[24] where the whole OTFS grid is used. Thus, this allows us to gain in terms of complexity in the CE step.
- To address the pilot contamination issue, an IC scheme is adopted in the proposed pilot pattern. This scheme is performed at each iteration after the channel estimation step to remove the effect of pilots on the data symbols. The iterative process containing channel estimation, IC scheme and symbol detection results in a good channel estimation and thus a good accuracy in data detection after a few iterations.
- Performance study is conducted, in terms of SE, BER, NMSE and complexity. The obtained results are compared to recently proposed benchmarks in the literature. We highlight the good compromise in terms of SE, NMSE, BER, and complexity achieved when compared to other solutions.

The remainder of the article is organized as follows. Section 2 describes the basic concepts of OTFS modulation by giving the input/output equations of an OTFS system. The problem formulation is also introduced in this section. Section 3 presents the SA-based pilot optimization algorithm for OTFS. In Section 4, the proposed CE and data detection algorithm is described. The performance and complexity analysis of the proposed scheme is evaluated in Section 5 by various experiments. Finally, conclusions are given in Section 6.

**Notations**: $a$, $\boldsymbol{a}$ and $\boldsymbol{A}$ denote a scalar, a vector and a matrix, respectively. $[.]_N$ denotes the modulo-$N$ operator. Symbols $\otimes$ and $\odot$ denote Kronecker and Hadamard products, respectively. Operator vec(.) represents the vectorization of a matrix with shape $(m, n)$ into an $mn$ vector, where the shape $(m, n)$ is given by the context. The circular symmetric complex Gaussian distribution with mean $m$ and variance $\sigma^2$ is denoted by $\mathcal{CN}(m, \sigma^2)$. The notation $\text{diag}[d_1, d_2, ..., d_N]$ denotes an $N \times N$ diagonal matrix with entry $(i, i)$ equal to $d_i$. The superscript $(.)^H$ denotes Hermitian transposition. $\mathbb{E}\{.\}$ represents the expectation operation and $\delta(.)$ is the Dirac-delta function. $\mathcal{U}_A$ represents uniform distribution over $A$. Finally, $\boldsymbol{I}_N$, $\boldsymbol{F}_n$, and $\boldsymbol{F}_n^H$ represent the $N \times N$ identity matrix, the $n$-point DFT and IDFT matrices.

## 2 | SYSTEM MODEL

In this section, we first review the basic concepts of OTFS. Next, we detail the input/output equations of the OTFS transmitter, channel response in the DD domain, and OTFS receiver. Then, we illustrate the proposed pilot pattern. Finally, we formulate the optimization of pilots, CE and data detection problems. Fig. 1 shows the block diagram of OTFS system.

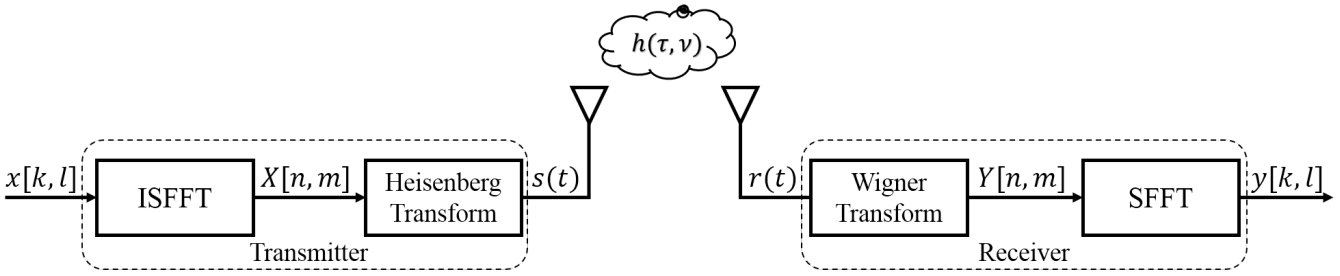


**FIGURE 1** OTFS modulation block diagram: transmitter, channel and receiver.

## 2.1 | OTFS Transmitter

Let $G_{DD}$, $G_{TF}$ and $A_{p_{tx},p_{rx}}$ denote the DD grid, the Time-Frequency (TF) grid and the cross-ambiguity function between transmit pulse $p_{tx}(t)$ and receive pulse $p_{rx}(t)$. $\frac{1}{M\Delta f}$, $\frac{1}{N\Delta t}$, $\Delta t$ $(s)$, and $\Delta f$ (Hz) denote the quantization steps along the delay and Doppler frequency axes, slot duration and subcarrier spacing, respectively.

$$G_{DD} = \left\{ \left( \frac{k}{N\Delta t}, \frac{l}{M\Delta f} \right) \right\} \text{ and } G_{TF} = \{(n\Delta t, m\Delta f)\}, \tag{1}$$

for $(n, k) \in \{0, ..., N-1\}^2$ and $(m, l) \in \{0, ..., M-1\}^2$. The cross-ambiguity function between $p_{tx}(t)$ and $p_{rx}(t)$ is given as follows:

$$A_{p_{tx},p_{rx}}(t, f) \triangleq \int p_{tx}^*(u-t) p_{rx}(u) e^{-j2\pi f(u-t)} \mathrm{d}u. \tag{2}$$

It is assumed that the *bi-orthonormal property* condition between $p_{tx}$ and $p_{rx}$ is satisfied if $\int p_{tx}^*(t) p_{rx}(t - n\Delta t) e^{j2\pi m\Delta f(t-n\Delta t)} \mathrm{d}t = \delta(n)\delta(m)$.

In OTFS modulation, the 2D information symbols $x[k, l]$, $(k, l) \in \{0, 1, ..., N-1\} \times \{0, 1, ..., M-1\}$ are arranged over $G_{DD}$. These symbols are firstly mapped onto $G_{TF}$ via the Inverse Symplectic Fast Fourier Transform (ISFFT) pre-coding as follows [1,5]:

$$X[n, m] = \mathrm{ISFFT}(x[k, l]) = \frac{1}{\sqrt{NM}} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x[k, l] e_{n,m}, \tag{3}$$

where $e_{n,m} = e^{j2\pi(\frac{nk}{N} - \frac{ml}{M})}$. Then, TF samples $X[n, m]$ are converted to waveform $s(t)$ using the Heisenberg transform as

$$s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X[n, m] \Phi_{n,m}(t), \tag{4}$$

where $\Phi_{n,m}(t) = p_{tx}(t - n\Delta t) e^{j2\pi m\Delta f(t-n\Delta t)}$. The output of the Heisenberg transform can be written in a vector form [25]:

$$\boldsymbol{s} = \mathrm{vec}(\boldsymbol{S}) = (\boldsymbol{F}_N^H \otimes \boldsymbol{P}_{tx})\boldsymbol{x}, \tag{5}$$

where $\boldsymbol{S} = \boldsymbol{P}_{tx}\boldsymbol{F}_M^H(\boldsymbol{F}_M \boldsymbol{X} \boldsymbol{F}_N^H) = \boldsymbol{P}_{tx}\boldsymbol{X}\boldsymbol{F}_N^H$, $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X})$ and $\boldsymbol{P}_{tx} = \mathrm{diag}[p_{tx}(0), p_{tx}(T/M), ..., p_{tx}(M-1)T/M] \in \mathbb{C}^{M\times N}$. Specifically, for a rectangular waveform, we get $\boldsymbol{P}_{tx} = \boldsymbol{I}_M$.

## 2.2 | Channel response in DD domain

The channel in the DD domain is sparse with a few paths, each with a specifically constant delay and Doppler during one frame [26]. The baseband impulse response can be represented as [1,27]

$$h(\tau, \nu) = \sum_{i=1}^{P} h_i \delta(\tau - \tau_i)\delta(\nu - \nu_i), \tag{6}$$

where $P$, $h_i$, $\nu_i$, and $\tau_i$ are the number of channel paths, complex channel gain, Doppler, and delay shifts of the $i^{th}$ path, respectively. The $i^{th}$ delay and Doppler taps $(l_i, k_i)$ can be written in the form

$$l_i = \tau_i M\Delta f,\ k_i = \nu_i N\Delta t. \tag{7}$$

The output of the doubly-selective fading channel can be expressed as[5]

$$r(t) = \iint h(\tau,\nu)s(t-\tau)e^{j2\pi\nu(t-\tau)}d\tau d\nu + w(t), \tag{8}$$

where $w(t) \sim \mathcal{CN}(0,\sigma^2)$ is the additive noise. After sampling $r(t)$ at $f_s = M/\Delta t = M\Delta f$, we form the vector $\boldsymbol{r}$. The entries of $\boldsymbol{r}$, regarding Eq. 7 and Eq. 8, are the samples

$$r(n) = \sum_{i=1}^{P} h_i e^{j2\pi\frac{k_i(n-l_i)}{MN}} s([n-l_i]_{MN}) + w(n). \tag{9}$$

Hence, Eq. 9 can be rewritten in a vector form:

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{w}, \tag{10}$$

where $\boldsymbol{H} = \sum_{i=1}^{P} h_i \boldsymbol{\Pi}^{l_i}\boldsymbol{\Delta}^{k_i} \in \mathbb{C}^{MN\times MN}$ is the channel matrix with $\boldsymbol{\Pi} \in \mathbb{C}^{MN\times MN}$, is the permutation matrix (forward cyclic shift) and $\boldsymbol{\Delta} = \text{diag}[\alpha^{(0)},\alpha^{(1)},...,\alpha^{(MN-1)}] \in \mathbb{C}^{MN\times MN}$, where $\alpha = e^{\frac{j2\pi}{MN}}$.

$$\boldsymbol{\Pi} = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{11}$$

We note that matrices $\boldsymbol{\Pi}$ and $\boldsymbol{\Delta}$ correspond to delay shift $\frac{1}{M\Delta f}$ and Doppler shift $\frac{1}{N\Delta t}$, respectively. Then, for the $i^{th}$ channel path, corresponding delay and Doppler shifts are accounted via matrices $\boldsymbol{\Pi}^{l_i}$ and $\boldsymbol{\Delta}^{k_i}$.

## 2.3 | OTFS Receiver

At the receiver side, we first perform a multicarrier demodulation for $r(t)$ through a matched filter bank to obtain the TF domain signal $Y(t,f)$. Then, the output $Y(t,f)$ of the matched filter is sampled, leading to $Y[n,m] = Y(t,f)|\{t = n\Delta t, f = m\Delta f\}$. This procedure is referred to as the Wigner transform. Then, the TF signal $Y[n,m]$ is transformed to the DD domain signal $y[k,l]$ using a Symplectic Finite Fourier Transform (SFFT) as follows[5]:

$$y[k,l] = \frac{1}{\sqrt{NM}} \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} Y[n,m]e_{k,l}^{-1}, \tag{12}$$

where $e_{k,l} = e^{j2\pi(\frac{nk}{N}-\frac{ml}{M})}$. The relationship between the received symbols $y[k,l]$ and transmitted symbols $x[k,l]$ can be expressed as a 2D circular convolution[28] as follows:

$$y[k,l] = \sum_{k'=-k_\nu}^{k_\nu}\sum_{l'=0}^{l_\tau} b_{k',l'}h_{k',l'}\beta_{k,l}x[[k-k']_N,[l-l']_M] + v[k,l], \tag{13}$$

where $b_{k',l'} \in \{0,1\}$ indicates the existence of path and $v[k,l] \sim \mathcal{CN}(0,\sigma^2)$ is an additive circular white noise with variance $\sigma^2$. $\beta_{k,l}$ is a known phase shift caused by imperfect bi-orthogonality of the rectangular waveform[25], given by

$$\beta_{k,l} = \begin{cases} e^{j2\pi(\frac{l-l'}{M})\frac{k'}{N}} & \text{if } l' \le l < M, \\ e^{j2\pi(\frac{l-l'}{M})\frac{k'}{N}}e^{-j2\pi\frac{k}{N}} & \text{if } 0 \le l < l'. \end{cases} \tag{14}$$

After substituting Eq. 5 in Eq. 10, the received signal in the DD domain can be written in vectorized form as

$$\begin{aligned} \boldsymbol{y} &= (\boldsymbol{F}_N \otimes \boldsymbol{P}_{rx})\boldsymbol{r}, \\ &= \boldsymbol{\mathcal{H}}\boldsymbol{x} + \tilde{\boldsymbol{w}}, \end{aligned} \tag{15}$$

where $\boldsymbol{P}_{rx} = \text{diag}[p_{rx}(0),p_{rx}(T/M),...,p_{rx}((M-1)T/M)] \in \mathbb{C}^{M\times M}$ (for rectangular waveforms, $\boldsymbol{P}_{rx} = \boldsymbol{I}_M$) represents the receiver filter using the pulse-shaping $p_{rx}(t)$. $\tilde{\boldsymbol{w}} = (\boldsymbol{F}_N \otimes \boldsymbol{P}_{rx})\boldsymbol{w}$ denote the noise vector. Since $(\boldsymbol{F}_N \otimes \boldsymbol{I}_M)$ is a unitary matrix

and $\boldsymbol{w} \sim \mathcal{CN}(0, \sigma^2 \boldsymbol{I}_M)$, $\widetilde{\boldsymbol{w}}$ and $\boldsymbol{w}$ share the same distributions. $\boldsymbol{\mathcal{H}} = (\boldsymbol{F}_N \otimes \boldsymbol{P}_{rx})\boldsymbol{H}(\boldsymbol{F}_N^H \otimes \boldsymbol{P}_{tx})$ is the effective channel matrix for rectangular pulses which is written as

$$\begin{aligned}\boldsymbol{\mathcal{H}} &= \sum_{i=1}^{P} h_i[(\boldsymbol{F}_M \otimes \boldsymbol{I}_M)\boldsymbol{M}^{l_i}(\boldsymbol{F}_N^H \otimes \boldsymbol{I}_M)][(\boldsymbol{F}_M \otimes \boldsymbol{I}_M)\boldsymbol{\Delta}^{k_i}(\boldsymbol{F}_N^H \otimes \boldsymbol{I}_M)], \\ &= \sum_{i=1}^{P} h_{(i)}\boldsymbol{P}^{(i)}\boldsymbol{Q}^{(i)}, \\ &= \sum_{i=1}^{P} h_{(i)}\boldsymbol{T}^{(i)},\end{aligned} \tag{16}$$

with $\boldsymbol{P}^{(i)} = (\boldsymbol{F}_M \otimes \boldsymbol{I}_M)\boldsymbol{M}^{l_i}(\boldsymbol{F}_N^H \otimes \boldsymbol{I}_M)$, $\boldsymbol{Q}^{(i)} = (\boldsymbol{F}_M \otimes \boldsymbol{I}_M)\boldsymbol{\Delta}^{k_i}(\boldsymbol{F}_N^H \otimes \boldsymbol{I}_M)$ and $\boldsymbol{T}^{(i)} = \boldsymbol{P}^{(i)}\boldsymbol{Q}^{(i)}$. $\boldsymbol{T}^{(i)}$ has only one non-zero element in each row. The position and value of this non-zero element depends on the Doppler and delay values. Entries of $\boldsymbol{T}^{(i)}$ write

$$\boldsymbol{T}^{(i)}(p,q) = \begin{cases} e^{-j2\pi\frac{n}{N}}e^{j2\pi\frac{k_i([m-l_i]_M)}{MN}}, & \text{if } q = [M-l_i]_M + M[n-k_i]_N \text{ and } m < l_i \\ e^{j2\pi\frac{k_i([m-l_i]_M)}{MN}}, & \text{if } q = [M-l_i]_M + M[n-k_i]_N \text{ and } m \geq l_i \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

## 2.4 | Proposed pilot pattern

In conventional OTFS structures, several pilot pattern schemes have been proposed to guarantee the accuracy of CE. Fig. 2a shows the first proposed pilot pattern. This scheme has a large GI around the pilot to avoid its interference with data symbols [12]. One major drawback of this scheme is the degradation of SE. This scheme also suffers from high Peak to Average Power Ratio (PAPR) because, the single pilot must be transmitted with high power to guarantee good CE. Another scheme has been proposed in Siqiang et al. [29] to improve the SE of the system (see Fig. 2b). In this pattern, there is no GI on the right side of the pilots (NGR). This improved configuration of pilots has been proposed considering that data symbols on the right side do not interfere with pilots. Moreover, the use of multiple pilots in this scheme reduces the PAPR effect, since the power is spread over several pilots, unlike the first scheme where all the power is assigned to a single pilot. In order to further improve the SE of the system, a SP scheme has been proposed in Mishra et al. [22] (see Fig. 2c). In this scheme, data symbols and pilots interfere with each other. Consequently, a powerful CE and data detection algorithm are required. The main issue of this approach is the computational load of the CE. The complexity increases significantly with the size of the data frame and thus does not scale for high bitrate systems where the data-pilot pattern involves high product $MN$. Another pilot pattern has been proposed in Yuan et al. [21], where only one superimposed pilot is adopted (see Fig. 2d). CE in this scheme can be done by a simple threshold method by calculating each time an adapted threshold that takes into account the interference as in Mishra et al. [22]. However, its great disadvantage is that it requires a large pilot-to-data power to get an accurate channel estimation. This causes a high PAPR. To solve the highlighted issues of the previous schemes, in this paper, we investigate a new pilot pattern, as shown in Fig. 2e. Compared to the schemes Raviteja et al. [12] and Siqiang et al. [29] (Fig. 2a and 2b), the proposed solution enjoy better SE, since no GI are used and data are mapped over the whole DD grid. Moreover, compared to scheme Mishra et al. [22] (Fig. 2c), we obtain good CE at lower computational complexity while preserving similar SE. According to Yuan et al. [21], this proposed scheme achieves a good trade-off between PAPR and computational complexity.

According to Fig. 2e, letting $\boldsymbol{X}_p$ the $N \times M$ matrix of pilots, its entries are 0, expect for $k \in [k_p - N_p, k_p + N_p]$ and $l \in [l_p, l_p + M_p - 1]$. Note that there are $L_p = (2N_p + 1)M_p$ pilots and $(k_p, l_p)$ denotes a reference position for pilots. $\boldsymbol{X}_d$ denotes the $N \times M$ matrix of data symbols. Letting $\boldsymbol{X} = \boldsymbol{X}_d + \boldsymbol{X}_p$, its entries write

$$X[k,l] = \begin{cases} X_p[k,l] + X_d[k,l], & k \in [k_p - N_p, k_p + N_p], \\ & \text{and } l \in [l_p, l_p + M_p - 1], \\ X_d[k,l] & \text{otherwise.} \end{cases} \tag{18}$$

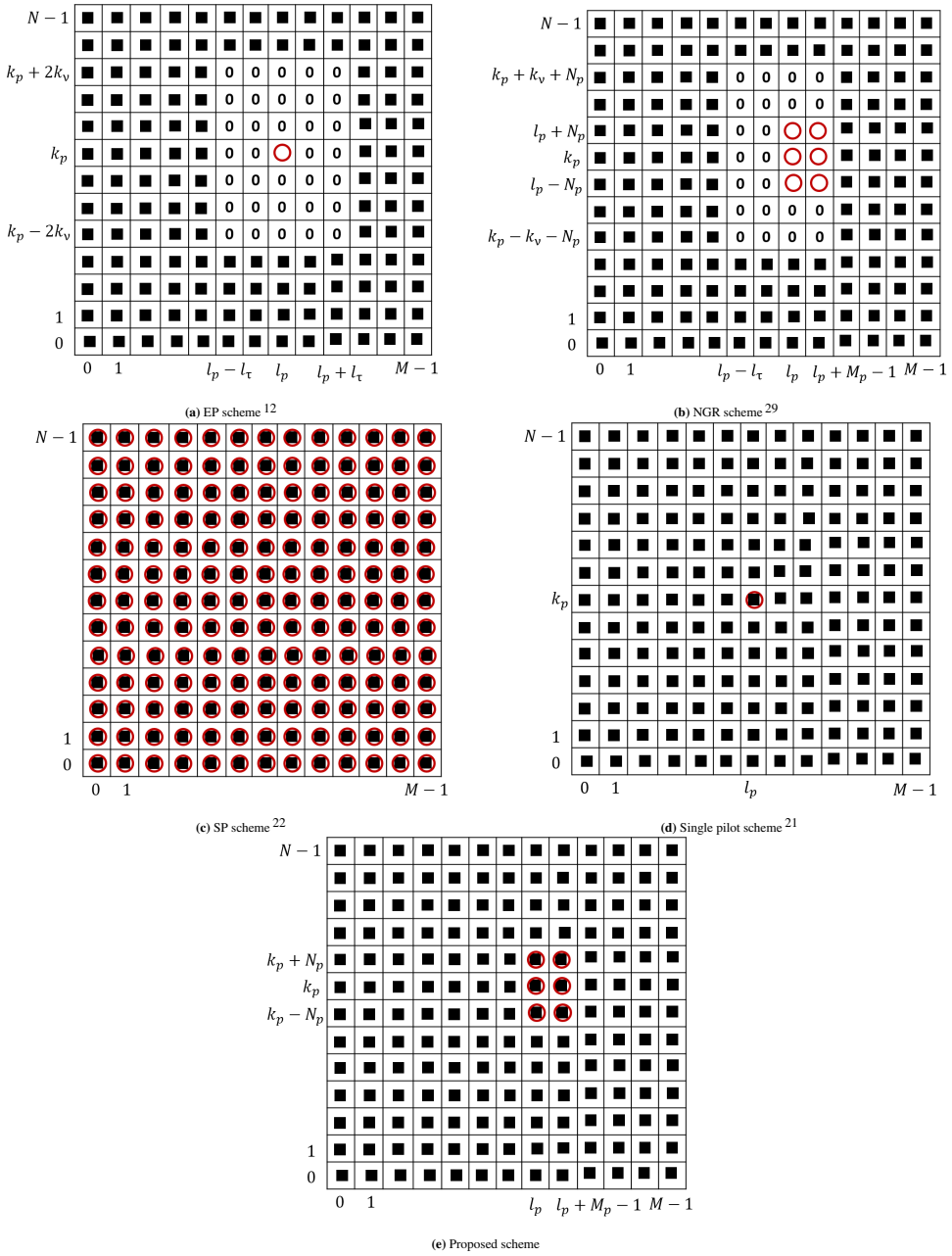


**FIGURE 2** Data pilot patterns in delay-Doppler domain (■: data, **0**: guard interval,∘: pilot).

## 2.5 | Problem formulation

For $k \in [k_p - N_p, k_p + N_p]$ and $l \in [l_p, l_p + M_p - 1]$, using Eq. 18, the received symbols in DD domain (Eq. 13) can be written in a vector form as follows:

$$\boldsymbol{y}_p = (\boldsymbol{S}_p \odot \boldsymbol{\Psi})\boldsymbol{h} + \boldsymbol{D}\tilde{\boldsymbol{h}} + \tilde{\boldsymbol{w}}, \tag{19}$$

where $\boldsymbol{y}_p \in \mathbb{C}^{L_p}$ is formed by the column wise rearrangement of 2D data $y[k,l]$ for $k \in [k_p - N_p, k_p + N_p]$ and for $l \in [l_p, l_p + M_P - 1]$: $\boldsymbol{y}_{p[l-l_p+(k+N_p)M_p]} = y[k,l]$. $\boldsymbol{\Psi} \in \mathbb{C}^{L_p \times L}$ is written as $\boldsymbol{\Psi}_{[l-l_p+(k+N_p)M_p, l'(2k_\nu+1)+k'+k_\nu]} = e^{-j2\pi k'(l-l')/NM}$. $\boldsymbol{S}_p \in \mathbb{C}^{L_p \times L}$ is the pilots matrix given by $\boldsymbol{S}_{p[l-l_p+(k+N_p)M_p, l'(2k_\nu+1)+k'+k_\nu]} = x_p[k-k', l-l']$. Considering (18), $\boldsymbol{S}_p$ is composed of pilots and zeros. $\boldsymbol{h} \in \mathbb{C}^L$ is a $P$-sparse channel vector and $L = (2k_\nu + 1)(l_\tau + 1)$. $\boldsymbol{D} \in \mathbb{C}^{L_p \times P}$ is a mini matrix of data symbols given by

$\boldsymbol{D}_{[l-l_p+(k+N_p)M_p,k']} = \beta_{k',l-l'} x_d[k-k', l-l']$ and $\tilde{\boldsymbol{h}} \in \mathbb{C}^P$ is the vector formed by the $P$ non-zero elements of $\boldsymbol{h}$. Let $\boldsymbol{A} = (\boldsymbol{S}_p \odot \boldsymbol{\Psi})$ and $\boldsymbol{v} = \boldsymbol{D}\tilde{\boldsymbol{h}} + \tilde{\boldsymbol{w}}$. Then, $\boldsymbol{v} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C}_v)$, $\boldsymbol{C}_{\boldsymbol{v}} = \mathbb{E}\{\boldsymbol{v}\boldsymbol{v}^{\boldsymbol{H}}\} = \left(\left(\sum_{i=1}^{P} \sigma_{h_i}^2\right)\sigma_d^2 + \sigma_w^2\right)\boldsymbol{I}_{L_p}$ (see appendix A). Therefore, from (19),

$$\boldsymbol{y}_p|\boldsymbol{h} \sim \mathcal{CN}(\boldsymbol{A}\boldsymbol{h}, \boldsymbol{C}_v). \tag{20}$$

In what follows, to achieve accurate CE, we first address the problem of pilots design. Then, we formulate an OTFS pilot sequence optimization problem based on the mutual incoherence property (MIP), and develop a (SA)-based algorithm for solving it.

## 3 | SA-BASED PILOT OPTIMIZATION FOR OTFS

### 3.1 | Pilot values optimization problem

To achieve accurate CE it is important to account for the sparsity of $\boldsymbol{h}$. There are two main properties to consider, the sparsity $P$ of a vector $\boldsymbol{h}$ and to what degree the pilots matrix $\boldsymbol{A}$ satisfies the restrict isometry property (RIP). This means, for an arbitrary level $\delta \in (0, 1)$, and any index set $\mathcal{I} \subset \{0, 1, ..., L\}$ such that $\text{card}(\mathcal{I}) \leq P$, and for all $\boldsymbol{h} \in \mathbb{C}^{\text{card}(\mathcal{I})}$, the following relation holds: $(1-\delta)||\boldsymbol{h}||^2 \leq ||\boldsymbol{A}_{\mathcal{I}}\boldsymbol{h}||^2 \leq (1+\delta)||\boldsymbol{h}||^2$. Where $\boldsymbol{A}_{\mathcal{I}}$ is the matrix formed by the columns of which the indices are elements of the set $\mathcal{I}$ (see Zhang et al.[30]). However, the RIP evaluation of a matrix is an NP-hard problem[31]. Another more practical and computable criterion is introduced in the optimization process of the $\boldsymbol{A}$ matrix instead of RIP namely the mutual incoherence property (MIP). The MIP of $\boldsymbol{A}$ is defined as the largest absolute and normalized inner product between its different columns, and it is given by[32,33]:

$$\mu(\boldsymbol{A}) = \max_{1 \leq i < j \leq L} \frac{|\boldsymbol{a}_i^H \boldsymbol{a}_j|}{||\boldsymbol{a}_i||_2 . ||\boldsymbol{a}_j||_2}, \tag{21}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_L]$.

It is worth noting that the mutual coherence reflects the correlation between the columns of $\boldsymbol{A}$ and higher correlation degrades the performance. The overall coherence cannot be measured from Eq. 21 because only the maximum absolute value is considered[29]. Thus, we chose pilot values so as to optimize an average MIP (A-MIP: $\mu_{avg}$) of $\boldsymbol{A}$:

$$\mu_{avg}(\boldsymbol{A}) = \frac{2}{L(L-1)} \sum_{1 \leq i < j \leq L} \frac{|\boldsymbol{a}_i^H \boldsymbol{a}_j|}{||\boldsymbol{a}_i||_2 . ||\boldsymbol{a}_j||_2}. \tag{22}$$

From Eq. 20 and Eq. 22, we formulate a pilot optimization problem for OTFS as in Wang et al.[29] as follows:

$$\underset{\boldsymbol{x}_p}{\text{minimize}} \quad \mu_{avg}(\boldsymbol{A}) \ \text{ subject to } \quad \max(|\boldsymbol{x}_p|) \leq P_p, \tag{23}$$

where $\boldsymbol{P}_p$ is the maximum power threshold. It is important to note that if the matrix $\boldsymbol{A}$ is orthogonal (e.g., $\boldsymbol{A}$ is a Fourier matrix), $\mu_{avg}$ achieves lower bound which is 0. In practice, the matrix $\boldsymbol{A}$ cannot be orthogonalized because this matrix is composed of pilots and zeros and the position of the pilots depends on the delay and Doppler taps of the channel, hence the high complexity of this optimization problem.

### 3.2 | SA-based optimization algorithm

The problem (23) is non-convex, which means that conventional convex optimization algorithms are not guaranteed to reach a global minimum. This led us to take advantage of the SA algorithm[34,35,36] to solve the problem (23). We divide this algorithm into four steps:

1. **Initialization step**: all the initializations are done in this step. The efficiency of the SA algorithm depends on the choice of its control parameters, hence, a careful choice of the parameters is mandatory. We first randomly generate an initial solution $\boldsymbol{x}_p^{(0)}$ which satisfies the constraint in (23). Then, the fitness value of this initial solution is calculated according to Eq. 22 as $F(\boldsymbol{x}_p^{(0)}) = \mu_{avg}(\boldsymbol{A}(\boldsymbol{x}_p^{(0)}))$. $\boldsymbol{x}_p^{best}$ at start is set at $\boldsymbol{x}_p^{(0)}$. A high initial temperature $T_0$ is chosen empirically. The neighborhood $V$ of a point is determined by a distance $r$ which is initialized at $r_0$. The decreasing temperature's constant $\alpha$ is initialized at $\alpha_0$.

2. **Inner loop step**: at each iteration of the algorithm, an elementary modification of the state is carried out. This modification consists in calculating a new solution to problem (23) as follows: $\boldsymbol{x}_p^{(i)} = \boldsymbol{x}_p^{(i-1)} + rd$, where $d$ is a random number that satisfies $d \in [-1, 1]$. This modification leads to a variation $\Delta F$ of the system's energy $\Delta F = F(\boldsymbol{x}_p^{(i)}) - F(\boldsymbol{x}_p^{(i-1)})$, where $F(\boldsymbol{x}) = \mu_{avg}(\boldsymbol{A}(\boldsymbol{x}))$. $\boldsymbol{x}_p^{(i)}$ is accepted as a new configuration with probability $\min\{1, \exp(-\Delta F/T)\}$ and if $F(\boldsymbol{x}_p^{(i)}) < F(\boldsymbol{x}_p^{best})$, $\boldsymbol{x}_p^{best}$ is set to $\boldsymbol{x}_p^{(i)}$. This choice of exponential for the probability is called the Metropolis rule. This process is then repeated $N_T$ times at a fixed temperature.

3. **Outer loop step**: In the previous step, the algorithm iterates keeping the temperature constant. When the system has reached a thermodynamic equilibrium (after a certain number of changes), the temperature of the system is reduced. This is referred to as a temperature step. After the temperature has been reduced, the previous step is repeated. In both cases, if the temperature has reached a low enough threshold set beforehand or $iter_{max}$ is reached, the algorithm stops.

4. **Temperature update**: temperature plays an important role. At high temperatures, the system is free to move in the state space ($\exp(-\Delta F/T)$ close to 1) by choosing configurations that do not necessarily minimize the energy of the system. At low temperatures, the system is sensitive to finer energy variations and will then favours more and more the descending movements (i.e. towards lower energy values), and will avoid those that go up. The temperature decay law we have chosen and which is often used in the literature is $T_j = \alpha T_{j-1}$.

The detailed SA-based pilot optimization procedure is summarized in Algorithm 1.

**Algorithm 1** SA-based pilot optimization algorithm for OTFS

**Input:** $T_0$, $\boldsymbol{x}_p^{(0)}$, $\boldsymbol{x}_p^{best}$, $\alpha_0$, $r_0$, $iter_{max}$, $N_T = 1000$, $\epsilon$, pilot matrix $\boldsymbol{A}$,
  **Initialization**: $\alpha = \alpha_0$, $iter_{max} = 10^6$, $\epsilon = 1$, $iter = 1$, $T = T_0$, $\boldsymbol{x}_p^{best} = \boldsymbol{x}_p^{(0)}$,
  **while** ($iter < iter_{max}$ & $\epsilon > 10^{-9}$) **do**
    **for** $i = 1 : N_T$ **do**
      generate $\tilde{\boldsymbol{x}}_p^{(i)} = \boldsymbol{x}_p^{(i-1)} + rd$ with $d \sim \mathcal{U}_{[-1,1]}$,
      evaluate $F(\tilde{\boldsymbol{x}}_p^{(i)}) = \mu_{avg}(\boldsymbol{A}(\tilde{\boldsymbol{x}}_p^{(i)}))$,
      compute difference $\Delta F = F(\tilde{\boldsymbol{x}}_p^{(i)}) - F(\boldsymbol{x}_p^{(i-1)})$,
      **if** $\Delta F < 0$ **then**
        $\boldsymbol{x}_p^{(i)} = \tilde{\boldsymbol{x}}_p^{(i)}$,
        **if** $F(\boldsymbol{x}_p^{(i)}) < F(\boldsymbol{x}_p^{best})$ **then**
          $\boldsymbol{x}_p^{best} = \boldsymbol{x}_p^{(i)}$,
        **end if**
      **else**
        $p \sim \mathcal{U}_{[0,1]}$,
        $\boldsymbol{x}_p^{(i)} = \boldsymbol{x}_p^{(i-1)} + (\tilde{\boldsymbol{x}}_p^{(i)} - \boldsymbol{x}_p^{(i-1)})\mathbf{1}_{p<\exp(-\Delta F/T)}$,
      **end if**
    **end for**
    $T = \alpha T$, $r = \alpha r$, $\epsilon = T/T_0$, $iter = iter + 1$,
  **end while**
**Output:** optimized $\boldsymbol{x}_p$.

## 4 | SUGGESTED ALGORITHM FOR CE AND SYMBOL DETECTION

Here, we feature the suggested algorithm for CE and data detection with the IC scheme. This algorithm iterates between data-aided CE and MP-assisted data detection. After each CE, the IC scheme is executed to minimize the effect of pilots on the data to achieve good detection of data symbols.

## 4.1 | Initial channel estimation

An initial channel estimate is computed at this current phase of the algorithm. Denoting by $\boldsymbol{a}_i$ the $i^{th}$ column of $\boldsymbol{A}$, we can rewrite the model (20) in the following form:

$$\boldsymbol{y}_p = \sum_{i=1}^{L} b_i g_i \boldsymbol{a}_i + \boldsymbol{v}, \tag{24}$$

where $h_i = b_i g_i$, i.e., $\boldsymbol{h} = \boldsymbol{b} \odot \boldsymbol{g}$, with $\boldsymbol{b} = [b_1, b_2, ..., b_L]^T$ denotes the support vector of channel ($b_i \in \{0, 1\}$) and $\boldsymbol{g} = [g_1, g_2, ..., g_L]^T$ is the channel gains vector. Therefore, $p(\boldsymbol{y}_p|\boldsymbol{g}, \boldsymbol{b}) = \mathcal{CN}(\boldsymbol{A}_b \boldsymbol{g}_b, \boldsymbol{C}_{\boldsymbol{v}})$, where $\boldsymbol{g}_b \in \mathbb{C}^P$ and $\boldsymbol{A}_b \in \mathbb{C}^{L_p \times P}$ are made up from $\boldsymbol{g}$ and $\boldsymbol{A}$ considering indices $i$ where $b_i \neq 0$.

To consider the sparsity of $\boldsymbol{h}$, its inputs are modeled using a Bernoulli-Gaussian (BG) model for which, $\boldsymbol{g}$ follows the probabilistic model

$$\boldsymbol{g}|\boldsymbol{b} \sim p(\boldsymbol{g}|\boldsymbol{b}) = \prod_{i=1}^{L} p(g_i|b_i) \text{ where } p(g_i|b_i) = \mathcal{CN}(0, \sigma^2_{b_i}), \tag{25}$$

with $\sigma_1^2 \gg \sigma_0^2$. Thus, the vector $\boldsymbol{y}_p$ can be seen as a noisy mixture of atoms specified by $\boldsymbol{b}$. The weights of the mixture are realizations of Gaussian distributions whose variances $\sigma^2_{g_i}$ are independent of $\boldsymbol{b}$. To detect the location of spikes, we note that $b_i$ are assumed to be independent Bernoulli random variables, such that $b_i = 1$ if a spike is present at $h_i$ and $b_i = 0$ otherwise. Unstructured sparsity can be modeled by a standard choice based on a product of the Bernoulli distributions, as follows:

$$\boldsymbol{b} \sim p(\boldsymbol{b}) = \prod_{i=1}^{L} p(b_i) \text{ where } p(b_i) = \text{Ber}(p_i), \tag{26}$$

where $p_i = p(b_i = 1) = 1 - p(b_i = 0)$.

Besides, in recent approaches, instead of trying to approximate the means of the posterior distributions via MCMC simulation, one tries to iteratively calculate an exact variational approximation of the target posterior distribution.

We derive an estimator based on a MAP criterion for the sparse channel parameters. It corresponds to the optimal Bayesian estimator using a Bayesian cost[37]. In this context, the estimation of $(\boldsymbol{b}, \boldsymbol{g})$ take the form

$$(\hat{\boldsymbol{b}}, \hat{\boldsymbol{g}}) = \arg\max_{\boldsymbol{b},\boldsymbol{g}} \log p(\boldsymbol{b}, \boldsymbol{g}|\boldsymbol{y}_p). \tag{27}$$

We first estimate the support vector $\boldsymbol{b}$. The decision that minimizes the probability of a wrong decision on support $\boldsymbol{b}$ is given as follows:

$$\hat{\boldsymbol{b}} = \arg\max_{\boldsymbol{b} \in \{0,1\}^L} \log p(\boldsymbol{b}|\boldsymbol{y}_p). \tag{28}$$

Solving problem (28) requires the evaluation of the function $\log(p(\boldsymbol{b}|\boldsymbol{y}_p)$ for all sequences of $\boldsymbol{b}$ in $\{0, 1\}^L$ ($2^L$ evaluation), this makes the maximization of $p(\boldsymbol{b}|\boldsymbol{y}_p)$ too complex. However, individual decisions for the entries of the support $\boldsymbol{b}$ can be made from a marginalized MAP, leading to

$$\hat{b}_i = \arg\max_{b_i \in \{0,1\}} \log p(b_i|\boldsymbol{y}_p). \tag{29}$$

Furthermore, the evaluation of $p(b_k|\boldsymbol{y}_p)$ is intractable due to the costly marginalization of $p(\boldsymbol{b}|\boldsymbol{y}_p)$ over the $b_l$'s, for $l \neq k$. To get around this issue, variational mean-field approximations compute a tractable surrogate $q(b_i)$ of $q(b_i|\boldsymbol{y}_p)$ (see Dremeau et al.[38] and Wainwright and Jordan[39]). The adopted procedure is to compute an approximation $q(b_i)$ of the posterior probability $p(b_i|\boldsymbol{y}_p)$, named the mean-field approximation, which is summarized in appendix B. In this case, problem (29) will be approximated by

$$\hat{b}_k = \arg\max_{b_k \in \{0,1\}} \log(q(b_k)). \tag{30}$$

Problem (30) can easily be solved by a thresholding method, i.e.

$$\hat{b}_i = \begin{cases} 1, & \text{if } q(b_i = 1) > \rho \\ 0, & \text{otherwise,} \end{cases} \tag{31}$$

with $\rho = 0.5$. This value minimizes the Bayes risk when uniform and equal costs are selected. After estimating the support vector $\boldsymbol{b}$, the channel gains vector $\boldsymbol{g}$ can be estimated by MAP estimate $\hat{\boldsymbol{g}} = \arg\max_{\boldsymbol{g}} \log p(\boldsymbol{g}|\hat{\boldsymbol{b}}, \boldsymbol{y}_p)$:

$$\begin{aligned} \hat{\boldsymbol{g}}_{\hat{\boldsymbol{b}}} = &\quad (\boldsymbol{A}_{\hat{\boldsymbol{b}}}^T \boldsymbol{A}_{\hat{\boldsymbol{b}}} + \boldsymbol{\Delta})^{-1} \boldsymbol{A}_{\hat{\boldsymbol{b}}}^T \boldsymbol{y}_p, \\ \text{and } \hat{g}_k = 0 &\quad \text{if } b_k = 0, \end{aligned} \tag{32}$$

where $\hat{\boldsymbol{g}}_{\hat{\boldsymbol{b}}}$ and $\boldsymbol{A}_{\hat{\boldsymbol{b}}}$ are the entries of $\boldsymbol{g}$ restricted to $\hat{\boldsymbol{b}}$ and the related columns of $\boldsymbol{A}$, respectively. $\boldsymbol{\Delta} = \text{diag}[\sigma^2/\sigma_{g_1}^2, \sigma^2/\sigma_{g_2}^2, ..., \sigma^2/\sigma_{g_L}^2]$. This solution reduces to the least-square solution when $\sigma^2 \lll \sigma_{g_i}^2$ and to matched filtering when $\sigma^2$ is large. The proposed algorithm for CE is summarized in Algorithm 2.

**Algorithm 2** The proposed CE algorithm for OTFS

**Input:** measurements $\boldsymbol{y}_p \in \mathbb{C}^{L_p}$, sensing matrix $\boldsymbol{A} \in \mathbb{C}^{L_p \times L}$,
  **Initialization**: $p(\boldsymbol{b}) = \prod_k p(b_k)$,
    prior mean for $\boldsymbol{g}$: $\boldsymbol{m} = \boldsymbol{0}_L$,
    probability $\boldsymbol{q}$: $\boldsymbol{q}^{(0)} \sim (\mathcal{U}_{[0,1]})_{1:L}$,
    $\boldsymbol{r}$: $\boldsymbol{r}^{(0)} = \boldsymbol{y}_p - \boldsymbol{A}(\boldsymbol{b} \odot \boldsymbol{m})$,
  **while** $k \leq K$ and $\forall i \in \{1 : L\}$, $|q(b_i^{(k)}) - q(b_i^{(k-1)})| < \epsilon$ **do**
    **for** $l = 1 : L$ **do**
      compute $\Sigma(b_k|\boldsymbol{y}_p) = \frac{\sigma_{g_k}^2 \sigma^2}{\sigma^2 + b_k \sigma_{g_k}^2 \boldsymbol{A}_k^T \boldsymbol{A}_k}$,
      compute $m(b_l|\boldsymbol{y}_p)^{(k)} = b_l \frac{\sigma_{g_l}^2}{\sigma^2 + b_l \sigma_{g_l}^2 \boldsymbol{A}_l^T \boldsymbol{A}_l} \boldsymbol{r}_l^T \boldsymbol{A}_l$,
      compute $q_l^{(k)} = q(b_l|\boldsymbol{y}_p)^{(k)} \propto \sqrt{\Sigma(b_l|\boldsymbol{y}_p)} e^{\left(\frac{1}{2} \frac{m(b_l|\boldsymbol{y}_p)^2}{\Sigma(b_l|\boldsymbol{y}_p)}\right)} p(b_l)$,
      update $\boldsymbol{r}^{(k)}$: $\boldsymbol{r}^{(k)} = \boldsymbol{r}^{(k)} - \boldsymbol{A}_l(b_l^{(k)} m(b_l|\boldsymbol{y}_p)^{(k)})$,
    **end for**
  **end while**
  estimate support $\boldsymbol{b}$: $\hat{\boldsymbol{b}} = (\boldsymbol{q} > 0.5)$,
  estimate $\boldsymbol{g}$ conditional to $\hat{\boldsymbol{b}}$: $\hat{\boldsymbol{g}}_{\hat{\boldsymbol{b}}} = (\boldsymbol{A}_{\hat{\boldsymbol{b}}}^T \boldsymbol{A}_{\hat{\boldsymbol{b}}} + \boldsymbol{\Delta})^{-1} \boldsymbol{A}_{\hat{\boldsymbol{b}}}^T \boldsymbol{y}_p$ and $\hat{g}_k = 0$ if $b_k = 0$,
**Output:** $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{g}}$.

## 4.2 | Interference cancellation

The absence of GI in the proposed pilot pattern improves SE, but, unfortunately the pilots will interfere with some of the data symbols as shown in the red box in Fig. 3. This interference is due to the delay and Doppler shifts of the pilots experiencing the channel.

The sampled signal in contaminated region in the DD grid is expressed as follows:

$$\begin{aligned} \boldsymbol{Y}_I = &\ y[k,l],\ k \in [k_p - k_\nu - N_p, k_p + k_\nu + N_p], \\ \text{and} &\quad l \in [l_p, l_p + l_\tau + M_p - 1]. \end{aligned} \tag{33}$$

We put $p = k - k_p + k_\nu + N_p$ and $q = l - l_p$. CE supplies the channel support $\{\hat{l}_i, \hat{k}_i\}$ as well as the corresponding path parameters $\{\hat{h}_i, \hat{\tau}_i, \hat{\nu}_i\}$. Then, for $i = 1 : P$, the interference at the $(p,q)^{th}$ element in the DD grid is expressed as

$$\boldsymbol{I}[p,q] = \sum_{i=1}^{P} \hat{h}_i e^{-j\frac{2\pi}{MN}\gamma_i} X_p[p + l_p - k_\nu - N_p - \hat{k}_i, q + l_p - \hat{l}_i], \tag{34}$$

for $p \in [0, 2(k_\nu + N_p)]$, $q \in [0, l_\tau + M_p - 1]$ and $\gamma_i = \hat{k}_i[q - \hat{l}_i + 2l_p + M_p + 1]_M$. The estimated interference is removed from $\boldsymbol{Y}_I$:

$$\boldsymbol{Y} = \begin{cases} \boldsymbol{Y}_I - \boldsymbol{I} & k \in [k_p - k_\nu - N_p, k_p + k_\nu + N_p], \\ & \text{and } l \in [l_p, l_p + l_\tau + M_p - 1], \\ y[k,l] & \text{otherwise.} \end{cases} \tag{35}$$

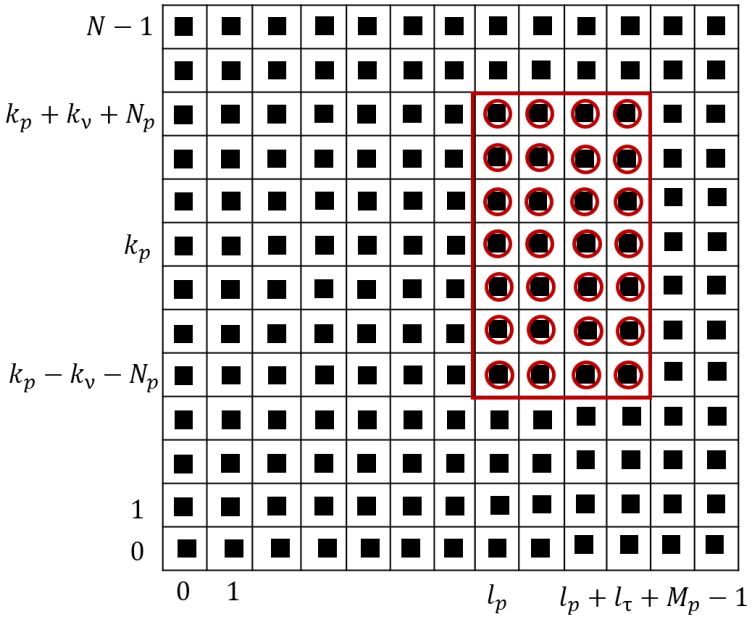


**FIGURE 3** Pilot contamination pattern (■: data symbol, ∘: pilot).

### 4.3 | Message passing (MP) data detection

After channel estimation and pilots interference removal, the vector $\boldsymbol{y}_d = \text{vec}(\boldsymbol{Y})$ will be used for data symbol detection. According to Eq. 15, $\boldsymbol{y}_d$ can be written as follows:

$$\boldsymbol{y}_d = \hat{\boldsymbol{\mathcal{H}}}\boldsymbol{x}_d + \boldsymbol{w}_e, \tag{36}$$

where $\boldsymbol{w}_e = \tilde{\boldsymbol{w}} + \boldsymbol{B}(\boldsymbol{h} - \hat{\boldsymbol{h}})$ consists of noise and CE error $\boldsymbol{h} - \hat{\boldsymbol{h}}$. The matrix $\boldsymbol{B} \in \mathbb{C}^{MN\times L}$ is designed in exactly the same way as the matrix $\boldsymbol{A}$ in subsection (2.5), using $\boldsymbol{X}_p$. $\boldsymbol{w}_e(n) \sim \mathcal{CN}(0, \sigma_w^2 + \sigma_p^2 B_h)$, where $B_h = \mathbb{E}\{||\boldsymbol{h} - \hat{\boldsymbol{h}}||^2\}$ is the MSE of channel estimation. The aim is to determine $\boldsymbol{x}_d$ from $(\hat{\boldsymbol{\mathcal{H}}}, \boldsymbol{y}_d)$ by using a low complexity MP algorithm[28] which is suitable for uncoded OTFS, taking advantage of channel sparsity.

### 4.4 | Data-aided channel estimation

After estimating the vector of data symbols $\boldsymbol{x}_d$, the model (20) used for CE becomes as follows:

$$\boldsymbol{y}_p = \boldsymbol{A}\boldsymbol{h} + \boldsymbol{v}_e, \tag{37}$$

where $\boldsymbol{v}_e = \hat{\boldsymbol{D}}\hat{\tilde{\boldsymbol{h}}} + \tilde{\boldsymbol{w}}$ consist of noise, channel estimate vector and the mini data symbol matrix. The MP data detection step produces independent values for the inputs of $\hat{\boldsymbol{D}}$. Therefore, for CE upgrade, we recommend to use the algorithm already used in subsection (4.1) with the following noise approximation in Eq. 19: $\boldsymbol{v}_e \sim \mathcal{N}\left(\mathbb{E}\{\hat{\boldsymbol{D}}\hat{\tilde{\boldsymbol{h}}}\}, \left(\left(\sum_{i=1}^{P}\sigma_{h_i}^2\right)\sigma_d^2 + \sigma_w^2\right)\boldsymbol{I}_{L_p}\right)$.

The proposed iterative algorithm for CE and data detection with the interference cancellation scheme is summarized in Algorithm 3.

## 5 | SIMULATIONS, DISCUSSIONS AND COMPLEXITY ASSESSMENT

Here, we first specify all the parameters used in the simulations. Then, we evaluate the performance of the proposed SA-based pilots' optimization algorithm and we make a comparison with several other optimization methods, mainly to a particle swarm optimization (PSO)-based reference method proposed for OTFS systems. Next, we study the performance of the proposed iterative CE and data detection algorithm that uses the previously optimized pilots along with using a pilot interference cancellation

**Algorithm 3** Proposed algorithm for CE and data detection

**Input:** measurements $\boldsymbol{y} \in \mathbb{C}^{MN}$, pilot matrix $\boldsymbol{A} \in \mathbb{C}^{L_p \times L}$, initial channel estimate $\hat{\boldsymbol{h}}^{(0)}$.

**repeat**

Compute $\hat{\boldsymbol{\mathcal{H}}}^{(i)} = (\boldsymbol{F}_n \otimes \boldsymbol{I}_M)\left(\sum_{n=1}^{P} \hat{h}_n^{(i)} \boldsymbol{\Pi}^{l_n} \boldsymbol{\Delta}^{k_n}\right)(\boldsymbol{F}_n^H \otimes \boldsymbol{I}_M)$

Compute $\boldsymbol{y}_d^{(i)}$, the output of interference cancellation scheme

Compute $\hat{\boldsymbol{x}}_d = \text{MPA}(\hat{\boldsymbol{\mathcal{H}}}^{(i)}, \boldsymbol{y}_d^{(i)})$ % MPA: MP Algorithm

Compute $\hat{\boldsymbol{h}}^{(i+1)}$ by feeding $\boldsymbol{y}_e^{(i)}$ and $\boldsymbol{A}$ to (37):

$\hat{\boldsymbol{h}}^{(i+1)} = \arg\max_{\boldsymbol{h}} \mathcal{CN}(\boldsymbol{y}_p; \boldsymbol{A}\boldsymbol{h} + \mathbb{E}\{\hat{\boldsymbol{D}}\hat{\tilde{\boldsymbol{h}}}\}, \boldsymbol{A}\boldsymbol{R}_{hh}\boldsymbol{A}^H + \left(\left(\sum_{i=1}^{P} \sigma_{h_i}^2\right)\sigma_d^2 + \sigma_w^2\right)\boldsymbol{I}_{L_p})$ % $\boldsymbol{R}_{hh} = \text{diag}(\sigma_{h_1}^2, \sigma_{h_2}^2, ..., \sigma_{h_L}^2)$

**until** stopping condition

**Output:** $\hat{\boldsymbol{h}}$, $\hat{\boldsymbol{x}}_d$.

(IC) scheme. The metrics used for the performance evaluation of this algorithm are NMSE, BER, and SE. A comparison will be made between the proposed algorithm and the techniques proposed in the literature. Finally, a complexity study is conducted.

## 5.1 | Simulation parameters

We consider an OTFS system with the following parameters: the carrier frequency $f_c$ and the spacing between sub-carriers $\Delta f$ are set to 4 GHz and 15 kHz, respectively. The size of one OTFS frame is $M = N = 32$. BPSK constellation is treated. The channel delay model parameters are displayed in Table 1 [22]. We suppose that each delay tap has only one Doppler shift [12]. Therefore, we generate Doppler values using the Jakes' formula as $\nu_i = \nu_{max}\cos(\theta_i)$, where $\theta_i \sim \mathcal{U}_{[0,2\pi]}$, and $\nu_{max}$ represents the maximum Doppler shift determined by the users' equipment (UE) speed [16]. The maximum delay shift $\tau_{max}$ is 20.8 $\mu s$ and has a corresponding delay tap $l_\tau = 10$. The maximum Doppler shift $\nu_{max}$ is 1851 Hz and has a corresponding Doppler tap $k_\nu = 4$. These values correspond to a high-speed scenario with a maximum UE speed of 500 km/h. To achieve good CE accuracy, the pilot size signals must cover the maximum delay and Doppler spread, i.e. $M_p \geq l_\tau$ and $2N_p + 1 \geq k_\nu$ [29]. Consequently, to achieve a better CE accuracy and minimize the influence of pilots on the data symbols, we set $M_p = 13$ and $N_p = 3$. In this case, the ratio of data symbols affected by pilots to the total data symbols is given by $\rho = \frac{(2(k_\nu + N_p)+1)(l_\tau + M_p)}{MN} \times 100 \simeq 34\%$.

For the parameters in the SA-based pilot optimization, from tryouts, it appears that setting $N_T = 1000$, $iter_{max} = 10^6$, $r_0 = \sqrt{P_p}/10$, $\alpha_0 = 0.99$ and $T_0 = 30°C$ is a judicious choice.

We assume that each symbol in the DD grid has a power of $\sigma_x^2$. The total power of each OTFS frame is $MN\sigma_d^2 + (2N_p+1)M_p\sigma_p^2$. Thus, by taking $\sigma_x^2 = \sigma_d^2 + \frac{(2N_p+1)M_p}{MN}\sigma_p^2$, the total power for each OTFS frame is fixed to $NM\sigma_x^2$. For a fair comparison with the other schemes, this total power must remain the same for each scheme. For the superimposed scheme shown in Fig 2c, the power per DD symbol is $\sigma_x^2 = \sigma_p^2 + \sigma_d^2$ for $k \in [0, N-1]$ and $l \in [0, M-1]$, where 30% of the total power is allocated to pilots and 70% of it is allocated to data. In Mishra et al., [22] these values allow to maximize the SE and minimize the BER of the proposed SP-aided design. For the EP scheme shown in Fig 2a, due to the insertion of GI between the pilots and data symbols, all the power that is assigned to the GI positions must be assigned to the pilot, i.e. $\sigma_p^2 = ((2l_\tau + 1)(4k_\nu + 1) - 1)\sigma_x^2$. Therefore, we have a total power per frame of $((2l_\tau + 1)(4k_\nu + 1) - 1)\sigma_x^2 + (NM - ((2l_\tau + 1)(4k_\nu + 1) - 1))\sigma_x^2 = NM\sigma_x^2$ which is the same as the proposed scheme and the superimposed scheme. From trials, it appears that choosing $|\hat{\boldsymbol{h}}^{(n)} - \hat{\boldsymbol{h}}^{(n-1)}| < 10^{-6}$ or a maximum number of iterations equal to 10 is a good stopping criterion for CE.

**TABLE 1** 5-tap delay-Doppler channel parameters [22].

| Channel tap no. | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\tau_i$ ($\mu s$) | 2.08 | 5.20 | 8.328 | 11.46 | 20.8 |
| Power of $h_i$ (dB) | 1 | -1.804 | -3.565 | -5.376 | -8.860 |

## 5.2 | Performance evaluation of the proposed SA-based pilot optimization

In this subsection, we first make a comparison in terms of A-MIP (Eq. 22) between the developed SA-based pilot optimization algorithm and the PSO-based pilot optimization algorithm suggested in Wang et al.[29] Other solutions based on Genetic algorithms (GA)[40] and deterministic Nelder Mead (NM) algorithm[41] is also tested and compared to the proposed algorithm. Then, we compare the NMSE and BER performance of CE and data detection with the optimized pilot using NM, PSO, and GA.

### 5.2.1 | A-MIP evaluation

Table 2 shows the obtained A-MIP of the optimized pilots using NM, PSO, GA, and SA algorithms. It can be seen that the highest A-MIP is obtained for NM-based optimization. For the PSO-based optimization and the GA-based optimization, the achieved A-MIPs are close. For the proposed algorithm, the A-MIP can be reduced to 0.00291 which is the minimum compared to state-of-the-art methods, showing then the superiority of the SA approach in terms of A-MIP when compared to NM, PSO and GA methods. Therefore, the suggested solution for pilot optimization for OTFS outperforms the benchmark proposed by Wang et al.[29]

**TABLE 2** Average-MIP comparison.

| Method | $\mu_{avg}$ | Method | $\mu_{avg}$ |
|---|---|---|---|
| Optimized pilot using NM | $0,157$ | Optimized pilot using PSO | 0.0292 |
| Optimized pilot using GA | $0,0278$ | Optimized pilot using SA | 0.00291 |

### 5.2.2 | Normalized Mean square error (NMSE)

Fig 4 shows a comparison in terms of NMSE for CE between the proposed SA-based optimization algorithm and the other three algorithms based on NM, PSO, and GA using highlighted iterative CE and data detection algorithm. From Fig 4, the achieved NMSE performance of the proposed SA-based pilot optimization algorithm outclasses the other three algorithms and particularly outperforms the PSO-based optimization algorithm by about 4 dB. This can be attributed to the significant achieved performance of A-MIP by the proposed solution.

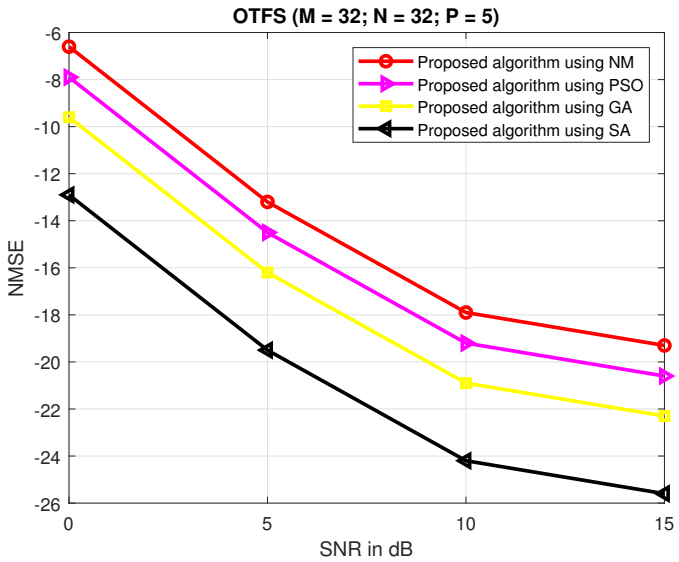


**FIGURE 4** NMSE comparison between proposed SA-based pilot optimization and the methods using NM, PSO and GA.

### 5.2.3 | Bit error rate (BER)

Fig 5 shows a comparison in terms of BER for data detection between the developed SA-based optimization algorithm and the other three techniques. For SNR $< 5$ dB, similar BER performance can be achieved by the four methods. As the SNR increases, the proposed algorithm outperforms the other three solutions. Particularly, it exceeds the PSO-based pilot optimization by about 1.2 dB at BER $= 10^{-4}$.

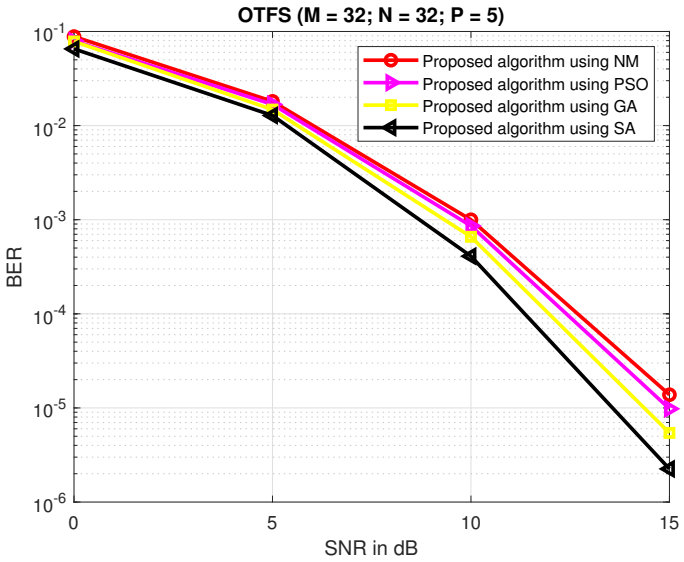


**FIGURE 5** BER comparison between proposed SA-based pilot optimization and the methods using NM, PSO and GA.

### 5.2.4 | Effect of the size of pilot on BER performance

We now investigate the effect of the size of the pilots on BER performance. Fig 6 shows the BER versus SNR of the proposed scheme with optimized pilots using SA, for different pilots sizes. We observe from this figure that each time the pilot size $L_p$ increases, the performance of the system in terms of BER improves. This is justified by the fact that as $L_p$ increases, the error recovery of the channel vector $\boldsymbol{h}$ from problem (19) decreases, thus the NMSE of the channel estimation decreases, which leads to an improvement of BER. It will be shown in section 5.4 that the computational complexity of the proposed channel estimation algorithm depends on $L_p$: as $L_p$ increases the BER improves and the computational complexity increases. Therefore, a judicious choice of $L_p$ achieves a good compromise between performance and complexity.

## 5.3 | Performance analysis of the proposed iterative algorithm with optimized pilots

In this subsection, we first compare the average SE of the proposed scheme with the conventional pilot aided (CPA) design[13], embedded pilots (EP)[12] (Fig. 2a), the scheme with none GI in the right side (NGR)[29] (Fig. 2b) and the superimposed scheme (SP-I)[22] (Fig. 2c). Afterwards, we give performance comparisons in terms of NMSE and BER between the proposed algorithm for CE and data detection against the EP-based design and the SP-I scheme.

The CPA design uses two OTFS frames. The first one is for CE and the second one for data detection. The EP-based scheme, shown in Fig 2a, uses a single pilot in the DD domain. This pilot is separated from the data symbols through GI. The number of zeros inserted for the guard interval depends on $l_\tau$ and $k_\nu$. This scheme uses enough GI to avoid interference between the pilot and the received data symbols. However, in the NGR design, unlike the EP-design, there are no guard symbols on the right side of the pilots. This is due to the fact that the data symbols on the right side will not interfere with the pilot. Hence, the SE is increased compared to the EP scheme. Concerning the SP-I scheme, the pilots and data symbols are arranged in a superimposed way in the DD domain without using GI. The proposed scheme benefits from the same advantages of the previous schemes like

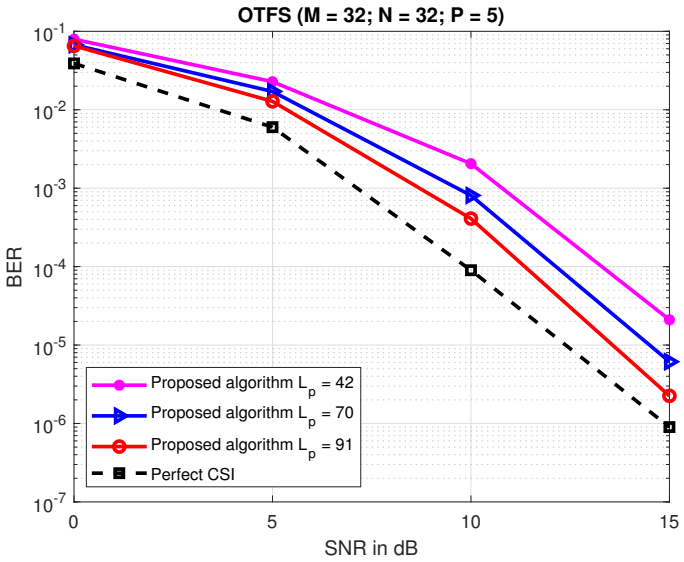


**FIGURE 6** BER performance of the proposed algorithm for different pilot sizes.

they are GI-free to maximize SE and the pilots are not spread over the entire DD grid in order to minimize the pilot interference on the data symbols. The SE expression of CPA, EP, NGR, and SP designs is given as[22]:

$$\mathcal{R}_s = (1-\eta_s)\log_2(1+\text{SINR}_s), \tag{38}$$

where $s \in \{\text{CPA, EP, NGR, SP}\}$ and $\eta_s$ is the pilot overhead relative to the scheme $s$. The pilot overhead $\eta_s$ can be calculated using the frame-structure for the scheme $s$ and it is given for all previous schemes as follows: $\eta_{CPA} = \frac{MN}{2MN} = \frac{1}{2}$, $\eta_{EP} = \frac{(2l_\tau+1)(4k_\nu+1)}{MN}$, $\eta_{NGR} = \frac{(2(k_\nu+N_p)+1)(l_\tau+M_p)}{MN}$ and $\eta_{SP} = 0$. For the CPA, EP and SP designs, the SINR is given in Mishra et al.[22] as:

$$\text{SINR}_s = \frac{(\sigma_h^2 - B_{h,s})\sigma_{d,s}^2}{\sigma_w^2 + \sigma_{d,s}^2 B_{h,s}}, \quad \text{SINR}_{\text{SPI}} = \frac{(\sigma_h^2 - B_{h,\text{SPI}})\sigma_{d,\text{opt}}^2}{\sigma_w^2 + \sigma_{d,\text{opt}}^2 B_{h,\text{SPI}} + \sigma_{p,\text{opt}}^2 P B_{h,\text{SPI}}} \tag{39}$$

where $s \in \{\text{CPA, EP}\}$, $B_{h,s}$ and $B_{h,\text{SPI}}$ denote the MSE of MMSE CE in the $s$ design and the MSE of CE in the SP-I design, respectively. Similarly, the SINR for the NGR scheme $\text{SINR}_{\text{NGR}}$ is given as

$$\text{SINR}_{\text{NGR}} = \frac{(\sigma_h^2 - B_{h,\text{NGR}})\sigma_{d,\text{NGR}}^2}{\sigma_w^2 + \sigma_{d,NGR}^2 B_{h,\text{NGR}}}, \tag{40}$$

where $B_{h,\text{NGR}}$ is the MSE of CE in the NGR design. For the proposed design, the SE is given as follows:

$$\mathcal{R} = (1-\eta)\log_2(1+\text{SINR}), \tag{41}$$

where $\eta = 0$, $B_h$ is the MSE of CE in the proposed algorithm and the SINR expression can be expressed in the same way as $\text{SINR}_{\text{SPI}}$ as follows:

$$\text{SINR} = \frac{(\sigma_h^2 - B_h)\sigma_d^2}{\sigma_w^2 + \sigma_d^2 B_h + \sigma_p^2\left(\frac{(2N_p+1)M_p}{MN}\right) P B_h}. \tag{42}$$

The difference between this expression and that of $\text{SINR}_{\text{SPI}}$ in Eq. 39 is the third term of the denominator. This term is equal to $B_{h,s}\text{Tr}(\tilde{\boldsymbol{x}}_p\tilde{\boldsymbol{x}}_p^H)$. For the SP-I algorithm, $B_{h,s}\text{Tr}(\tilde{\boldsymbol{x}}_p\tilde{\boldsymbol{x}}_p^H) = B_{h,\text{SPI}} P\sigma_p^2$. This is because $\tilde{\boldsymbol{x}}_p \in \mathbb{C}^P$ is a vector full of pilots and $\mathbb{E}\{|\tilde{x}_p(i)|^2\} = \sigma_p^2$. In our case, $\tilde{\boldsymbol{x}}_p$ is a combination between pilots and zeros. The ratio of pilots to zeros is $\frac{(2N_p+1)M_p}{MN}$. Therefore, $B_{h,s}\text{Tr}(\tilde{\boldsymbol{x}}_p\tilde{\boldsymbol{x}}_p^H) = \left(\frac{(2N_p+1)M_p}{MN}\right)\sigma_p^2 P B_h$.

### 5.3.1 | Spectral efficiency (SE)

Fig 7 shows the average SE comparison of the proposed scheme with the SP-I, EP, NGR, and CPA schemes for $M = N = 32$ and for a given $\{l_\tau, k_\nu\}$. We see that the proposed scheme and the SP-I algorithm have almost the same average SE. We also see that the average SE of the proposed scheme is larger than those of EP-based scheme, NGR scheme and the CPA design. In addition, note that as $l_\tau$ and $k_\nu$ increase, in contrast to the proposed scheme, the average SE of the EP design and NGR scheme degrades and the difference with the proposed scheme becomes larger. This is because the pilot overhead for the EP-based design and the NGR scheme increases with $l_\tau$ and $k_\nu$.

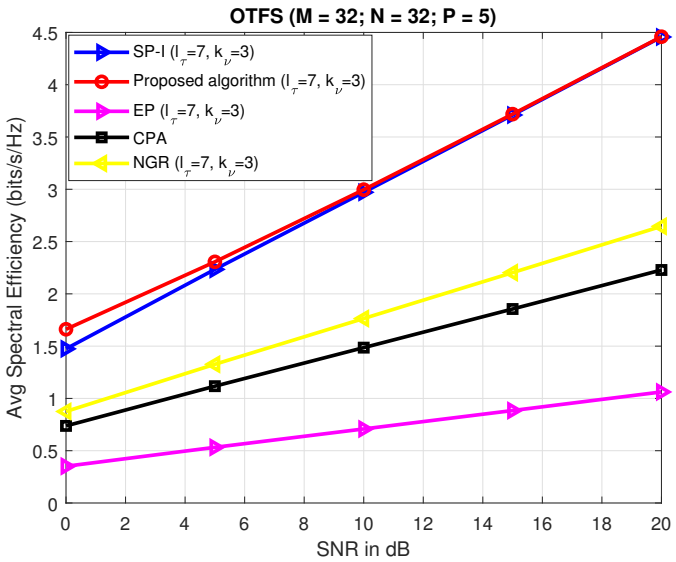


**FIGURE 7** Average SE comparison between the proposed scheme and SP-I, EP, NGR, and CPA designs.

### 5.3.2 | Normalized mean square error (NMSE)

Fig 8 shows an NMSE comparison for an OTFS system using the suggested scheme against the SP-I algorithm and the EP-based scheme. According to the obtained results, the proposed scheme performs slightly better than the SP-I design. Nevertheless, as the EP-based algorithm associates higher power for pilots, this latter performs better than the proposed algorithm.

### 5.3.3 | Bit error rate (BER)

Fig 9 shows a BER comparison between, on one hand, an OTFS system using the proposed iterative algorithm for CE and data detection associated with the SA-based pilot optimization scheme, and on the other hand, a known channel state information (CSI), the SP-I algorithm, and the EP designs. We first observe that the proposed algorithm and the SP-I algorithm have almost the same BER at low SNRs. However, when the SNR increases, the benefits of using the pilot cancellation scheme appear since the difference in residual symbols' contamination power becomes reduced. We also see that the EP scheme in Raviteja et al.[12] has a slightly better BER compared to the proposed algorithm. This is, as highlighted before, in the EP scheme, the pilot has higher power, unlike the other schemes where the power is equally distributed over all symbols. However, it is worth pointing out that this high power associated with the pilot causes the PAPR problem, unlike the proposed scheme and the SP-I where this problem does not occur. Moreover, as mentioned before, the cost to pay for this slight BER improvement is a slight decrease in SE for EP schemes (Fig. 2a).

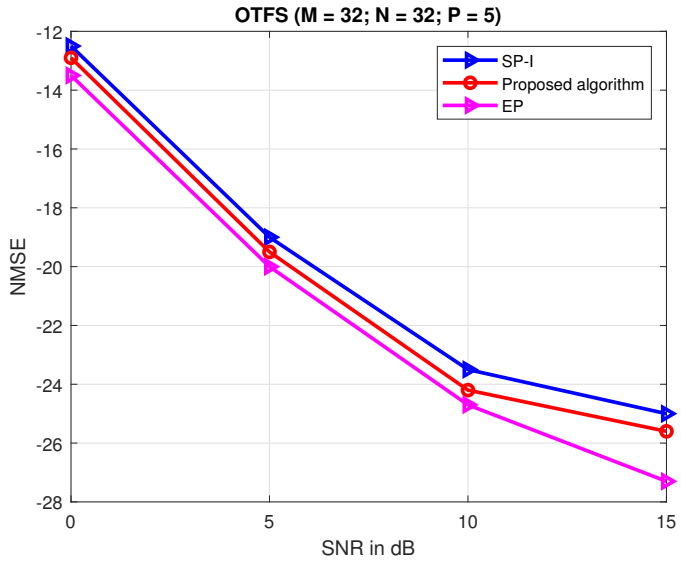


**FIGURE 8** NMSE comparison between the proposed algorithm against SP-I and EP schemes.

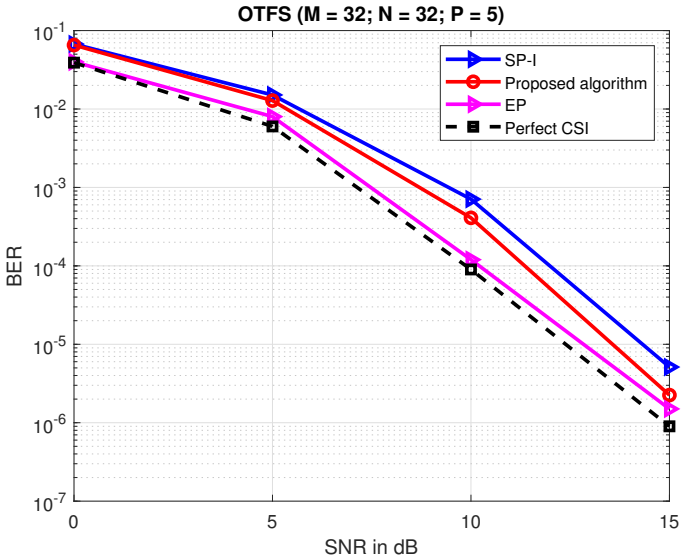


**FIGURE 9** BER comparison between the proposed algorithm against SP-I and EP schemes.

## 5.4 | Complexity analysis

The SA-based pilot optimization algorithm can run offline. Therefore, its complexity will not be taken into account in practice. The proposed algorithm for CE and data detection is divided into three steps; the initial channel estimate step, the CE step and the data detection step. For the CE step, the computational cost per iteration is $\mu_e = (4P^2+6(2k_v+1)(l_\tau+1))(2N_p+1)M_p+(P^3+(2k_v+1)(l_\tau+1))$. For the MP detection step, the overall cost for $n_{iter}$ iterations is $\mu_d = n_{iter}NMSP$. The total operations number required by the proposed method is $\mu = \mu_e + \mu_d + \mu_i$, where $\mu_i = \mu_e$ is the computational cost of the first channel estimate. The overall complexity of the proposed method is $C = (N_{iter}+1)\mathcal{O}(\mu_e)+N_{iter}\mathcal{O}(\mu_d)$. In practice we have, $l_\tau, k_v, P, N_p, M_p \ll MN$, and $C = N_{iter}\mathcal{O}(\mu_d) = N_{iter}\mathcal{O}(n_{iter}NMPS)$. Under the same conditions i.e. $l_\tau, k_v, P, N_p, M_p \ll MN$, the complexity of SP-I scheme is $(N_{SPI}+1)\mathcal{O}(MN)+(N_{SPI}+1)\mathcal{O}(n_{iter}NMPS) \approx N_{SPI}\mathcal{O}(n_{iter}NMPS)$ and that of the EP-based design is dominated by $\mathcal{O}(n_{iter}(MN-((2l_\tau+1)(4k_v+1)))PS) \approx \mathcal{O}(n_{iter}(MN)$ since $l_\tau, k_v \ll MN$. It should be noted that $N_{iter}$

and $N_{\text{SPI}}$ at convergence are of the same order of magnitude for both algorithms. Thus, we observe that there is a marginal difference between the three schemes. We note that the suggested algorithm has a lower complexity when compared to SP-I. This difference is manifested in the CE step because both methods use the MP algorithm for data symbol detection. Finally, It is worth noting that we manage to estimate the channel without any a priori information on the channel compared to the SP-I algorithm in Mishra et al.[22] where the channel support is supposed to be known.

## 6 | CONCLUSION

In this manuscript, we developed a new scheme to rearrange pilots and data symbols in the DD domain for OTFS systems. To achieve better CE accuracy in terms of MIP, we formulated a pilot optimization problem that we solved via an SA-based pilot optimization algorithm. We have also developed an iterative channel estimation and data detection algorithm that benefits from channel sparsity in the DD domain. Also, we formulated the sparse CE step as a specific marginalization of MAP criterion by providing a Bayesian approach via the variational MF approximation and the VB-EM scheme. Detection of data symbols is done using a low complexity MP algorithm. An interference cancellation scheme has also been proposed to minimize the effect of pilots on data symbols and detect them properly. In addition, to the gain in terms of spectral efficiency that offers the proposed scheme compared to embedded pilots-based design, it surpasses the SP-I scheme in terms of BER and NMSE. Also, the proposed scheme outperforms the SP-I algorithm in terms of complexity and without any prior knowledge of channel taps.

**How to cite this article:** Ouchikh R, A. Aïssa-El-Bey, T. Chonavel, and M. Djeddou (2022), Joint channel estimation and data detection for high rate OTFS systems, *Int J Commun Syst.*, *2022;xx:1–23*.

## APPENDIX

## A CALCULATING $\mu_V$ AND $C_V$

We have $\boldsymbol{v} = \boldsymbol{D}\tilde{\boldsymbol{h}} + \tilde{\boldsymbol{w}}$. The elements of a random matrix $\boldsymbol{D}$ are centred, statistically independent of each other and statistically independent from the noise $\tilde{\boldsymbol{w}}$ and $\mathbb{E}\{|\boldsymbol{D}_{(i,j)}|^2\} = \sigma_d^2$. Therefore, $\boldsymbol{D}$ has $\mathbb{E}\{\boldsymbol{D}\} = \boldsymbol{0}_{L_p \times P}$ and $\mathbb{E}\{\boldsymbol{D}\boldsymbol{D}^H\} = \sigma_d^2 P \boldsymbol{I}_{L_p}$. For $\tilde{\boldsymbol{h}}$, $\mathbb{E}\{\tilde{\boldsymbol{h}}\} = \boldsymbol{0}_{L_p}$ and $\mathbb{E}\{\boldsymbol{h}\boldsymbol{h}^H\} = \text{diag}[\sigma_{h_1}^2, \sigma_{h_2}^2, ..., \sigma_{h_P}^2]$. Since $\mathbb{E}\{\boldsymbol{D}\} = \boldsymbol{0}_{L_p \times P}$ and $\mathbb{E}\{\tilde{\boldsymbol{h}}\} = \boldsymbol{0}_{L_p}$, the mean of $\boldsymbol{v}$ is

$$\mu_{\boldsymbol{v}} = \mathbb{E}\{\boldsymbol{v}\} = \boldsymbol{0}_{L_p}, \tag{A1}$$

and its covariance matrix is given by

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{v}} &= \mathbb{E}\{\boldsymbol{v}\boldsymbol{v}^H\}, \\ &= \mathbb{E}\{(\boldsymbol{D}\tilde{\boldsymbol{h}} + \tilde{\boldsymbol{w}})(\boldsymbol{D}\tilde{\boldsymbol{h}} + \tilde{\boldsymbol{w}})^H\}, \\ &= \mathbb{E}\{\boldsymbol{D}\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^H\boldsymbol{D}^H\} + \mathbb{E}\{\boldsymbol{D}\tilde{\boldsymbol{h}}\tilde{\boldsymbol{w}}^H\} + \mathbb{E}\{\tilde{\boldsymbol{w}}\tilde{\boldsymbol{h}}^H\boldsymbol{D}^H\} + \mathbb{E}\{\tilde{\boldsymbol{w}}\tilde{\boldsymbol{w}}^H\}. \end{aligned} \tag{A2}$$

Since $\boldsymbol{D}$ is statistically independent of $\tilde{\boldsymbol{w}}$, we have $\mathbb{E}\{\boldsymbol{D}\tilde{\boldsymbol{h}}\tilde{\boldsymbol{w}}^H\} = \mathbb{E}\{\tilde{\boldsymbol{w}}\tilde{\boldsymbol{h}}^H\boldsymbol{D}^H\} = \boldsymbol{0}_{L_p \times L_p}$ and (Eq. A2) is reduced to

$$\boldsymbol{C}_{\boldsymbol{v}} = \mathbb{E}\{\boldsymbol{D}\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^H\boldsymbol{D}^H\} + \mathbb{E}\{\tilde{\boldsymbol{w}}\tilde{\boldsymbol{w}}^H\}. \tag{A3}$$

The expression $\mathbb{E}\{\boldsymbol{D}\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^H\boldsymbol{D}^H\}$ can be evaluated based on the following property[42]: 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 $\boldsymbol{Y}$, $\mathbb{E}\{\boldsymbol{X}\boldsymbol{Y}\boldsymbol{X}^H\} = \frac{\text{Tr}(\boldsymbol{Y})}{L}\mathbb{E}\{\boldsymbol{X}\boldsymbol{X}^H\}$. Therefore,

$$\begin{aligned} \mathbb{E}\{\boldsymbol{D}\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^H\boldsymbol{D}^H\} &= \frac{\text{Tr}(\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^H)}{P}\mathbb{E}\{\boldsymbol{D}\boldsymbol{D}^H\}, \\ &= \left(\sum_{i=1}^{P} \sigma_{h_i}^2\right)\sigma_d^2 \boldsymbol{I}_{L_p}. \end{aligned} \tag{A4}$$

Note also that $\tilde{\boldsymbol{w}}_n \sim \mathcal{CN}(0, \sigma_w^2 \boldsymbol{I}_{L_p})$ and putting all pieces together yields

$$\boldsymbol{C}_{\boldsymbol{v}} = \left(\left(\sum_{i=1}^{P} \sigma_{h_i}^2\right)\sigma_d^2 + \sigma_w^2\right)\boldsymbol{I}_{L_p}. \tag{A5}$$

## B COMPUTATION OF THE MF APPROXIMATION

The objective is to compute an approximation $q(b_k)$ of the posterior probability $p(b_k|\boldsymbol{y}_p)$. For this purpose, a methodology called Mean Field (MF) approximation is adopted. Letting $\boldsymbol{\theta} = (\boldsymbol{b}, \boldsymbol{g})$ and $p(\boldsymbol{\theta}|\boldsymbol{y}_p)$ its posterior distribution, the MF approximation [38,43] of $p(\boldsymbol{\theta}|\boldsymbol{y}_p)$ is the surrogate distribution $q^\star(\boldsymbol{\theta})$ which satisfies

$$q^\star(\boldsymbol{\theta}) = \arg\min_{q(\boldsymbol{\theta})} \left\{ \int_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) \log\left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\boldsymbol{y}_p)}\right) \mathrm{d}\boldsymbol{\theta} \right\}, \tag{B6}$$

subject to

$$q(\boldsymbol{\theta}) = \prod_{k=1}^{K} q(\boldsymbol{\theta}_k), \; \int_{\boldsymbol{\theta}_k} q(\boldsymbol{\theta}_k)\mathrm{d}\boldsymbol{\theta}_k = 1 \;\; \forall k \in [1, K]\,. \tag{B7}$$

Successive minimizations of the Kullback-Leibler divergence [44] with respect to the parameters of factors $q(\boldsymbol{\theta}_i)$ can solve the problem (B6), (B7) [45]. The procedure given in Dremeau et al., [38] named VB-EM algorithm [46,47,48], is ensured to converge to a saddle point or a (local or global) maximum of problem (B6), (B7) under mild conditions [38]. The relation between this procedure and the algorithm EM [45,49] results from imposing the constraint $q(\boldsymbol{\theta}_i) = \delta(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)$ on some $q(\boldsymbol{\theta}_i)'s$.

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

$$\begin{aligned} p(\boldsymbol{\theta}_i|\boldsymbol{y}_p) &= \textstyle\int_{\boldsymbol{\theta}_{-i}} p(\boldsymbol{\theta}|\boldsymbol{y}_p)\mathrm{d}\boldsymbol{\theta}_{-i}, \\ &\simeq \textstyle\int_{\boldsymbol{\theta}_{-i}} q(\boldsymbol{\theta}|\boldsymbol{y}_p)\mathrm{d}\boldsymbol{\theta}_{-i}, \\ &\simeq q(\boldsymbol{\theta}_i|\boldsymbol{y}_p), \end{aligned} \tag{B8}$$

where the last equality stems from (B7).

Here, we consider the particular case where the MF approximation $q(\boldsymbol{g}, \boldsymbol{b})$ of $p(\boldsymbol{g}, \boldsymbol{b}|\boldsymbol{y}_p)$ simply writes $q(\boldsymbol{g}, \boldsymbol{b}) = \prod_k q(g_k, b_k)$. Together with models (25), (26), the corresponding VB-EM update is given as follows:

$$q(g_k, b_k|\boldsymbol{y}_p) = q(g_k|b_k, \boldsymbol{y}_p)q(b_k|\boldsymbol{y}_p), \tag{B9}$$

where

$$q(g_k|b_k, \boldsymbol{y}_p) = \mathcal{N}(m(b_k), \Sigma(b_k)). \tag{B10}$$

$$q(b_k|\boldsymbol{y}_p) \propto \sqrt{\Sigma(b_k)} e^{\left(\frac{1}{2}\frac{m(b_k)^2}{\Sigma(b_k)}\right)} p(b_k). \tag{B11}$$

and

$$\Sigma(b_k|\boldsymbol{y}_p) = \frac{\sigma^2_{g_k}\sigma^2}{\sigma^2 + b_k\sigma^2_{g_k}\boldsymbol{A}_k^T\boldsymbol{A}_k}, \tag{B12}$$

$$m(b_k|\boldsymbol{y}_p) = b_k \frac{\sigma^2_{g_k}}{\sigma^2 + b_k\sigma^2_{g_k}\boldsymbol{A}_k^T\boldsymbol{A}_k} \boldsymbol{r}_k^T\boldsymbol{A}_k, \tag{B13}$$

$$\boldsymbol{r}_k = \boldsymbol{y}_p - \sum_{l\neq k} q(b_l = 1)m(b_l = 1)\boldsymbol{A}_l. \tag{B14}$$

From (Eq. B8), an approximation of $p(b_k|\boldsymbol{y}_p)$ therefore follows simply from the relations

$$p(b_k|\boldsymbol{y}_p) \simeq \int q(g_k, b_k|\boldsymbol{y}_p)\mathrm{d}g_k = q(b_k). \tag{B15}$$

## References

1. Khan IA, Mohammed SK. Low Complexity Channel Estimation for OTFS Modulation with Fractional Delay and Doppler. *arXiv preprint arXiv:2111.06009* 2021.

2. Mohammed SK. Derivation of OTFS modulation from first principles. *IEEE transactions on vehicular technology* 2021; 70(8): 7619–7636.

3. Series M. IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond. *Recommendation ITU* 2015; 2083: 21.

4. Monk A, Hadani R, Tsatsanis M, Rakib S. OTFS-orthogonal time frequency space: a novel modulation technique meeting 5G high mobility and massive MIMO challenges. *arXiv preprint arXiv:1608.02993* 2016; 9.

5. Hadani R, Rakib S, Tsatsanis M, et al. Orthogonal time frequency space modulation. In: IEEE. ; 2017: 1–6.

6. Wei Z, Yuan W, Li S, et al. Orthogonal time-frequency space modulation: A promising next-generation waveform. *IEEE wireless communications* 2021; 28(4): 136–144.

7. NAIKOTI A, CHOCKALINGAM A. Signal Detection and Channel Estimation in OTFS. *ZTE Communications* 2022; 19(4): 16–33.

8. Zhang M, Wang F, Yuan X, Chen L. 2D structured turbo compressed sensing for channel estimation in OTFS systems. In: IEEE. ; 2018: 45–49.

9. Rasheed OK, Surabhi G, Chockalingam A. Sparse delay-Doppler channel estimation in rapidly time-varying channels for multiuser OTFS on the uplink. In: IEEE. ; 2020: 1–5.

10. Gómez-Cuba F. Compressed Sensing Channel Estimation for OTFS Modulation in Non-Integer Delay-Doppler Domain. *arXiv preprint arXiv:2111.12382* 2021.

11. Murali K, Chockalingam A. On OTFS modulation for high-Doppler fading channels. In: IEEE. ; 2018: 1–10.

12. Raviteja P, Phan KT, Hong Y. Embedded pilot-aided channel estimation for OTFS in delay–Doppler channels. *IEEE Transactions on Vehicular Technology* 2019; 68(5): 4906–4917.

13. Ramachandran MK, Chockalingam A. MIMO-OTFS in high-Doppler fading channels: Signal detection and channel estimation. In: IEEE. ; 2018: 206–212.

14. Shen W, Dai L, An J, Fan P, Heath RW. Channel estimation for orthogonal time frequency space (OTFS) massive MIMO. *IEEE Transactions on Signal Processing* 2019; 67(16): 4204–4217.

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

16. Zhao L, Gao WJ, Guo W. Sparse Bayesian learning of delay-Doppler channel for OTFS system. *IEEE communications letters* 2020; 24(12): 2766–2769.

17. Qu H, Liu G, Zhang L, Imran MA, Wen S. Low-dimensional subspace estimation of continuous-doppler-spread channel in otfs systems. *IEEE transactions on communications* 2021; 69(7): 4717–4731.

18. Zhao L, Yang J, Liu Y, Guo W. Block Sparse Bayesian Learning-based Channel Estimation for MIMO-OTFS Systems. *IEEE Communications Letters* 2022: 1-1. doi: 10.1109/LCOMM.2022.3144674

19. Guo G, Jin Z, Zhang X, Wei J. Joint Iterative Channel Estimation and Symbol Detection for Orthogonal Time Frequency Space Modulation. In: IEEE. ; 2021: 1–6.

20. Liu R, Huang Y, He D, Xu Y, Zhang W. Optimizing Channel Estimation Overhead for OTFS with Prior Channel Statistics. In: IEEE. ; 2021: 1–6.

21. Yuan W, Li S, Wei Z, Yuan J, Ng DWK. Data-aided channel estimation for OTFS systems with a superimposed pilot and data transmission scheme. *IEEE wireless communications letters* 2021; 10(9): 1954–1958.

22. Mishra HB, Singh P, Prasad AK, Budhiraja R. OTFS Channel Estimation And Data Detection Designs With Superimposed Pilots. *IEEE Transactions on Wireless Communications* 2021: 1-1. doi: 10.1109/TWC.2021.3110659

23. Yuan W, Wei Z, Li S, Yuan J, Ng DWK. Integrated sensing and communication-assisted orthogonal time frequency space transmission for vehicular networks. *IEEE Journal of Selected Topics in Signal Processing* 2021; 15(6): 1515–1528.

24. Ouchikh R, Aïssa-El-Bey A, Chonavel T, Djeddou M. Iterative channel estimation and data detection algorithm for OTFS modulation. In: IEEE. ; 2022: 5263–5267.

25. Raviteja P, Hong Y, Viterbo E, Biglieri E. Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS. *IEEE Transactions on Vehicular Technology* 2018; 68(1): 957–961.

26. ZHANG C, XING W, YUAN J, ZHOU Y. Performance of LDPC Coded OTFS Systems over High Mobility Channels. *ZTE Communications* 2022; 19(4): 45–53.

27. Bello P. Characterization of randomly time-variant linear channels. *IEEE transactions on Communications Systems* 1963; 11(4): 360–393.

28. Raviteja P, Phan KT, Hong Y, Viterbo E. Interference cancellation and iterative detection for orthogonal time frequency space modulation. *IEEE Transactions on Wireless Communications* 2018; 17(10): 6501–6515.

29. Wang S, Guo J, Wang X, Yuan W, Fei Z. Pilot design and optimization for OTFS modulation. *IEEE Wireless Communications Letters* 2021; 10(8): 1742–1746.

30. Zhang Y, Venkatesan R, Dobre OA, Li C. Novel compressed sensing-based channel estimation algorithm and near-optimal pilot placement scheme. *IEEE Transactions on Wireless Communications* 2015; 15(4): 2590–2603.

31. Donoho DL. Compressed sensing. *IEEE Transactions on information theory* 2006; 52(4): 1289–1306.

32. Donoho DL, Elad M. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences* 2003; 100(5): 2197–2202.

33. Tropp JA. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information theory* 2004; 50(10): 2231–2242.

34. Kirkpatrick S, Gelatt Jr CD, Vecchi MP. Optimization by simulated annealing. *science* 1983; 220(4598): 671–680.

35. Zhang K, Duan C, Jia H. Genetic simulated annealing-based coverage-enhancing algorithm for multimedia directional sensor networks. *International Journal of Communication Systems* 2015; 28(9): 1598–1609.

36. Al-Khaled FS. Optimal radio channel assignment through a new binary dynamic simulated annealing algorithm. *International Journal of Communication Systems* 1998; 11(5): 327–335.

37. Levy BC. *Principles of signal detection and parameter estimation*. Springer Science & Business Media . 2008.

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

39. Wainwright MJ, Jordan MI. Variational inference in graphical models: The view from the marginal polytope. In: . 41. The University; 1998. ; 2003: 961–971.

40. Wright AH. Genetic algorithms for real parameter optimization. In: . 1. Elsevier. 1991 (pp. 205–218).

41. Glaudell R, Garcia RT, Garcia JB. Nelder-mead simplex method. *Computer Journal* 1965; 7(4): 308–313.

42. Singh P, Mishra HB, Jagannatham AK, Vasudevan K. Semi-blind, training, and data-aided channel estimation schemes for MIMO-FBMC-OQAM systems. *IEEE Transactions on Signal Processing* 2019; 67(18): 4668–4682.

43. Beal MJ. *Variational algorithms for approximate Bayesian inference*. PhD thesis. UCL (University College London), 2003.

44. Van Erven T, Harremos P. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory* 2014; 60(7): 3797–3820.

45. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 1977; 39(1): 1–22.

46. Minka TP. *A family of algorithms for approximate Bayesian inference*. PhD thesis. Massachusetts Institute of Technology, 2001.

47. Bernardo J, Bayarri M, Berger J, et al. The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian statistics* 2003; 7(453-464): 210.

48. Bishop CM, Nasrabadi NM. *Pattern recognition and machine learning*. 4. Springer . 2006.

49. Neal RM, Hinton GE. A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Springer. 1998 (pp. 355–368).

## AUTHOR BIOGRAPHY

**Rabah Ouchikh.** was born in 1993, in Boumerdes, Algeria. He received the State Engineering degree (Ingéniorat d'État) in electrical engineering (Telecommunications) in 2018 from the Military Polytechnic School (EMP), Algiers, Algeria, and the Master degree in electronics (telecommunications) in 2018 from the Military Polytechnic School (EMP), Algiers. Currently, he is a PhD student at the telecommunications laboratory, EMP, under the supervision of Prof. Mustapha Djeddou. His research interests are in statistical signal processing with applications to communications.

**Thierry Chonavel.** received the master's degree from IMT Atlantique (Télécom Bretagne), Brest, France, in 1987, and the Ph.D. degree in signal and image processing from Télécom ParisTech, Paris, France, in 1992. After working for Alcatel, he joined IMT. Since 1993 he has been with IMT Atlantique as an Associate Professor and then as a Professor (since 2004). His research activities are mainly related to statistical signal processing and machine learning, with applications in the fields of transmissions, networks, radar, sonar and array processing for radio and underwater acoustics.

**Abdeldjalil Aïssa-El-Bey.** received the State Engineering degree from Ecole Nationale Polytechnique (ENP), Algiers, Algeria, in 2003, the M.S. degree in signal processing from Supelec and Paris XI University, Orsay, France, in 2004, and the Ph.D. degree in signal and image processing from ENST Paris, France, in 2007. In 2007, he joined the Signal and Communications department of IMT Atlantique (Telecom Bretagne), Brest, France as an Associate Professor, and then Professor, since 2015. He was a Visiting Researcher at Fujitsu Laboratories, Japan and the department of Electrical and Electronic Engineering of The University of Melbourne, Australia in 2010 and 2015, respectively. His research interests are blind source separation, blind system identification and equalization, compressed sensing, sparse signal processing, statistical signal processing, wireless communications, and adaptive filtering.

**Mustapha Djeddou.** received his MS degree in 1998 and his PhD with highest honor in electrical engineering in 2005, both from the National Polytechnic School, Algiers. Currently, he is a senior lecturer at the National Polytechnic School, Algiers. His research interests include statistical signal processing, applied signal processing for digital communication, and wireless communication.