OTFS Channel Estimation and Data Detection Designs With Superimposed Pilots

Himanshu B. Mishra, *Member, IEEE*, Prem Singh, Abhishek K. Prasad, and Rohit Budhiraja

***Abstract*— We propose a superimposed pilot (SP)-based channel estimation and data detection framework for orthogonal time-frequency space (OTFS) systems, which superimposes low-powered pilots on to data symbols in the delay-Doppler domain. We propose two channel estimation and data detection designs for SP-OTFS systems which, unlike the existing OTFS designs, do not designate any slots for pilots, and consequently have higher spectral efficiency (SE). The first SP design estimates channel by treating data as interference, which degrades its performance at high signal to noise ratio. The second SP design alleviates this problem by iterating between channel estimation and data detection. Both these designs detect data using message passing algorithm which exploits OTFS channel sparsity, and consequently has low computational complexity. We also derive a lower bound on the signal-to-interference-plus-noise ratio of the proposed designs and maximize it by optimally allocating power between data and pilot symbols. We numerically validate the derived analytical results, and show that the proposed designs have superior SE than the existing OTFS channel estimation and data detection designs.**

***Index Terms*— Message passing algorithm, orthogonal time-frequency space (OTFS) scheme, superimposed pilots.**

I. Introduction

A PRACTICAL wireless channel, due to multi-path propagation and Doppler shift, is both time- and frequency-selective. The cyclic prefix (CP)-aided orthogonal frequency division multiplexing (OFDM) combats the channel frequency-selectivity [1]. A large Doppler shift, due to high-speed relative movement between the transmitter and receiver, disturbs inter-subcarrier orthogonality in an OFDM system, which significantly degrades its performance [1]. Orthogonal time frequency and space (OTFS) scheme improves the performance in high-Doppler scenarios by multiplexing transmit symbols in the delay-Doppler domain [2]–[7]. This is unlike OFDM which multiplexes them in the time-frequency domain. The OTFS scheme, by using inverse symplectic finite Fourier transform (ISFFT) and Heisenberg transform at the transmitter, and their inverse at the receiver, converts a doubly-selective channel into an almost time-invariant one, in the delay-Doppler domain. OTFS symbols, thus, experience almost-constant channel gain, which can be exploited to reduce pilot overhead for estimating rapidly time-varying channel. Further, the delay-Doppler domain channel, due to small number of clusters, is sparse. This can also be exploited to reduce channel estimation and data detection complexity [4]–[8].

G. D. Surabhi *et al.* in [2] derived the diversity of single-input single-output (SISO)/multiple-input multiple-output (MIMO)-OTFS systems using maximum-likelihood (ML) decoding with perfect receive channel state information (CSI). The authors in [3] developed linear equalizers for SISO OTFS systems. Raviteja *et al.* in [4] proposed a computationally-efficient detector based on message passing algorithm, which exploits OTFS channel sparsity. These works assumed perfect receive CSI which, however, needs to be estimated in practice. The perfect CSI assumption also simplifies their message calculations. References [5], [6] designed pilot-based OTFS channel estimators in time-frequency domain which, due to non-sparse time-frequency channel, are computationally complex. Also, the time-frequency channel estimate needs to be transformed into the delay-Doppler domain for incorporating low-complexity message passing receiver. Reference [9] proposed orthogonal matching pursuit (OMP) and modified subspace pursuit based delay-Doppler uplink channel estimation algorithm for ideal-pulse-shaped multi-user OTFS systems. The authors in [7], [8] developed channel estimators in the delay-Doppler domain. In particular, Ramachandran *et al.* in [7] first estimated the OTFS channel using a frame consisting entirely of pilot symbols, and used it in the subsequent frames to detect data. This degrades the system spectral efficiency (SE), as an entire OTFS frame is used for estimating channel.

Raviteja *et al.* in [8] proposed an embedded pilot (EP)-based OTFS channel estimator for the frame structure which, as shown in Fig. 1(a), consists of both pilot and data. Here (l_p, k_p) is the pilot location at the l_pth delay, and the k_pth Doppler tap. This design first estimates the channel using the pilot symbols, and subsequently detects data using the estimated channel. This frame structure necessitates the

Manuscript received November 7, 2020; revised March 21, 2021 and July 20, 2021; accepted August 30, 2021. Date of publication September 15, 2021; date of current version April 11, 2022. This work was supported in part by the Science and Engineering Research Board, Department of Science and Technology, Government of India, under Project SRG/2019/000861, in part by the Faculty Research Scheme (FRS), IIT (ISM) Dhanbad, under Grant (118)/2017-18/ECE, and in part by Visvesvaraya Ph.D. Scheme, MeitY, Government of India, under Grant MEITY-PHD-2721. This article was presented in part at the 2021 IEEE International Conference on Communications (ICC) Workshop [DOI: 10.1109/ICCWorkshops50388.2021.9473544]. The associate editor coordinating the review of this article and approving it for publication was C.-K. Wen. *(Corresponding author: Prem Singh.)*

Himanshu B. Mishra and Abhishek K. Prasad are with the Department of Electronics Engineering, IIT Dhanbad, Dhanbad 826004, India (e-mail: himanshu@iitism.ac.in; akp.1462@gmail.com).

Prem Singh is with IIIT Bangalore, Bengaluru 560100, India (e-mail: prem.singh@iiitb.ac.in).

Rohit Budhiraja is with the Department of Electrical Engineering, IIT Kanpur, Kanpur 208016, India (e-mail: rohitbr@iitk.ac.in).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3110659.

Digital Object Identifier 10.1109/TWC.2021.3110659

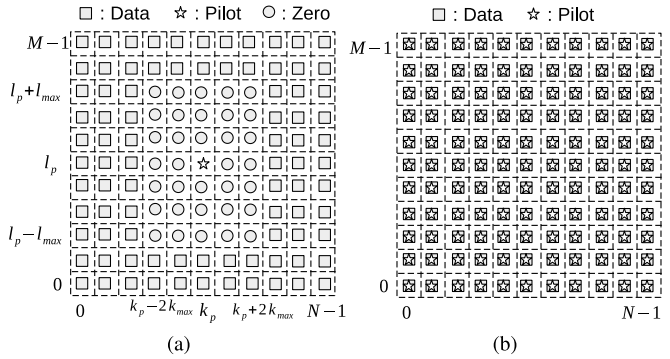


Fig. 1. Frame structure in OTFS systems: (a) EP-based design in [8]; and (b) proposed SP-aided designs.

insertion of zeros between the pilots and data symbols to avoid mutual interference between them [8]. The number of inserted zeros depends on the values of l_{max} and k_{max}, which are the delay and Doppler taps corresponding to their maximum value. The insertion of zeros reduces the SE, especially when the maximum delay value and the Doppler shift are high, a fact we will numerically verify in the sequel.

This SE loss, due to conventional pilot-based channel estimation [7], [8], can be reduced by superimposing pilots on to information symbols. Reference [10] proposed a superimposed training (ST)-aided method for estimating doubly-selective channels in the time-frequency domain for a single carrier system. References [11]–[13] used idea of ST for single-/multi-antenna channel estimation in the time-frequency domain. These works *crucially* considered a nearly time-invariant channel, and used a periodic superimposed pilot (SP) with zero-mean data symbols. This enabled them to mitigate mutual interference between data and pilots using the first-order statistics of the received signal. The authors in [14], [15] proposed affine-precoder-based SP channel estimator, which, unlike [11]–[13], completely removes the interference between data and pilots at the cost of reduced SE. The designs in [11]–[15] also assumed *time-invariant channel*, which degrades their performance for rapidly time-varying channels.

The data-dependent superimposed training (DDST) scheme superimposes, in addition to pilots, the arithmetic mean of transmit data on the transmitted information symbols [16]. The DDST scheme, as explained in [17], has data identifiability problem. Reference [17] resolved it by proposing a hybrid solution, known as joint mean removal SP and pilot-aided training for detection and channel estimation in MIMO system over flat-fading quasi-static channels. The design in [17], similar to [11]–[13], requires channel to be time-invariant for a large number of frames, to compute the time average of received signals.

The SP-aided time-frequency domain designs in [11]–[15] can be seamlessly integrated with OFDM systems. Their extension to estimate delay-Doppler domain OTFS channel systems is non-trivial. This is because the channel gain in the delay-Doppler domain rapidly varies across frames [8], which is not desirable for time-frequency domain SP designs. For extending SP framework to OTFS systems, it is therefore necessary to estimate channel and detect data within a frame, and that too in presence of mutual interference between data and pilots in the delay-Doppler domain [8]. An OTFS system, unlike its OFDM counterpart processes the complete transmit frame of size $M \times N$ where M and N, as shown in Fig. 1, denote the number of delay and Doppler bins [4]. This considerably increases the OTFS system complexity as its channel matrix is of $MN \times MN$ size. *Both channel estimator and data detector in OTFS need to be designed by exploiting delay-Doppler channel sparsity which will reduce their complexities.* The conventional data detectors do not exploit the OTFS sparsity, and are therefore not computationally-efficient [4]. Further, to reduce the bit error rate (BER) and to increase the SE, the transmit power need to be optimally allocated between the data and pilot symbols.

Given the aforementioned challenges, we develop SP-aided channel estimators and data detectors in the delay-Doppler domain for OTFS systems. The SP-aided OTFS systems, to the best of our knowledge, have not yet been investigated. The **main contributions** of this paper, which help in extending the SP framework to SISO OTFS systems, are summarized below. 1) OTFS scheme processes symbols in the delay-Doppler domain, wherein a rapidly time-varying channel is transformed in to a near-stationary channel [2]. This motivates us to propose spectrally-efficient SP-based framework for OTFS which, as shown in Fig. 1(b), consists of a frame which, unlike [8] in Fig. 1(a), does not i) insert zeros between data and pilots; and ii) a require dedicated delay-Doppler slot to transmit pilots. The proposed framework, unlike [7], does not require a dedicated pilot frame to estimate channel. The proposed framework, will thus have significantly higher SE than [7], [8], with a minor BER degradation, a fact we will numerically validate later. 2) We propose two SP-aided channel estimation and data detection designs – SP-non-iterative (SP-NI) and SP-iterative (SP-I). The SP-NI design exploits OTFS channel sparsity by performing minimum mean square error (MMSE) channel estimation in the delay-Doppler domain. It then detects data using computationally-efficient message passing algorithm [4], which again exploits the delay-Doppler domain channel sparsity. Its BER and SE, however, due to the mutual interference between data and pilots, degrades at high signal to noise ratio (SNR) values. The SP-I design mitigates this interference by iterating between channel estimation and data detection in the delay-Doppler domain, and has better BER and SE. 3) We maximize a lower bound on the signal-to-interference-plus-noise-ratio (SINR) of the two proposed designs by deriving a closed-form expression to optimally allocate transmit power between data and pilot symbols. We also evaluate the computational complexities of the proposed and existing schemes. We numerically: i) validate the tightness of the theoretical mean squared error (MSE) and the optimal power allocation; and ii) investigate the effect of power distribution between data and pilot symbols on the MSE, BER, and SE of the proposed designs; and iii) show that the proposed SP-I design, to achieve a SE of 3 bps/Hz,

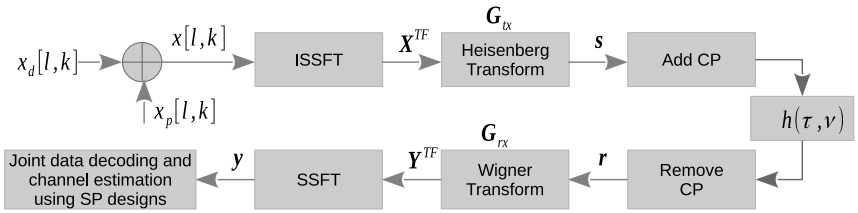


Fig. 2. OTFS system model with superimposed pilot (SP) sequence.

require approximately 15 dB and 5 dB lower SNR than the designs in [7] and [8], respectively.

A. Notations

The superscript $(\cdot)^H$ denotes Hermitian transposition operator and the operation $\mathbf{A} \otimes \mathbf{B}$ denotes Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$. The operator vec$(\mathbf{A})$ vectorises the matrix $\mathbf{A}$, and $\mathbb{E}\{\cdot\}$ represents the exception operation. The operation $\mathbf{A}(p,q)$ extracts the (p,q)th element of the matrix $\mathbf{A}$. The notation diag$\{a_0, a_1, \ldots, a_{N-1}\}$ denotes an $N \times N$ diagonal matrix. The notations $\mathbf{I}_N$ and $\mathbf{0}_{M\times N}$ represents $N \times N$ identity and zero matrices, respectively. The operation Tr$(\mathbf{A})$ computes the trace of the matrix $\mathbf{A}$ and $(\cdot)_M$ denotes modulo M operation. The operators Var$[x]$ and $\mathbb{E}_{\mathbf{a}|\mathbf{b}}$ define the variance and conditional expectation operations, respectively.

II. OTFS System Model With Superimposed Pilot

A. Transmit Vector Generation

We consider, as shown in Fig. 2, a single-antenna SP-OTFS system which transmits a data symbol $x_d[l,k]$ at the lth delay and the kth Doppler location, with $l = 0, 1, \cdots, M-1$ and $k = 0, 1, \cdots N-1$. Here M and N denote the number of delay and Doppler bins, respectively. The data symbol $x_d[l,k]$ is superimposed on to the pilot symbol $x_p[l,k]$ in the delay-Doppler domain by arithmetical addition as follows $x[l,k] = x_d[l,k] + x_p[l,k]$. This equation is next re-expressed in matrix a form as

$$\mathbf{X} = \mathbf{X}_d + \mathbf{X}_p, \tag{1}$$

by arranging $x_d[l,k]$, $x_p[l,k]$, and $x[l,k]$ in matrix form as $\mathbf{X}_d \in \mathbb{C}^{M\times N}$, $\mathbf{X}_p \in \mathbb{C}^{M\times N}$ and $\mathbf{X} \in \mathbb{C}^{M\times N}$, respectively. The zero-mean independent and identically distributed (i.i.d.) elements of the matrices $\mathbf{X}_d$ and $\mathbf{X}_p$ are assumed to have $\mathbb{E}\{|x_d[l,k]|^2\} = \sigma_d^2$ and $\mathbb{E}\{|x_p[l,k]|^2\} = \sigma_p^2$, respectively. We also impose the following constraint that $\mathbb{E}\{|x[l,k]|^2\} = \sigma_d^2 + \sigma_p^2 = 1$.

The OTFS scheme, as shown in Fig. 2, first maps the MN symbols in the delay-Doppler domain to the time-frequency domain by using ISFFT. Let $\mathbf{X}^{\text{TF}} \in \mathbb{C}^{M\times N}$ be the superimposed symbol matrix in the time-frequency domain whose (m,n)th element is given as

$$x^{\text{TF}}[m,n] = \frac{1}{\sqrt{NM}} \sum_{l=0}^{M-1} \sum_{k=0}^{N-1} x[l,k]\, e^{j2\pi\left(\frac{ml}{M} - \frac{nk}{N}\right)}, \tag{2}$$

where $m = 0, 1, \ldots, M-1$ and $n = 0, 1, \ldots, N-1$. The time-frequency frame has a duration NT, and bandwidth $M\Delta f$. Here T and Δf, with $T\Delta f = 1$ [4], are the sampling intervals along the time and frequency axis, respectively. The time-frequency symbol matrix $\mathbf{X}^{\text{TF}}$, using (1) and (2), can be represented as the function of the delay-Doppler matrix $\mathbf{X}$ as $\mathbf{X}^{\text{TF}} = \mathbf{F}_M \mathbf{X} \mathbf{F}_N^H$ [18]. Here $\mathbf{F}_M \in \mathbb{C}^{M\times M}$ and $\mathbf{F}_N \in \mathbb{C}^{N\times N}$ are the normalized discrete Fourier transform (DFT) matrices with $\mathbf{F}_M(p,q) = \left(1/\sqrt{M}\right)\exp\left(-j2\pi pq/M\right)$ and $\mathbf{F}_N(p,q) = \left(1/\sqrt{N}\right)\exp\left(-j2\pi pq/N\right)$. The time-frequency domain samples $x^{\text{TF}}[m,n]$ are pulse-shaped using a transmit pulse $g_{\text{tx}}(t)$ to generate a continuous-time signal $s(t)$ by using the Heisenberg transform [4]. The signal $s(t)$, sampled at a rate $f_s = M\Delta f = M/T$, can be expressed in the matrix form as [18] $\mathbf{S} = \mathbf{G}_{\text{tx}} \mathbf{F}_M^H \mathbf{X}^{\text{TF}} = \mathbf{G}_{\text{tx}} \mathbf{X} \mathbf{F}_N^H$. The matrix $\mathbf{S} \in \mathbb{C}^{M\times N}$ consists of MN samples of the signal $s(t)$. We consider, similar to [4], [18], symbol spaced sampling (SSS) approach wherein the sampling interval $T_s = 1/M\Delta f = T/M$ [19]. This lead to M-sample-length transmit and receive pulses. The diagonal matrix $\mathbf{G}_{\text{tx}} \in \mathbb{C}^{M\times M}$ is obtained by uniformly sampling the transmit pulse $g_{\text{tx}}(t)$ at the time instants $\frac{mT}{M}$ with $m = 0, 1, \ldots, M-1$. The time-domain transmit vector $\mathbf{s} = \text{vec}(\mathbf{S}) \in \mathbb{C}^{MN\times 1}$. By substituting $\mathbf{X}$ from (1), $\mathbf{s}$ can be expressed using the identity vec$(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)vec(\mathbf{B})$ [20] and the property $\mathbf{F}_N^T = \mathbf{F}_N$ of the DFT matrix $\mathbf{F}_N$ as

$$\mathbf{s} = \left(\mathbf{F}_N^H \otimes \mathbf{G}_{\text{tx}}\right)\mathbf{x}. \tag{3}$$

Here $\mathbf{x} = \text{vec}(\mathbf{X}) = \mathbf{x}_p + \mathbf{x}_d$ with $\mathbf{x}_d \in \mathbb{C}^{NM\times 1}$ and $\mathbf{x}_p \in \mathbb{C}^{NM\times 1}$ being the data and pilot vectors in the delay-Doppler domain, respectively. To mitigate inter-frame interference in the time domain, as shown in Fig. 2, a cyclic prefix (CP) of length l_{max} samples is appended to the transmit signal vector $\mathbf{s}$, where l_{max} is the tap corresponding to the maximum delay τ_{max}.

B. Channel Impulse Response

A delay-Doppler domain channel coefficient is characterized by a cluster, which ideally consists of infinite many reflectors. We assume that there are Q such clusters in the channel, where the ith cluster has a dominant component h_i with associated delay τ_i and Doppler shift ν_i [2]. Typically, there are small number of clusters in a sub-6 GHz delay-Doppler domain channel, which makes it sparse [2], [4]. The impulse response of the wireless channel $h(\tau,\nu)$ in the delay-Doppler domain is [4]: $h(\tau,\nu) = \sum_{i=1}^{Q} h_i \delta(\tau - \tau_i)\,\delta(\nu - \nu_i)$. Here $\tau \in [0, \tau_{max}]$ and $\nu \in [-\nu_{max}, \nu_{max}]$ are the delay and

Doppler shifts respectively, with τ_{max} and ν_{max} being the maximum delay and the maximum Doppler shift among all channel paths. The quantity Q is the number of propagation paths, h_i is the complex path gain for the ith path, which is distributed as $\mathcal{CN}(0, \sigma_{h_i}^2)$ [8]. The delay τ_i and the Doppler ν_i, associated with the ith path, are expressed as $\tau_i = \frac{l_i}{M\Delta f}$ and $\nu_i = \frac{k_i}{NT}$, respectively. Here the integers l_i and k_i respectively denote the delay and Doppler taps for the ith path. Since h_i is not a function of (l_i, k_i) [18], its probability distribution is independent of the parameters (l_i, k_i). We, similar to [4], assume l_i to be an integer. This is because in a typical wideband system, the sampling time resolution $\frac{1}{M\Delta f}$ is sufficient to approximate the path delays to the nearest sampling point [4]. For the sampling rate $f_s = M/T$, the maximum channel delay is assumed to be $\tau_{max} = (l_{max})\, T/M$, which implies that $l_i \in [0, l_{max}]$. We also, similar to [2], do not consider the effect of fractional Doppler. This is because the Doppler resolution $1/NT$ progressively reduces with increasing number of Doppler bins N [2]. However, the proposed designs are also valid for fractional Doppler scenario, as shown later in the simulation section.

C. Receive Processing

The discrete baseband received signal vector $\mathbf{r} \in \mathbb{C}^{MN\times 1}$, after removing CP, is [18] $\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{w}$. Let the matrix

$$\mathbf{\Pi} \triangleq \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \tag{4}$$

be the $MN \times MN$ forward cyclic-shift (permutation) matrix, and $\mathbf{\Delta} = \text{diag}\left\{z^0, z^1, \ldots, z^{MN-1}\right\}$, is a diagonal matrix with $z = \exp(2\pi j/MN)$, where $j = 0, 1, \ldots, MN-1$. The matrices $\mathbf{\Pi}$ and $\mathbf{\Delta}$ model the delays and Doppler shifts, respectively. The channel $\mathbf{H} \in \mathbb{C}^{MN\times MN}$ is *sparse*, and is now defined as $\mathbf{H} = \sum_{i=1}^{Q} h_i \mathbf{\Pi}^{l_i} \mathbf{\Delta}^{k_i}$ [18]. The noise vector $\mathbf{w} \in \mathbb{C}^{MN\times 1}$ has i.i.d. complex Gaussian entries with zero mean and variance σ_w^2. The time-frequency domain receive signal matrix $\mathbf{Y}^{\text{TF}} \in \mathbb{C}^{M\times N}$, as shown in Fig. 2, is derived from the received signal $\mathbf{r}$ using the Wigner transformation (inverse of Heisenberg transform) as $\mathbf{Y}^{\text{TF}} = \mathbf{F}_M \mathbf{G}_{\text{rx}} \mathbf{R}$ [18]. Here $\mathbf{R} = \text{vec}^{-1}(\mathbf{r}) \in \mathbb{C}^{M\times N}$, and the diagonal matrix $\mathbf{G}_{\text{rx}} \in \mathbb{C}^{M\times M}$ is obtained by sampling the receive pulse $g_{\text{rx}}(t)$ at the time instants $\frac{mT}{M}$ with $m = 0, 1, \ldots, M-1$. The delay-Doppler receive signal matrix $\mathbf{Y} \in \mathbb{C}^{M\times N}$ is obtained from $\mathbf{Y}^{\text{TF}}$ using the SFFT operation as [18]

$$\mathbf{Y} = \mathbf{F}_M^H \mathbf{Y}^{\text{TF}} \mathbf{F}_N = \mathbf{G}_{\text{rx}} \mathbf{R} \mathbf{F}_N. \tag{5}$$

The receive vector $\mathbf{y} = \text{vec}(\mathbf{Y})$, using the identity $\text{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)\text{vec}(\mathbf{B})$ [20], is

$$\mathbf{y} = (\mathbf{F}_N \otimes \mathbf{G}_{\text{rx}})\, \mathbf{r} = (\mathbf{F}_N \otimes \mathbf{G}_{\text{rx}})\, (\mathbf{H}\mathbf{s} + \mathbf{w}). \tag{6}$$

By substituting (3) in (6), we get

$$\mathbf{y} = (\mathbf{F}_N \otimes \mathbf{G}_{\text{rx}})\, \mathbf{H}\, \left(\mathbf{F}_N^H \otimes \mathbf{G}_{\text{tx}}\right) \mathbf{x} + \tilde{\mathbf{w}} = \mathbf{H}_{\text{eff}}\, (\mathbf{x}_p + \mathbf{x}_d) + \tilde{\mathbf{w}}, \tag{7}$$

where the vector $\tilde{\mathbf{w}} = (\mathbf{F}_N \otimes \mathbf{G}_{\text{tx}})\, \mathbf{w}$, and the effective channel matrix $\mathbf{H}_{\text{eff}} \in \mathbb{C}^{MN\times MN}$ is

$$\mathbf{H}_{\text{eff}} = (\mathbf{F}_N \otimes \mathbf{G}_{\text{rx}})\, \mathbf{H}\, \left(\mathbf{F}_N^H \otimes \mathbf{G}_{\text{tx}}\right). \tag{8}$$

We assume rectangular transmit and receive pulse shapes of duration T at both transmitter and receiver, which implies that $\mathbf{G}_{\text{rx}} = \mathbf{G}_{\text{tx}} = \mathbf{I}_M$. The system model in (7) adds one CP per OTFS frame. This suppresses inter-frame-interference but not inter-symbol-interference (ISI) and the inter-carrier-interference (ICI) between intra-frame OTFS symbols. This causes phase shifts in the delay-Doppler domain [4], which can be included in the effective delay-Doppler channel matrix $\mathbf{H}_{\text{eff}}$ as [18]:

$$\begin{aligned} \mathbf{H}_{\text{eff}} &= (\mathbf{F}_N \otimes \mathbf{I}_M) \left(\textstyle\sum_{i=1}^{Q} h_i \mathbf{\Pi}^{l_i} \mathbf{\Delta}^{k_i}\right) \left(\mathbf{F}_N^H \otimes \mathbf{I}_M\right) \\ &= \mathbf{B}_{\text{rx}} \left(\textstyle\sum_{i=1}^{Q} h_i \mathbf{\Theta}_i\right) \mathbf{B}_{\text{tx}}. \end{aligned} \tag{9}$$

The matrices $\mathbf{B}_{\text{tx}} = \left(\mathbf{F}_N^H \otimes \mathbf{I}_M\right)$, $\mathbf{B}_{\text{rx}} = (\mathbf{F}_N \otimes \mathbf{I}_M)$ and $\mathbf{\Theta}_i = \left(\mathbf{\Pi}^{l_i} \mathbf{\Delta}^{k_i}\right)$ have size $MN \times MN$.

Remark 1: The channel matrix for an OFDM system over highly time-varying channel is $\mathbf{H}_{\text{ofdm}} = \mathbf{F}_M \mathbf{H}_{\text{TF}} \mathbf{F}_M^H$ [4], where $\mathbf{F}_M \in \mathbb{C}^{M\times M}$ is the FFT matrix and the (m, n)th element of the time-frequency domain channel matrix $\mathbf{H}_{\text{TF}} \in \mathbb{C}^{M\times M}$ is given as

$$\mathbf{H}_{\text{TF}}(m, n) = \sum_{i=1}^{Q} h_i \delta \left[\left(m - n - \frac{\tau_i M}{T}\right)_M\right] e^{\frac{j2\pi(n-1)\nu_i}{M}}, \tag{10}$$

where $1 \le m, n \le M$. We observe from (9) and (10) that the:

- OFDM channel matrix $\mathbf{H}_{\text{TF}}$ in (10) is in the time-frequency domain, and is not sparse. The OTFS channel matrix $\mathbf{H}_{\text{eff}}$ in (9) is in the delay-Doppler domain, and is sparse. *The channel models are thus radically different for the two systems, which completely changes signal processing for the SP-aided OTFS systems.*
- size of OFDM and OTFS channel matrices are $M \times M$ and $MN \times MN$, respectively [4]. The large dimensionality of OTFS channel matrix radically increases its receiver complexity.

The received signal vector $\mathbf{y}$, by substituting $\mathbf{H}_{\text{eff}}$ from (9) in (7), can be recast as follows

$$\mathbf{y} = \left(\sum_{i=1}^{Q} h_i \mathbf{\Gamma}_i\right) (\mathbf{x}_p + \mathbf{x}_d) + \tilde{\mathbf{w}}. \tag{11}$$

The matrix $\mathbf{\Gamma}_i = \mathbf{B}_{\text{rx}} \mathbf{\Theta}_i \mathbf{B}_{\text{tx}} \in \mathbb{C}^{MN\times MN}$. The received vector in (11) for the SP-based OTFS system can be succinctly written as

$$\mathbf{y} = \mathbf{\Omega}_p \mathbf{h} + \mathbf{\Omega}_d \mathbf{h} + \tilde{\mathbf{w}}. \tag{12}$$

The concatenated matrices $\mathbf{\Omega}_p \in \mathbb{C}^{MN\times Q}$ and $\mathbf{\Omega}_d \in \mathbb{C}^{MN\times Q}$ corresponding to the pilot vector $\mathbf{x}_p$ and data vector $\mathbf{x}_d$ are obtained as follows

$$\mathbf{\Omega}_p = [\mathbf{\Gamma}_1 \mathbf{x}_p \;\; \mathbf{\Gamma}_2 \mathbf{x}_p \;\; \cdots \;\; \mathbf{\Gamma}_Q \mathbf{x}_p] \tag{13}$$

$$\mathbf{\Omega}_d = [\mathbf{\Gamma}_1 \mathbf{x}_d \;\; \mathbf{\Gamma}_2 \mathbf{x}_d \;\; \cdots \;\; \mathbf{\Gamma}_Q \mathbf{x}_d]. \tag{14}$$

The matrix $\mathbf{\Gamma}_i$ in the $\mathbf{\Omega}_d$ and $\mathbf{\Omega}_p$, as observed from Appendix A and Appendix D, significantly complicates their mathematical characterization. The delay-Doppler domain channel vector $\mathbf{h} = [h_1, h_2, \cdots, h_Q]^T \in \mathbb{C}^{Q\times 1}$ has zero mean and covariance matrix

$$\mathbf{C}_h = \mathbb{E}\left\{\mathbf{h}\mathbf{h}^H\right\} = \text{diag}\left\{\sigma_{h_1}^2, \cdots, \sigma_{h_Q}^2\right\}. \tag{15}$$

We assume that the covariance matrix is known at the receiver [21]. the proposed designs estimate channel in the delay-Doppler domain whose statistical characteristics, and consequently the covariance matrix, remains constant over multiple frames [22]. By exploiting this property, the existing techniques in the literature can be employed to estimate the covariance matrix. For example, reference [23] proposed advanced algorithms to estimate covariance matrix of channel using very few sample observations. This work estimated its eigenvalues and eigenvectors using eigen-projection matrices, which are shown to be consistent in finite sample size situations. We next state the following result related to random matrix $\mathbf{\Omega}_d$, which is proved in Appendix-A.

Lemma 1: The random matrix $\mathbf{\Omega}_d$ has $\mathbb{E}\{\mathbf{\Omega}_d\} = \mathbf{0}_{MN\times Q}$ and $\mathbb{E}\left\{\mathbf{\Omega}_d\mathbf{\Omega}_d^H\right\} = \sigma_d^2 Q\mathbf{I}_{MN}$.

The noise vector $\tilde{\mathbf{w}} = (\mathbf{F}_N \otimes \mathbf{I}_M)\,\mathbf{w}$ of size $MN \times 1$ has

$$\mathbb{E}\{\tilde{\mathbf{w}}\} = \mathbf{0}_{MN\times 1} \quad \text{and} \quad \mathbf{C}_{\tilde{\mathbf{w}}} = \mathbb{E}\left\{\tilde{\mathbf{w}}\tilde{\mathbf{w}}^H\right\} = \sigma_w^2\mathbf{I}_{MN}. \tag{16}$$

Remark 2: The input-output relations in (7) and (12) are same, with different mathematical formulations. They will, therefore, be used interchangeably for developing the proposed designs.

Remark 3: Regarding interference in pilot and data: The superimposed-pilot scheme will cause a series of problems in decoupling the interference cancellation and data demodulation. The proposed design handles this by i) iterating between channel estimation and data detection; and ii) optimally allocating power between data and pilot symbols.

III. Superimposed Pilot -Based OTFS Channel Estimator and Data Detector

This section proposes the SP-NI and SP-I designs for estimating channel and detecting data in OTFS systems. We begin with the SP-NI design which first calculates the MMSE channel estimate in the delay-Doppler domain by treating data as interference. It later detects data using the reduced-complexity message passing algorithm [4], which exploits the delay-Doppler domain channel sparsity. This algorithm, with a significantly lower computational complexity, yields a BER which is very close to the maximum-a-posterior probability (MAP) detector [4].

A. Proposed SP-NI Design

The proposed SP-NI design exploits the fact that very few parameters are needed to model delay-Doppler channel, and estimates it in the same domain by considering data as an interference. Furthermore, to incorporate message passing algorithm, which exploits the OTFS channel sparsity for data detection, the proposed design evaluates both first- and second-order statistical characteristics of the SP-aided channel estimation error. The statistics are then used to calculate the probability mass functions (pmfs) for the message passing algorithm.

1) Channel Estimation in SP-NI Design: The received signal vector in (12) is

$$\mathbf{y} = \mathbf{\Omega}_p\mathbf{h} + \underbrace{\mathbf{\Omega}_d\mathbf{h} + \tilde{\mathbf{w}}}_{\tilde{\mathbf{w}}_d} = \mathbf{\Omega}_p\mathbf{h} + \tilde{\mathbf{w}}_d, \tag{17}$$

where the mean $\boldsymbol{\mu}_{\tilde{\mathbf{w}}_d}$ and covariance matrix $\mathbf{C}_{\tilde{\mathbf{w}}_d}$ of the noise-plus-interference vector $\tilde{\mathbf{w}}_d$ are derived in the following lemma, whose proof is relegated to Appendix-B.

Lemma 2: Since $\mathbb{E}\{\mathbf{\Omega}_d\} = \mathbf{0}_{MN\times Q}$ and $\mathbb{E}\{\mathbf{h}\} = \mathbf{0}_{Q\times 1}$, the mean of the vector $\tilde{\mathbf{w}}_d$ is

$$\boldsymbol{\mu}_{\tilde{\mathbf{w}}_d} = \mathbb{E}\{\tilde{\mathbf{w}}_d\} = \mathbf{0}_{MN\times 1}, \tag{18}$$

and its covariance matrix using *Lemma* 1 is

$$\mathbf{C}_{\tilde{\mathbf{w}}_d} = \mathbb{E}\left\{\tilde{\mathbf{w}}_d\tilde{\mathbf{w}}_d^H\right\} = \left(\left(\sum_{i=1}^{Q}\sigma_{h_i}^2\right)\sigma_d^2 + \sigma_w^2\right)\mathbf{I}_{MN}. \tag{19}$$

The proof also uses the following lemma from [24].

Lemma 3: If $\mathbf{A} \in \mathbb{C}^{m\times n}$ is a random matrix such that $\mathbb{E}\left\{\mathbf{A}\mathbf{A}^H\right\} = \sigma^2\mathbf{I}_m$, then for any hermitian matrix $\mathbf{B} \in \mathbb{C}^{n\times n}$, $\mathbb{E}\left\{\mathbf{A}\mathbf{B}\mathbf{A}^H\right\} = \dfrac{\text{Tr}\,(\mathbf{B})}{n}\mathbb{E}\left\{\mathbf{A}\mathbf{A}^H\right\}$.

The MMSE estimate $\hat{\mathbf{h}}_{\text{NI}}$ of the channel vector $\mathbf{h}$ using the SP-NI design and the linear observation model in (17), is given as [25]

$$\hat{\mathbf{h}}_{\text{NI}} = \left(\mathbf{\Omega}_p^H\mathbf{C}_{\tilde{\mathbf{w}}_d}^{-1}\mathbf{\Omega}_p + \mathbf{C}_{\mathbf{h}}^{-1}\right)^{-1}\mathbf{\Omega}_p^H\mathbf{C}_{\tilde{\mathbf{w}}_d}^{-1}\mathbf{y}. \tag{20}$$

The covariance matrices $\mathbf{C}_{\mathbf{h}}$ and $\mathbf{C}_{\tilde{\mathbf{w}}_d}$ are given in (15) and (19), respectively. *We observe that the channel estimator in* (20) *exploits OTFS channel sparsity by computing the inverse of a $Q \times Q$ matrix, where $Q \ll MN$ denotes the number delay-Doppler taps.* The error covariance matrix of $\hat{\mathbf{h}}_{\text{NI}}$ can next be calculated as [26]

$$\mathbf{\Sigma}_{\text{NI}} = \left(\mathbf{\Omega}_p^H\mathbf{C}_{\tilde{\mathbf{w}}_d}^{-1}\mathbf{\Omega}_p + \mathbf{C}_{\mathbf{h}}^{-1}\right)^{-1}. \tag{21}$$

We see that, unlike pilot-aided channel estimation methods in [7], [8], the covariance matrix $\mathbf{C}_{\tilde{\mathbf{w}}_d}$ in the proposed design, due to superimposed pilots, includes the effect of mutual interference between data and pilot symbols. The MSE of channel estimate is

$$B_{h,\text{NI}} = \mathbb{E}\{\|\mathbf{h} - \hat{\mathbf{h}}_{\text{NI}}\|^2\} = \text{Tr}\,[\mathbf{\Sigma}_{\text{NI}}]. \tag{22}$$

After SP-based channel estimation, data vector $\mathbf{x}_d$ is detected using message passing algorithm [4].

2) Message Passing-Aided Data Detection in SP-NI Design: The low-complexity message passing algorithm exploits sparsity of the delay-Doppler OTFS channel while detecting data [4], [27]. The signal used for detecting data is obtained by subtracting the pilot signal from the received signal in (7) as follows

$$\mathbf{y}_d = \mathbf{y} - \widehat{\mathbf{H}}_{\text{eff-NI}}\mathbf{x}_p = \mathbf{H}_{\text{eff}}\mathbf{x}_d + \tilde{\mathbf{w}}_e. \tag{23}$$

The vector $\tilde{\mathbf{w}}_e = \mathbf{e}_p + \tilde{\mathbf{w}}$ consists of channel estimation error and noise, with the error vector $\mathbf{e}_p = \left(\mathbf{H}_{\text{eff}} - \widehat{\mathbf{H}}_{\text{eff-NI}}\right)\mathbf{x}_p$.

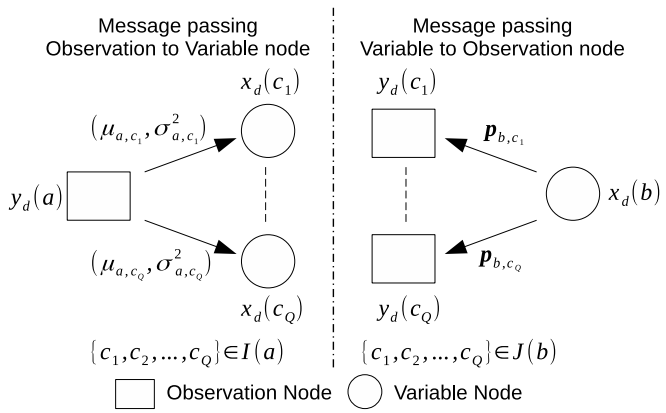


Fig. 3. Graphical representation of the message passing algorithm for data detection in the SP-NI design.

The estimate of effective channel matrix $\mathbf{H}_{\text{eff}}$, defined in (9), is obtained by substituting the expression of channel estimate $\hat{\mathbf{h}}_{\text{NI}}$ in (9) as

$$\widehat{\mathbf{H}}_{\text{eff-NI}} = \mathbf{B}_{\text{rx}} \left(\sum_{i=1}^{Q} \hat{h}_{\text{NI},i} \boldsymbol{\Theta}_i \right) \mathbf{B}_{\text{tx}}. \tag{24}$$

We now use the system model in (23) to develop message-passing receiver. To begin with, let $\mathcal{I}(a)$ and $\mathcal{J}(b)$ be the sets containing indices of non-zero elements in the ath row and the bth column of the matrix $\widehat{\mathbf{H}}_{\text{eff-NI}}$, respectively. This implies that $|\mathcal{I}(a)| = |\mathcal{J}(b)| = Q$. Figure 3 shows the factor-graph for the message-passing detection algorithm which consists of observation and variable nodes [28], which are defined next. Before doing that, it is worth noting that due to the OTFS channel sparsity, an observation node $y_d(a)$, for $1 \leq a \leq MN$, is connected to only Q number of variable nodes $x_d(c_1), \ldots, x_d(c_Q)$, where $Q \ll MN$ and $\{c_1, \ldots, c_Q\} \in \mathcal{I}(a)$. Similarly, a variable $x_d(b)$, for $1 \leq b \leq MN$, is connected to Q number of observation nodes $y_d(c_1), \ldots, y_d(c_Q)$, where $\{c_1, \ldots, c_Q\} \in \mathcal{J}(b)$.

We now use (23) to write the input-output relation for the ath observation node $\mathbf{y}_d(a)$ with the bth variable nodes $\mathbf{x}_d(b)$, for $b \in \mathcal{I}(a)$ as $\mathbf{y}_d(a) = \mathbf{x}_d(b) \mathbf{H}_{\text{eff}}(a,b) + \gamma_{a,b}$, where $1 \leq a \leq MN$ and the interference-plus-noise term

$$\gamma_{a,b} = \sum_{c \in \mathcal{I}(a), c \neq b} \mathbf{x}_d(c) \mathbf{H}_{\text{eff}}(a,c) + \tilde{\mathbf{w}}_e(a). \tag{25}$$

The first part of the expression, by using central-limit theorem [4], can be approximated as a Gaussian random variable. Since $\tilde{\mathbf{w}}_e$, defined below (23), is also Gaussian distributed, the quantity $\gamma_{a,b}$ can also be approximated as a Gaussian random variable. The message passing algorithm, in the ith iteration, passes messages in terms of mean $\mu^{(i)}_{a,b}$ and variance $\left(\sigma^{(i)}_{a,b}\right)^2$ of the parameter $\gamma_{a,b}$, from the observation node $y_d(a)$ to the variable node $x_d(b)$, where $b \in \mathcal{I}(a)$.

Remark 4: We crucially note that, in contrast to [8], [29], the interference-plus-noise term $\gamma_{a,b}$ is a function of $\tilde{\mathbf{w}}_e$ which, due to imperfect channel estimate and as shown in (23), is a function of the noise $\tilde{\mathbf{w}}$ and the interference $\mathbf{e}_p$. The superimposed pilots are only partially canceled while detecting data. This is unlike the message passing i) algorithm in [4], [29], which assumes perfect channel; ii) receiver in [8] which, due to insertion of zeros, need not consider the mutual interference between data and pilot symbols. The mutual interference in the proposed SP-aided designs significantly changes the message calculations. We have to, in contrast, derive the first- and second-order statistical characteristics of the vector $\mathbf{e}_p$ for evaluating the mean $\mu^{(i)}_{a,b}$ and the variance $\left(\sigma^{(i)}_{a,b}\right)^2$ of $\gamma_{a,b}$. We see from (24) that the estimated channel matrix $\widehat{\mathbf{H}}_{\text{eff-NI}}$ in the $\mathbf{e}_p$ expression is a function of the estimate $\hat{h}_{\text{NI},i}$ and matrices $\mathbf{B}_{\text{tx}}$, $\mathbf{B}_{\text{rx}}$, $\boldsymbol{\Theta}_i$. The estimate $\hat{h}_{\text{NI},i}$, as shown in (20), is a function of pilot matrix $\boldsymbol{\Omega}_p$, covariance matrix $\mathbf{C}_{\tilde{\mathbf{w}}_d}$ of data interference and the observation vector $\mathbf{y}$. The statistical characterization of $\mathbf{e}_p$, as shown next, is thus non-trivial.

Since $\mathbb{E}\{h_i\} = 0$, it follows from (9) that $\mathbb{E}\{\mathbf{H}_{\text{eff}}\} = \mathbf{0}_{MN \times MN}$. We, consequently, have from (7), $\mathbb{E}\{\mathbf{y}\} = \mathbf{0}_{MN \times 1}$, which also implies $\mathbb{E}\{\hat{\mathbf{h}}_{\text{NI}}\} = \mathbf{0}_{Q \times 1}$. We finally have, by using the property $\mathbb{E}\{\hat{\mathbf{h}}_{\text{NI}}\} = \mathbf{0}_{Q \times 1}$, that $\mathbb{E}\{\widehat{\mathbf{H}}_{\text{eff-NI}}\} = \mathbf{0}_{MN \times MN}$. Exploiting the above results, we have

$$\mathbb{E}\{\mathbf{e}_p\} = \mathbf{0}_{MN \times 1}. \tag{26}$$

It can be readily verified that the $(l + Mk)$th element of the vector $\mathbf{e}_p$ is

$$\mathbf{e}_p(l + Mk) = \sum_{i=1}^{Q} \left(h_i - \hat{h}_{\text{NI},i} \right) \alpha_i(l + Mk) \times x_p\left[(l - l_i)_M, (k - k_i)_N \right]. \tag{27}$$

The factor $\alpha_i(l + Mk)$ models the time-frequency domain ISI and IC) [18], and is given as

$$\alpha_i(l + Mk) = \begin{cases} e^{-j2\pi \frac{k}{N} z^{k_i}\left[(l - l_i)_M\right]}, & \text{if } \; l < l_i \\ z^{k_i}\left[(l - l_i)_M\right], & \text{if } \; l \geq l_i \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

This expression in (27), along with the fact that all pilot symbols have an equal power of σ_p^2, is used to compute the variance of the $(l + Mk)$th element of the vector $\mathbf{e}_p$ as

$$\begin{aligned} \text{var}\left[\mathbf{e}_p(l + Mk)\right] &= \sum_{i=1}^{Q} \mathbb{E}\left\{ \left| h_i - \hat{h}_{\text{NI},i} \right|^2 \right\} \sigma_p^2 \\ &= \sigma_p^2 \mathbb{E}\left\{ \left\| \mathbf{h} - \hat{\mathbf{h}}_{\text{NI}} \right\|^2 \right\} \overset{(a)}{=} \sigma_p^2 B_{h,\,\text{NI}}. \end{aligned} \tag{29}$$

Equality in (a) is due to (22). Since the vectors $\mathbf{e}_p$ and $\tilde{\mathbf{w}}$ are statistically independent, the mean and the variance of ith element of $\tilde{\mathbf{w}}_e$ using (16), (26) and (29), can be computed as

$$\mathbb{E}\{\tilde{\mathbf{w}}_e(i)\} = \mathbb{E}\{\mathbf{e}_p(i)\} + \mathbb{E}\{\tilde{\mathbf{w}}(i)\} = 0 \tag{30}$$

$$\text{var}\left[\tilde{\mathbf{w}}_e(i)\right] = \text{var}\left[\mathbf{e}_p(i)\right] + \text{var}\left[\tilde{\mathbf{w}}(i)\right] = \sigma_p^2 B_{h,\text{NI}} + \sigma_w^2. \tag{31}$$

Using (30), and the channel estimate $\widehat{\mathbf{H}}_{\text{eff-NI}}$, the mean $\mu^{(i)}_{a,b}$ of the interference-plus-noise term $\gamma_{a,b}$ in (25)

can be evaluated as follows

$$\mu_{a,b}^{(i)} = \sum_{c\in\mathcal{I}(a),c\neq b}\sum_{j=1}^{S} p_{c,a}^{(i-1)}(\alpha_j)\,\alpha_j \widehat{\mathbf{H}}_{\text{eff-NI}}(a,c) + \mathbb{E}\{\tilde{\mathbf{w}}_e(a)\}. \quad (32)$$

The quantity $p_{c,a}(\alpha_j)$, which is calculated later in this section, denotes the pmf of the jth constellation symbol α_j between the observation node a and the variable node c. The parameter S denotes the constellation size. The variance of noise-plus-interference term $\gamma_{a,b}$ for the ith iteration is next computed as follows

$$\left(\sigma_{a,b}^{(i)}\right)^2 = \sum_{c\in\mathcal{I}(a),c\neq b}\sum_{j=1}^{S} p_{c,a}^{(i-1)}(\alpha_j)\,|\alpha_j|^2\left|\widehat{\mathbf{H}}_{\text{eff-NI}}(a,c)\right|^2 - \left|\mu_{a,b}^{(i)}\right|^2 + \text{var}\left[\tilde{\mathbf{w}}_e(a)\right]. \quad (33)$$

The variance $\text{var}[\tilde{\mathbf{w}}_e(a)] = \sigma_p^2 B_{h,\text{NI}} + \sigma_w^2$ is derived in (31). We see that, unlike [8], [4, Eq. (31)], $\left(\sigma_{a,b}^{(i)}\right)^2$ is also a function of the MSE of the SP-NI channel estimator and the pilot power σ_p^2.

We now calculate the elements of the pmf vector $\mathbf{p}_{b,a} \in \mathbb{R}^{S\times 1}$, which is passed from the variable nodes $\mathbf{x}_d(b)$ to the observation node $\mathbf{y}_d(a)$, where $a \in \mathcal{J}(b)$. The messages in terms of pmfs can be computed at each iteration using various methods e.g., belief propagation (BP) based algorithm [28], double loop methods [30] and damping method [31]. As explained in [28], the damping method has better convergence behavior. We, therefore, use the damping method from [31] to recursively update the pmf $p_{b,a}^{(i)}$ in the ith iteration as follows [4]

$$p_{b,a}^{(i)}(\alpha_j) = \Delta\tilde{p}_{b,a}^{(i)}(\alpha_j) + (1-\Delta)\,p_{b,a}^{(i-1)}(\alpha_j). \quad (34)$$

Here $\Delta \in (0,1]$ is the damping factor, which controls the convergence of the message passing algorithm, and $1 \le j \le S$. The probability $\tilde{p}_{b,a}^{(i)}(\alpha_j)$ is computed as [4]

$$\tilde{p}_{b,a}^{(i)}(\alpha_j) \approx \prod_{c\in\mathcal{J}(b),c\neq a}\frac{\beta_{c,b,j}^{(i)}}{\sum_{s=1}^{S}\beta_{c,b,s}^{(i)}}, \quad (35)$$

where $\beta_{c,b,s}^{(i)} = \exp\left(\frac{-\left|\mathbf{y}_d(c)-\mu_{c,b}^{(i)}-\widehat{\mathbf{H}}_{\text{eff-NI}}(c,b)\alpha_s\right|}{\left(\sigma_{c,b}^{(i)}\right)^2}\right)$. To terminate the message passing algorithm, the convergence indicator $\zeta^{(i)}$ is given as [4]

$$\zeta^{(i)} = \frac{1}{MN}\sum_{b=1}^{MN}\mathbb{I}\left(\max_{a_j\in\mathbb{A}} p_b^{(i)}(\alpha_j) \ge 1-\epsilon\right), \quad (36)$$

where $\mathbb{I}(\cdot)$ is the indicator function, $\epsilon > 0$ and $p_b^{(i)}(\alpha_j) = \prod_{c\in\mathcal{J}(b)}\frac{\beta_{c,b,j}^{(i)}}{\sum_{s=1}^{S}\beta_{c,b,s}^{(i)}}$. If the convergence indicator $\zeta^{(i)} > \zeta^{(i-1)}$, decision on the delay-Doppler data symbol in the SP-NI design for the ith iteration of the message passing algorithm is updated as

$$\hat{\mathbf{x}}_{d,\text{NI}}(b) = \arg\max_{1\le j\le S} p_b^{(i)}(\alpha_j). \quad (37)$$

Algorithm 1: Message Passing Data Detection Algorithm for the Proposed SP-NI Design

Input: $\mathbf{y}_d$, $\widehat{\mathbf{H}}_{\text{eff-NI}}$
Output: $\hat{\mathbf{x}}_{d,\text{NI}}$
1 **Initialization:** $\mathbf{p}_{b,a}^{(0)} = 1/S$, for $1 \le b \le MN$ and $a \in \mathcal{J}(b)$, Counter $i = 1$.
2 **repeat**
3 Pass the mean $\mu_{a,b}^{(i)}$ in (32) and variance $(\sigma_{a,b}^{(i)})^2$ in (33) from observation nodes $\mathbf{y}_d$ to variable nodes $\mathbf{x}_d(c)$ for $c \in \mathcal{I}(a)$
4 At variable nodes $\mathbf{x}_d(b)$, update the pmf using (34), and pass it to the observation nodes $\mathbf{y}_d(a)$ for $a \in \mathcal{J}(b)$
5 Update the decision on the transmitted symbols $\mathbf{x}_d$, using (37), and obtain $\hat{\mathbf{x}}_{d,\text{NI}}$
6 $i = i + 1$
7 **until** *Stopping criteria*
8 **return**: $\hat{\mathbf{x}}_{d,\text{NI}}$

Algorithm 1 summarizes the message-passing-aided detection for the proposed SP-NI design. Step 3 passes mean $\mu_{a,b}^{(i)}$ and variance $(\sigma_{a,b}^{(i)})^2$ from observation to variable nodes. Step 4 updates the pmfs and passes them from variable to observation nodes. Step 5 detects the data.

B. Proposed Superimposed Pilot-Iterative (SP-I) Design

The SP-NI design, discussed earlier, estimates channel by treating data as interference, which has shown later in Section-VI, degrades its performance at high SNR values. The SP-I design, proposed next, handles this problem by iterating between channel estimation and message passing-aided data detection. This design takes an initial data estimate $\hat{\mathbf{x}}_d^{(0)} = \hat{\mathbf{x}}_{d,\text{NI}}$ from the SP-NI design, and uses it along with the superimposed pilot vector $\mathbf{x}_p$ to perform data-aided channel estimation.

1) Channel Estimation in SP-I Design: Let $\widehat{\mathbf{\Omega}}_d^{(0)}$ be the initial estimate of the data matrix $\mathbf{\Omega}_d$. It is obtained from SP-NI design by substituting the data estimate $\mathbf{x}_d = \hat{\mathbf{x}}_d^{(0)} = \hat{\mathbf{x}}_{d,\text{NI}}$ in (14) as

$$\widehat{\mathbf{\Omega}}_d^{(0)} = \left[\mathbf{\Gamma}_1\hat{\mathbf{x}}_d^{(0)}\ \ \mathbf{\Gamma}_2\hat{\mathbf{x}}_d^{(0)}\ \ \cdots\ \ \mathbf{\Gamma}_Q\hat{\mathbf{x}}_d^{(0)}\right]. \quad (38)$$

The received signal vector in (12) can now be re-expressed as follows

$$\mathbf{y} = \left(\mathbf{\Omega}_p + \widehat{\mathbf{\Omega}}_d^{(0)}\right)\mathbf{h} + \left(\mathbf{\Omega}_d - \widehat{\mathbf{\Omega}}_d^{(0)}\right)\mathbf{h} + \tilde{\mathbf{w}} = \mathbf{\Omega}_{\mathbf{x}_p\hat{\mathbf{x}}_d}^{(0)}\mathbf{h} + \boldsymbol{\xi}_{\tilde{\mathbf{w}}}^{(0)}. \quad (39)$$

The data-aided pilot matrix $\mathbf{\Omega}_{\mathbf{x}_p\hat{\mathbf{x}}_d}^{(0)}$ and the noise-plus-interference vector $\boldsymbol{\xi}_{\tilde{\mathbf{w}}}^{(0)}$ are defined as

$$\mathbf{\Omega}_{\mathbf{x}_p\hat{\mathbf{x}}_d}^{(0)} \triangleq \mathbf{\Omega}_p + \widehat{\mathbf{\Omega}}_d^{(0)} \ \text{ and } \ \boldsymbol{\xi}_{\tilde{\mathbf{w}}}^{(0)} \triangleq \mathbf{\Xi}_{\mathbf{x}_d}^{(0)}\mathbf{h} + \tilde{\mathbf{w}}, \quad (40)$$

with the estimation error matrix $\mathbf{\Xi}_{\mathbf{x}_d}^{(0)} \triangleq \mathbf{\Omega}_d - \widehat{\mathbf{\Omega}}_d^{(0)}$. Using the above initial data estimate, the SP-I design performs

data-aided estimation of the channel vector $\mathbf{h}$ in (39) in the nth iteration as

$$\hat{\mathbf{h}}^{(n)} = \left(\left(\mathbf{\Omega}^{(n-1)}_{\mathbf{x}_p\hat{\mathbf{x}}_d}\right)^H \left(\mathbf{C}^{(n-1)}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}}\right)^{-1} \mathbf{\Omega}^{(n-1)}_{\mathbf{x}_p\hat{\mathbf{x}}_d} + \mathbf{C}^{-1}_{\mathbf{h}} \right)^{-1} \times \left(\mathbf{\Omega}^{(n-1)}_{\mathbf{x}_p\hat{\mathbf{x}}_d}\right)^H \left(\mathbf{C}^{(n-1)}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}}\right)^{-1} \mathbf{y}. \quad (41)$$

Since $\mathbb{E}\{\mathbf{h}\} = \mathbf{0}_{Q\times 1}$, we have from (16) that $\mathbb{E}\left\{\boldsymbol{\xi}^{(n)}_{\tilde{\mathbf{w}}}\right\} = \mathbf{0}_{MN\times 1}$. The covariance matrix $\mathbf{C}^{(n)}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}}$ of the vector $\boldsymbol{\xi}^{(n)}_{\tilde{\mathbf{w}}}$ in the nth iteration is computed in the next lemma, which is proved in Appendix-C.

Lemma 4: The covariance matrix $\mathbf{C}^{(n)}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}} = \mathbb{E}\{\boldsymbol{\xi}^{(n)}_{\tilde{\mathbf{w}}}(\boldsymbol{\xi}^{(n)}_{\tilde{\mathbf{w}}})^H\}$ of the vector $\boldsymbol{\xi}^{(n)}_{\tilde{\mathbf{w}}}$ is

$$\mathbf{C}^{(n)}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}} = 2\left(\sum_{i=1}^{Q}\sigma^2_{h_i}\right)\sigma^2_d\mathbf{I}_{MN} + \sigma^2_w\mathbf{I}_{MN}. \quad (42)$$

The MSE of the data-aided channel estimator is [26]:

$$B^{(n)}_h = \mathrm{Tr}\left(\mathbf{\Sigma}^{(n)}_{\mathbf{h}}\right), \quad (43)$$

where $\mathbf{\Sigma}^{(n)}_{\mathbf{h}} = \left(\left(\mathbf{\Omega}^{(n-1)}_{\mathbf{x}_p\hat{\mathbf{x}}_d}\right)^H \left(\mathbf{C}^{(n-1)}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}}\right)^{-1} \mathbf{\Omega}^{(n-1)}_{\mathbf{x}_p\hat{\mathbf{x}}_d} + \mathbf{C}^{-1}_{\mathbf{h}} \right)^{-1}$ is the covariance matrix of the error $\left(\mathbf{h} - \hat{\mathbf{h}}^{(n)}\right)$. Let $\widehat{\mathbf{H}}^{(n)}_{\text{eff}}$ be the estimate of the effective channel matrix corresponding to the data-aided channel estimate $\hat{\mathbf{h}}^{(n)}$ in the nth iteration. It is obtained by substituting $\hat{\mathbf{h}}^{(n)}$ in (9):

$$\widehat{\mathbf{H}}^{(n)}_{\text{eff}} = \mathbf{B}_{\text{rx}}\left(\sum_{i=1}^{Q}\hat{h}^{(n)}_i\mathbf{\Theta}_i\right)\mathbf{B}_{\text{tx}}. \quad (44)$$

2) Message Passing-Aided Data Detection in SP-I Design: We now use the message passing algorithm to detect data by using data-aided channel estimate. The receive signal $\tilde{\mathbf{y}}^{(n)}_d \in \mathbb{C}^{MN\times 1}$ used for detecting data in the nth SP-I iteration is obtained from the received vector in (7) as

$$\tilde{\mathbf{y}}^{(n)}_d = \mathbf{y} - \widehat{\mathbf{H}}^{(n)}_{\text{eff}}\mathbf{x}_p = \mathbf{H}_{\text{eff}}\mathbf{x}_d + \tilde{\mathbf{e}}^{(n)}_p + \tilde{\mathbf{w}} = \mathbf{H}_{\text{eff}}\mathbf{x}_d + \tilde{\mathbf{w}}^{(n)}_{\tilde{e}}, \quad (45)$$

where the error vector $\tilde{\mathbf{e}}^{(n)}_p = \left(\mathbf{H}_{\text{eff}} - \widehat{\mathbf{H}}^{(n)}_{\text{eff}}\right)\mathbf{x}_p$ and the noise-plus-estimation error vector $\tilde{\mathbf{w}}^{(n)}_{\tilde{e}} = \tilde{\mathbf{e}}^{(n)}_p + \tilde{\mathbf{w}}$. *We see that statistics $\tilde{\mathbf{e}}^{(n)}_p$ in each iteration is also a function of data estimate. This is because the channel estimate, as shown in* (41)*, is a function of data estimate. The first- and second-order characteristics of $\tilde{\mathbf{e}}^{(n)}_p$ now also need the first- and second-order statistics of the data estimate.* The mean and variance of the jth element of $\tilde{\mathbf{e}}^{(n)}_p$, similar to the derivation of (26) and (29), can be obtained as $\mathbb{E}\left\{\tilde{\mathbf{e}}^{(n)}_p(j)\right\} = 0$ and $\mathrm{Var}\left[\tilde{\mathbf{e}}^{(n)}_p(j)\right] = \sigma^2_p B^{(n)}_h$, respectively. Furthermore, since $\tilde{\mathbf{e}}^{(n)}_p$ and $\tilde{\mathbf{w}}$ are statistically independent random vectors, we get

$$\mathbb{E}\left\{\tilde{\mathbf{w}}^{(n)}_{\tilde{\mathbf{e}}}(j)\right\} = 0 \text{ and } \mathrm{Var}\left[\tilde{\mathbf{w}}^{(n)}_{\tilde{\mathbf{e}}}(j)\right] = \sigma^2_p B^{(n)}_h + \sigma^2_w. \quad (46)$$

For detecting data with the channel estimate $\widehat{\mathbf{H}}^{(n)}_{\text{eff}}$, the mean $\mu^{(i)}_{a,b}$ in the ith iteration of the message passing algorithm, similar to (32), is calculated as follows

$$\mu^{(i)}_{a,b} = \sum_{c\in\mathcal{I}(a),c\neq b}\sum_{j=1}^{S} p^{(i-1)}_{c,a}(\alpha_j)\,\alpha_j\widehat{\mathbf{H}}^{(n)}_{\text{eff}}(a,c) + \mathbb{E}\left\{\tilde{\mathbf{w}}^{(n)}_{\tilde{e}}(a)\right\}. \quad (47)$$

Here, as shown in (46), $\mathbb{E}\left\{\tilde{\mathbf{w}}^{(n)}_{\tilde{e}}(a)\right\} = 0$. The variance $\left(\sigma^{(i)}_{a,b}\right)^2$ for the ith iteration of the message passing design, using (46) and the estimate $\widehat{\mathbf{H}}^{(n)}_{\text{eff}}$, is given as follows

$$\left(\sigma^{(i)}_{a,b}\right)^2 = \sum_{c\in\mathcal{I}(a),c\neq b}\sum_{j=1}^{S} p^{(i-1)}_{c,a}(\alpha_j)\,|\alpha_j|^2\left|\widehat{\mathbf{H}}^{(n)}_{\text{eff}}(a,c)\right|^2 - \left|\mu^{(i)}_{a,b}\right|^2 + \sigma^2_p B^{(n)}_h + \sigma^2_w. \quad (48)$$

The pmf $p^{(i)}_{c,a}$ is calculated from (34) and (35) by using the channel estimate $\widehat{\mathbf{H}}^{(n)}_{\text{eff}}$, and the above computed mean $\mu^{(i)}_{a,b}$ and variance $\left(\sigma^{(i)}_{a,b}\right)^2$. We once again observe that, unlike [4], [8], the variance $\left(\sigma^{(i)}_{a,b}\right)^2$ is also a function of the MSE $B^{(n)}_h$ of the proposed SP-I channel estimator and the pilot power σ^2_p. The computed mean and variance, along with the vector $\tilde{\mathbf{y}}^{(n)}_d$ and estimate $\widehat{\mathbf{H}}^{(n)}_{\text{eff}}$, are fed to Algorithm 1 for calculating the data estimate $\hat{\mathbf{x}}^{(n)}_d$. Algorithm 2 summarizes the proposed SP-I design. It is initialized with data detected by the SP-NI design. The parameters in Steps $3-4$ are computed using data detected in the $(n-1)$th iteration. Step 5 describes the iterative data-aided MMSE channel estimation in nth iteration. The next step feeds data-aided channel estimate $\widehat{\mathbf{H}}^{(n)}_{\text{eff}}$, along with the vector $\tilde{\mathbf{y}}^{(n)}_d$, to Algorithm-1 for calculating $\hat{\mathbf{x}}^{(n)}_d$.

C. Estimation of l_i and k_i

For a doubly-underspread (DU) channel [22], as shown later in Section-VI, the parameters l_i and k_i remain constant over multiple OTFS frames and the complex channel gain h_i varies across frames. We propose the following super-frame architecture to estimate l_i and k_i. The first frame in each super-frame consists of non-superimposed EP delay-Doppler frame in Fig. 1(a). Remaining frames in the super-frame consist of the proposed superimposed frames in Fig. 1(b).

1) First frame has dedicated slots for pilot and data symbols. This frame suppresses interference between data and pilots by inserting zeros between them. For this frame structure, we estimate l_i and k_i using threshold-based method given in [8, Eq. (6)].
2) For a DU channel, the parameters l_i and k_i remain constant over multiple OTFS frames in the super-frame, whereas the channel gain varies across frames. We thus need to estimate only channel gains h_i in the subsequent frames, which carry both data and pilot symbols according to the proposed SP-aided frame structure in Fig. 1(b).

We assume, similar to [2], [18], one dominant path per cluster. Since there are Q clusters in the channel, the number of

Algorithm 2: Iterative Channel and Data Detection for the SP-I Design

Input: Observation vector $\mathbf{y}$, pilot matrix $\mathbf{\Omega}_p$, $\mathbb{E}\{\mathbf{h}\}$, $\mathbf{C}_{\mathbf{h}} = \mathbb{E}\{\mathbf{h}\mathbf{h}^H\}$ and the initial data estimate $\hat{\mathbf{x}}_d^{(0)} = \hat{\mathbf{x}}_{d,\text{NI}}$ from Algorithm-1.

Output: Channel estimate $\hat{\mathbf{h}}$ and data estimate $\hat{\mathbf{x}}_d$

Initialize: Counter $n = 1$

repeat

- Compute $\widehat{\mathbf{\Omega}}_d^{(n-1)}$ as given in (38). Compute $\mathbf{\Omega}_{\mathbf{x}_p\hat{\mathbf{x}}_d}^{(n-1)}$ and $\mathbf{C}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}}^{(n-1)}$ as derived in (40) and (42), respectively
- Compute $\hat{\mathbf{h}}^{(n)}$ and $\widehat{\mathbf{H}}_{\text{eff}}^{(n)}$ as derived in (41) and (44), respectively
- Feed the estimate $\widehat{\mathbf{H}}_{\text{eff}}^{(n)}$ and the vector $\tilde{\mathbf{y}}_d^{(n)}$ to the message passing receiver Algorithm 1 to calculate the data estimate $\hat{\mathbf{x}}_d^{(n)}$. Note that the messages $\mu_{a,b}^{(i)}$ and $\left(\sigma_{a,b}^{(i)}\right)^2$ for the ith iteration of the message passing algorithm are computed from (47) and (48), respectively. Corresponding pmf $p_{c,a}^{(i)}$ is evaluated by using $\widehat{\mathbf{H}}_{\text{eff}}^{(n)}$ and the messages $\mu_{a,b}^{(i)}$ and $\left(\sigma_{a,b}^{(i)}\right)^2$ in (34) and (35).
- $n = n + 1$

until *Stopping criteria*

return: $\hat{\mathbf{x}}_d$

channel paths in the delay-Doppler domain equals Q. We see from the expression $\mathbf{H} = \sum_{i=1}^{Q} h_i \mathbf{\Pi}^{l_i} \mathbf{\Delta}^{k_i}$ that there is one (l_i, k_i) for the ith path. Thus, the proposed methodology discussed above for estimating (l_i, k_i) also estimates the parameter Q.

IV. Optimal Power Allocation Between Data and Pilot Symbols

The proposed designs superimpose pilots on to data symbols. Given the total power constraint, we now optimally allocate power between data and pilots to maximize SINR, which will minimize the BER and maximize the SE. To achieve this objective, we first derive the SINR using the SP-NI channel estimate $\hat{\mathbf{h}}_{\text{NI}}$, which is then maximized to calculate the optimal pilot power $\sigma^2_{p,\text{opt}}$ and the data power $\sigma^2_{d,\text{opt}} = 1 - \sigma^2_{p,\text{opt}}$. We show that this optimal power allocation maximizes the SE and minimizes the BER of SP-I design also. The SINR derivation, as shown next, is significantly challenging, since it is a function of the MSE of the channel estimate, which in turn is a function of the matrices $\mathbf{\Omega}_d$ and $\mathbf{\Omega}_p$. Further, $\mathbf{\Omega}_d$ and $\mathbf{\Omega}_p$ are a function of the matrix $\mathbf{\Gamma}_i = \mathbf{B}_{\text{rx}} \mathbf{\Theta}_i \mathbf{B}_{\text{tx}}$. The noise-plus-interference vector $\check{\mathbf{w}}$, as shown below, is also a function of $\mathbf{\Gamma}_i$ through $\mathbf{\Omega}_d$ and $\mathbf{\Omega}_p$. By using ideas/inequalities from random matrix theory and linear algebra, we derive the SINR lower bound. We calculate the effective SINR by using the channel estimate $\hat{\mathbf{h}}_{\text{NI}}$ from (20). The observation vector $\mathbf{y}_d$ in (23), using (12), can be rewritten as

$$\begin{aligned}\mathbf{y}_d &= \mathbf{y} - \mathbf{\Omega}_p \hat{\mathbf{h}}_{\text{NI}} = \mathbf{\Omega}_d \mathbf{h} + \left(\mathbf{h} - \hat{\mathbf{h}}_{\text{NI}}\right) \mathbf{\Omega}_p + \tilde{\mathbf{w}} \\ &= \mathbf{\Omega}_d \hat{\mathbf{h}}_{\text{NI}} + (\mathbf{\Omega}_d + \mathbf{\Omega}_p)\, \tilde{\mathbf{h}} + \tilde{\mathbf{w}} = \mathbf{\Omega}_d \hat{\mathbf{h}}_{\text{NI}} + \check{\mathbf{w}}. \end{aligned} \tag{49}$$

The error vector $\tilde{\mathbf{h}}$ and the noise-plus-interference vector $\check{\mathbf{w}}$ are defined as follows

$$\tilde{\mathbf{h}} = \mathbf{h} - \hat{\mathbf{h}}_{\text{NI}} \text{ and } \check{\mathbf{w}} = (\mathbf{\Omega}_d + \mathbf{\Omega}_p)\, \tilde{\mathbf{h}} + \tilde{\mathbf{w}}. \tag{50}$$

The received symbol at delay-Doppler location (l, k), using (49), can be expressed as [18]

$$\begin{aligned} y_d[l,k] = &\sum_{i=1}^{Q} \hat{h}^{\alpha}_{\text{NI},i} x_d\left[(l-l_i)_M, (k-k_i)_N\right] \\ &+ \sum_{i=1}^{Q} \tilde{h}^{\alpha}_{\text{NI},i} \Big(x_d\left[(l-l_i)_M, (k-k_i)_N\right] \\ &+ x_p\left[(l-l_i)_M, (k-k_i)_N\right] \Big) + \tilde{w}[l,k]. \end{aligned} \tag{51}$$

Here $\hat{h}^{\alpha}_{\text{NI},i} = \hat{h}_{\text{NI},i} \alpha_i(l,k)$ and $\tilde{h}^{\alpha}_{\text{NI},i} = \left(h_i - \hat{h}_{\text{NI},i}\right) \alpha_i(l,k)$, where $\alpha_i(l,k)$ is given in (28). Equation (51) can be re-written as

$$\begin{aligned} y_d[l,k] &= \tilde{\mathbf{x}}_d^T \hat{\mathbf{h}}^{\alpha}_{\text{NI}} + (\tilde{\mathbf{x}}_d + \tilde{\mathbf{x}}_p)^T \tilde{\mathbf{h}}^{\alpha}_{\text{NI}} + \tilde{w}[l,k] \\ &= \tilde{\mathbf{x}}_d^T \hat{\mathbf{h}}^{\alpha}_{\text{NI}} + v[l,k]. \end{aligned} \tag{52}$$

The noise-plus-interference term $v[l,k] = (\tilde{\mathbf{x}}_d + \tilde{\mathbf{x}}_p)^T \tilde{\mathbf{h}}^{\alpha}_{\text{NI}} + \tilde{w}[l,k]$, and $\tilde{\mathbf{x}}_d \in \mathbb{C}^{Q\times 1}$ and $\tilde{\mathbf{x}}_p \in \mathbb{C}^{Q\times 1}$ are data and pilot vectors. Their ith element is given as $x_d\left[(l-l_i)_M, (k-k_i)_N\right]$ and $x_p\left[(l-l_i)_M, (k-k_i)_N\right]$, respectively. The scalers $\hat{h}^{\alpha}_{\text{NI},i}$ and $\tilde{h}^{\alpha}_{\text{NI},i}$ denote the ith element of $\hat{\mathbf{h}}^{\alpha}_{\text{NI}}$ and $\tilde{\mathbf{h}}^{\alpha}_{\text{NI}}$, respectively. The effective SINR for (l,k)th delay-Doppler domain symbol is given as

$$\text{SINR}_{\text{eff}}(l,k) = \frac{\mathbb{E}\left\{\left|\tilde{\mathbf{x}}_d^T \hat{\mathbf{h}}^{\alpha}_{\text{NI}}\right|^2\right\}}{\mathbb{E}\left\{\left|v[l,k]\right|^2\right\}}. \tag{53}$$

Its numerator and denominator are a function of the channel estimate and estimation error, respectively. For a linear MMSE estimator, the estimation error is orthogonal to the observations [26]. The noise is also independent of data symbols. These properties imply that the numerator and denominator of (53) are independent. The numerator of (53) is simplified next.

$$\mathbb{E}\left\{\left|\tilde{\mathbf{x}}_d^T \hat{\mathbf{h}}^{\alpha}_{\text{NI}}\right|^2\right\} \overset{(b)}{=} \sigma_d^2 \mathbb{E}\left\{\left\|\hat{\mathbf{h}}_{\text{NI}}\right\|^2\right\}.$$

Equality (b) is because $\mathbb{E}\left\{\tilde{\mathbf{x}}_d^* \tilde{\mathbf{x}}_d^T\right\} = \sigma_d^2 \mathbf{I}_Q$ and $\alpha_i(l,k)$ is a phase factor. The quantity $\mathbb{E}\left\{\left\|\hat{\mathbf{h}}_{\text{NI}}\right\|^2\right\} = \text{Tr}\left(\mathbb{E}\left\{\hat{\mathbf{h}}_{\text{NI}} \left(\hat{\mathbf{h}}_{\text{NI}}\right)^H\right\}\right)$ can be calculated using (50) as follows

$$\text{Tr}\left(\mathbb{E}\left\{\hat{\mathbf{h}}_{\text{NI}} \left(\hat{\mathbf{h}}_{\text{NI}}\right)^H\right\}\right) = \sigma_h^2 - B_{h,\text{NI}}. \tag{54}$$

Here $\sigma_h^2 = \text{Tr}(\mathbf{C}_{\mathbf{h}}) = \sum_{i=1}^{Q} \sigma_{h_i}^2$ and $B_{h,\text{NI}} = \text{Tr}\left(\mathbf{\Sigma}_{\text{NI}}\right)$ is derived in (22). The above equation implies that the numerator of SINR expression in (53) is

$$\mathbb{E}\left\{\left|\tilde{\mathbf{x}}_d^T \hat{\mathbf{h}}^{\alpha}_{\text{NI}}\right|^2\right\} = \sigma_d^2 \left(\sigma_h^2 - B_{h,\text{NI}}\right). \tag{55}$$

We next simplify the denominator of SINR expression in (53) as follows

$$\mathbb{E}\left\{|v[l,k]|^2\right\} = \mathrm{Tr}\left(\mathbb{E}\left\{\tilde{\mathbf{h}}_{\mathrm{NI}}^{\alpha}\left(\tilde{\mathbf{h}}_{\mathrm{NI}}^{\alpha}\right)^H\right\}\tilde{\mathbf{x}}_p^*\tilde{\mathbf{x}}_p^T\right) + \sigma_d^2\mathbb{E}\left\{\left\|\tilde{\mathbf{h}}_{\mathrm{NI}}\right\|^2\right\}\sigma_w^2. \quad (56)$$

We now simplify the second term of the above expression as

$$\mathrm{Tr}\left(\mathbb{E}\left\{\tilde{\mathbf{h}}_{\mathrm{NI}}^{\alpha}\left(\tilde{\mathbf{h}}_{\mathrm{NI}}^{\alpha}\right)^H\right\}\tilde{\mathbf{x}}_p^*\tilde{\mathbf{x}}_p^T\right) \leq \mathrm{Tr}\left(\mathbb{E}\left\{\tilde{\mathbf{h}}_{\mathrm{NI}}^{\alpha}\left(\tilde{\mathbf{h}}_{\mathrm{NI}}^{\alpha}\right)^H\right\}\right)\mathrm{Tr}\left(\tilde{\mathbf{x}}_p^*\tilde{\mathbf{x}}_p^T\right) = Q\sigma_p^2 B_{h,\mathrm{NI}}. \quad (57)$$

The above inequality is because for two positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$, the following holds i.e., $\mathrm{Tr}(\mathbf{AB}) \leq \mathrm{Tr}(\mathbf{A})\,\mathrm{Tr}(\mathbf{B})$ [32]. Using (57), we simplify $\mathbb{E}\left\{|v[l,k]|^2\right\}$ as

$$\mathbb{E}\left\{|v[l,k]|^2\right\} \leq \sigma_d^2 B_{h,\mathrm{NI}} + Q\sigma_p^2 B_{h,\mathrm{NI}} + \sigma_w^2. \quad (58)$$

The SINR in (53) is computed on substitution of (55) and (58) as

$$\mathrm{SINR}_{\mathrm{eff}} \geq \frac{\sigma_d^2\left(\sigma_h^2 - B_{h,\mathrm{NI}}\right)}{\sigma_d^2 B_{h,\mathrm{NI}} + Q\sigma_p^2 B_{h,\mathrm{NI}} + \sigma_w^2}. \quad (59)$$

We have dropped the subscript (l,k) as the SINR expression is independent of them. To evaluate (59), we now calculate the MSE $B_{h,\mathrm{NI}}$ of the SP-aided channel estimation. It follows from (22) that $B_{h,\mathrm{NI}} = \mathrm{Tr}[\boldsymbol{\Sigma}_{\mathrm{NI}}]$. We next propose a lemma, whose proof is relegated to Appendix-D.

Lemma 5: The MSE $B_{h,\mathrm{NI}}$ is given as $B_{h,\mathrm{NI}} \approx Q^2/\left(\frac{QMN\sigma_p^2}{\sigma_h^2\sigma_d^2+\sigma_w^2} + \tilde{\sigma}_h^2\right)$.

By substituting $B_{h,\mathrm{NI}}$ from Lemma 5, and by substituting $\sigma_d^2 = 1-\sigma_p^2$, the effective SINR is

$$\mathrm{SINR}_{\mathrm{eff}} \geq \frac{\sigma_p^4 N_1 + \sigma_p^2 N_2 + N_3}{\sigma_p^4 D_1 + \sigma_p^2 D_2 + D_3}. \quad (60)$$

Here $N_1 = \sigma_h^2 QMN - \sigma_h^2\tilde{\sigma}_h^2 + \sigma_h^2 Q^2, N_2 = \sigma_h^2 QMN - 2\sigma_h^2\tilde{\sigma}_h^2 + \sigma_h^2 Q^2 - \tilde{\sigma}_h^2\sigma_w^2 + Q\sigma_h^2 + Q^2\sigma_w^2$ and $N_3 = \sigma_h^2\tilde{\sigma}_h^2 + \tilde{\sigma}_h^2\sigma_w^2 - Q^2\sigma_h^2 - Q^2\sigma_w^2$. And, $D_1 = \sigma_h^2 Q^2 - \sigma_h^2 Q^3, D_2 = \sigma_h^2 Q^3 + \sigma_w^2 Q^3 - 2\sigma_h^2 Q^2 - \sigma_w^2 Q^2 + \sigma_w^2 QMN - \sigma_h^2\tilde{\sigma}_h^2\sigma_w^2$ and $D_3 = Q^2\sigma_h^2 + Q^2\sigma_w^2 + \sigma_h^2\tilde{\sigma}_h^2\sigma_w^2 + \tilde{\sigma}_h^2\sigma_w^4$. To calculate the optimal pilot power, we take the derivative of the lower bound on $\mathrm{SINR}_{\mathrm{eff}}$, and equate it to zero:

$$\frac{\partial \mathrm{SNR}_{\mathrm{eff}}}{\partial\sigma_p^2} = \sigma_p^4(D_2N_1 - D_1N_2) + \sigma_p^2(2D_3N_1 - 2D_1N_3) + (D_3N_2 - D_2N_3) = 0. \quad (61)$$

After solving the above expression, the optimal pilot power is obtained as $\sigma_{p,\mathrm{opt}}^2 = \left|\frac{-b+\sqrt{b^2-4ac}}{2a}\right|$, where $a = D_2N_1 - D_1N_2$, $b = 2D_3N_1 - 2D_1N_3$ and $c = D_3N_2 - D_2N_3$. Using the constraint on total power per symbol, i.e., $\sigma_p^2 + \sigma_d^2 = 1$, the optimal data power is obtained as $\sigma_{d,\mathrm{opt}}^2 = 1 - \sigma_{p,\mathrm{opt}}^2$.

Remark 5: We note that the SINR is commonly calculated in the SP literature by averaging over estimated channel [12]. This is because a major aim of deriving SINR is to optimally allocate power between data and pilot symbols to minimize BER. Since pilots are used to estimate channel, their power thus cannot depend on instantaneous channel [12]. The SINR determines the allocated pilot power and consequently, cannot be a function of the instantaneous channel.

V. Computational Complexity Analysis

We count multiplication/division and addition/subtraction as operations [24]. We see from Table-I that the total number of operations required by the proposed SP-NI scheme is $\mu_{\mathrm{SPNI}} = \left(2Q^2+2Q\right)MN + 3Q^2 + \mathcal{O}(Q^3) + \mathcal{O}(N_I MNQS)$. We see that the complexity of the proposed SP-I scheme varies as $\mu_{\mathrm{SPI}} = \left(\left(2Q^2+2Q+1\right)MN + 3Q^2 + \mathcal{O}(Q^3) + \mathcal{O}(N_I MNQS)\right)N_{\mathrm{SPI}} + \mu_{\mathrm{SPNI}}$, where N_{SPI} denotes the number of iteration required by the SP-I scheme and μ_{SPNI} operations are needed for the initial estimate. In practice, $Q \ll MN$, we see that the complexity of the proposed SP-NI scheme is $\mathcal{O}(MN) + \mathcal{O}(N_I MNQS)$. The complexity of SP-I scheme is $\mathcal{O}(MN)(N_{\mathrm{SPI}}+1) + \mathcal{O}(N_I MNQS)(N_{\mathrm{SPI}}+1)$. We see from Table-I that the EP scheme requires $\mu_{\mathrm{EP}} = (2k_{\max}+1)(l_{\max}+1) + 6Q + \mathcal{O}(N_I(MN - \left((2l_{\max}+1)(4k_{\max}+1)\right)QS)$ operations, where $l_{\max}$ and $k_{\max}$ denote the tap corresponding to maximum delay and Doppler shift, respectively. Since typically, $k_{\max}, l_{\max}, Q \ll MN$, the last complexity term of the EP scheme dominates.

VI. Simulation Results

We now numerically validate the derived analytical results, and the performance of the proposed designs. For this study, we consider an OTFS system with the number of delay bins $M = 16$, the Doppler bins $N \in [12, 128]$, and set the carrier frequency and subcarrier spacing as 4 GHz and 15 KHz, respectively. The system uses a rectangular pulse and symbols are drawn from BPSK, 4-QAM and 16-QAM constellations. We terminate the i) message passing algorithm, similar to [4], using the convergence indicator in (36); ii) proposed SP-I algorithm when $||\hat{\mathbf{h}}^{(n)} - \hat{\mathbf{h}}^{(n+1)}||^2 < 10^{-6}$ or when the number of iteration 10, whichever is achieved earlier. We define the SNR as $1/\sigma_w^2$. We use, similar to [7], a 5-tap delay-Doppler channel with parameters specified in Table II. We generate Doppler values, similar to [4], using the Jakes' formula as $\nu_i = \nu_{\max} cos(\theta_i)$. Here ν_i is the Doppler shift of the ith delay-Doppler propagation path, $\nu_{max} = 1851$ Hz is the maximum Doppler shift corresponding to the maximum relative speed of 500 kmph, and the angle θ_i is uniformly distributed in the range $[-\pi, \pi]$. We consider 10^4 different channel realizations for each SNR in the Monte-Carlo simulations.

A. Doubly Under-Spread Channel

We first show that the channel with parameters given in Table II is a DU channel i.e., it satisfies the following condition [22, Eq. (31)]: $\Delta\tau_{\max}\Delta\nu_{\max} \ll \tau_{\max}\nu_{\max} \ll 1$, where $\tau_{\max}$ and $\nu_{\max}$ are the maximum delay and Doppler shifts, respectively. Parameters $\Delta\tau_{\max}$ and $\Delta\nu_{\max}$ denote the

TABLE I

COMPUTATIONAL COMPLEXITIES OF THE PROPOSED AND THE EXISTING EP SCHEMES

Proposed SP-NI Scheme			
Operation	Complex Multiplications	Complex Additions	Total number of operations
$\mathbf{A}=\mathbf{\Omega}_p^H\mathbf{C}_{\tilde{\mathbf{w}}_d}^{-1}\mathbf{\Omega}_p+\mathbf{C}_h^{-1}$	Q^2MN+Q^2+Q	$Q^2(MN-1)+Q^2$	$2Q^2MN+Q^2+Q$
$\mathbf{A}^{-1}$	-	-	$\mathcal{O}(Q^3)$
$\mathbf{b}=\mathbf{\Omega}_p^H\mathbf{C}_{\tilde{\mathbf{w}}_d}^{-1}\mathbf{y}$	$Q(MN+1)$	$Q(MN-1)$	$2QMN$
$\mathbf{A}^{-1}\mathbf{b}$	Q^2	$Q(Q-1)$	$2Q^2-Q$
Data Detection using MP	-	-	$\mathcal{O}(N_I MNQS)$ [4]

Proposed SP-I Scheme (per iteration)			
Operation	Complex Multiplications	Complex Additions	Total number of operations
$\mathbf{\Omega}_{\mathbf{x}_p\hat{\mathbf{x}}_d}^{(n)}$	-	MN	MN
$\mathbf{A}^{(n)}=(\mathbf{\Omega}_{\mathbf{x}_p\hat{\mathbf{x}}_d}^{(n)})^H(\mathbf{C}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}}^{(n)})^{-1}\mathbf{\Omega}_{\mathbf{x}_p\hat{\mathbf{x}}_d}^{(n)}+\mathbf{C}_h^{-1}$	Q^2MN+Q^2+Q	$Q^2(MN-1)+Q^2$	$2Q^2MN+Q^2+Q$
$(\mathbf{A}^{(n)})^{-1}$	-	-	$\mathcal{O}(Q^3)$
$\mathbf{b}^{(n)}=(\mathbf{\Omega}_{\mathbf{x}_p\hat{\mathbf{x}}_d}^{(n)})^H(\mathbf{C}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}}^{(n)})^{-1}\mathbf{y}$	$Q(MN+1)$	$Q(MN-1)$	$2QMN$
$(\mathbf{A}^{(n)})^{-1}\mathbf{b}^{(n)}$	Q^2	$Q(Q-1)$	$2Q^2-Q$
Data Detection using MP	-	-	$\mathcal{O}(N_I MNQS)$ [4]

Existing EP scheme			
Operation	Complex Multiplications	Complex Additions	Total number operations
$\lvert y(k,l)\rvert$	$(2k_{\max}+1)(l_{\max}+1)$	-	$(2k_{\max}+1)(l_{\max}+1)$
$\hat{h}_{\text{EP},i}$ for $1\le i\le Q$	$5Q$	Q	$6Q$
Data Detection using MP	-	-	$\mathcal{O}(N_I QS(MN-(2l_{\max}+1)(4k_{\max}+1)))$

TABLE II

DELAY-DOPPLER CHANNEL PARAMETERS FOR RELATIVE SPEED OF 500 KMPH

Channel tap no.	1	2	3	4	5
Delay (μs)	2.08	5.20	8.328	11.46	20.8
Channel tap power (dB)	1	-1.804	-3.565	-5.376	-8.860

TABLE III

OPTIMAL TRAINING AND DATA POWERS WITH $Q=5$, $M=N=16$ AND σ_h^2 (TAKEN FROM TABLE II)

SNR (dB)	0	5	10	15	20
$\sigma_{p,\text{opt}}^2$	0.3020	0.3153	0.3322	0.3479	0.3600
$\sigma_{d,\text{opt}}^2=1-\sigma_{p,\text{opt}}^2$	0.6980	0.6847	0.6678	0.6521	0.6400

maximum delay and the Doppler correlation lag, respectively. As explained in Section-VI-B of [22], $\Delta\tau_{\max}$ and $\Delta\nu_{\max}$ are calculated as follows $\Delta\tau_{\max}=w/c$ and $\Delta\nu_{\max}=2\nu_{\max}\sin(\delta/2)$. Here w and δ are the maximum spatial extension and the maximum angular spread of the scatterers respectively and c denotes the speed of light. We have from Table II: $\tau_{\max}=20.8\mu s$ and $\nu_{\max}=1851$ Hz, which implies that $\tau_{\max}\nu_{\max}\ll 1$. Assuming $w=30m$ and $\delta=3$ degrees [22], we have $\Delta\tau_{\max}\Delta\nu_{\max}=9.6\times10^{-6}\ll\tau_{\max}\nu_{\max}\ll 1$. The channel is thus DU. We now calculate the number of frames over which the parameters l_i and k_i remain approximately constant. Let the maximum relative speed between transmitter and receiver be 100 kmph. Thus, for 4 GHz carrier frequency, $\nu_{\max}\approx 370$ Hz and $\Delta\nu_{\max}\approx 9.68$ Hz. The stationary time T_s, which denotes the duration over which l_i and k_i remains approximately same, is calculated as $T_s=1/\Delta\nu_{\max}=103$ ms. For $N=16$, the OTFS frame duration $N/\Delta f=16/(15\times10^3)\approx 1$ ms. The number of OTFS frames over which l_i and k_i remain approximately same are $N_f\approx 103$. Similarly, for a speed 500 kmph, l_i and k_i remain same for $N_f=10$ OTFS frames, and N_f increases with decrease in velocity.

For a fair comparison between the proposed and existing designs, we assume that for each scheme, similar to [14], total power per delay-Doppler frame is same, and we fix it as MN. It implies that each delay-Doppler bin (l,k), for $0\le l\le M-1$ and $0\le k\le N-1$, is assigned a unit power. For the proposed designs, the power per delay-Doppler superimposed symbol in the frame, as shown in Fig. 1(b), is $\sigma_d^2+\sigma_p^2=1$, where σ_d^2 and σ_p^2 denote data symbol and pilot symbol power respectively. This allows us to define SNR per superimposed symbol as $1/\sigma_w^2$, where σ_w^2 is the noise variance. The EP design, as shown in Fig. 1(a), inserts zeros to avoid interference between data and pilot symbols. Total power of MN units in this scheme is therefore distributed as follows. Its power per data symbol is $\sigma_{d,\text{EP}}^2=\sigma_d^2+\sigma_p^2=1$. This also implies that the SNR per data symbol is $1/\sigma_w^2$, which is equal to that of the proposed SP-aided designs. The pilot power in the EP scheme, however, due to insertion of $(2l_{max}+1)(4k_{max}+1)-1$ zeros around the pilot symbol, is $\sigma_{p,\text{EP}}^2=(2l_{max}+1)(4k_{max}+1)$ [8]. The CPA design employs an entire frame for transmitting pilots, its total pilot power is therefore MN. Each of its delay-Doppler data symbol in the subsequent frame has a power $\sigma_{\text{CPA}}^2=1$.

We recall from Section IV that the optimal pilot and data powers are a function of SNR, number of channel taps Q, channel delay-Doppler profile parameter σ_h^2, number of delay bins M and the number of Doppler bins N. Table-III summarizes the optimal pilot power $\sigma_{p,\text{opt}}^2$ and the optimal data power $\sigma_{d,\text{opt}}^2$ for different SNR values obtained using the expression derived in the paragraph below (61). We see from this table that for the given simulation parameters, the optimal pilot and data powers are approximately 0.3 and 0.7, respectively. To maximize the SE and to minimize the BER of the proposed SP-aided designs, 30% of the total power should thus be allocated to pilots and 70% to data, as also numerically verified in the sequel.

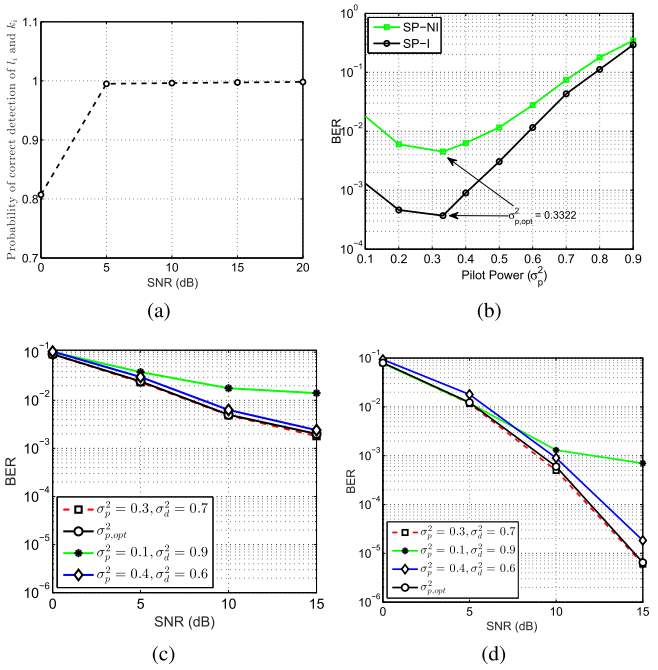


Fig. 4. (a) Probability of correct detection of delay-Doppler taps (l_i, k_i) using the frame structure given in Fig. 1(a) with $M = N = 16$. (b) BER versus pilot power σ_p^2 for the proposed schemes with $M = N = 16$ and SNR = 10 dB. Effect of power distribution among pilot and data on the BER with $N = M = 16$ (c) SP-NI and (d) SP-I.

B. Effect of Power Allocation on BER of the Proposed Designs

Figure 4(a) shows the probability of correct detection of the delay-Doppler taps (l_i, k_i) for the ith path. Recall that l_i and k_i, as discussed at the end of Section-III, are estimated using the EP-aided frame in Fig. 1(a). We see that the probability of correct detection is close to one for SNR ≥ 5 dB. Figure 4(b) shows the BER of the proposed SP-I and SP-NI designs as a function of the pilot power σ_p^2 at a of SNR = 10 dB. We see that both the schemes yield minimum BER when pilot power $\sigma_p^2 = \sigma_{p,\text{opt}}^2 = 0.3322$. This numerically verifies the results shown in Table-III. Since $\sigma_p^2 + \sigma_d^2 = 1$, we also observe that the BER of the proposed designs degrades for i) $\sigma_p^2 < \sigma_{p,\text{opt}}^2$, due to the poor channel estimate; ii) for $\sigma_p^2 > \sigma_{p,\text{opt}}^2$, due to reduced data power.

We plot in Fig. 4(c) and Fig. 4(d) the BER of the SP-NI and SP-I designs respectively, for different pilot and data powers. We observe from these plots that the proposed designs, with optimal power allocation, have minimum BER, which also overlaps with BER with $\sigma_p^2 = 0.3$ and $\sigma_d^2 = 0.7$. This validates the optimal pilot and data powers in Table-III, which are calculated analytically. We further observe that the SP-I design has a much lower BER than the SP-NI design for all SNR values. The BER gap between the two widens at high SNR values. For example, the SP-I design achieves a BER of 10^{-3} at 8.5 dB, whereas the SP-I design is not able to attain this BER in the operating SNR range. This is because the SP-NI design experiences mutual interference between data and pilot symbols, which degrades its performance at high SNR values.

C. SE and BER Comparison of the Proposed and Existing Designs

This section first compares the SE of the proposed SP-NI and SP-I designs with the state-of-the-art embedded-pilot (EP) design in [8] and conventional pilot aided (CPA) design in [7]. The EP design, as shown in Fig. 1(a), insert zeros between data and pilot, whose number depends on both l_{max} and k_{max} i.e., the tap corresponding to the maximum delay and Doppler shifts, respectively. The CPA design uses an entire frame for OTFS channel estimation. The proposed SP-aided designs, as shown in Fig. 1(b), neither transmit separate pilots nor insert guard symbols (zeros). This radically increase their SE, which we compare next. Before doing that, we calculate the SE of both EP and CPA designs. The SE of EP design is $\mathcal{R}_{\text{EP}} = (1-\eta)\log_2(1+\text{SINR}_{\text{EP}})$. The pilot overhead η can be calculated using its frame-structure in Fig. 1(a), and is given as

$$\eta = \frac{(2l_{max}+1)(4k_{max}+1)}{MN}. \tag{62}$$

The SINR of EP scheme, as derived in Appendix E, is $\text{SINR}_{\text{EP}} = \dfrac{(\sigma_h^2 - B_{h,\text{EP}})\sigma_{d,\text{EP}}^2}{\sigma_w^2 + \sigma_{d,\text{EP}}^2 B_{h,\text{EP}}}$. The CPA design in [7] uses the first OTFS frame for estimating channel, and the subsequent frame for transmitting data. The value of the pilot overhead parameter $\eta = MN/(2MN) = 0.5$. Its SE is, therefore, $\mathcal{R}_{\text{CPA}} = \frac{1}{2}\log_2(1+\text{SINR}_{\text{CPA}})$ with $\text{SINR}_{\text{CPA}} = \dfrac{(\sigma_h^2 - B_{h,\text{CPA}})\sigma_{d,\text{CPA}}^2}{\sigma_w^2 + \sigma_{d,\text{CPA}}^2 B_{h,\text{CPA}}}$. Here $B_{h,\text{CPA}}$ is the MSE of MMSE channel estimator in the CPA design.

Recall that the proposed designs estimate l_i and k_i by using the super-frame architecture discussed in Section-III, wherein the first frame is non-superimposed EP from Fig. 1(a) and the remaining frames are proposed SP frames in Fig. 1(b). The SE of SP-NI scheme for each SP frame is $\mathcal{R}_{\text{SPNI}} = \log_2(1+\text{SINR}_{\text{SPNI}})$, where $\text{SINR}_{\text{SPNI}}$ is given in (59). The SE of SP-I design for each SP frame is $\mathcal{R}_{\text{SPI}} = \log_2(1+\text{SINR}_{\text{SPI}})$, where SINR_{SPI} is computed using (59) as

$$\text{SINR}_{\text{SPI}} = \frac{\left(\sigma_h^2 - B_h\right)\sigma_{d,\text{opt}}^2}{B_h\sigma_{d,\text{opt}}^2 + B_h Q\sigma_{p,\text{opt}}^2 + \sigma_w^2}. \tag{63}$$

Here B_h is the MSE of channel estimation in the SP-I design, which is obtained from (43), once the Algorithm 2 converges. We calculate the SINR of the proposed designs using optimal powers $\sigma_{p,\text{opt}}^2$ and $\sigma_{d,\text{opt}}^2$. This is because these powers maximize both SINR and SE of the proposed designs, as verified next. The average SE $R_{\text{A, avg}}$ for the proposed SP-aided designs, for A $\in$ {SPI, SPNI}, can now be obtained as $R_{\text{A, avg}} = \dfrac{R_{\text{EP}} + (N_f - 1)R_{\text{A}}}{N_f}$.

Figure 5(a) shows that the optimal power allocation between pilot and data symbols maximizes the average SE of both SP-I and SP-NI designs. The SP-I design, due to its lower channel estimation MSE, significantly outperforms the SP-NI design, whose SE degrades at high SNR values due to interference between data and pilots.

Figure 5(b) compares the average SE of the proposed SP-NI and SP-I designs with the EP and CPA designs for

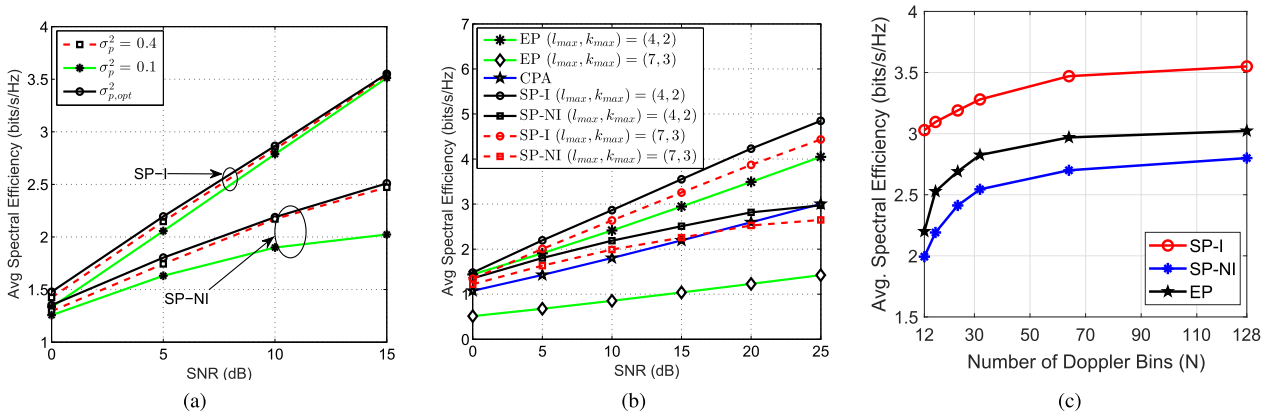


Fig. 5. (a) Effect of power allocation between data and pilots on the SE of SP-NI and SP-I designs with $M = N = 16$; (b) SE of the proposed designs with $\sigma_p^2 = \sigma_{p,\text{opt}}^2$ and the EP scheme [8] and CPA [7] designs with $M = N = 16$; and (c) SE comparison of EP and proposed schemes for $M = 16$, $Q = 5$ and varying N.

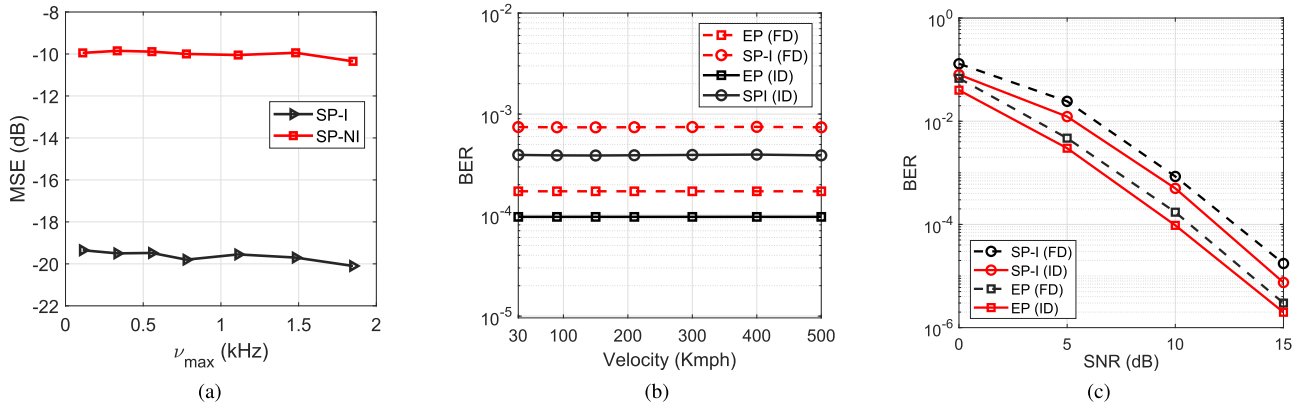


Fig. 6. (a) MSE comparison of SP-I and SP-NI schemes for varying maximum Doppler spread $\nu_{\max}$ at SNR $= 10$ dB. BPSK BER comparison of the SP-I and EP schemes for integer Doppler (ID) and fractional Doppler (FD) shifts (b) versus velocity at SNR $= 10$ dB. (c) versus SNR. For all the figures $M = N = 16$, $\sigma_p^2 = \sigma_{p,\text{opt}}^2$.

$M = N = 16$, where M and N are the number of delay and Doppler bins, respectively. We observe that the SE of all designs increases with SNR, which is not surprising. We crucially observe that the average SE of the proposed SP-I design for given l_{max} and k_{max} is significantly higher than its EP counterpart and the CPA design. We see that for large values of l_{max} and k_{max}, the average SE of the EP design degrades so much that even the SP-NI design outperforms it. This is because, as shown in (62), its pilot overhead increases with l_{max} and k_{max}. The proposed designs, in contrast, avoid this overhead.

Figure 5(c) compares the average SE of the SP and EP designs by varying the number of Doppler bins N. We consider $(\tau_{\max}, \nu_{\max}) = (21\ \mu s, 1851\ \text{Hz})$. This corresponds to $l_{\max} = 5$, and the $k_{\max}$ varies with N as $\nu_{\max} = \frac{k_{\max}}{NT}$. We see that the SP-NI design has a lower SE than the EP scheme. This is because it experiences data interference. The SP-I design significantly outperforms the EP scheme.

Figure 6(a) shows the MSE of the proposed SP-I and SP-NI schemes with increasing maximum Doppler $\nu_{\max}$. We see that the MSE of both the schemes remains almost constant with $\nu_{\max}$. This is because we assume, similar to several existing studies [4], [8], that the number of delay-Doppler channel taps Q is constant with the Doppler spread and the velocity. This is because the increase in these parameters do not imply the increase in the number of clusters in the delay-Doppler domain i.e., Q [4], [8]. The proposed schemes therefore, irrespective of the velocity, estimate only Q unknown parameters, and hence retain their performance. [4], [8].

Figure 6(b) plots the BER of the message-passing aided receiver for the SP-NI and EP schemes for fractional Doppler (FD) and integer Doppler (ID) scenarios by varying the velocity. With a fractional Doppler κ_i, the ith Doppler shift is calculated as $\nu_i = \frac{k_i + \kappa_i}{NT}$. We see that the BER of both schemes remains constant with velocity. This is because, as mentioned earlier, the two schemes, irrespective of the velocity, estimate only Q unknown parameters. A similar behavior of message passing algorithm can also be observed in [Fig. 5 and Fig. 7, 4], which considered the mobile user speed in the range of 30 to 500 Kmph. We also see that the

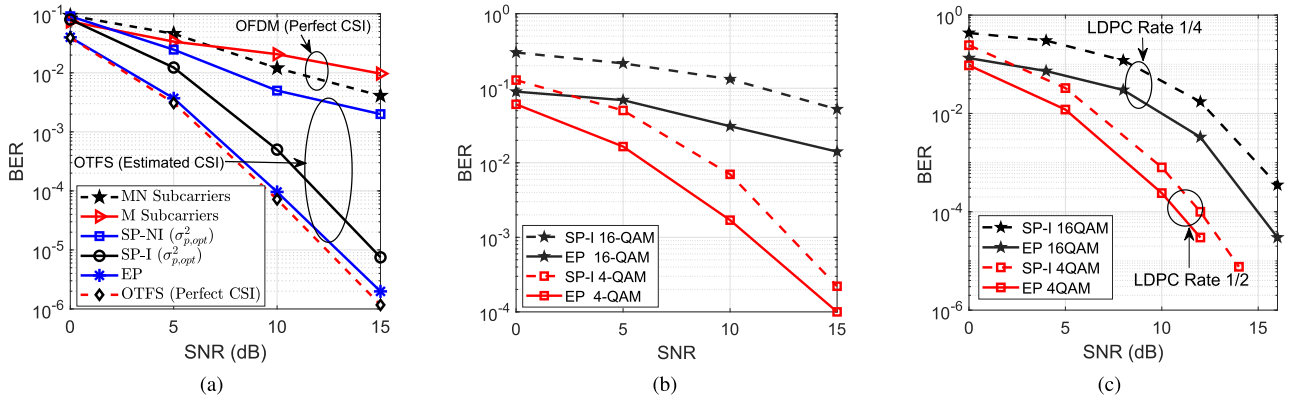


Fig. 7. (a) BER comparison of OFDM and OTFS systems with $M = N = 16$ and BPSK modulation. BER of the EP and SP-I scheme for 4/16- QAM for $M = N = 16$ and $Q = 5$. (b) Uncoded. (c) Coded.

FD slightly degrades the BER of both the schemes.[1] This is because, as explained in [18], FD adds virtual integer taps in the delay-Doppler domain, which increases the number of delay-Doppler taps Q. The algorithm thus needs to estimate a higher number of parameters for FD than ID shift for the same number of observations [8], which degrades the BER. We next investigate the BER of SP-I and EP schemes for FD and ID shifts by varying the SNR. We again observe that FD shift slightly degrades the BER of both the schemes.

Figure 7(a) compares the BER of SP-NI, SP-I and EP designs based OTFS systems with an OFDM system with MN and N subcarriers with perfect receive CSI. We consider BPSK constellation for this study. We first observe that the MN-subcarrier OFDM system has a slightly lower BER than its M-subcarrier counterpart. We also observe that the SP-NI, SP-I and EP designs based OTFS systems has lower BER than both these OFDM systems. This happens because the large Doppler shift, which occurs at a speed of 500 kmph, disturbs inter-subcarrier orthogonality in an OFDM system, and causes ICI. This significantly degrades its BER [1]. We also see that EP-based design in [8] is approximately 1.5 dB superior than the SP-I design. This is because, unlike the EP scheme, the power per symbol in the proposed designs is divided between pilot and data symbols. The power allocated to a data symbol in the proposed designs, as shown in Table III, is approximately 30% lower than the power allocated to data symbols in the EP scheme. *We, however, crucially note that the proposed designs, as shown earlier, has significantly higher SE than the EP and CPA schemes.*

Figure 7(b) shows the BER of the proposed SP-I and the existing EP schemes for higher order constellations. We see that for 4-QAM, the SP-I scheme has slightly inferior BER than the EP scheme. For 16-QAM, the BER gap between the two schemes increases. This is because the i) power allocated to a data symbol in the SP-I design is approximately 70% of the total power; and ii) increased error propagation in the SP-I scheme. We know that practical wireless systems employ error-control coding to improve performance. Figure 7(c) compares BER of two schemes for 4- and 16-QAM constellation by using low density parity check (LDPC) coding with $1/2$ and $1/4$ rate, respectively. We observe that the gap between two schemes reduces for both constellations. This is due to the reduced error propagation in the SP-I scheme.

[1]Similar to existing OTFS studies [2], [4], [8], [18], we consider a wide-band system with $M\Delta f \gg 1$. The sampling time resolution $\frac{1}{M\Delta f}$ is, therefore, sufficient to approximate the path delays to the nearest sampling points [2], [4], [8], [18]. The design of SP-aided OTFS schemes with off-grid delays requires a separate study, which can be taken up as future work.

VII. Conclusion

We proposed superimposed pilot (SP)-aided non-iterative (SP-NI) and iterative (SP-I) designs for estimating channel and detecting data. These designs superimpose pilots on to data symbols and, contrary to the existing OTFS channel estimation and data detection designs, do not incur the consequent SE loss. Another key advantage of the proposed designs is that they exploit the OTFS channel sparsity in the delay-Doppler domain by using computationally-efficient message passing algorithm for detecting data. We demonstrated that the proposed SP-I design, with a negligible BER degradation, has a significantly higher SE than the existing state-of-the-art designs. This work did not consider data-dependent SP schemes and mean-removal based techniques. Future lines of this work may consider these ideas.

Appendix A

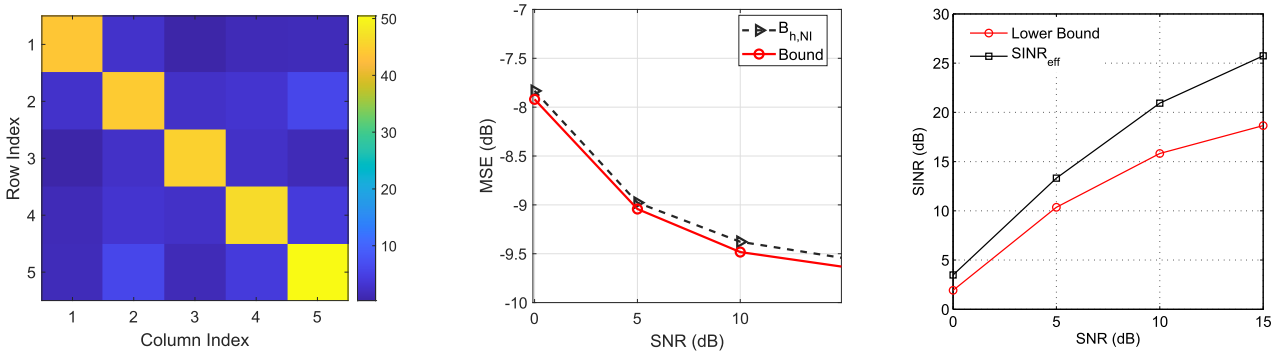
Using (14), the covariance matrix of $\mathbf{\Omega}_d$ can be evaluated as

$$\mathbb{E}\left\{\mathbf{\Omega}_d\mathbf{\Omega}_d^H\right\} = \sum_{i=1}^{Q}\mathbf{\Gamma}_i\mathbb{E}\left\{\mathbf{x}_d\mathbf{x}_d^H\right\}\mathbf{\Gamma}_i^H = \sigma_d^2\sum_{i=1}^{Q}\mathbf{\Gamma}_i\mathbf{\Gamma}_i^H,$$

where we have used: $\mathbb{E}\left\{\mathbf{x}_d\mathbf{x}_d^H\right\} = \sigma_d^2\mathbf{I}_{MN}$. We have

$$\mathbf{\Gamma}_i\mathbf{\Gamma}_i^H = \mathbf{B}_{\text{rx}}\mathbf{\Theta}_i\mathbf{B}_{\text{tx}}\mathbf{B}_{\text{tx}}^H\mathbf{\Theta}_i^H\mathbf{B}_{\text{rx}}^H = \mathbf{B}_{\text{rx}}\mathbf{\Pi}^{l_i}\left(\mathbf{\Pi}^{l_i}\right)^H\mathbf{B}_{\text{rx}}^H. \tag{64}$$

By singular valued decomposition, it can be easily shown that the cyclic-shift matrix $\mathbf{\Pi}$ satisfies the following property



(a) (b) (c)

Fig. 8. (a) Absolute values of elements of the inverse of $\mathbf{\Sigma}_{\text{NI}}$ for number of paths $Q = 5$. (b) Comparison of $B_{h,\text{NI}} = \text{Tr}(\mathbf{\Sigma}_{\text{NI}})$ and its lower bound below with $M = N = 16$ and $\sigma_p^2 = \sigma_{p,\text{opt}}^2$. (c) Comparison of the average SINR and its lower bound given in (60) with $M = N = 16$.

$\mathbf{\Pi}^{l_i}\left(\mathbf{\Pi}^{l_i}\right)^H = \mathbf{I}_{MN}$. Using this property, we get $\mathbf{\Gamma}_i\mathbf{\Gamma}_i^H = \mathbf{B}_{\text{rx}}\mathbf{B}_{\text{rx}}^H = \mathbf{I}_{MN}$. Using the last equation, we get

$$\mathbb{E}\left\{\mathbf{\Omega}_d\mathbf{\Omega}_d^H\right\} = \sigma_d^2\sum_{i=1}^{Q}\mathbf{I}_{MN} = \sigma_d^2 Q\mathbf{I}_{MN}. \tag{65}$$

APPENDIX B

Since $\mathbf{x}_d$ is statistically independent of $\mathbf{w}$, the covariance matrix of $\tilde{\mathbf{w}}_d$ can be evaluated as

$$\mathbb{E}\left\{\tilde{\mathbf{w}}_d\tilde{\mathbf{w}}_d^H\right\} = \mathbb{E}\left\{\mathbf{\Omega}_d\mathbf{h}\mathbf{h}^H\mathbf{\Omega}_d^H\right\} + \mathbb{E}\left\{\tilde{\mathbf{w}}\tilde{\mathbf{w}}^H\right\}. \tag{66}$$

The first part of the above expression can be evaluated as

$$\mathbb{E}\left\{\mathbf{\Omega}_d\mathbf{h}\mathbf{h}^H\mathbf{\Omega}_d^H\right\} \overset{(a)}{=} \left(\sum_{i=1}^{Q}\sigma_{h_i}^2\right)\sigma_d^2\mathbf{I}_{MN}. \tag{67}$$

Equality (a) follows from Lemma 3 and (65). Substituting (67) and (16) in (66), we get $\mathbb{E}\left\{\tilde{\mathbf{w}}_d\tilde{\mathbf{w}}_d^H\right\} = \left(\left(\sum_{i=1}^{Q}\sigma_{h_i}^2\right)\sigma_d^2 + \sigma_w^2\right)\mathbf{I}_{MN}$.

APPENDIX C

The covariance matrix of the vector $\boldsymbol{\xi}_{\tilde{\mathbf{w}}}^{(n)}$ can be evaluated using (40) as

$$\mathbf{C}_{\boldsymbol{\xi}_{\tilde{\mathbf{w}}}}^{(n)} \overset{(b)}{=} \mathbb{E}\left\{\mathbf{\Xi}_{\mathbf{x}_d}^{(n)}\mathbf{h}\mathbf{h}^H\left(\mathbf{\Xi}_{\mathbf{x}_d}^{(n)}\right)^H\right\} + \mathbb{E}\left\{\tilde{\mathbf{w}}\tilde{\mathbf{w}}^H\right\}, \tag{68}$$

where the equality (b) holds since the data symbols, channel parameters and noise samples are statistically independent. Evaluating $\mathbb{E}\left\{\mathbf{\Xi}_{\mathbf{x}_d}^{(n)}\mathbf{h}\mathbf{h}^H\left(\mathbf{\Xi}_{\mathbf{x}_d}^{(n)}\right)^H\right\}$ using Lemma-1, we get

$$\mathbb{E}\left\{\mathbf{\Xi}_{\mathbf{x}_d}^{(n)}\mathbf{h}\mathbf{h}^H\left(\mathbf{\Xi}_{\mathbf{x}_d}^{(n)}\right)^H\right\} = \mathbb{E}\left\{\mathbf{\Omega}_d\mathbf{C}_{\mathbf{h}}\mathbf{\Omega}_d^H\right\} + \mathbb{E}\left\{\widehat{\mathbf{\Omega}}_d^{(n)}\mathbf{C}_{\mathbf{h}}\left(\widehat{\mathbf{\Omega}}_d^{(n)}\right)^H\right\}. \tag{69}$$

Substituting (67) in (69), we get $\mathbb{E}\{\mathbf{\Xi}_{\mathbf{x}_d}^{(n)}\mathbf{h}\mathbf{h}^H(\mathbf{\Xi}_{\mathbf{x}_d}^{(n)})^H\} = 2(\sum_{i=1}^{Q}\sigma_{h_i}^2)\sigma_d^2\mathbf{I}_{MN}$. Finally, substitution of this result and (16) in (68) yields the desired result in (42).

APPENDIX D

We have by using (13)

$$\text{Tr}\left(\mathbf{\Omega}_p\mathbf{\Omega}_p^H\right) = \sum_{i=1}^{Q}\text{Tr}\left(\mathbf{\Gamma}_i\mathbf{x}_p\mathbf{x}_p^H\mathbf{\Gamma}_i^H\right) = QMN\sigma_p^2. \tag{70}$$

We next use the following result from [32]: For any positive definite matrix $\mathbf{A} \in \mathbb{C}^{m\times m}$, we have $\text{Tr}\left(\mathbf{A}^{-1}\right) \geq \frac{m^2}{\text{Tr}(\mathbf{A})}$. Note that the equality holds if the matrix $\mathbf{A}$ is a scaled identity matrix [32]. Using $\mathbf{A}^{-1} = \mathbf{\Sigma}_{\text{NI}}$, we get from (21) that $B_{h,\text{NI}} = \text{Tr}(\mathbf{\Sigma}_{\text{NI}}) \geq \frac{Q^2}{\text{Tr}\left(\mathbf{\Omega}_p^H\mathbf{C}_{\tilde{\mathbf{w}}_d}^{-1}\mathbf{\Omega}_p + \mathbf{C}_{\mathbf{h}}^{-1}\right)}$. As shown in Fig. 8(a), the matrix $\mathbf{\Sigma}_{\text{NI}}^{-1}$ is very close to a scaled identity matrix, which implies that $B_{h,\text{NI}} = \text{Tr}(\mathbf{\Sigma}_{\text{NI}}) \approx \frac{Q^2}{\text{Tr}\left(\mathbf{\Omega}_p^H\mathbf{C}_{\tilde{\mathbf{w}}_d}^{-1}\mathbf{\Omega}_p + \mathbf{C}_{\mathbf{h}}^{-1}\right)}$. To further validate it, we also plot in Fig. 8(b) the theoretical MSE $B_{h,\text{NI}}$ and its lower bound by varying the SNR. We see that the MSE lower bound is very close to its theoretical values. The denominator of $B_{h,\text{NI}}$ can be simplified as

$$\begin{aligned}\text{Tr}\left(\mathbf{\Omega}_p^H\mathbf{C}_{\tilde{\mathbf{w}}_d}^{-1}\mathbf{\Omega}_p + \mathbf{C}_{\mathbf{h}}^{-1}\right) &\overset{(a)}{=} \frac{\text{Tr}\left(\mathbf{\Omega}_p^H\mathbf{\Omega}_p\right)}{\sigma_h^2\sigma_d^2 + \sigma_w^2} + \sum_{i=1}^{Q}\frac{1}{\sigma_{h_i}^2} \\ &= \frac{QMN\sigma_p^2}{\sigma_h^2\sigma_d^2 + \sigma_w^2} + \tilde{\sigma}_h^2. \end{aligned}\tag{71}$$

Here (a) is due to (19) and Lemma 3, and $\tilde{\sigma}_h^2 = \sum_{i=1}^{Q}\left(1/\sigma_{h_i}^2\right)$. We get the result by substituting (71) in $B_{h,\text{NI}}$. We use this result to derive the lower bound of SINR in (60). Fig. 8(c) compares the average SINR and its lower bound in (60). We see that the average SINR always below the simulated SINR.

APPENDIX E

Equation (2) of [8], similar to (51), can be written as

$$\begin{aligned} y[l,k] = &\sum_{i=1}^{Q}\hat{h}_{\text{EP},i}\alpha_i(l,k)\,x_{d,\text{EP}}\left[(l-l_i)_M,(k-k_i)_N\right] \\ &+\sum_{i=1}^{Q}\left(h_i - \hat{h}_{\text{EP},i}\right)\alpha_i(l,k)\,x_{d,\text{EP}}\left[(l-l_i)_M,(k-k_i)_N\right] \\ &+\tilde{w}[l,k] = \tilde{\mathbf{x}}_{d,\text{EP}}^T\hat{\mathbf{h}}_{\text{EP}}^{\alpha} + \tilde{v}[l,k]. \end{aligned}\tag{72}$$

Here $\hat{h}_{\text{EP},i}\alpha_i$ is the MMSE channel estimate corresponding to the ith delay-Doppler path, which is obtained using the EP scheme proposed in [8]. The noise plus interference term $\tilde{v}\left[l,k\right] = \tilde{\mathbf{x}}_d^T\tilde{\mathbf{h}}_{\text{EP}}^{\alpha} + \tilde{w}\left[l,k\right]$. Here, the ith element of $\hat{\mathbf{h}}_{\text{EP}}^{\alpha} \in \mathbb{C}^{Q\times 1}$ and $\tilde{\mathbf{h}}_{\text{EP}}^{\alpha} \in \mathbb{C}^{Q\times 1}$, are $\hat{h}_{\text{EP},i}\alpha_i\left(l,k\right)$ and $\left(h_i - \hat{h}_{\text{EP},i}\right)\alpha_i\left(l,k\right)$, respectively. The SINR of the (l,k)th received symbol in the delay-Doppler domain can be formulated from the above expression as $\text{SINR}_{\text{EP},l,k} = \frac{\mathbb{E}\left\{\left|\tilde{\mathbf{x}}_{d,\text{EP}}^T\hat{\mathbf{h}}_{\text{EP}}^{\alpha}\right|^2\right\}}{\mathbb{E}\left\{|\tilde{v}[l,k]|^2\right\}}$. We first simplify the numerator of this SINR as

$$\mathbb{E}\left\{\left|\tilde{\mathbf{x}}_{d,\text{EP}}^T\hat{\mathbf{h}}_{\text{EP}}^{\alpha}\right|^2\right\} \overset{(a)}{=} \sigma_{d,\text{EP}}^2\mathbb{E}\left\{\left\|\hat{\mathbf{h}}_{\text{EP}}\right\|^2\right\}. \tag{73}$$

Equality (a) is because $\mathbb{E}\left\{\tilde{\mathbf{x}}_{d,\text{EP}}^*\tilde{\mathbf{x}}_{d,\text{EP}}^T\right\} = \sigma_{d,\text{EP}}^2\mathbf{I}_Q$.The expression $\mathbb{E}\{\|\hat{\mathbf{h}}_{\text{EP}}\|^2\}$ is next evaluated using (54) as $\mathbb{E}\{\|\hat{\mathbf{h}}_{\text{EP}}\|^2\} = \sigma_h^2 - B_{h,\text{EP}}$, where $B_{h,\text{EP}}$ denotes the MSE of the MMSE channel estimator in EP-based design in [8], which is calculated as [26]: $B_{h,\text{EP}} = \sum_{i=1}^{Q}\frac{\sigma_w^2\sigma_{h_i}^2}{\sigma_{p,\text{EP}}^2\sigma_{h_i}^2+\sigma_w^2}$. The numerator of $\text{SINR}_{\text{EP},l,k}$ expression is simplified using the above equation as $\mathbb{E}\{|\tilde{\mathbf{x}}_{d,\text{EP}}^T\hat{\mathbf{h}}_{\text{EP}}^{\alpha}|^2\} = \sigma_{d,\text{EP}}^2\left(\sigma_h^2 - B_{h,\text{EP}}\right)$. The denominator of the SINR in is given as $\mathbb{E}\left\{|\tilde{v}\left[l,k\right]|^2\right\} = \sigma_{d,\text{EP}}^2B_{h,\text{EP}} + \sigma_w^2$. Substituting these terms in SINR, yields the desired expression below (62).

REFERENCES

[1] T. Wang, J. G. Proakis, E. Masry, and J. R. Zeidler, "Performance degradation of OFDM systems due to Doppler spreading," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1422–1432, Jun. 2006.

[2] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "On the diversity of uncoded OTFS modulation in doubly-dispersive channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3049–3063, Jun. 2019.

[3] G. D. Surabhi and A. Chockalingam, "Low-complexity linear equalization for OTFS modulation," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 330–334, Feb. 2020.

[4] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501–6515, Oct. 2018.

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

[6] A. Fish, S. Gurevich, R. Hadani, A. M. Sayeed, and O. Schwartz, "Delay-Doppler channel estimation in almost linear complexity," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7632–7644, Nov. 2013.

[7] M. Kollengode Ramachandran and A. Chockalingam, "MIMO-OTFS in high-Doppler fading channels: Signal detection and channel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 206–212.

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

[9] O. K. Rasheed, G. D. Surabhi, and A. Chockalingam, "Sparse delay-Doppler channel estimation in rapidly time-varying channels for multi-user OTFS on the uplink," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, Antwerp, Belgium, May 2020, pp. 1–5.

[10] R. Carrasco-Alvarez, R. Parra-Michel, A. G. Orozco-Lugo, and J. K. Tugnait, "Time-varying channel estimation using two-dimensional channel orthogonalization and superimposed training," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4439–4443, Aug. 2012.

[11] J. K. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Lett.*, vol. 7, no. 9, pp. 413–415, Sep. 2003.

[12] S. He, J. K. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3007–3021, Jun. 2007.

[13] S. He and J. K. Tugnait, "On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3214–3228, Jul. 2008.

[14] N. N. Tran, D. H. Pham, H. D. Tuan, and H. H. Nguyen, "Orthogonal affine precoding and decoding for channel estimation and source detection in MIMO frequency-selective fading channels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1151–1162, Mar. 2009.

[15] N. Tran, H. Tuan, and H. Nguyen, "Superimposed training designs for spatially correlated MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 876–880, Mar. 2010.

[16] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 226–229, Mar. 2005.

[17] O. Longoria-Gandara *et al.*, "Iterative MIMO detection and channel estimation using joint superimposed and pilot-aided training," *Mobile Inf. Syst.*, vol. 2016, pp. 3723862:1–3723862:11, Jan. 2016.

[18] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Tech.*, vol. 68, no. 1, pp. 957–961, Jan. 2019.

[19] Y. Ge, Q. Deng, P. C. Ching, and Z. Ding, "Receiver design for OTFS with a fractionally spaced sampling approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4072–4086, Jul. 2021.

[20] K. B. Petersen and M. S. Pedersen. (Nov. 2012). *The Matrix Cookbook*. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php

[21] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[22] G. Matz, "On non-WSSUS wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2465–2478, Sep. 2005.

[23] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.

[24] P. Singh, H. B. Mishra, A. K. Jagannatham, and K. Vasudevan, "Semi-blind, training, and data-aided channel estimation schemes for MIMO-FBMC-OQAM systems," *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4668–4682, Sep. 2019.

[25] P. Singh, R. Budhiraja, and K. Vasudevan, "Probability of error in MMSE detection for MIMO-FBMC-OQAM systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8196–8200, Aug. 2019.

[26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[27] P. Raviteja, K. T. Phan, Q. Jin, Y. Hong, and E. Viterbo, "Low-complexity iterative detection for orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Barcelona, Spain, Apr. 2018, pp. 1–6.

[28] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-complexity detection in large-dimension MIMO-ISI channels using graphical models," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1497–1511, Dec. 2011.

[29] Y. Ge, Q. Deng, P. C. Ching, and Z. Ding, "OTFS signaling for uplink NOMA of heterogeneous mobility users," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3147–3161, May 2021.

[30] A. L. Yuille, "CCCP algorithms to minimize the Bethe and Kikuchi free energies: Convergent alternatives to belief propagation," *Neural Comput.*, vol. 14, no. 7, pp. 1691–1722, 2002.

[31] M. Pretti, "A message-passing algorithm with damping," *J. Stat. Mech., Theory Exp.*, vol. 2005, no. 11, Nov. 2005, Art. no. P11008.

[32] Z. Ulukök and R. Türkmen, "On some matrix trace inequalities," *J. Inequalities Appl.*, vol. 2010, pp. 1–8, Dec. 2010.

Himanshu B. Mishra (Member, IEEE) received the M.Tech. degree in electronics and communication engineering from the National Institute of Technology, Rourkela, India, in 2012, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology Kanpur, Kanpur, India, in 2016. He is currently working as an Assistant Professor with the Electronics Engineering Department, Indian Institute of Technology (Indian School of Mines) Dhanbad, India. His current research interests include signal processing, next generation of wireless communication systems, estimation, and detection theory.

Prem Singh received the B.Tech. degree in electronics and communication engineering from U.P. Technical University, Lucknow, in 2006, and the M.Tech. and Ph.D. degrees in electrical engineering from the Indian Institute of Technology Kanpur, India, in 2011 and 2020, respectively. He worked as a Project Executive Officer on the Indigenous 5G TestBed Project, where he developed hardware and software algorithms for an end-to-end 3GPP compliant 5G-NR testbed. Since July 2021, he has been working as an Assistant Professor with the International Institute of Information Technology (IIIT) Bangalore, India. His research interests include parameter estimation and transceiver design for 5G and next-generation wireless technologies, including filter bank multicarrier (FBMC), orthogonal time-frequency space (OTFS), and massive MIMO and millimeter-wave. His interests also include designing practical 4G/5G wireless systems using 3GPP standards.

Abhishek K. Prasad received the B.E. degree in electronics and telecommunication engineering from Chhatrapati Shivaji Institute of Technology, Durg, India, in 2016, and the M.Tech. degree in communication and signal processing from IIT (ISM), Dhanbad. His research interests include MIMO systems, deep learning, and orthogonal time frequency space modulation.

Rohit Budhiraja received the M.S. degree in electrical engineering and the Ph.D. degree from IIT Madras in 2004 and 2015, respectively. From 2004 to 2011, he worked for two start-ups, where he designed both hardware and software algorithms, from scratch, for physical layer processing of WiMAX- and LTE-based cellular systems. He is currently an Assistant Professor with IIT Kanpur, where he is also leading an effort to design a 5G research testbed. His current research interests include design of energy-efficient transceiver algorithms for 5G massive MIMO and full-duplex systems, robust precoder design for wireless relaying, machine learning methods for channel estimation in mm-wave systems, and spatial modulation system design. His paper was shortlisted as one of the finalists for the Best Student Paper Awards at the IEEE International Conference on Signal Processing and Communications, Bengaluru, India, in 2014. He also received IIT Madras Research Award for the quality and quantity of research work done in the Ph.D., the Early Career Research Award, and the Teaching Excellence Certificate at IIT Kanpur.