Sparse Superimposed Pilot Based Channel Estimation in OTFS Systems

Fathima Jesbin, Sandesh Rao Mattu, and A. Chockalingam
Department of ECE, Indian Institute of Science, Bangalore

***Abstract*—Traditional orthogonal time frequency space (OTFS) channel estimation schemes dedicate an entire frame or a part of the frame for accommodating pilot and guard symbols to avoid pilot-data interference, which compromises spectral efficiency. This spectral efficiency loss can be avoided using superimposed pilots, where delay-Doppler (DD) bins in the OTFS frame carries both data and pilot symbols. In this paper, we propose a *sparse superimposed pilot scheme* for channel estimation, where all the DD bins in a frame carry data symbols and pilot symbols are superimposed over some of them, sparsely. The proposed scheme does not suffer spectral efficiency loss due to pilot/guard symbols. It also has the advantage of more localized pilot-data interference profile that leads to better performance. We derive the minimum mean square error (MMSE) channel estimator for the proposed scheme. We obtain optimum number of pilot symbols per frame and power distribution among data and pilot symbols through simulations. Simulation results show that the proposed scheme achieves better performance at a lesser complexity compared to existing superimposed pilot scheme. An iterative scheme that further improves performance is also proposed.**

***Index Terms*—OTFS modulation, delay-Doppler channel estimation, superimposed pilots, pilot arrangement, MMSE estimate.**

I. Introduction

Orthogonal time frequency space (OTFS) modulation is a robust modulation scheme for achieving reliable communication in doubly-selective channels [1]-[6]. Two key features of OTFS modulation are: 1) multiplexing of information symbols in the delay-Doppler (DD) domain, and 2) representing the doubly-selective channel in the DD domain. The advantage of representing the time-varying channel in the DD domain is that the channel response is sparse in the DD domain and it remains time invariant for long observation times. Thus, only a few parameters are required to model the channel in the DD domain, making channel estimation less complex [5]-[10].

In order to facilitate channel estimation at the receiver, pilot symbols are transmitted along with data symbols. The number of pilot symbols and their placement in an OTFS frame can impact performance and spectral efficiency. Several existing OTFS channel estimation schemes exploit the DD channel sparsity and time-invariance, and use various pilot placement schemes. The exclusive pilot scheme in [7] dedicates an entire frame for pilot, where a single pilot symbol is placed at the middle of the frame and the rest of the frame are filled with zeros. In the embedded pilot scheme in [8], a pilot symbol is placed at the middle of the frame, which is surrounded by guard symbols (zeros), which are surrounded by data symbols. While pilot-data interference is avoided in both these schemes, spectral efficiency is compromised. This drawback can be alleviated using superimposed pilots [9],[10].

In superimposed pilot schemes, data symbols are placed in all the bins with pilots superimposed on few or all data symbols, thus acquiring full rate. However, this introduces interference/leakage between pilot and data symbols. In the superimposed pilot scheme in [9], all data symbols are superimposed with pilot symbols. We call this scheme as *SP-Full* scheme. For this scheme, a minimum mean square error (MMSE) channel estimator followed by an iterative scheme to improve estimation accuracy is presented in [9]. In the superimposed pilot scheme in [10], only the data symbol in the middle of the frame is superimposed with a pilot symbol. We call this scheme as *SP-One* scheme. For this scheme, a threshold based estimator followed by an iterative interference cancellation scheme is presented in [10]. We note that these two schemes represent two ends of a pilot placement numerology that achieves full rate, i.e., while a pilot is superimposed in only one bin in SP-One scheme, pilots are superimposed in all the bins in SP-Full scheme. While both schemes are attractive for their full rate and simplicity in pilot placement, they are not optimum. For example, while the SP-Full scheme has the advantage of having multiple measurements due to multiple pilots in a frame, it suffers from significant pilot-data interference over the entire frame and a weak per pilot symbol energy for a given total pilot energy per frame. On the other hand, the SP-One scheme has the benefit of a more localized pilot-data interference and a strong pilot symbol. But it suffers from lack of inadequate measurements due to only one pilot. This observation indicates the possibility of better pilot placement numerology in between the two end cases, and this motivates the new contribution in this paper. Specifically, we propose a full-rate superimposed pilot scheme where pilots are superimposed on few data symbols in a sparse manner, striking a balance between adequate number of measurements and localized pilot-data interference, leading to better performance compared to existing superimposed pilot schemes. The new contributions in this paper can be summarized as follows.

- We propose a full rate *sparse superimposed pilot scheme* for DD channel estimation in OTFS, where all the bins in a frame carry data symbols and pilot symbols are superimposed over some of them, sparsely. The proposed scheme strikes a good balance between pilot-data interference and adequate number of measurements. We call the proposed scheme as *SP-Sparse* scheme.
- We derive the MMSE channel estimator for the proposed

This work was supported in part by the J. C. Bose National Fellowship, Department of Science and Technology, Government of India.

scheme. We obtain optimum number of pilot symbols per frame and power distribution among the data and pilot symbols through simulations.

- Simulation results show that the proposed scheme achieves better performance at a lesser complexity compared to the scheme in [9]. While the better performance is due to the better pilot-data interference profile in the proposed scheme, the lesser complexity of the estimator is because of the need to process only a smaller number of bins in the frame in the proposed scheme.
- An iterative scheme that further improves the performance is also proposed.

The rest of the paper is organized as follows. The OTFS system model is presented in Sec. II. The proposed SP-Sparse scheme, its MMSE estimator derivation, and the iterative scheme are presented in Sec. III. Results and discussions are presented in Sec. IV. Conclusions are presented in Sec. V.

II. OTFS System Model

The block diagram of OTFS modulation scheme is shown in Fig. 1. At the transmitter, information symbols are multiplexed in the DD domain. This is transformed to the time-frequency (TF) domain using the inverse symplectic finite Fourier transform (ISFFT), followed by Heisenberg transform to convert to time domain for transmission. At the receiver, the time domain signal is converted back to TF domain using Wigner transform and subsequently transformed to DD domain using the symplectic finite Fourier transform (SFFT) where data detection is done.

Information symbols, $\mathbf{X}^{\text{dd}}[k,l]$s, drawn from a modulation alphabet $\mathbb{A}$ are multiplexed in the DD grid, given by $\left\{\left(\frac{k}{NT},\frac{l}{M\Delta f}\right), k=0,\cdots,N-1, l=0,\cdots,M-1\right\}$, where M and N denote the number of delay and Doppler bins, respectively, Δf is the subcarrier spacing, and $\mathbf{X}^{\text{dd}} \in \mathbb{A}^{M\times N}$ denotes the information symbol matrix. The DD domain symbols are mapped to TF domain symbols using ISFFT operation as

$$\mathbf{X}^{\text{tf}}[n,m] = \frac{1}{\sqrt{MN}} \sum_{k=0}^{N-1}\sum_{l=0}^{M-1} \mathbf{X}^{\text{dd}}[k,l] e^{j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)}, \quad (1)$$

where $n = 0,\cdots,N-1$ and $m = 0,\cdots,M-1$. The TF frame has duration NT and bandwidth $M\Delta f$, where T and Δf are the sampling intervals along time and frequency axis, respectively, satisfying $T\Delta f = 1$. These TF domain symbols arranged in matrix form is denoted by $\mathbf{X}^{\text{tf}} \in \mathbb{C}^{M\times N}$, and (1) can alternatively be expressed in matrix form as $\mathbf{X}^{\text{tf}} = \mathbf{F}_M \mathbf{X}^{\text{dd}} \mathbf{F}_N^H$, where $\mathbf{F}_M \in \mathbb{C}^{M\times M}$ and $\mathbf{F}_N \in \mathbb{C}^{N\times N}$ are the normalized discrete Fourier transform (DFT) matrices with $\mathbf{F}_M[a,b] = (1/\sqrt{M})\exp(-j2\pi ab/M)$, $\mathbf{F}_N[a,b] = (1/\sqrt{N})\exp(-j2\pi ab/N)$, and $(\cdot)^H$ denotes the Hermitian operation. The TF domain samples, $\mathbf{X}^{\text{tf}}[n,m]$s, are pulse shaped using transmit pulse $g_{\text{t}}(t)$ to generate the time domain signal $x(t)$, which can be expressed in matrix form as $\mathbf{X}^t = \mathbf{G}_{\text{t}}\mathbf{F}_M^H\mathbf{X}^{\text{tf}} = \mathbf{G}_{\text{t}}\mathbf{X}^{\text{dd}}\mathbf{F}_N^H$, where $\mathbf{X}^t \in \mathbb{C}^{M\times N}$ contains MN samples of $x(t)$, obtained using the sampling

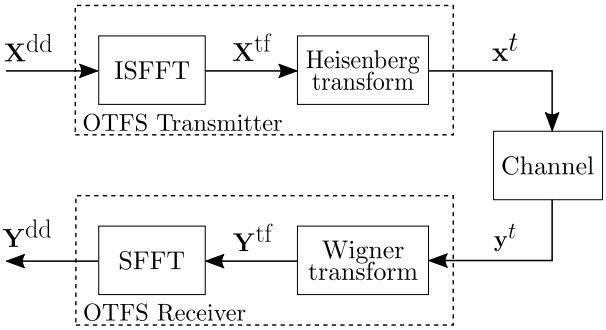


Fig. 1: OTFS modulation scheme.

rate $f_s = M\Delta f$. Further, for the transmit and receive pulses, the sampling interval is set to $T_s = 1/M\Delta f = T/M$ as per symbol spaced sampling approach, which results in M-length samples of the transmit and receive pulses. $\mathbf{G}_{\text{t}} \in \mathbb{C}^{M\times M}$ is a diagonal matrix, whose diagonal entries are obtained by uniformly sampling the transmit pulse $g_{\text{t}}(t)$ at time instants $mT/M, m = 0,1,\cdots,M-1$. Using the relation $\text{vec}(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X})\text{vec}(\mathbf{Y})$, where $\otimes$ denotes Kronecker product, the time domain vector can be written as

$$\mathbf{x}^t = \text{vec}(\mathbf{X}^t) = \text{vec}(\mathbf{G}_{\text{t}}\mathbf{X}^{\text{dd}}\mathbf{F}_N^H) = (\mathbf{F}_N^H \otimes \mathbf{G}_{\text{t}})\mathbf{x}, \quad (2)$$

where $\mathbf{x} = \text{vec}(\mathbf{X}^{\text{dd}})$ and the operation $\text{vec}(\mathbf{Z})$ vectorizes matrix $\mathbf{Z}$. Let $h(\tau,\nu)$ denote the complex baseband channel impulse response of the time-varying channel in the DD domain, where τ and ν represent the delay and Doppler variables, respectively. Then,

$$h(\tau,\nu) = \sum_{i=1}^{P} h_i \delta(\tau-\tau_i)\delta(\nu-\nu_i), \quad (3)$$

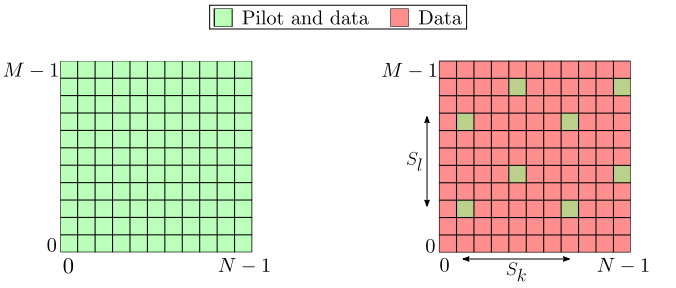
where P is the number of channel paths in the DD domain, δ is the Kronecker delta function, and h_i, τ_i, and ν_i denote the complex channel gain, delay, and Doppler, respectively, corresponding to the ith path. For integer delays and Dopplers, $\tau_i = \frac{l_i}{M\Delta f}$ and $\nu_i = \frac{k_i}{NT}$, where l_i and k_i denote the corresponding delay and Doppler taps, respectively. The received time domain signal, $y(t)$, is given by

$$y(t) = \int_\nu \int_\tau h(\tau,\nu) x(t-\tau) e^{j2\pi\nu(t-\tau)} \text{d}\tau \text{d}\nu + w(t), \quad (4)$$

where $w(t)$ is the additive noise. A forward cyclic shift matrix defined as

$$\mathbf{\Pi} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \in \{0,1\}^{MN\times MN}, \quad (5)$$

and $\mathbf{\Delta} = \text{diag}\{s^0, s^1, \cdots, s^{MN-1}\}$ with $s = \exp(2\pi j/MN)$ model the delays and Dopplers, respectively, so that the channel matrix $\mathbf{H} \in \mathbb{C}^{MN\times MN}$ can be defined as $\mathbf{H} = \sum_{i=1}^{P} h_i \mathbf{\Pi}^{l_i}\mathbf{\Delta}^{k_i}$ [11]. The received signal vector can be written as $\mathbf{y}^t = \mathbf{H}\mathbf{x}^t + \mathbf{n} \in \mathbb{C}^{MN\times 1}$, where $\mathbf{y}^t \in \mathbb{C}^{MN\times 1}$ is obtained by sampling $y(t)$. The TF matrix $\mathbf{Y}^{\text{tf}} \in \mathbb{C}^{M\times N}$ is derived from $\mathbf{y}^t$ using the Wigner transform, i.e., $\mathbf{Y}^{\text{tf}} = \mathbf{F}_M\mathbf{G}_{\text{r}}\mathbf{Y}^t$, where $\mathbf{Y}^t = \text{vec}^{-1}(\mathbf{y}^t) \in \mathbb{C}^{M\times N}$, and $\mathbf{G}_{\text{r}}$ is a



(a) Superimposed pilot frame in [9]

(b) Sparse superimposed pilot frame (Proposed)

Fig. 2: Pilot and data symbol placements in the SP-Full scheme in [9] and the proposed SP-Sparse scheme.

diagonal matrix obtained by sampling the receive pulse $g_{\rm r}(t)$. The DD signal matrix $\mathbf{Y}^{\rm dd} \in \mathbb{C}^{M\times N}$ is obtained from $\mathbf{Y}^{\rm tf}$ as

$$\mathbf{Y}^{\rm dd} = \mathbf{F}_M^H \mathbf{Y}^{\rm tf} \mathbf{F}_N = \mathbf{G}_{\rm r} \mathbf{Y}^t \mathbf{F}_N. \tag{6}$$

This can be vectorized to obtain

$$\mathbf{y} = (\mathbf{F}_N \otimes \mathbf{G}_{\rm r})\mathbf{y}^t = (\mathbf{F}_N \otimes \mathbf{G}_{\rm r})(\mathbf{H}\mathbf{x}^t + \mathbf{n}). \tag{7}$$

Substituting (2) in (7), we get

$$\begin{aligned} \mathbf{y} &= (\mathbf{F}_N \otimes \mathbf{G}_{\rm r})\mathbf{H}(\mathbf{F}_N^H \otimes \mathbf{G}_{\rm t})\mathbf{x} + \mathbf{n}' \\ &= \mathbf{H}_{\rm eff}\mathbf{x} + \mathbf{n}'. \end{aligned} \tag{8}$$

For rectangular pulses at the transmitter and receiver, $\mathbf{G}_{\rm t}$ = $\mathbf{G}_{\rm r}$ = $\mathbf{I}_M$, where $\mathbf{I}_M$ is $M \times M$ identity matrix. Also, $\mathbf{n}'$ = $(\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{n}$, and $\mathbf{H}_{\rm eff} = (\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{H}(\mathbf{F}_N^H \otimes \mathbf{I}_M) \in \mathbb{C}^{MN\times MN}$ is the effective channel matrix.

III. Proposed Sparse Superimposed Pilot Scheme

In this section, we present the proposed sparse superimposed pilot (SP-Sparse) scheme for DD channel estimation in OTFS. The scheme aims to alleviate the problem of inter-pilot interference and combat the effect of pilot-data interference, which is inevitable for a superimposed pilot scheme. We achieve this by 1) appropriate pilot placement in an OTFS frame, 2) optimizing the power distribution between the pilot and data symbols, 3) optimizing the number of pilot symbols in a frame, and 4) iterative interference cancellation.

Let x_d denote the data symbols with $\mathbb{E}\{|x_d[k,l]|^2\} = \sigma_d^2, \forall\ k,l$. Let x_p denote the pilot symbols and N_p denote the number of pilot symbols in a frame. We fix the average symbol energy per bin (and, therefore, the average energy of the frame) to be one as in [9]. Let $\sigma_p^2 = 1 - \sigma_d^2$. Then the total energy in the pilot symbols in a frame is $MN\sigma_p^2$, which is distributed uniformly among the N_p pilot symbols.

A. Pilot placement

Accurate channel estimation is a requisite for reliable data detection. Pilot symbols are transmitted along with data to obtain an estimate of the DD channel matrix $\mathbf{H}_{\rm eff}$ (see (8)). Depending on the channel characteristics, the transmitted symbols spread into the adjacent bins. The extent of spread is determined by delay and Doppler spreads of the channel. The pilot arrangement in a frame has to be suitably chosen taking into account 1) the spread due to the channel which causes interference that influences performance, and 2) the spectral efficiency. The popular pilot placement schemes like exclusive pilot [7] and embedded pilot [8] schemes take a toll on spectral efficiency, in spite of having better bit error performance by virtue of their pilot arrangement that avoids pilot-data interference. Use of superimposed pilots, on the other hand, does not cause spectral efficiency loss at the cost of interference between symbols.

Figure 2 shows two frame structures that achieve full rate by superimposing pilot symbols over data symbols. Figure 2a shows the full superimposed pilot (SP-Full) frame proposed in [9], where pilot symbols are superimposed on data symbols in all the bins. In contrast, in the proposed sparse superimposed pilot (SP-Sparse) frame shown in Fig. 2b, the pilot symbols are interleaved in the DD grid with a separation of S_l and S_k bins along the delay and Doppler axis, respectively, in a lattice-type arrangement. S_l and S_k are chosen so as to accommodate the optimal number of pilots and to minimize the inter-pilot interference. As mentioned earlier, the proposed scheme has the benefit of higher energy per pilot symbol which is favourable for channel estimation (as will be shown later).

B. MMSE channel estimator for the proposed scheme

In this subsection, we derive the MMSE channel estimate for the proposed SP-Sparse scheme. As the MMSE estimator is optimal in the MSE sense and a closed-form expression can be derived for the proposed setting, we choose this estimator for its simplicity. For estimating the channel coefficients, N_p pilot symbols are superimposed on the data symbols at the transmitter. At the receiver, the knowledge of these pilot symbols are used for deriving an MMSE channel estimator, as detailed below.

Let (k_{p_i}, l_{p_i}), $i = 1, \cdots, N_p$ denote the pilot symbol locations in the DD grid, with the separations S_l and S_k, in a lattice-type arrangement. For the ith pilot, the received symbols $\mathbf{Y}^{\rm dd}[k,l]$, with $k_{p_i} - k_{max} \le k \le k_{p_i} + k_{max}$, $l_{p_i} \le l \le l_{p_i} + l_{max}$, are obtained. These can be expressed in terms of the DD domain transmitted symbols as [6]

$$\mathbf{Y}^{\rm dd}[k,l] = \sum_{j=1}^{P} h_j \beta_j^{(i)} \mathbf{X}^{\rm dd}[k-k_j, l-l_j] + \mathbf{N}[k,l], \tag{9}$$

where $\beta_j^{(i)} = \exp\left(j2\pi\frac{l_{p_i}k_j}{MN}\right)$, for $i = 1, \cdots, N_p$, and $\mathbf{N}[k,l] \sim \mathcal{CN}(0,\sigma^2)$ is the additive noise sample. We note that the effect of each transmitted symbol is seen in P received symbols at the receiver. For the DD bin containing the ith pilot symbol, let $y_j^{(i)}$ with $j = 1, \cdots, P$, denote the corresponding P received symbols. Each $y_j^{(i)}$ is due to a combination of the effect of the channel on P transmit DD symbols, denoted by $\{x_q^{(i,j)}, q = 1, \cdots, P\}$. Among the P transmit symbols $x_q^{(i,j)}$s, one of them, say $x_j^{(i,j)}$, contains both pilot and data symbols ($x_{d_q}^{(i,j)}$s). Then,

$$x_q^{(i,j)} = \begin{cases} x_p + x_{d_q}^{(i,j)}, & \text{for } q = j \\ x_{d_q}^{(i,j)}, & \text{for } q \neq j. \end{cases}$$

Using (9), each $y_j^{(i)}$ can be expressed as

$$y_j^{(i)} = h_j \beta_j^{(i)} (x_p + x_{d_j}^{(i,j)}) + \sum_{q=1, q\neq j}^{P} h_q \beta_q^{(i)} x_{d_q}^{(i,j)} + n_j^{(i)}, \quad (10)$$

for $j = 1, \cdots, P$ and $i = 1, \cdots, N_p$. (10) can be vectorized to obtain the received vector corresponding to ith pilot as

$$\mathbf{y}^{(i)} = \mathbf{X}^{(i)} \boldsymbol{\beta}^{(i)} \mathbf{h} + \mathbf{n}^{(i)}, \quad (11)$$

where $\mathbf{y}^{(i)} \in \mathbb{C}^{P\times 1}$, $\mathbf{n}^{(i)} \in \mathbb{C}^{P\times 1}$, $\mathbf{h} = [h_1, \cdots, h_P]^T \in \mathbb{C}^{P\times 1}$ is the DD domain channel vector, $\boldsymbol{\beta}^{(i)} = \text{diag}(\beta_1^{(i)}, \cdots, \beta_P^{(i)}) \in \mathbb{C}^{P\times P}$, and

$$\mathbf{X}^{(i)} = \begin{bmatrix} x_p + x_{d_1}^{(i,1)} & x_{d_2}^{(i,1)} & \cdots & x_{d_P}^{(i,1)} \\ x_{d_1}^{(i,2)} & x_p + x_{d_2}^{(i,2)} & \cdots & x_{d_P}^{(i,2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d_1}^{(i,P)} & x_{d_2}^{(i,P)} & \cdots & x_p + x_{d_P}^{(i,P)} \end{bmatrix} \in \mathbb{C}^{P\times P}. \quad (12)$$

For N_p pilot symbols, (11) can be written as

$$\mathbf{y}_{\text{obs}} = \mathbf{X}\boldsymbol{\beta}\mathbf{h} + \mathbf{n}_{\text{obs}}, \quad (13)$$

where $\mathbf{y}_{\text{obs}} = [\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(N_p)}]^T \in \mathbb{C}^{PN_p\times 1}$, $\mathbf{n}_{\text{obs}} = [\mathbf{n}^{(1)}, \cdots, \mathbf{n}^{(N_p)}]^T \in \mathbb{C}^{PN_p\times 1}$, $\boldsymbol{\beta} = [\boldsymbol{\beta}^{(1)}, \cdots, \boldsymbol{\beta}^{(N_p)}]^T \in \mathbb{C}^{PN_p\times P}$, and $\mathbf{X} = \text{diag}(\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(N_p)}) \in \mathbb{C}^{PN_p\times PN_p}$ is a block diagonal matrix. Further, $\mathbf{X}$ can be expressed as

$$\begin{aligned} \mathbf{X} &= \text{diag}\big(x_p\mathbf{I}_P + \mathbf{X}_d^{(1)}, \cdots, x_p\mathbf{I}_P + \mathbf{X}_d^{(N_p)}\big) \\ &= x_p\mathbf{I}_{PN_p} + \mathbf{X}_d, \end{aligned} \quad (14)$$

where $\mathbf{X}_d = \text{diag}\big(\mathbf{X}_d^{(1)}, \cdots, \mathbf{X}_d^{(N_p)}\big) \in \mathbb{C}^{PN_p\times PN_p}$ and

$$\mathbf{X}_d^{(i)} = \begin{bmatrix} x_{d_1}^{(i,1)} & x_{d_2}^{(i,1)} & \cdots & x_{d_P}^{(i,1)} \\ x_{d_1}^{(i,2)} & x_{d_2}^{(i,2)} & \cdots & x_{d_P}^{(i,2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d_1}^{(i,P)} & x_{d_2}^{(i,P)} & \cdots & x_{d_P}^{(i,P)} \end{bmatrix} \in \mathbb{C}^{P\times P}, \quad (15)$$

for $i = 1, \cdots, N_p$. Substituting (14) in (13), we have

$$\begin{aligned} \mathbf{y}_{\text{obs}} &= x_p\boldsymbol{\beta}\mathbf{h} + \mathbf{X}_d\boldsymbol{\beta}\mathbf{h} + \mathbf{n}_{\text{obs}} \\ &= \boldsymbol{\beta}_p\mathbf{h} + \boldsymbol{\beta}_d\mathbf{h} + \mathbf{n}_{\text{obs}}, \end{aligned} \quad (16)$$

where $\boldsymbol{\beta}_p = x_p\boldsymbol{\beta} \in \mathbb{C}^{PN_p\times P}$ and $\boldsymbol{\beta}_d = \mathbf{X}_d\boldsymbol{\beta} \in \mathbb{C}^{PN_p\times P}$. The simplified expression for the observation vector becomes

$$\mathbf{y}_{\text{obs}} = \boldsymbol{\beta}_p\mathbf{h} + \mathbf{n}_d, \quad (17)$$

where $\mathbf{n}_d = \boldsymbol{\beta}_d\mathbf{h} + \mathbf{n}_{\text{obs}} \in \mathbb{C}^{PN_p\times 1}$ is the noise plus interference vector, whose mean and covariance matrix are denoted by $\boldsymbol{\mu}_{n_d}$ and $\mathbf{C}_{n_d}$, respectively. Note that noise vector has mean $\mathbb{E}[\mathbf{n}_{\text{obs}}] = \mathbf{0}_{PN_p\times 1}$ and covariance matrix $\mathbf{C}_{n_{\text{obs}}} = \mathbb{E}[\mathbf{n}_{\text{obs}}\mathbf{n}_{\text{obs}}^H] = \sigma^2\mathbf{I}_{PN_p}$. Also, the channel vector $\mathbf{h}$ has zero mean and covariance matrix $\mathbf{C}_h = \mathbb{E}[\mathbf{h}\mathbf{h}^H] = \text{diag}(\sigma_{h_1}^2, \cdots, \sigma_{h_P}^2)$, where $\sigma_{h_i}^2$ represent the power of ith channel gain. Since $\mathbb{E}[\boldsymbol{\beta}_d] = \mathbf{0}_{PN_p\times P}$ and $\mathbb{E}[\mathbf{h}] = \mathbf{0}_{P\times 1}$, $\boldsymbol{\mu}_{n_d} = \mathbb{E}[\mathbf{n}_d] = \mathbf{0}_{PN_p\times 1}$, we can write

$$\begin{aligned} \mathbf{C}_{n_d} &= \mathbb{E}\{\mathbf{n}_d\mathbf{n}_d^H\} \\ &= \mathbb{E}\{\boldsymbol{\beta}_d\mathbf{h}\mathbf{h}^H\boldsymbol{\beta}_d^H\} + \sigma^2\mathbf{I}_{PN_p} \\ &= \frac{\text{Tr}(\mathbf{C}_h)}{P}\mathbb{E}\{\boldsymbol{\beta}_d\boldsymbol{\beta}_d^H\} + \sigma^2\mathbf{I}_{PN_p}, \end{aligned} \quad (18)$$

where $\text{Tr}(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$ and $\text{Tr}(\mathbf{C}_h) = \sum_{i=1}^{P}\sigma_{h_i^2}$. Further,

$$\mathbb{E}\{\boldsymbol{\beta}_d\boldsymbol{\beta}_d^H\} = \mathbb{E}\{\mathbf{X}_d\mathbf{X}_d^H\} = P\sigma_d^2\mathbf{I}_{PN_p}. \quad (19)$$

Substituting (19) in (18) gives

$$\mathbf{C}_{n_d} = \Big(\sum_{i=1}^{P}\sigma_{h_i}^2\sigma_d^2 + \sigma^2\Big)\mathbf{I}_{PN_p}. \quad (20)$$

Using (17), the MMSE estimate, $\hat{\mathbf{h}}$, of the channel vector $\mathbf{h}$ can be obtained as [12]

$$\hat{\mathbf{h}} = (\boldsymbol{\beta}_p^H\mathbf{C}_{n_d}^{-1}\boldsymbol{\beta}_p + \mathbf{C}_h^{-1})^{-1}\boldsymbol{\beta}_p^H\mathbf{C}_{n_d}^{-1}\mathbf{y}_{\text{obs}}. \quad (21)$$

The MMSE estimate $\hat{\mathbf{h}} = \{\hat{h}_i, i = 1\cdots, P\}$ is used to construct the channel matrix estimate $\hat{\mathbf{H}} = \sum_{i=1}^{P}\hat{h}_i\boldsymbol{\Pi}^{l_i}\boldsymbol{\Delta}^{k_i}$ and the effective channel matrix estimate, $\hat{\mathbf{H}}_{\text{eff}} = (\mathbf{F}_N \otimes \mathbf{I}_M)\hat{\mathbf{H}}(\mathbf{F}_N^H \otimes \mathbf{I}_M)$. This channel matrix estimate is used for cancelling the effect of sparsely placed pilots in the as

$$\mathbf{y}_{\text{d}} = \mathbf{y} - \hat{\mathbf{H}}_{\text{eff}}\mathbf{x}_{\text{p}}, \quad (22)$$

where

$$\mathbf{x}_p[kM + l] = \begin{cases} x_p, & \text{if } (k,l) \in \{(k_{p_1}, l_{p_1}), \cdots, (k_{p_{N_p}}, l_{p_{N_p}})\} \\ 0, & \text{else,} \end{cases}$$

and $\mathbf{y} = \mathbf{H}_{\text{eff}}\mathbf{x} + \mathbf{n}' = \mathbf{H}_{\text{eff}}(\mathbf{x}_p + \mathbf{x}_d) + \mathbf{n}'$ (see (8)). The message passing (MP) detection algorithm in [6] is employed on this vector for data detection to obtain $\hat{\mathbf{x}}_d$, the estimate of the transmitted data symbols, $\mathbf{x}_d$. The estimate for the channel coefficients obtained in (21) is in the presence of interference from data symbols. To further improve the accuracy of the estimates, we follow an iterative scheme as outlined below.

C. Iterative scheme for improving performance

Using the initial data estimates, $\hat{\mathbf{x}}_d$, a new frame is constructed by removing the effect of data symbols as

$$\mathbf{y}_{\text{e}} = \mathbf{y} - \hat{\mathbf{H}}_{\text{eff}}\hat{\mathbf{x}}_d = \mathbf{H}_{\text{eff}}\mathbf{x}_p + (\mathbf{H}_{\text{eff}}\mathbf{x}_d - \hat{\mathbf{H}}_{\text{eff}}\hat{\mathbf{x}}_d) + \mathbf{n}'. \quad (23)$$

Assuming that cancellation is perfect, the second term on the right hand side of (23) can be omitted to obtain

$$\mathbf{y}_{\text{e}}{}' = \mathbf{H}_{\text{eff}}\mathbf{x}_p + \mathbf{n}'. \quad (24)$$

Following the same procedure as before, we pick P symbols corresponding to each pilot location using $\mathbf{y}_{\text{e}}{}'$ as the received vector. For $\mathbf{y}_{\text{e}}{}'$, (16) can be approximated as

$$\mathbf{y}'_{\text{obs}} = \boldsymbol{\beta}_p\mathbf{h} + \mathbf{n}_{\text{obs}}. \quad (25)$$

The MMSE estimate now becomes

$$\hat{\mathbf{h}}' = (\boldsymbol{\beta}_p^H\mathbf{C}'^{-1}_{n_d}\boldsymbol{\beta}_p + \mathbf{C}_h^{-1})^{-1}\boldsymbol{\beta}_p^H\mathbf{C}'^{-1}_{n_d}\mathbf{y}'_{\text{obs}}, \quad (26)$$

where $\mathbf{C}'_{n_d} = \sigma^2\mathbf{I}_{PN_p}$. Using the estimate $\hat{\mathbf{h}}'$, the effective channel matrix is computed again. This is used for data detection using (22). This completes one iteration of the scheme employed for combating interference between pilot and data symbols. The iterations are carried out for a maximum of N_{iter} iterations or until a convergence criterion is met. The iterations are stopped at the Lth iteration if $\|\hat{\mathbf{h}}'^{(L)} - \hat{\mathbf{h}}'^{(L-1)}\|^2 < \epsilon$,

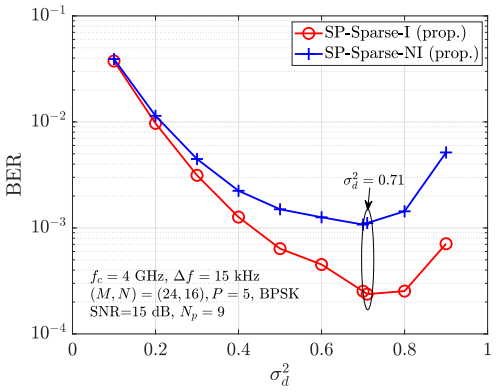


Fig. 3: BER as function of the data energy, σ_d^2.

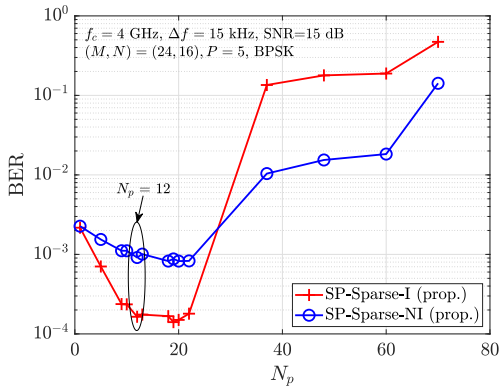


Fig. 4: BER as a function of number of pilots, N_p.

i.e., the squared norm of the difference between the channel estimate vector for the Lth iteration and $(L-1)$th is less than a small value ϵ. The estimate of the data symbols obtained after the iterations is used for evaluating the bit error performance for the iterative scheme.

IV. Results and Discussions

In this section, we present the performance of the proposed SP-Sparse scheme. We also present the performance of the SP-Full scheme in [9] for comparison. For all the simulation results presented below, we consider an OTFS system with delay bins $M = 24$, Doppler bins $N = 16$, carrier frequency $f_c = 4$ GHz, and subcarrier spacing $\Delta f = 15$ kHz. We consider a channel having 5 taps with power delay profile as in [9]. The Doppler spreads are generated using Jakes' formula, $\nu_i = \nu_{\max} \cos\theta_i$ where $\nu_{\max} = 1851$ Hz corresponding to a maximum speed of $v_{\max} = 500$ kmph, and θ_i is uniformly distributed in $[-\pi, \pi]$. Data symbols are drawn from BPSK alphabet and a constant +1 is chosen as the pilot symbol. For the iterative scheme, the maximum number of iterations, N_{iter}, is chosen to be 10, and $\epsilon = 10^{-6}$. Similar to the SP-Full scheme in [9], the total energy per bin is fixed to be one, i.e, for each bin $\sigma_d^2 + \sigma_p^2 = 1$, for fair comparison. The optimal power allocation for the SP-Full scheme in [9] is $\sigma_{d_{opt}}^2 = 0.6678$. For each SNR, 10^4 different channel realizations are considered.

A. Choosing the optimal energy distribution

The BER performance of the proposed SP-Sparse scheme as a function of data energy (σ_d^2) at an SNR of 15 dB is shown in Fig. 3. For the performance, number of pilots, N_p, is assumed to be 9, which is the maximum number of pilots that can be accommodated in the frame without inter-pilot interference. At low σ_d^2 values, data symbols have low energy and the pilot energy is high. Even though this results in good accuracy of the channel estimates (due to high pilot energy), the data detection has high error rate (due to low energy in data symbols). As σ_d^2 increases, the BER performance improves due to increase in the data symbol energy. In the high data energy regime, the accuracy of the channel estimates is poor due to weak pilot symbols, resulting in subpar BER performance. The BER performance is found to be minimum at $\sigma_d^2 = 0.71$, which gives $\sigma_{d_{opt}}^2 = 0.71$. For the rest of the simulations, we fix the data symbol energy to be 0.71.

B. Choosing the optimal number of pilots

The BER performance of the SP-Sparse scheme as a function of number of pilots, N_p, at an SNR of 15 dB is shown in Fig. 4. At low N_p values, the number of observations that are available is less (see (21)), which results in poor accuracy of channel estimates. This is reflected in the BER performance in this regime. As N_p is increased, the accuracy of the estimates improves, and consequently the BER performance also improves. When N_p is increased beyond 20, the pilot energy per pilot symbol ($MN\sigma_p^2/N_p$) reduces. This, in addition to the reduced separation between the pilot symbols, negatively affects the channel estimate performance and the BER shoots up. We note that increasing N_p also increases the complexity of the MMSE estimate. As a good balance between the BER performance and the estimation complexity, we choose $N_p = 12$, with $S_l = 7$ and $S_k = 2$, in the simulations presented below.

C. MSE performance

The MSE performance of the proposed scheme, with and without iterations, as a function of SNR is shown in Fig. 5. The MSE performance for the non-iterative (SP-Full-NI) and iterative (SP-Full-I) schemes in [9] are also plotted for comparison. In addition, the performance of the iterative SP-One (SP-One-I) scheme in [10] is also shown. The proposed non-iterative scheme (SP-Sparse-NI) outperforms the corresponding SP-Full-NI scheme by a good margin. For example, at 10 dB SNR, the proposed SP-Sparse-NI scheme achieves an MSE of about 3.5×10^{-2} while the SP-Full-NI scheme in [9] achieves an MSE of about 2×10^{-1}, which is roughly an order of difference. Next, the proposed iterative scheme (SP-Sparse-I) performs similarly to that of the SP-Full-I scheme for SNRs between 0 and 10 dB. The performance of the SP-Sparse-I scheme is better than that of the SP-Full-I scheme at 15 dB SNR. Also, the performance of the SP-Sparse-NI scheme is similar to that of the SP-One-I scheme, while the performance of the proposed SP-Sparse-I scheme is much better.

D. BER performance

The BER performance of the proposed scheme as a function of SNR is shown in Fig. 6. Performance of the SP-Full-NI and SP-Full-I schemes, SP-One-I scheme, and performance

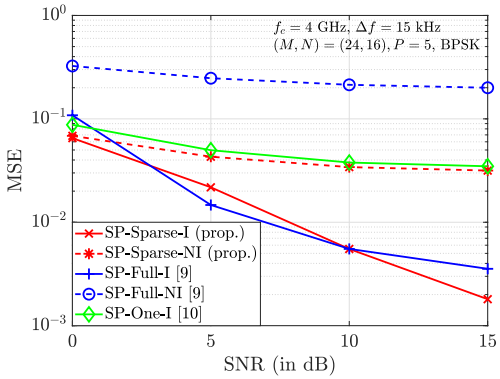


Fig. 5: MSE performance as a function of SNR.

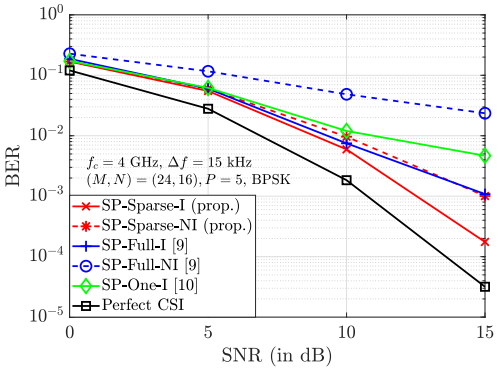


Fig. 6: BER performance as a function of SNR.

with perfect channel knowledge are also plotted for comparison. It can be seen that the proposed SP-Sparse-NI scheme outperforms the SP-Full-NI scheme. For example, a BER of 6×10^{-2} is achieved at 5 dB with SP-Sparse-NI scheme, while SP-Full-NI scheme achieves the same BER at around 9 dB, i.e., a performance advantage of 4 dB is observed in favor of the proposed scheme. The SP-Sparse-I scheme performs closely to that of the SP-Full-I scheme till 5 dB, after which the performance is observed to be better than that of the SP-Full-I scheme. This performance is also close to the performance obtained using perfect channel knowledge. This performance advantage can be attributed to the sparse pilot placements in the DD grid as opposed to pilots placed in every bin in [9]. It is also seen that the performance of SP-One-I scheme is worse compared to the proposed SP-Sparse-NI and SP-Sparse-I schemes, with the performance gap being larger in the latter case. It is observed in Fig. 5 that MSE performance of SP-Full-I is better compared to SP-Sparse-NI. This is owing to the iterative interference cancellation in the former case. However, there is no significant BER improvement in Fig. 6, which is due to increased pilot-data interference resulting due to pilot placement in the SP-Full-I scheme.

E. Complexity

Here, we present the run time complexity of the proposed scheme for generating the initial channel estimate and compare it with that of the SP-Full scheme, where both the schemes use MMSE channel estimator. For the SP-Full scheme, the average run time required to compute an estimate is about 3.56×10^{-3} s, whereas for the proposed SP-Sparse scheme is only about 7.74×10^{-5} s. This is also reflected in the number of floating point operations (FLOPs), where the SP-Full-NI requires 3×10^{6} whereas SP-Sparse-NI requires only 7.6×10^{4} FLOPs. This complexity advantage is because the proposed scheme needs to process only a part of the frame to obtain an estimate, whereas the SP-Full scheme processes the entire frame to obtain an estimate.

V. Conclusions

We proposed a full rate sparse superimposed pilot scheme for DD channel estimation in OTFS systems. The sparsity allowed for reduced interference among pilot symbols and higher energy per pilot symbol. To obtain the channel estimates in the presence of interference from data symbols, we derived MMSE channel estimator for the proposed scheme. We obtained the optimum number of pilot symbols and energy distribution between pilot and data symbols through simulations. The proposed scheme along with the proposed MMSE estimator showed better MSE and BER performance compared to superimposed pilot scheme in the literature. This was achieved at a lower complexity because of the fewer number of symbols involved in computing the proposed estimate. The performance was further improved using iterative interference cancellation. Analytical solutions for obtaining the optimum number and location of pilot symbols, and optimal power allocation between data and pilot symbols can be considered for future research.

References

[1] R. Hadani *et al.*, "Orthogonal time frequency space modulation," *Proc. IEEE WCNC'2017*, pp. 1-6, Mar. 2017.

[2] S. S. Das and R. Prasad, *OTFS: Orthogonal Time Frequency Space Modulation A Waveform for 6G*, River Publishers, 2021.

[3] Y. Hong, T. Thaj, and E. Viterbo, *Delay-Doppler Communications: Principles and Applications*, Academic Press, 2022.

[4] Z. Wei *et al.*, "Orthogonal time-frequency space modulation: a promising next-generation waveform," *IEEE Wireless Commun. Mag.*, vol. 28, no. 4, pp. 136-144, Aug. 2021.

[5] K. R. Murali and A. Chockalingam, "On OTFS modulation for high-Doppler fading channels," *Proc. ITA*, pp. 1-10, Feb. 2018.

[6] P. Raviteja *et al.*, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501-6515, Aug. 2018.

[7] M. K. Ramachandran and A. Chockalingam, "MIMO-OTFS in high-Doppler fading channels: signal detection and channel estimation," *Proc. IEEE GLOBECOM'2018*, pp. 206-212, Dec. 2018.

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

[9] H. B. Mishra *et al.*, "OTFS channel estimation and data detection designs with superimposed pilots," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2258-2274, Apr. 2022.

[10] W. Yuan *et al.*, "Data-aided channel estimation for OTFS systems with a superimposed pilot and data transmission scheme," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1954-1958, Sep. 2021.

[11] P. Raviteja *et al.*, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Tech.*, vol. 68, pp. 957-961, Jan. 2019.

[12] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge Univ. Press, 2005.