Optimal pilot design for OTFS in linear time-varying channels

van der Werf, Ids; Heusdens, Richard; Hendriks, Richard C.; Leus, Geert

DOI
10.1016/j.sigpro.2025.110223

Publication date
2026

Document Version
Final published version

Published in
Signal Processing

Citation (APA)
van der Werf, I., Heusdens, R., Hendriks, R. C., & Leus, G. (2026). Optimal pilot design for OTFS in linear time-varying channels. *Signal Processing*, *239*, Article 110223. https://doi.org/10.1016/j.sigpro.2025.110223

Signal Processing 239 (2026) 110223

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Optimal pilot design for OTFS in linear time-varying channels

Ids van der Werf [a,*], Richard Heusdens [a,b], Richard C. Hendriks [a], Geert Leus [a]

[a] Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
[b] Netherlands Defence Academy (NLDA), Het Nieuwe Diep 8, 1781 AC Den Helder, The Netherlands

ARTICLE INFO

Keywords:
Doubly selective channels
Optimal pilot design
Modulation
OTFS
OSDM

ABSTRACT

This paper investigates the positioning of the pilot symbols, as well as the power distribution between the pilot and the communication symbols for the orthogonal time frequency space (OTFS) modulation scheme. We analyze the pilot placements that minimize the mean squared error (MSE) in estimating the channel taps. This allows us to identify two new pilot allocations for OTFS that save approximately 50% of the pilot overhead compared to existing allocations. In addition, we optimize the average channel capacity by adjusting the power distribution. We show that this leads to a significant increase in average capacity. The results provide valuable guidance for designing the OTFS parameters to achieve maximum capacity. Numerical simulations are performed to validate the findings.

1. Introduction

To address the growing need for data, it is important to judiciously consider the design of the modulation scheme of a communication system. In recent wireless communication standards, orthogonal frequency division multiplexing (OFDM) has been widely adopted as the preferred choice [1]. However, a known drawback of OFDM is its sensitivity to Doppler shifts, which cause intercarrier interference. In contrast, single-carrier modulation (SCM) is more robust to Doppler effects but suffers from intersymbol interference due to multipath delays. As communication scenarios increasingly involve dynamic environments with both delay and Doppler spread, known as doubly selective channels, new modulation schemes have been proposed to improve robustness against such conditions. Lately, orthogonal time frequency space (OTFS) [2,3] modulation has received a lot of attention. OTFS defines symbols in the DD domain and then transforms the signal into the time domain using the Zak transform [4]. OTFS has been shown to have improved performance compared to OFDM [3–6], which is attributed to the fact that OTFS can benefit from diversity in both time and frequency. Throughout this paper, we will use the name OTFS; however, it is worth mentioning that the (older) modulation schemes vector OFDM (V-OFDM) [7], asymmetric OFDM (A-OFDM) [8] and orthogonal sequence division multiplexing (OSDM) [5] were shown to be equivalent to OTFS [9–11]. Therefore, our analysis and conclusions also apply to these modulation schemes. We will in general use the abbreviation OTFS to denote the schemes V-OFDM, A-OFDM, OSDM and OTFS, except when referencing papers written specifically for one of these modulation schemes.

To get the most out of the OTFS modulation, careful design of the pilot symbols is required. Although many pilot allocations have been proposed for OTFS (see, e.g., [12–15]), a comparison and (mathematical) analysis of the optimality of these allocations is lacking. We therefore start with an overview of the work on optimal pilot design for linear time-invariant (LTI) and linear time-variant (LTV) channels, and of the work on specific pilot designs for the related modulation schemes (Section 3). After this, we continue with our proposed work which has the following main contributions:

- We reformulate the effect of an LTV channel on OTFS modulation (Section 4). This reformulation provides insight into how to choose the OTFS parameters N and M to design two new pilot allocations that reduce the pilot overhead by approximately 50%. These pilot allocations also achieve the MMSE on the estimation of the channel taps (Sections 5.2 and 5.3) .
- We derive a closed form expression for a lower bound on the channel capacity. The lower bound can be maximized with respect to the power distribution between the pilot and the communication symbols. We show that our new pilot design in conjunction with the optimal power distribution can significantly improve the system's achievable data rate.
- Finally, our findings can be used as guidance for designing the OTFS parameters based on the delay and Doppler spread to increase the channel capacity.

* Corresponding author.
E-mail addresses: i.vanderwerf@tudelft.nl (I. van der Werf), r.heusdens@tudelft.nl, r.heusdens@mindef.nl (R. Heusdens), r.c.hendriks@tudelft.nl (R.C. Hendriks), g.j.t.leus@tudelft.nl (G. Leus).

https://doi.org/10.1016/j.sigpro.2025.110223
Received 15 November 2024; Received in revised form 25 July 2025; Accepted 28 July 2025
Available online 6 August 2025

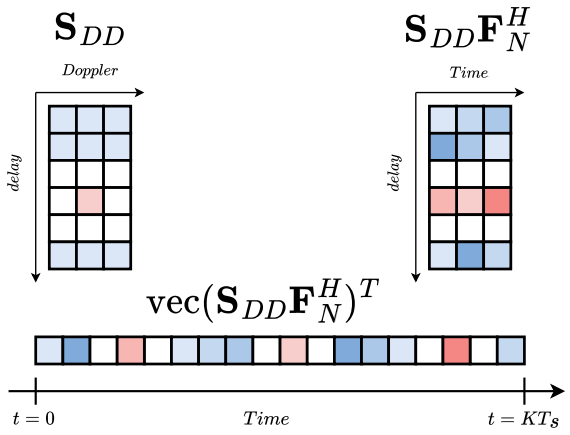


Fig. 1. Visualization of the OTFS transmitted signal, $\{K, N, M\} = \{18, 3, 6\}$.

Notation. In what follows, $\otimes$, $\odot$, $\circ$ and $*$ are used to denote the Kronecker product, the Khatri–Rao product, the element-wise multiplication and the linear convolution, respectively. Let K be a positive integer and let $\mathbf{P}_K$ denote a $K \times K$ cyclic permutation matrix given by,

$$\mathbf{P}_K = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}.$$

Further, let $\mathbf{P}_K^l$ denote the l'th power of $\mathbf{P}_K$ for some integer l. Let N and M be positive integers. If $K = NM$, we can rewrite the cyclic permutation matrix as $\mathbf{P}_K = \mathbf{I}_N \otimes \mathbf{L}_M + \mathbf{P}_N \otimes \mathbf{U}_M$, where $\mathbf{L}_M$ is a cyclic permutation matrix of size M except for the top right element, which is zero, and $\mathbf{U}_M$ is a zero matrix of size M except for the top right element, which is one (thus $\mathbf{P}_M = \mathbf{U}_M + \mathbf{L}_M$). Similarly, provided $l < M$, we can write $\mathbf{P}_K^l = \mathbf{I}_N \otimes \mathbf{L}_M^{(l)} + \mathbf{P}_N \otimes \mathbf{U}_M^{(l)}$, where the matrices $\mathbf{L}_M^{(l)}$ and $\mathbf{U}_M^{(l)}$ denote the lower and upper part of $\mathbf{P}_M^l$, respectively. Lastly, we define the matrix $\mathbf{\Lambda}_K = \mathrm{diag}(e^{j2\pi\, 0/K}, \dots, e^{j2\pi(K-1)/K})$.

Before we elaborate on the related work, we start by a brief discussion of the signal model in the next section.

2. Signal model

In this section, we discuss the OTFS modulation scheme and the LTV channel model.

2.1. OTFS transmitter

For OTFS, the transmitter characterizes $K = NM$ symbols in the DD domain (M by N symbols in the delay and Doppler dimension respectively), collected in the matrix $\mathbf{S}_{\mathrm{DD}} \in \mathbb{C}^{M \times N}$, after which the symbols are consecutively transformed to the time-frequency and the time domain by the inverse symplectic finite Fourier transform (ISFFT) and the Heisenberg transform [3], respectively. Assuming a rectangular transmit pulse is used [16], the transmitted signal in discrete-time baseband is given by [9,11],

$$\mathbf{x} = \mathrm{vec}(\mathbf{S}_{\mathrm{DD}}\mathbf{F}_N^H) = (\mathbf{F}_N^H \otimes \mathbf{I}_M)\mathbf{s}, \tag{1}$$

where $\mathbf{F}_N^H$ denotes the inverse discrete Fourier transform matrix, $\mathbf{I}_M$ is an identity matrix of size M and $\mathbf{s} = \mathrm{vec}(\mathbf{S}_{\mathrm{DD}})$. In Fig. 1, a visual representation of the operation at the transmitter, i.e. (1), is shown.

2.2. Channel model

Let $r(k)$, $h(k,l)$, $x(k)$, and $n(k)$ denote the discrete-time received signal, channel impulse response, transmitted signal, and additive channel noise, respectively. We assume Nyquist-rate sampling, such that the signals are observed at time instances $t_k = kT_s$ and time delays $\tau_l = lT_s$, with T_s denoting the sampling period. The received signal is then described by

$$r(k) = \sum_{l=0}^{L} h(k,l)x(k-l) + n(k).$$

Note that an LTI channel is subsumed by this model; if $h(k,l) = h(l)$, the model is time-invariant. Suppose we collect a total of K samples over time, then estimating all the $K(L+1)$ channel coefficients is an ill-posed problem, as in general $K < K(L+1)$. In order to decrease the number of coefficients that need to be estimated, the basis expansion model (BEM) was introduced [17].

The BEM is a model that represents how the channel changes over time. It approximates the varying channel taps by expressing them using a lower order basis of size $Q+1$ (with $Q+1 < K$),

$$h(k,l) = \sum_{q=-Q/2}^{Q/2} c_{l,q} b_q(k). \tag{2}$$

Here $b_q(k) \in \mathbb{C}$ is the function representing the basis and $c_{l,q} \in \mathbb{C}$, for $l = 0, \dots, L$ and $q = -Q/2, \dots, Q/2$, are the BEM channel coefficients. In general $Q + 1 \ll K$, and thus the total number of coefficients to be estimated is now lower, casting the problem well posed if $K \geq (Q+1)(L+1)$.

The BEM can be used with different bases and thus functions $b_q(k)$. Over the past decades, many various bases have been proposed. The most well-known is the complex exponential BEM (CE-BEM) [17], for which $b_q(k) = e^{j\omega_q k}$, $\omega_q = 2\pi q/K$. Note that the CE-BEM is quite comprehensible; the coefficients of the CE-BEM each represent a unique pair of Doppler shift ω_q and time delay τ_l. Some other popular BEMs are the generalized CE-BEM (GCE-BEM) [18], for which $b_q(k) = e^{j\omega_q k}$, $\omega_q = 2\pi q/(KR)$, $R \geq 1$, the polynomial BEM (P-BEM) [19], the discrete Karhuen–Loève BEM (DKL-BEM) [20] and the discrete prolate spheroidal BEM (DPS-BEM) [21]. Of course, the modeling accuracy of each choice differs per application.

Remark 1. In the OTFS literature, the so-called delay-Doppler channel is often used instead of a BEM. However, assuming that the time delays and Doppler shifts fall on the Nyquist grid, the delay-Doppler channel coincides with the (conventional) CE-BEM.

In this work, we adopt the CE-BEM due to its intuitive structure and widespread use in the literature. Assuming that we use a cyclic prefix and that we model the channel with additive Gaussian noise with zero mean and covariance $\mathbf{R}_{\mathbf{n}}$, the received signal is given by

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n} = \left(\sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \mathbf{\Lambda}_K^q \mathbf{P}_K^l \right) \mathbf{x} + \mathbf{n},$$

where the vectors $\mathbf{r}$, $\mathbf{x}$ and $\mathbf{n}$ collect the samples of $r(k)$, $x(k)$ and $n(k)$, respectively, for $k = 0, 1, \dots, K-1$. Note that when $Q = 0$, the LTV channel becomes time-invariant and thus coincides with an LTI channel.

Remark 2. The parameters L and Q denote the maximum time delay and Doppler shift in the channel, i.e. $\tau_{\max} = LT_s$ and $f_{\max} = Q/(2KT_s)$, respectively. Although specific channel taps vary, these parameters remain constant over a long time. Upper bounds for these values are typically well-known for particular environments, allowing L and Q to be set to these bounds before communication.

Table 1
Comparison of pilot overhead across different modulation schemes and channel models. The minimum pilot overhead values (for all L and Q) are indicated in bold.

Channel	Paper	Modulation	MMSE optimal	Pilot overhead K_p
LTI	[26,27]	SCM	Yes	$2L+1$
	[28,29]	OFDM	Yes	$\mathbf{L+1}$
	[5]	OSDM	Yes	M
LTV	[30–32]	SCM	Yes	$(2L+1)(Q+1)$
	[31,32]	OFDM	Yes	$(L+1)(2Q+1)$
	[12]	OTFS	Yes	$(2L+1)(2Q+1)$
	[13]	OTFS	Yes	$(2L+1)N$
	[15]	OTFS	No	$(2L+1)N$
	[33]	OSDM	No	$M(2Q+1)$
	This work	OTFS	Yes	$\mathbf{\min\{(L+1)(2Q+1),(2L+1)(Q+1)\}}$

2.3. OTFS receiver

At the receiver, the demodulated signal is given by,

$$\mathbf{y} = (\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{r}.$$

To obtain the signal received in the DD domain, the vector $\mathbf{y}$ is reshaped into an $M \times N$ matrix, i.e., $\mathbf{Y} = \text{vec}^{-1}(\mathbf{y})$. Typically, we choose $N \geq Q+1$ and $M \geq L+1$ to ensure that the channel can be properly resolved in both delay and Doppler dimensions, as will become clear shortly. The literature on OTFS receivers is rich; see, e.g. [6,22–24]. However, the most straightforward way is to apply a linear MMSE estimator to estimated the transmitted signal.

3. Related work on pilot design

In this section, we examine the existing literature on the development of (optimal) pilot designs.

Designing a pilot signal can be divided into two main components: pilot allocation and power distribution, which we will treat separately. Due to its practical relevance, we also briefly discuss pilot design in relation to peak-to-average power ratio (PAPR). We will focus exclusively on designs where pilot and communication symbols do not overlap, an assumption that underpins the remainder of this paper. For studies on superimposed pilot designs, we refer readers to [25] and references therein.

Throughout the first two subsections, we discuss various pilot allocations for different modulation schemes and different channel models. A comparison of the pilot overhead of the discussed allocations is provided in Table 1.

3.1. Pilot allocation for SCM and OFDM

In this subsection, we discuss the existing pilot allocation strategies for SCM and OFDM. First, we examine literature on LTI channels, followed by a discussion on LTV channels. Note that OTFS reduces to SCM when $N = 1$ and $M = K$, and to OFDM when $N = K$ and $M = 1$.

Research on pilot allocation dates back to the late 1990s, considering metrics such as MSE, channel capacity, and the Cramér-Rao bound (CRB), which often led to similar optimal designs. For LTI channels (i.e., $Q = 0$), a BEM is unnecessary, and typically, the system is overdetermined ($K > L+1$). In the SCM setting, [26] showed that clustering at least $K_\text{p} = 2L+1$ pilot symbols, zeroing the leading and trailing L symbols, yields capacity-optimal results. This structure was also found optimal under the CRB on the channel tap estimator in [27], reflecting the equivalence of MSE and CRB under additive white Gaussian noise.

For OFDM, optimal pilot placement involves only $K_\text{p} = L+1$ equispaced tones. This was shown to minimize the MSE on the channel taps [28] and maximize a capacity lower bound [26].

Unlike the LTI case, studies on LTV channels generally avoid assuming a specific modulation scheme. However, pilot and data symbols are often separated in either time or frequency, naturally aligning with SCM and OFDM, respectively. Similar to this work, existing works on pilot design typically model LTV channels using a CE-BEM.

In [30], assuming time-separated pilots and data, equispaced and equipowered pilot clusters of length $2L+1$ (with L leading and trailing zeros) are shown to maximize a lower bound on average capacity. A total of $Q+1$ such clusters are placed equidistantly in time, reflecting the Nyquist criterion for Doppler sampling (sampling at twice the maximum Doppler frequency, $Q/2$ in normalized baseband). This leads to a total pilot overhead of $K_\text{p} = (2L+1)(Q+1)$. This pilot design effectively probes the channel $Q+1$ times using impulse-like signals, as visualized in Fig. 2(a). Notably, for $Q = 0$, this reduces to the optimal SCM-based design for LTI channels.

Independently, [31] proposed three MSE-optimal pilot designs. The first mirrors [30] (time-separated). The second uses frequency separation, inserting guard bands to prevent pilot-data overlap post-channel. As visualized in Fig. 2(b), it probes the channel $L+1$ times in frequency, using equispaced, equipowered pilots surrounded by Q zeros, generalizing the OFDM-based design when $Q = 0$. This second design has a pilot overhead of $K_\text{p} = (L+1)(2Q+1)$. The third, based on linear chirps [31,32], has lower spectral efficiency and is therefore not considered further here.

Interestingly, frequency separation outperforms time separation when $L > Q$, and vice versa when $L < Q$ in terms of pilot overhead K_p [31,32].[1] Moreover, for both cases, it suffices that pilot clusters are equispaced; their absolute positions in time or frequency are irrelevant for MSE performance.

Finally, [34,35] address scenarios where the Doppler spread is overestimated. In such cases, changing zero-valued pilot symbols into nonzero-valued pilot symbols can improve estimation, though this yields sub-optimal designs because the pilot/data structure is not fully adapted to the actual channel.

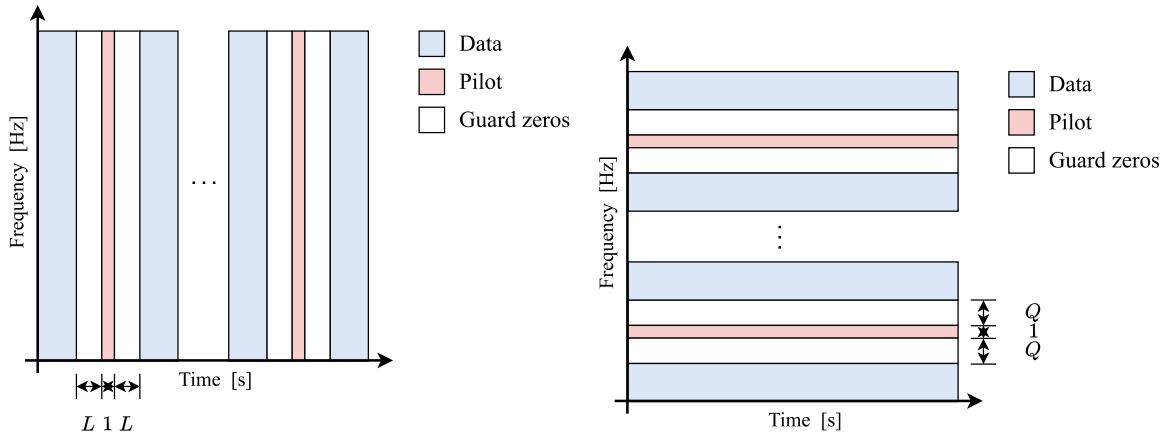
3.2. Pilot allocation for OSDM and OTFS

In this section, we discuss the most relevant existing pilot allocations for OSDM and OTFS. Note that since the two modulation schemes are equivalent [10,11], pilot designs proposed for one scheme can be directly applied to the other.

In [5], the authors approach the (OSDM) modulation from a sequence-based perspective. They propose allocating one out of the N available sequences – each of length M – entirely to pilot symbols, employing a shift-orthogonal sequence. When applied to an LTI channel, no guard intervals are inserted between pilot and communication symbols in this design. This idea was later extended to a modulation scheme termed Doppler-resilient OSDM (D-OSDM) [33], which introduces guard intervals to enhance robustness in LTV channels. A visual representation of this latter scheme is provided in Fig. 3.

In [12,13], two pilot allocations are proposed for OTFS. The first approach places a single pilot in the DD domain, surrounded by zeros

[1] Note that when $L > Q$ we have $(2L+1)(Q+1) > (L+1)(2Q+1)$ and that when $L < Q$ we have $(2L+1)(Q+1) < (L+1)(2Q+1)$.



(a) Separation in time, $Q+1$ pilot cluster repetitions.

(b) Separation in frequency, $L+1$ pilot cluster repetitions.

Fig. 2. Two methods, as proposed by (a) [30–32] and (b) [31,32].

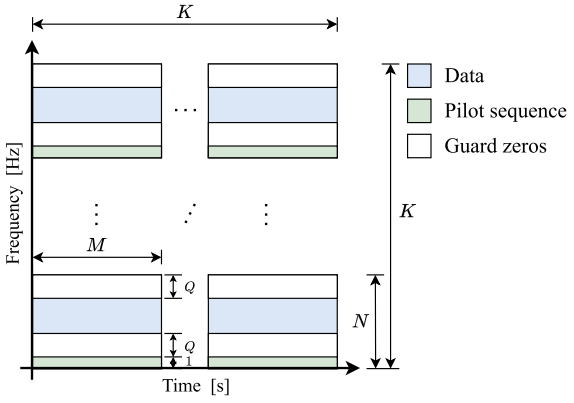


Fig. 3. Time-frequency plot for D-OSDM [33].

to prevent interference with communication symbols at the receiver. The pilot overhead is thus $K_p = (2Q+1)(2L+1)$, as visualized in Fig. 4(a). The second scheme accounts for fractional Doppler shifts, which occur when the actual Doppler shifts do not align perfectly with the grid. In this case, all neighboring Doppler bins are set to zero to mitigate interference. The pilot overhead of this second approach is therefore $K_p = N(2L+1)$, as illustrated in Fig. 4(b). Note that both pilot allocations have a higher overhead than the allocations proposed for SCM and OFDM by [30–32]. To reduce the high pilot overhead of the fractional Doppler scheme in [12], a different method is introduced in [24]. The work adopts the GCE-BEM and proposes a two-step estimation process. In the first step, the BEM order Q is fixed and $R = 1$ is used. In the second step, Q is increased by adjusting R from 1 to 2. Both the channel coefficients and communication symbols are iteratively refined, incorporating initial demodulation results. By starting with a reduced BEM order, pilot overhead is significantly lowered compared to the one-step method in [12], while improving estimation accuracy through model refinement. However, besides the higher pilot overhead, the pilot designs for OTFS by [9,12,13,24] did not provide insight into the design by discussing the pilot signal in the time-frequency domain, in contrast to the works on SCM and OFDM.

Interestingly, it is possible to characterize pilot designs for OTFS in the time-frequency domain. In our earlier work [11], we identified the equivalence of OTFS and OSDM, showing that symbols in the DD domain appear as repeated structures in the time-frequency domain. More specifically, *each OTFS symbol in the DD domain is repeated N times in time and M times in frequency.*

This phenomenon, also appearing in Fig. 1 (where one red symbol in the DD domain is repeated $N = 3$ times in time), is further visualized in Fig. 4(c), which shows the time-frequency plot of the delay-Doppler signal of Fig. 4(b).

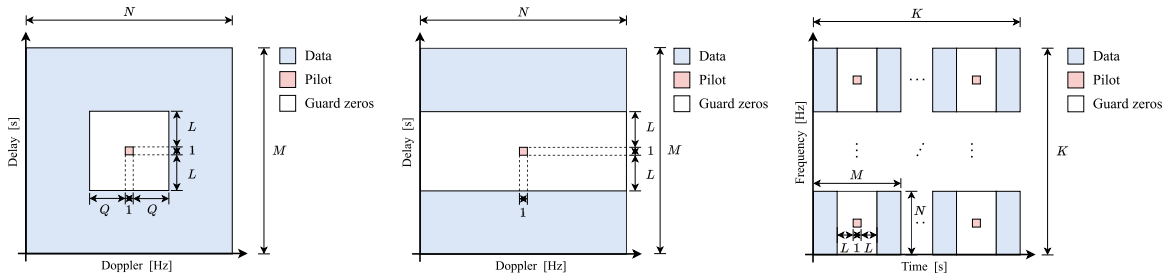
Note that Fig. 4(c), where the signal exhibits N repeated pilot clusters in time, each with L leading and trailing symbols set to zero, closely resembles the time-separated pilot design in [30–32], shown in Fig. 2(a), where $Q+1$ repeated pilot clusters also have L leading and trailing symbols equal to zero. This raises the question: how many repetitions are optimal?

Given $L+1$ delay taps and $Q+1$ Doppler taps, we require at least $Q+1$ time samples and $L+1$ frequency samples of the pilot signal. Since OTFS exhibits N repetitions in time and M in frequency, this requirement can be met by imposing $N \geq Q+1$ and $M \geq L+1$, and using one pilot symbol in the DD domain. Alternatively, it is also possible to meet the sampling requirement for $N < Q+1$ and/or $M < L+1$, however, this requires multiple equispaced pilots in the delay and/or Doppler dimension, respectively, each with appropriate guard bands. This, however, complicates the pilot design, does not reduce de pilot overhead, which is why we will not further investigate this option. With $K = NM$, we can flexibly choose either N or M (as long as K/N or K/M remains an integer). In this work, we will exploit this flexibility to reduce the pilot overhead by approximately 50%, from $K_p = (2L+1)(2Q+1)$ to $K_p = \min\{(L+1)(2Q+1), (2L+1)(Q+1)\}$.

Remark 3. A notable difference between SCM and OTFS is that the repetitions of the OTFS pilot impulses in time have different phases. As a result, the pilots are not active on all frequencies, whereas the frequency characteristics of pilot clusters in [30–32] was not explicitly analyzed. The authors of [30] did, however, observe that periodic repetition in time implicitly induces repetition in frequency.[2]

Furthermore, while [30–32] do not specify *how* communication symbols are allocated, OTFS inherently provides a form of precoding for these symbols. In fact, [30] does not consider the structure or coding of data symbols at all.

[2] Quote: *"We wish to show in this subsection that our optimal PSAM (...) enables 2-D sampling and estimation of our time-frequency selective channel. Intuitively thinking, the Kronecker deltas (...) surrounded by zero-guards implement time-domain sampling with pilot symbols; furthermore, the fact that these deltas are periodically inserted implies that they are also equivalent to Kronecker deltas in the frequency-domain and thus serve as pilot tones as well".*



(a) Integer Doppler case [12], [13]. (b) Fractional Doppler case [12], [13]. (c) Time-frequency plot of (b).

Fig. 4. Pilot allocations in the delay-Doppler, (a) and (b), and in the time-frequency domain, (c).

3.3. Power distribution

For distributing power, one can consider how to allocate power across multiple pilot *symbols* (if one has multiple) and/or how to divide the total available transmitter power between the pilot *signal* and the communication *signal*.

For LTI channels, employing OFDM, [29] proved that $L+1$ pilot tones with equal energy is optimal with respect to the MSE on the channel taps. The optimal power distribution between pilot and communication signals was derived in [26,29]. These results were extended to stochastic channels in [36]. A similar derivation for LTV channels was provided by [26,29], yielding significant gains.

In most OTFS pilot designs, only a single pilot symbol is specified in the delay-Doppler domain, which focuses the concern on the power allocation between the pilot and the communication signal. While it is recognized that assigning different power levels to pilot and communication signals impacts the performance [9,12,13], no optimality analysis has been provided in these works. Additionally, to the best of the authors' knowledge, an optimal power distribution has (also) not been derived in the existing literature on OSDM.

This work aims to fill this gap and provides a closed-form expression to identify the optimal power allocation.

3.4. Pilot designs mitigating PAPR for OTFS

A PAPR analysis of OTFS is conducted in [37], revealing that increasing pilot power significantly raises the PAPR. This contradicts the claim in [12] that the "spread-spectrum nature of OTFS" allows for higher pilot power without PAPR increase. Unfortunately, the pilot structure details (e.g., guard size) are not fully disclosed in [37]. Moreover, while the optimal power distribution is found empirically, the lack of theoretical analysis limits the generalizability of the findings.

An alternative scheme aimed at PAPR reduction is proposed in [15]. Here, the pilot from [12] is modified by extending it along the delay axis using a Zadoff-Chu (ZC) sequence and inserting a cyclic prefix. This configuration, shown in Fig. 5, significantly reduces PAPR, albeit at the cost of increased BER.

It is worth noting that both [15,33] make use of shift-orthogonal sequences, yet they differ in their pilot-data separation strategy: [33] achieves separation in the frequency domain, while [15] does so in the time domain. This suggests that D-OSDM may exhibit improved PAPR performance as well, compared to the impulse pilot scheme for OTFS [13], although this comes at the cost of reduced BER performance.

Another approach is presented in [14], where multiple pilots are placed at the corners of the DD domain, each with surrounding guard symbols. However, the rationale for this placement is unclear, and although a comparison with [12] is included, the use of different estimators per pilot design render the comparison inconclusive.

4. OTFS modulation through the LTV channel

In this section, we explore the interaction of the OTFS modulation scheme with a CE-BEM channel with L temporal delays and Q Doppler shifts. It will be observed that the modulation scheme transforms the time-varying channel into a time-invariant channel in the DD domain. Furthermore, the received signal, in the DD domain, can be represented as a circularly shifted form of the transmitted signal (in the same domain).

If we receive K samples over time, where K is factored as $K = NM$, the received signal can be expressed as,

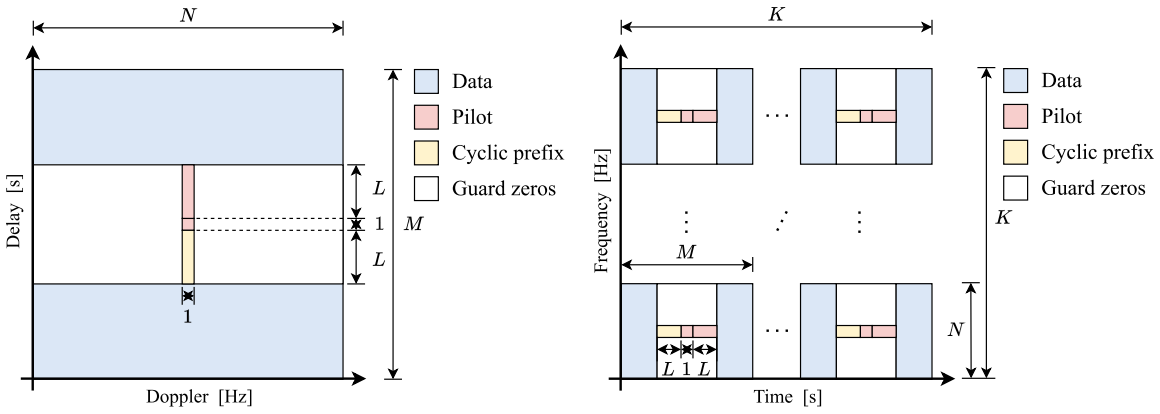
$$
\begin{aligned}
\mathbf{r} &= \mathbf{H}(\mathbf{F}_N^H \otimes \mathbf{I}_M)\mathbf{s} + \mathbf{n} \\
&= \left(\sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \boldsymbol{\Lambda}_K^q \mathbf{P}_K^l \right) (\mathbf{F}_N^H \otimes \mathbf{I}_M)\mathbf{s} + \mathbf{n} \\
&= \sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} (\boldsymbol{\Lambda}_N^q \otimes \boldsymbol{\Lambda}_M^{q/N}) \overbrace{\left[(\mathbf{I}_N \otimes \mathbf{L}_M^{(l)}) + (\mathbf{P}_N \otimes \mathbf{U}_M^{(l)}) \right]}^{\mathbf{P}_K^l} \\
&\quad \times (\mathbf{F}_N^H \otimes \mathbf{I}_M)\mathbf{s} + \mathbf{n} \\
&= \sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \left[(\boldsymbol{\Lambda}_N^q \mathbf{I}_N \otimes \boldsymbol{\Lambda}_M^{q/N} \mathbf{L}_M^{(l)}) + (\boldsymbol{\Lambda}_N^q \mathbf{P}_N \otimes \boldsymbol{\Lambda}_M^{q/N} \mathbf{U}_M^{(l)}) \right] \\
&\quad \times (\mathbf{F}_N^H \otimes \mathbf{I}_M)\mathbf{s} + \mathbf{n} \\
&= \sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \left[(\boldsymbol{\Lambda}_N^q \mathbf{F}_N^H \otimes \boldsymbol{\Lambda}_M^{q/N} \mathbf{L}_M^{(l)}) + (\boldsymbol{\Lambda}_N^q \mathbf{P}_N \mathbf{F}_N^H \otimes \boldsymbol{\Lambda}_M^{q/N} \mathbf{U}_M^{(l)}) \right] \mathbf{s} + \mathbf{n}.
\end{aligned}
\tag{3}
$$

At the receiver side we first demodulate the signal,

$$
\begin{aligned}
\mathbf{y} &= (\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{r} \\
&= \sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \left[(\mathbf{F}_N \boldsymbol{\Lambda}_N^q \mathbf{F}_N^H \otimes \boldsymbol{\Lambda}_M^{q/N} \mathbf{L}_M^{(l)}) + (\mathbf{F}_N \boldsymbol{\Lambda}_N^q \mathbf{P}_N \mathbf{F}_N^H \otimes \boldsymbol{\Lambda}_M^{q/N} \mathbf{U}_M^{(l)}) \right] \\
&\quad \times \mathbf{s} + (\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{n}.
\end{aligned}
\tag{4}
$$

Then, if the vector $\mathbf{y}$ is reshaped into an $M \times N$ matrix we can write,

$$
\begin{aligned}
\mathbf{Y} &= \mathrm{vec}^{-1}(\mathbf{y}) \\
&= \sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \boldsymbol{\Lambda}_M^{q/N} \left[\mathbf{L}_M^{(l)} \mathbf{S} \mathbf{F}_N^H \boldsymbol{\Lambda}_N^q \mathbf{F}_N + \mathbf{U}_M^{(l)} \mathbf{S} \mathbf{F}_N^H \mathbf{P}_N \boldsymbol{\Lambda}_N^q \mathbf{F}_N \right] + \mathbf{N}\mathbf{F}_N \\
&= \sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \boldsymbol{\Lambda}_M^{q/N} \left[\mathbf{L}_M^{(l)} \mathbf{S} \mathbf{P}_N^{-q} + \mathbf{U}_M^{(l)} \mathbf{S} \mathbf{F}_N^H \mathbf{P}_N \boldsymbol{\Lambda}_N^q \mathbf{F}_N \right] + \mathbf{N}\mathbf{F}_N \\
&= \sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \boldsymbol{\Lambda}_M^{q/N} \left[\mathbf{L}_M^{(l)} \mathbf{S} \mathbf{P}_N^{-q} + \mathbf{U}_M^{(l)} \mathbf{S} \mathbf{P}_N^{-q} \mathrm{diag} \right. \\
&\quad \left. \times \left(e^{-j2\pi(0-q)/N}, \ldots, e^{-j2\pi((N-1)-q)/N} \right) \right] + \mathbf{N}\mathbf{F}_N
\end{aligned}
$$



(a) Delay-Doppler domain. (b) Time-frequency domain.

Fig. 5. Pilot allocation as proposed by [15].

$$= \sum_{q=-Q/2}^{Q/2} \sum_{l=0}^{L} c_{l,q} \mathbf{W}_{l,q} \circ \left(\mathbf{P}_M^l \mathbf{S} \mathbf{P}_N^{-q}\right) + \mathbf{N}\mathbf{F}_N, \tag{5}$$

where

$$[\mathbf{W}_{l,q}]_{m,n} = \begin{cases} e^{j2\pi qm/K}, & \text{if } m \geq l, \\ e^{j2\pi qm/K} e^{-j2\pi(n-q)/N}, & \text{if } m < l, \end{cases} \tag{6}$$

for $m = 0, 1, \ldots, M-1$, $n = 0, 1, \ldots, N-1$. From (5), we learn that the delay tap l (circularly) shifts the rows of $\mathbf{S}$. Similarly, the Doppler tap q (circularly) shifts the columns in $\mathbf{S}$. Therefore, the first and second dimensions of $\mathbf{S}$ are often called the delay and Doppler dimensions, respectively. The channel coefficient $c_{l,q}$ and the matrix $\mathbf{W}_{l,q}$ apply a change in amplitude and phase.

5. Optimal pilot design for OTFS

In this section, we will propose a new pilot design for OTFS modulation, taking into account the relationship between the channel and the transmitted signal, as derived in the previous section. The following steps will be followed:

(1) We begin by establishing a criterion that ensures no overlap between pilot and communication symbols at the receiver (and consequently at the transmitter), and analyze the pilot allocation schemes that satisfy this condition. (Section 5.1)
(2) Then, we derive the allocations with minimum number of pilot overhead. (Section 5.2)
(3) We show that for fixed pilot power, these allocations achieve a channel estimate that minimizes the MSE on channel taps. (Section 5.3)
(4) Finally, we optimize the power distribution between the pilot and the communication symbols with respect to the channel capacity. (Section 5.4)

5.1. Step (1) analyzing possible pilot allocations

We will assign the available symbols to pilot and communication symbols. Let K_c and K_p denote the number of pilot and communication symbols, such that $K = K_\mathrm{c} + K_\mathrm{p}$. In matrix vector notation, this division can be written as $\mathbf{s} = (\boldsymbol{\Phi}_\mathrm{c}\mathbf{s}_\mathrm{c} + \boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p})$. Here, $\boldsymbol{\Phi}_\mathrm{c} \in \{0,1\}^{K \times K_\mathrm{c}}$ and $\boldsymbol{\Phi}_\mathrm{p} \in \{0,1\}^{K \times K_\mathrm{p}}$ are selection matrices containing only K_c and K_p columns of $\mathbf{I}_K$, indexed by $\mathbf{p}_\mathrm{c} \in \{0,1\}^{K\times 1}$ and $\mathbf{p}_\mathrm{p} \in \{0,1\}^{K\times 1}$. We have $\mathbf{1}_{K\times 1}^T \mathbf{p}_\mathrm{c} = K_\mathrm{c}$ and $\mathbf{1}_{K\times 1}^T \mathbf{p}_\mathrm{p} = K_\mathrm{p}$, and a symbol is either used for the pilot or communication, hence $\mathbf{p}_\mathrm{c} + \mathbf{p}_\mathrm{p} = \mathbf{1}_{K\times 1}$.

After demodulation we can write our received signal as,

$$\begin{aligned}\mathbf{y} &= (\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{H}(\mathbf{F}_N^H \otimes \mathbf{I}_M)(\boldsymbol{\Phi}_\mathrm{c}\mathbf{s}_\mathrm{c} + \boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p}) + (\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{n} \\ &= \mathbf{H}_{\mathrm{DD}}(\boldsymbol{\Phi}_\mathrm{c}\mathbf{s}_\mathrm{c} + \boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p}) + \mathbf{w}.\end{aligned}$$

Here $\mathbf{H}_{\mathrm{DD}} = (\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{H}(\mathbf{F}_N^H \otimes \mathbf{I}_M)$ and $\mathbf{w} = (\mathbf{F}_N \otimes \mathbf{I}_M)\mathbf{n}$. At the receiver, we can divide the received signal into a communication part and a pilot part, where we denote the communication part by $\mathbf{y}_\mathrm{c} = \boldsymbol{\Psi}_\mathrm{c}^H \mathbf{y}$ and the pilot part by $\mathbf{y}_\mathrm{p} = \boldsymbol{\Psi}_\mathrm{p}^H \mathbf{y}$. Here, the matrices $\boldsymbol{\Psi}_\mathrm{c} \in \mathbb{C}^{K\times R_\mathrm{c}}$ and $\boldsymbol{\Psi}_\mathrm{p} \in \mathbb{C}^{K\times R_\mathrm{p}}$ denote selection matrices, similar to the definition of $\boldsymbol{\Phi}_\mathrm{c}$ and $\boldsymbol{\Phi}_\mathrm{p}$, and contain R_c and R_p columns of $\mathbf{I}_K$ indexed by $\tilde{\mathbf{p}}_\mathrm{c} \in \{0,1\}^{K\times 1}$ and $\tilde{\mathbf{p}}_\mathrm{p} \in \{0,1\}^{K\times 1}$, respectively. It is important to mention that the $\boldsymbol{\Psi}$ matrices are selecting all the received symbols that include a communication symbol (for $\boldsymbol{\Psi}_\mathrm{c}$) or a pilot symbol (for $\boldsymbol{\Psi}_\mathrm{p}$). Thus, if $\boldsymbol{\Phi}$ is designed, $\boldsymbol{\Psi}$ follows automatically. Besides, note that the selection of $\boldsymbol{\Psi}_\mathrm{p}$ is larger than that of $\boldsymbol{\Phi}_\mathrm{p}$, i.e. $R_\mathrm{c} \geq K_\mathrm{c}$, because the transmitted communication symbols are spread out by the channel. A block diagram of the pipeline is shown in Fig. 6.

In order to facilitate a low-complexity receiver, we design the pilot and communication signals such that they do not interfere at the receiver. This configuration permits independent processing of channel estimation and symbol estimation. Hence, we have the following criterion:

C1 There is no overlap between pilot and communication symbols at the receiver side.

It follows from **C1** that we should have $\boldsymbol{\Psi}_\mathrm{c}^H \mathbf{H}_{\mathrm{DD}} \boldsymbol{\Phi}_\mathrm{p} = \mathbf{0}$ and $\boldsymbol{\Psi}_\mathrm{p}^H \mathbf{H}_{\mathrm{DD}} \boldsymbol{\Phi}_\mathrm{c} = \mathbf{0}$. From the interaction between the channel and the modulation, as derived in (5), we know that this is possible as long as the pilot and communication symbols are guarded by zeros accordingly, so that the shift operations of the channel do not mix them.

The communication part is given by

$$\mathbf{y}_\mathrm{c} = \boldsymbol{\Psi}_\mathrm{c}^H \mathbf{y} = \boldsymbol{\Psi}_\mathrm{c}^H \mathbf{H}_{\mathrm{DD}}(\boldsymbol{\Phi}_\mathrm{c}\mathbf{s}_\mathrm{c} + \boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p}) + \boldsymbol{\Psi}_\mathrm{c}^H \mathbf{w} = \boldsymbol{\Psi}_\mathrm{c}^H \mathbf{H}_{\mathrm{DD}} \boldsymbol{\Phi}_\mathrm{c}\mathbf{s}_\mathrm{c} + \mathbf{w}_\mathrm{c}, \tag{7}$$

and the pilot part is given by

$$\mathbf{y}_\mathrm{p} = \boldsymbol{\Psi}_\mathrm{p}^H \mathbf{y} = \boldsymbol{\Psi}_\mathrm{p}^H \mathbf{H}_{\mathrm{DD}}(\boldsymbol{\Phi}_\mathrm{c}\mathbf{s}_\mathrm{c} + \boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p}) + \boldsymbol{\Psi}_\mathrm{p}^H \mathbf{w} = \boldsymbol{\Psi}_\mathrm{p}^H \mathbf{H}_{\mathrm{DD}} \boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p} + \mathbf{w}_\mathrm{p}. \tag{8}$$

Recognizing that the LTV channel (circularly) shifts the rows and columns of the transmitted symbols $\mathbf{S}_{\mathrm{DD}}$, we can deduce which type of pilot allocation respects **C1**. We identify that every pilot allocation must consist of one of the three "basic" cases. Strictly speaking, a combination of multiple such cases is also possible. This requires multiple equispaced pilots in the delay and/or Doppler dimension, each with appropriate guard bands. This, however, complicates the pilot design, does not reduce the pilot overhead and does not improve the channel estimation, which is why we will not further investigate this option. The "basic" cases are:

(1) "Island case" - in case the pilot symbol "area" is embedded both in the delay ánd Doppler direction by communication symbols, in order to have nonzero pilot symbols, the "area" should be at

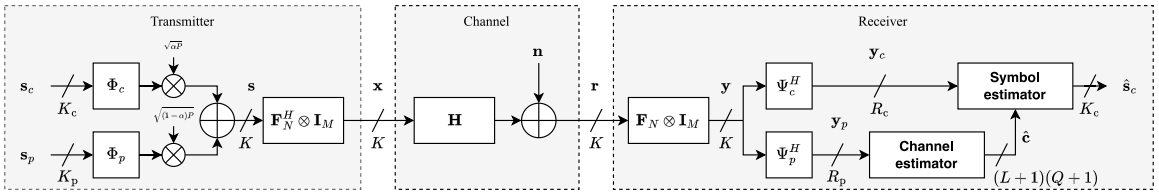
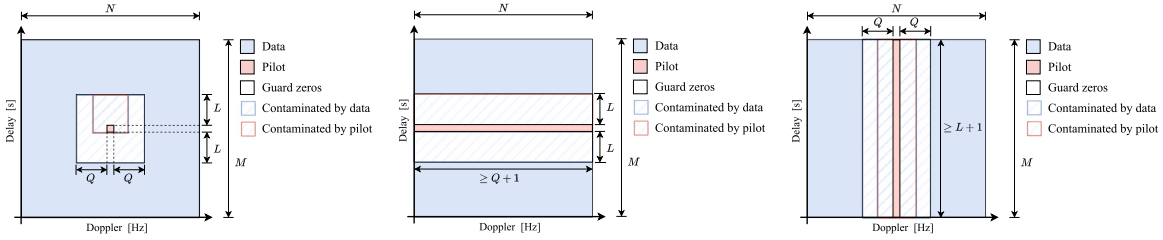


Fig. 6. Block diagram of the OTFS communication pipeline. Insertion and deletion of the cyclic prefix is omitted.



(a) $M \geq 2L+1$ and $N \geq 2Q+1$. (b) $M \geq 2L+1$ and $N \geq Q+1$. (c) $M \geq L+1$ and $N \geq 2Q+1$.

Fig. 7. DD configurations satisfying **C1**.

least $(2Q+1)\times(2L+1)$, thus $K_\mathrm{p} \geq (2Q+1)(2L+1)$. See Fig. 7(a) for a visualization. Note that this pilot allocation requires $N \geq 2Q+1$ and $M \geq 2L+1$.

(2) "Doppler slab" - in case the pilot symbol "area" is embedded only in the delay direction but not in the Doppler direction, the "area should be at least $N \times (2L+1)$, thus $K_\mathrm{p} \geq N(2L+1)$. See Fig. 7(b) for a visualization. Note that this pilot allocation requires $N \geq Q+1$ and $M \geq 2L+1$.

(3) "Delay slab" - in case the pilot symbol "area" is embedded only in the Doppler direction but not in the delay direction, the "area should be at least $(2Q+1)\times M$, thus $K_\mathrm{p} \geq (2Q+1)M$. See Fig. 7(c) for a visualization. Note that this pilot allocation requires $N \geq 2Q+1$ and $M \geq L+1$.

Having determined the possibilities for the pilot design, we are interested in its symbol overhead and the estimation performance.

5.2. Step (2) determining the pilot allocations with the lowest overhead

The minimum number of pilot symbols for each case is given by

(1) "Island case" - $K_\mathrm{p} = (2Q+1)(2L+1)$, consequently only one pilot symbol is nonzero.

(2) "Doppler slab" - $K_\mathrm{p} = (Q+1)(2L+1)$, thus choosing $N = Q+1$, consequently only one row containing $Q+1$ pilot symbols is nonzero.

(3) "Delay slab" - $K_\mathrm{p} = (2Q+1)(L+1)$, thus choosing $M = L+1$, consequently only one column containing $L+1$ pilot symbols is nonzero.

It is clear that case (2) and case (3) will require less pilot overhead compared to case (1), for every choice of L and Q. Therefore, it is crucial to examine whether there are any performance distinctions among these pilot designs. In the subsequent section, we will demonstrate that all these designs (1, 2, and 3) achieve the same minimum MSE.

5.3. Step (3) showing MMSE optimality

The pilot part at the receiver side can be rewritten as

$$\mathbf{y}_\mathrm{p} = \boldsymbol{\Psi}_\mathrm{p}^H \mathbf{H}_\mathrm{DD} \boldsymbol{\Phi}_\mathrm{p} \mathbf{s}_\mathrm{p} + \mathbf{w}_\mathrm{p} = \mathbf{Z}\mathbf{c} + \mathbf{w}_\mathrm{p}, \tag{9}$$

where $\mathbf{c} \in \mathbb{C}^{(L+1)(Q+1)\times 1}$ contains the coefficients $c_{l,q}$ of the BEM and where the columns of $\mathbf{Z} \in \mathbb{C}^{R_\mathrm{p}\times(L+1)(Q+1)}$ are given by,

$$\mathbf{z}_{l+q(L+1)} = \boldsymbol{\Psi}_\mathrm{p}^H (\mathbf{F}_N \otimes \mathbf{I}_M) \left(\boldsymbol{\Lambda}_K^q \mathbf{P}_K^l\right) (\mathbf{F}_N^H \otimes \mathbf{I}_M) \boldsymbol{\Phi}_\mathrm{p} \mathbf{s}_\mathrm{p}. \tag{10}$$

Let $\hat{\mathbf{c}}$ represent a channel estimator. Our objective is to minimize the MSE of this estimator, that is, $\mathbb{E}[(\mathbf{c}-\hat{\mathbf{c}})^H(\mathbf{c}-\hat{\mathbf{c}})]$. Note that we have a linear model, and consequently the MSE can be minimized by the linear MMSE (LMMSE) estimator. Assume the channel taps are independently distributed following a Gaussian distribution with a mean of zero and potentially varying variances, such that $\mathbb{E}[\mathbf{c}\mathbf{c}^H] = \mathrm{diag}([\sigma^2_{c_{0,0}}, \ldots, \sigma^2_{c_{L,Q}}]) = \mathbf{R}_\mathbf{c}$ and let $\mathbb{E}[\mathbf{w}_\mathrm{p}\mathbf{w}_\mathrm{p}^H] = \mathbf{R}_{\mathbf{w}_\mathrm{p}}$, then, the expression for the LMMSE estimator is

$$\hat{\mathbf{c}} = \left(\mathbf{R}_{\mathbf{w}_\mathrm{p}}^{-1}\mathbf{Z}(\mathbf{R}_\mathbf{c}^{-1} + \mathbf{Z}^H\mathbf{R}_{\mathbf{w}_\mathrm{p}}^{-1}\mathbf{Z})^{-1}\right)^H \mathbf{y}_\mathrm{p}. \tag{11}$$

If we assume that the noise is white, i.e. $\mathbf{R}_\mathbf{n} = \sigma_\mathbf{n}^2\mathbf{I}_K$, then $\mathbf{R}_\mathbf{w} = \mathbb{E}[(\mathbf{F}_N^H \otimes \mathbf{I}_M)\mathbf{n}\mathbf{n}^H(\mathbf{F}_N \otimes \mathbf{I}_M)] = \sigma_\mathbf{n}^2\mathbf{I}_K$, and consequently,

$$\mathbf{R}_{\mathbf{w}_\mathrm{p}} = \mathbb{E}[\boldsymbol{\Psi}_\mathrm{p}^H\mathbf{w}\mathbf{w}^H\boldsymbol{\Psi}_\mathrm{p}] = \boldsymbol{\Psi}_\mathrm{p}^H\mathbf{R}_\mathbf{w}\boldsymbol{\Psi}_\mathrm{p} = \sigma_\mathbf{n}^2\mathbf{I}_{R_\mathrm{p}}. \tag{12}$$

As a result, we can rewrite the LMMSE estimator as

$$\hat{\mathbf{c}} = \left(\sigma_\mathbf{n}^2\mathbf{R}_\mathbf{c}^{-1} + \mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H\left(\mathbf{Z}\mathbf{c} + \mathbf{w}_\mathrm{p}\right). \tag{13}$$

The MSE of this estimator is given by,

$$\begin{aligned}\mathbb{E}[(\mathbf{c}-\hat{\mathbf{c}})^H(\mathbf{c}-\hat{\mathbf{c}})] &= \mathbb{E}\left[\mathrm{tr}\left((\mathbf{c}-\hat{\mathbf{c}})(\mathbf{c}-\hat{\mathbf{c}})^H\right)\right] \\ &= \mathrm{tr}\left(\left(\mathbf{R}_\mathbf{c}^{-1} + \frac{1}{\sigma_\mathbf{n}^2}\mathbf{Z}^H\mathbf{Z}\right)^{-1}\right).\end{aligned} \tag{14}$$

To reach the minimum MSE, the pilot symbols should be placed such that $\mathbf{Z}^H\mathbf{Z}$ is diagonal ([36], Lemma 1). Note that the columns of $\mathbf{Z}$ are given by

$$\begin{aligned}\mathbf{z}_{l+q(L+1)} &= \boldsymbol{\Psi}_\mathrm{p}^H(\mathbf{F}_N \otimes \mathbf{I}_M)\left(\boldsymbol{\Lambda}_K^q\mathbf{P}_K^l\right)(\mathbf{F}_N^H \otimes \mathbf{I}_M)\boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p} \\ &= \boldsymbol{\Psi}_\mathrm{p}^H \mathrm{vec}\left(\mathbf{W}_{l,q} \circ \left(\mathbf{P}_M^l \mathrm{vec}^{-1}(\boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p})\mathbf{P}_N^{-q}\right)\right).\end{aligned} \tag{15}$$

Let $\mathbf{S}_\mathrm{p} = \mathrm{vec}^{-1}(\boldsymbol{\Phi}_\mathrm{p}\mathbf{s}_\mathrm{p})$. The elements of the matrix $\mathbf{Z}^H\mathbf{Z}$ can be rewritten[3] as in (16).

$$\begin{aligned}&\mathbf{z}_{l_1+q_1(L+1)}^H\mathbf{z}_{l_2+q_2(L+1)} \\ &= \mathrm{vec}^H\left(\mathbf{W}_{l_1,q_1} \circ \left(\mathbf{P}_M^{l_1}\mathbf{S}_\mathrm{p}\mathbf{P}_N^{-q_1}\right)\right)\boldsymbol{\Psi}_\mathrm{p}\boldsymbol{\Psi}_\mathrm{p}^H\mathrm{vec}\left(\mathbf{W}_{l_2,q_2} \circ \left(\mathbf{P}_M^{l_2}\mathbf{S}_\mathrm{p}\mathbf{P}_N^{-q_2}\right)\right) \\ &\overset{(a)}{=} \left[\mathrm{vec}^*\left(\mathbf{W}_{l_1,q_1} \circ \left(\mathbf{P}_M^{l_1}\mathbf{S}_\mathrm{p}\mathbf{P}_N^{-q_1}\right)\right) \circ \mathrm{vec}\left(\mathbf{W}_{l_2,q_2} \circ \left(\mathbf{P}_M^{l_2}\mathbf{S}_\mathrm{p}\mathbf{P}_N^{-q_2}\right)\right)\right]^T \\ &\quad \times \mathrm{diag}\left(\boldsymbol{\Psi}_\mathrm{p}\boldsymbol{\Psi}_\mathrm{p}^H\right)\end{aligned}$$

[3] Where (a) uses the fact that if $\mathbf{a}_1$ and $\mathbf{a}_2$ are two column vectors and $\mathbf{X}$ is a diagonal matrix, then $\mathbf{a}_1^H\mathbf{X}\mathbf{a}_2 = \mathrm{vec}(\mathbf{a}_2^T \odot \mathbf{a}_1^H)\mathrm{diag}(\mathbf{X}) = (\mathbf{a}_1^* \circ \mathbf{a}_2)^T\mathrm{diag}(\mathbf{X})$.

$$= \text{vec}^T\left(\mathbf{W}^*_{l_1,q_1} \circ \mathbf{W}_{l_2,q_2} \circ \left(\mathbf{P}_M^{l_1}\mathbf{S}_\text{p}^*\mathbf{P}_N^{-q_1}\right) \circ \left(\mathbf{P}_M^{l_2}\mathbf{S}_\text{p}\mathbf{P}_N^{-q_2}\right)\right) \times \text{diag}\left(\boldsymbol{\Psi}_\text{p}\boldsymbol{\Psi}_\text{p}^H\right) \tag{16}$$

For the diagonal elements, i.e. $(l_1, q_1) = (l_2, q_2) \in \{0, 1, \ldots, L\} \times \{0, 1, \ldots, Q\}$, we have,

$$\begin{aligned}\mathbf{z}_i^H\mathbf{z}_i &= \text{vec}^T\left(\mathbf{W}^*_{l,q} \circ \mathbf{W}_{l,q} \circ \left(\mathbf{P}_M^{l}\left(\mathbf{S}_\text{p}^* \circ \mathbf{S}_\text{p}\right)\mathbf{P}_N^{-q}\right)\right)\text{diag}\left(\boldsymbol{\Psi}_\text{p}\boldsymbol{\Psi}_\text{p}^H\right)\\ &= \text{vec}^T\left(\left(\mathbf{P}_M^{l}\left(\mathbf{S}_\text{p}^* \circ \mathbf{S}_\text{p}\right)\mathbf{P}_N^{-q}\right)\right)\text{diag}\left(\boldsymbol{\Psi}_\text{p}\boldsymbol{\Psi}_\text{p}^H\right)\\ &= \text{vec}^T\left(\left(\mathbf{P}_M^{l}\left(\mathbf{S}_\text{p}^* \circ \mathbf{S}_\text{p}\right)\mathbf{P}_N^{-q}\right)\right)\tilde{\mathbf{p}}_\text{p}\\ &\overset{(b)}{=} \left(\mathbf{s}_\text{p}^* \circ \mathbf{s}_\text{p}\right)^T\mathbf{1}_{K_\text{p}\times 1} = \mathbf{s}_\text{p}^H\mathbf{s}_\text{p} = P_\text{p},\end{aligned} \tag{17}$$

where in (b) we have used the fact that $\tilde{\mathbf{p}}_\text{p}$ is selecting all the symbols that include a pilot symbol after passing through the channel. Thus, the diagonal of $\mathbf{Z}^H\mathbf{Z}$ only contains the total power P_p of the pilot symbols. Note that the diagonal elements do not depend on K_p. If $\mathbf{Z}^H\mathbf{Z}$ must be diagonal, then the off-diagonal elements of $\mathbf{Z}^H\mathbf{Z}$ must be zero, that is,

$$\mathbf{z}^H_{l_1+q_1(L+1)}\mathbf{z}_{l_2+q_2(L+1)} = 0, \quad (l_1, q_1) \neq (l_2, q_2). \tag{18}$$

By inspecting (16) for $(l_1, q_1) \neq (l_2, q_2)$, we see that, left of the element-wise product, the pilot symbols are shifted by l_1 in the delay direction and q_1 in the Doppler direction, while on the right of the element-wise product, the pilot symbols are shifted by l_2 in the delay direction and q_2 in the Doppler direction. From this observation we can deduce that, to have the outcome equal to zero, *the pilot symbols in the delay-Doppler domain, that is the matrix* $\text{vec}^{-1}(\boldsymbol{\Phi}_\text{p}\mathbf{s}_\text{p})$*, should have shift orthogonal rows and columns.*

We can draw some conclusions:

(1) "Island case" - For a fixed power P_p, this pilot allocation, with $K_\text{p} = (2Q+1)(2L+1)$ and with only one nonzero pilot symbol in the middle, achieves the minimum MSE. However, it uses more pilot symbols compared to the "Doppler slab" and "Delay slab" case.

(2) "Doppler slab" - For a fixed power P_p, this pilot allocation, with $K_\text{p} = (Q+1)(2L+1)$ and with $Q+1$ nonzero pilot symbols, achieves the minimum MSE if and only if the nonzero pilots are shift orthogonal in both delay and Doppler direction at the same time. The only option adhering to this orthogonality is to have only one nonzero pilot symbol. This is visualized in Fig. 8(a). Although in the figure the pilot symbol is placed in the middle, the symbol could be placed anywhere along the Doppler direction (the delay position is fixed).

(3) "Delay slab" - For a fixed power P_p, this pilot allocation, with $K_\text{p} = (2Q+1)(L+1)$ and with $L+1$ nonzero pilot symbols, achieves the minimum MSE if and only if the nonzero pilots are shift orthogonal in both delay and Doppler direction at the same time. The only option adhering to this orthogonality is to have only one nonzero pilot symbol. This is visualized in Fig. 8(b). Although in the figure the pilot symbol is placed in the middle, the symbol could be placed anywhere along the delay direction (the Doppler position is fixed).

To recap, given **C1**, the pilot allocations with the lowest pilot overhead are (also) allocations that attain the minimum MSE. The latter result is supported by simulation experiments, as shown in Fig. 8(c), where the average MSE of the channel taps is computed over 100 Monte Carlo runs. In each run, a new channel and noise realization is drawn from their respective distributions, with $\sigma^2_{c_{l,q}} = 1/((L+1)(Q+1))$ for all l and q.

Note that the pilot allocations using the ZC sequence does not achieve the MMSE performance. This is because the pilot symbols in the DD domain do not produce shift-orthogonal rows and columns. For the current parameters, $\{K, L, Q\} = \{411, 8, 8\}$, the resulting performance gap is approximately 3 dB, although this difference depends on the specific simulation parameters.

Moreover, the three cases, the island case, Doppler slab, and delay slab, do achieve the same MMSE; however, our proposed Doppler slab and delay slab do so with less pilot overhead. As a result, the proposed allocations can fit more communication symbols. When the three allocations would operate at the same data rate, the proposed allocations will have a better BER than the island case, since in practice, the latter will have to use higher-order constellations, or less error correction coding. Therefore, our proposed Doppler slab and delay slab allocations are the preferred options. In order to use the least overhead at all times, we must choose the Doppler slab (set $N = Q+1$) in case $Q < L$, and we must choose the delay slab (set $M = L+1$) in case $Q > L$. With this decision rule, we can attain a pilot overhead of $K_\text{p} = \min\{(L+1)(2Q+1), (2L+1)(Q+1)\}$.

As the optimal pilot allocation is now determined, the next step is to optimize the power distribution between the pilot and the communication symbols.

5.4. Step (4) optimizing the power balance

Suppose we have a total power budget P, which can be distributed between the pilot and communication symbols. Let $P_\text{c} = \alpha P$ and $P_\text{p} = (1-\alpha)P$, where $\alpha \in [0, 1]$ controls the power allocation between communication and pilot symbols, while the total power constraint $P_\text{c} + P_\text{p} = P$ is satisfied. We can then optimize a performance measure with respect to this power balance.

Since pilot overhead is critical in time-varying communication systems, we first (step 2 of our approach) focused on reducing this overhead as much as possible. This naturally also maximizes the time-frequency occupancy of the communications symbols, and hence implicitly focuses on capacity.[4] Therefore, in this step 4, it makes sense to also optimize the power distribution in terms of the channel capacity. Moreover, channel capacity is the primary performance metric in communication systems. Unlike the BER, which depends on modulation order and channel coding, channel capacity provides a modulation- and coding-independent measure of link performance. It also has a closed-form expression, making it suitable for system analysis and optimization. Finally, systems that transmit large amounts of data and employ adaptive modulation and coding can approach the channel capacity in practice.

The received communication part is given by,

$$\mathbf{y}_\text{c} = \boldsymbol{\Psi}_\text{c}^H\mathbf{H}_\text{DD}\boldsymbol{\Phi}_\text{c}\mathbf{s}_\text{c} + \mathbf{w}_\text{c} = \tilde{\mathbf{H}}_\text{c}\mathbf{s}_\text{c} + \mathbf{w}_\text{c}. \tag{19}$$

The capacity of the channel, averaged over the random channel $\tilde{\mathbf{H}}_\text{c}$, induced by the coefficients in $\mathbf{c}$, is given by (see [30,38]),

$$C = \frac{1}{K+L}\mathbb{E}\left[\max_{p(\mathbf{s}_\text{c}), P_\text{c}=\mathbb{E}[\mathbf{s}_\text{c}^H\mathbf{s}_\text{c}]} \mathcal{I}(\mathbf{y}_\text{c}; \mathbf{s}_\text{c}|\hat{\mathbf{c}})\right]. \tag{20}$$

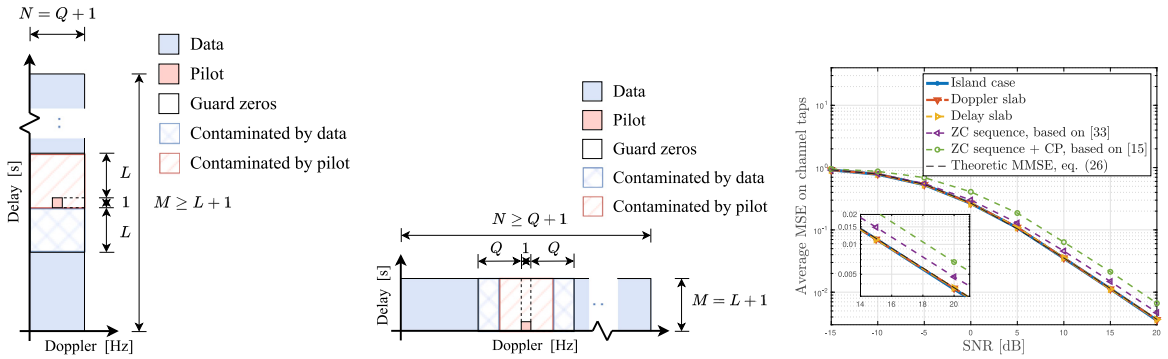
Here, $\mathcal{I}(\mathbf{y}_\text{c}; \mathbf{s}_\text{c}|\hat{\mathbf{c}})$ is the conditional mutual information between the received signal $\mathbf{y}_\text{c}$ and the transmitted symbols $\mathbf{s}_\text{c}$, given the channel coefficient estimate $\hat{\mathbf{c}}$, $p(\mathbf{s}_\text{c})$ is the probability distribution of $\mathbf{s}_\text{c}$ with fixed energy P_c. Note that we divide by $K+L$ to account for the cyclic prefix of length L.

Let the channel estimate be given by $\hat{\mathbf{c}}$, and let the *estimated* channel matrix be denoted by $\hat{\tilde{\mathbf{H}}}_\text{c}$, then the received communication part can be rewritten as,

$$\mathbf{y}_\text{c} = \hat{\tilde{\mathbf{H}}}_\text{c}\mathbf{s}_\text{c} + (\tilde{\mathbf{H}}_\text{c} - \hat{\tilde{\mathbf{H}}}_\text{c})\mathbf{s}_\text{c} + \mathbf{w}_\text{c} = \hat{\tilde{\mathbf{H}}}_\text{c}\mathbf{s}_\text{c} + \mathbf{v}, \tag{21}$$

where $\mathbf{v} = (\tilde{\mathbf{H}}_\text{c} - \hat{\tilde{\mathbf{H}}}_\text{c})\mathbf{s}_\text{c} + \mathbf{w}_\text{c}$. Now, since no knowledge is available at the transmitter about the channel, it is reasonable to assign equal energy to all communication symbols, i.e. $\mathbf{R}_{\mathbf{s}_\text{c}} = \frac{P_\text{c}}{K_\text{c}}\mathbf{I}_{K_\text{c}}$. With a fixed (i.e. equal)

[4] Interestingly, a similar connection between channel capacity and channel MMSE has been noticed in [30].



(a) Doppler slab, $M \geq 2L+1$, $N = Q+1$. (b) Delay slab, $M = L+1$, $N \geq 2Q+1$. (c) Channel estimation performance.

Fig. 8. (a) and (b): Pilot allocations satisfying (i) **C1**, (ii) achieving the MMSE ($\mathbf{Z}^H\mathbf{Z}$ diagonal), and (iii) having the lowest overhead. (c): Performance compared to existing pilot schemes, for $\{K, L, Q\} = \{441, 8, 8\}$.

communication symbol power, a lower bound on the channel capacity is given by [30,38],

$$C \geq \frac{1}{K+L}\mathbb{E}\left[\log\det\left(\mathbf{I}_{R_c\times R_c} + \frac{P_c}{K_c}\mathbf{R}_v^{-1}\hat{\mathbf{H}}_c\hat{\mathbf{H}}_c^H\right)\right]. \tag{22}$$

Here,

$$\mathbf{R}_v = \mathbb{E}[\mathbf{v}\mathbf{v}^H] = \frac{P_c}{K_c}\mathbb{E}[(\tilde{\mathbf{H}}_c - \hat{\tilde{\mathbf{H}}}_c)(\tilde{\mathbf{H}}_c - \hat{\tilde{\mathbf{H}}}_c)^H] + \sigma_{\mathbf{n}}^2\mathbf{I}_{R_c\times R_c}. \tag{23}$$

The goal is to determine the appropriate power distribution, that is, select α, to maximize the lower bound on the channel capacity C. To express the lower bound in terms of the power distribution parameter α, we proceed with some derivations. First of all, we can write (see Appendix A for a detailed derivation),

$$\mathbb{E}[(\tilde{\mathbf{H}}_c - \hat{\tilde{\mathbf{H}}}_c)(\tilde{\mathbf{H}}_c - \hat{\tilde{\mathbf{H}}}_c)^H] \preceq \mathbb{E}\left[\mathrm{tr}\left((\mathbf{c}-\hat{\mathbf{c}})(\mathbf{c}-\hat{\mathbf{c}})^H\right)\right]\mathbf{I}_{R_c}. \tag{24}$$

Secondly, because the pilot allocation makes $\mathbf{Z}^H\mathbf{Z}$ diagonal, we can write

$$\begin{aligned}\left(\mathbf{R}_c^{-1} + \frac{1}{\sigma_{\mathbf{n}}^2}\mathbf{Z}^H\mathbf{Z}\right)^{-1} &= \left(\mathrm{diag}([\sigma_{c_{0,0}}^2,\dots,\sigma_{c_{L,Q}}^2])^{-1} + \frac{1}{\sigma_{\mathbf{n}}^2}P_p\mathbf{I}_{(L+1)(Q+1)}\right)^{-1} \\ &= \mathrm{diag}\left(\frac{\sigma_{c_{0,0}}^2\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{n}}^2+\sigma_{c_{0,0}}^2P_p},\dots,\frac{\sigma_{c_{L,Q}}^2\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{n}}^2+\sigma_{c_{L,Q}}^2P_p}\right),\end{aligned} \tag{25}$$

so that, the channel MSE (14) is given by,

$$\mathbb{E}\left[\mathrm{tr}(\mathbf{c}-\hat{\mathbf{c}})(\mathbf{c}-\hat{\mathbf{c}})^H\right] = \sum_{l=0}^{L}\sum_{q=0}^{Q}\frac{\sigma_{c_{l,q}}^2\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{n}}^2+\sigma_{c_{l,q}}^2P_p}. \tag{26}$$

We can then combine (24) and (26) to rewrite $\mathbf{R}_v$ as

$$\begin{aligned}\mathbf{R}_v &\preceq \frac{P_c}{K_c}\mathbb{E}\left[\mathrm{tr}(\mathbf{c}-\hat{\mathbf{c}})(\mathbf{c}-\hat{\mathbf{c}})^H\right]\mathbf{I}_{R_c} + \sigma_{\mathbf{n}}^2\mathbf{I}_{R_c} \\ &= \left[\frac{P_c}{K_c}\sum_{l=0}^{L}\sum_{q=0}^{Q}\frac{\sigma_{c_{l,q}}^2\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{n}}^2+\sigma_{c_{l,q}}^2P_p} + \sigma_{\mathbf{n}}^2\right]\mathbf{I}_{R_c}.\end{aligned} \tag{27}$$

Finally we can derive a looser lower bound on the capacity from (22) as

$$\begin{aligned}C &\geq \frac{1}{K+L}\mathbb{E}\left[\log\det\left(\mathbf{I}_{R_c} + \frac{P_c}{K_c}\mathbf{R}_v^{-1}\hat{\mathbf{H}}_c\hat{\mathbf{H}}_c^H\right)\right] \\ &\geq \frac{1}{K+L}\mathbb{E}\left[\log\det\left(\mathbf{I}_{R_c} + \frac{P_c}{K_c}\left[\frac{P_c}{K_c}\sum_{l=0}^{L}\sum_{q=0}^{Q}\frac{\sigma_{c_{l,q}}^2\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{n}}^2+\sigma_{c_{l,q}}^2P_p} + \sigma_{\mathbf{n}}^2\right]^{-1}\hat{\mathbf{H}}_c\hat{\mathbf{H}}_c^H\right)\right] = \underline{C}.\end{aligned} \tag{28}$$

Now lastly, we rewrite $\hat{\mathbf{H}}_c\hat{\mathbf{H}}_c^H$.

We can show that (refer to Appendix B for the derivation),

$$\mathrm{tr}\left(\mathbb{E}\left[\hat{\mathbf{H}}_c\hat{\mathbf{H}}_c^H\right]\right) = K_c\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\sigma_{\hat{c}_{l,q}}^2 = K_c\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\frac{P_p\sigma_{c_{l,q}}^4}{\sigma_{\mathbf{n}}^2+P_p\sigma_{c_{l,q}}^2}. \tag{29}$$

This motivates us to normalize the channel matrix $\hat{\hat{\mathbf{H}}}_c$ as,

$$\hat{\hat{\mathbf{H}}}_c = \sqrt{K_c\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\sigma_{\hat{c}_{l,q}}^2}\,\hat{\hat{\mathbf{H}}}_c', \tag{30}$$

where $\hat{\hat{\mathbf{H}}}_c'$ is the normalized channel matrix. We substitute this normalization and the expression obtained in (27) in the lower bound on the capacity ($\underline{C}$ in (28)) and obtain,

$$\begin{aligned}\underline{C} &= \frac{1}{K+L}\mathbb{E}\left[\log\det\left(\mathbf{I}_{R_c} + \frac{P_c}{K_c}\left[\frac{P_c}{K_c}\sum_{l=0}^{L}\sum_{q=-Q/2}^{Q/2}\frac{\sigma_{c_{l,q}}^2\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{n}}^2+\sigma_{c_{l,q}}^2P_p} + \sigma_{\mathbf{n}}^2\right]^{-1}\right.\right. \\ &\quad \left.\left.\times\, K_c\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\sigma_{\hat{c}_{l,q}}^2\hat{\hat{\mathbf{H}}}_c'\hat{\hat{\mathbf{H}}}_c'^H\right)\right] \\ &= \frac{1}{K+L}\mathbb{E}\left[\log\det\left(\mathbf{I}_{R_c} + \rho\hat{\hat{\mathbf{H}}}_c'\hat{\hat{\mathbf{H}}}_c'^H\right)\right],\end{aligned} \tag{31}$$

where, using the expression for $\sigma_{\hat{c}_{l,q}}^2$ in (29),

$$\rho = P_c\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\frac{P_p\sigma_{c_{l,q}}^4}{\sigma_{\mathbf{n}}^2+P_p\sigma_{c_{l,q}}^2}\left[\frac{P_c}{K_c}\sum_{l=0}^{L}\sum_{q=-Q/2}^{Q/2}\frac{\sigma_{c_{l,q}}^2\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{n}}^2+\sigma_{c_{l,q}}^2P_p} + \sigma_{\mathbf{n}}^2\right]^{-1}.$$

Remark 4. Note that ρ is essentially a signal-to-interference-plus-noise ratio (SINR). The numerator contains the energy related to the (target) signal, i.e. $\hat{\mathbf{H}}_c\mathbf{s}_c$, while the denominator contains the interference energy due to the channel estimation error, related to $(\tilde{\mathbf{H}}_c - \hat{\tilde{\mathbf{H}}}_c)\mathbf{s}_c$, and the noise energy.

We can maximize ρ with respect to $P_c = \alpha P$ and $P_p(1-\alpha)P$ to optimize $\underline{C}$ in (31) (since the normalized channel matrix is independent of the power distribution). Let P be the total transmitter power, and set $P_c = \alpha P$ and $P_p = (1-\alpha)P$; thus, ρ becomes a function of α. The optimal power distribution is given by

$$\alpha^* = \arg\max_{\alpha}\rho. \tag{32}$$

Before proceeding to the simulations in the next section, we make some remarks with respect to the complexity and implementation of our proposed pilot design.

Note that using a different pilot *allocation* does not increase or decrease complexity of the communication system. Determining the optimal *power distribution* requires maximizing a one-dimensional (concave) function, which has negligible complexity compared to the channel or symbol estimation steps, which typically involve matrix inversions or iterative algorithms. The complexity of our pilot design is therefore comparable to that of existing designs.

Table 2
Parameters for the simulations in Fig. 9.

Parameters	Channel 1	Channel 2	Channel 3
K	441	441	441
Q	6	8	2
L	6	2	8
SNR_{tx}	20 dB	20 dB	20 dB
$\sigma^2_{c_{l,q}}$	$1/((Q+1)(L+1)), \quad \forall\{l,q\}$	$1/((Q+1)(L+1)), \quad \forall\{l,q\}$	$1/((Q+1)(L+1)), \quad \forall\{l,q\}$
Island case - $\{N, M, \alpha_{\text{opt}}\}$	$\{21, 21, 0.7015\}$	$\{21, 21, 0.7834\}$	$\{21, 21, 0.7834\}$
Doppler slab - $\{N, M, \alpha_{\text{opt}}\}$	$\{7, 63, 0.7270\}$	$\{9, 49, 0.7922\}$	$\{3, 147, 0.7910\}$
Delay slab - $\{N, M, \alpha_{\text{opt}}\}$	$\{63, 7, 0.7270\}$	$\{147, 3, 0.7910\}$	$\{49, 9, 0.7922\}$

Implementation of our pilot design in practical systems is not significantly different from existing approaches. All discussed pilot designs assume that L and Q are known, which is reasonable in practice as noted in Remark 2, and that the noise level $\sigma^2_{\mathbf{n}}$ is available. In modern communication systems, the noise level is typically tracked by the receiver, since both channel and symbol estimators depend on it.

The only additional requirement of our method is knowledge of the channel tap distribution, i.e., the variances $\sigma^2_{c_{l,q}}$. These can be estimated using a probe signal, or a preamble, and in practice, this distribution tends to remain stable over multiple communication packets as well.

6. Simulations

In this section, we validate the theoretical findings and compare our framework with related work.

6.1. Numerical validation

We simulate three baseband channels (i.e., frequency shifts between 0 and 1 and time delays of multiples of the sampling frequency) with different parameters, which are specified in Table 2.

We call a channel 'Doppler dominant' if $Q > L$ and 'delay dominant' if $Q < L$. Thus, Channel 2 and Channel 3 are Doppler and delay dominant, respectively. We define the signal-to-noise ratio with respect to the transmitted signal $\mathbf{x}$ as,

$$\text{SNR}_{\text{tx}} = \frac{\mathbb{E}[\mathbf{x}^H\mathbf{x}]}{\mathbb{E}[\mathbf{n}^H\mathbf{n}]} = \frac{P}{K\sigma^2_{\mathbf{n}}}.$$

We set $P = 1$, and change the value of $\sigma^2_{\mathbf{n}}$ according to the desired SNR_{tx}. In all simulations, the (pilot and communication) symbols are uncoded QPSK symbols. The channel coefficients $c_{l,q}$ are realizations of a (independent) complex Gaussian process with zero mean and variance $\sigma^2_{c_{l,q}} = 1/((Q+1)(L+1))$. For all three channels we draw ten noise and channel realizations, and calculate the average capacity. The results are shown in Fig. 9. We can draw the following conclusions. First of all, the Doppler slab and delay slab, which alter the modulation parameters N and M according to the channel, exhibit higher capacity in all three channels, compared to the 'island case' proposed in [12,13]. Moreover, we note that either the Doppler slab or delay slab performs the best, according to whether the channel is Doppler or delay dominant. This is to be expected, since the Doppler (delay) slab has the lowest pilot overhead for a Doppler (delay) dominant channel. Finally, we can see that all three pilot allocations benefit from the optimal power allocation. We see that, indeed, the maximum is reached at the power distribution α_{opt}. In the simulation for Fig. 9, we considered only those pilot allocations for OTFS that achieve the channel MMSE. For allocations that do not attain the MMSE, a general expression for the channel estimation error is not available, making it nontrivial to derive a meaningful lower bound on the channel capacity.

In Fig. 10, we illustrate (with lines without markers) how the capacity optimal power distribution, obtained from (32), evolves with different noise levels. We set $\{K, Q, L\} = \{441, 2, 6\}$ and consider two different channels: in Fig. 10(a), we use channel taps with identical distributions, while in Fig. 10(b), we let the variance of the channel taps decay exponentially. The capacity-optimal power distribution evolves differently per channel, but both curves saturate to $\alpha = 0.5$ for high noise levels. A similar trend was found for SCM in [30].

Additionally, we plot (with lines with markers) the power distribution that is optimal in terms of BER. The minimum BER over 50 Monte Carlo runs has been found by simulating every combination of noise level and power distribution. Results at extremely high and low noise levels are omitted, as the BER saturates to 0 (low noise level) and 0.5 (high noise level) for all power distributions, so no meaningful optimum can be identified in those cases.

Fig. 10 shows that the BER optimal power distribution closely follows the capacity-optimal power distribution, even for D-OSDM, which employs a pilot design not suited for our capacity optimization. This suggests that (32) can be used not only to optimize channel capacity but also to improve the BER.

6.2. Comparison with related work

So far, we have compared our proposed pilot designs to the 'island case' pilot allocation from [12,13] using our capacity optimal power distribution for all cases. Here, we explicitly compare the capacity of our power distribution with the one originally used in [13]. To ensure fairness, we align the parameters used in [13] with those discussed in this paper.

Let s_{p} denote a pilot symbol and let s_{c} denote a communication symbol, then in [13] the pilot and communication SNR were defined per *symbol*, that is,

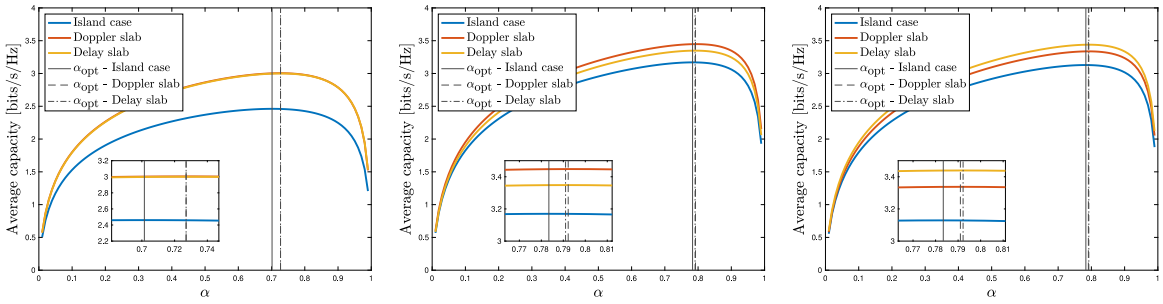
$$\text{SNR}_{\text{p}} = \frac{|s_{\text{p}}|^2}{\sigma^2_{\mathbf{n}}}, \quad \text{SNR}_{\text{c}} = \frac{\mathbb{E}[|s_{\text{c}}|^2]}{\sigma^2_{\mathbf{n}}}.$$

Note that the relation to our SNR of the transmitted signal (thus pilot and communication signal together) is given by $\text{SNR}_{\text{tx}} = \frac{1}{K}\text{SNR}_{\text{p}} + \frac{K_{\text{c}}}{K}\text{SNR}_{\text{c}}$. We can relate the notion of SNR per *symbol* of [13] to our power distribution parameter,

$$\alpha = \frac{K_{\text{c}}\text{SNR}_{\text{c}}}{K_{\text{c}}\text{SNR}_{\text{c}} + \text{SNR}_{\text{p}}}.$$

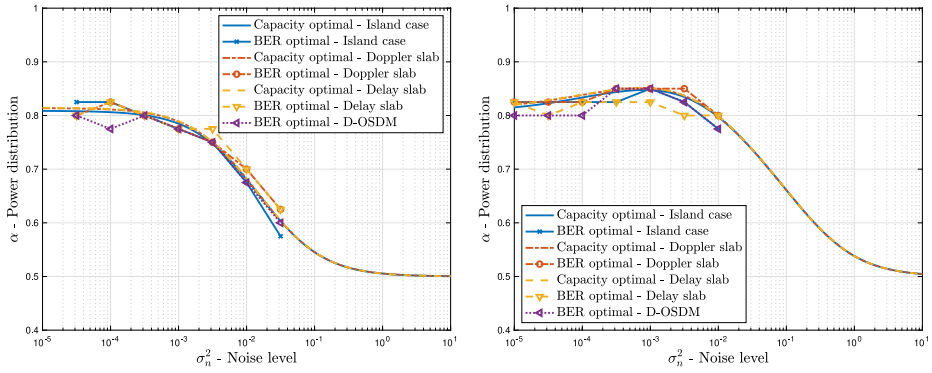
Then, in Table 3 we list: the parameters that were used in [13] (first four columns), "our" parameters that follow from that (fifth to ninth column), and the parameters for the optimal pilot design (last four columns). We can see that the α that follows from [13] differs a lot from the optimal distribution α^*. In fact, we see that modifying the power distribution can lead to a substantial improvement in performance (compare seventh and ninth column). Furthermore, selecting the appropriate values for M and N can further enhance performance as we can see in the eleventh column.

Note that an increase in SNR_{tx} from 21.6 dB to 27.7 dB and from 25.5 dB to 29.0 dB is more than double the amount of power. It is important to highlight that this increase in power leads to only a minor BER improvement in [13] as well as a minor capacity improvement as seen in the seventh column. By contrasting the optimal power allocation with the actual power distribution, the slight increase in BER becomes more understandable; the power increase is counteracted by the bad power distribution.



(a) Channel 1: $Q = 6$, $L = 6$. (b) Channel 2: $Q = 8$, $L = 2$. (c) Channel 3: $Q = 2$, $L = 8$.

Fig. 9. Performance of the pilot allocation schemes with respect to the power distribution for three different channels.



(a) Identically distributed channel taps, $\sigma^2_{c_{l,q}} = 1/(L+1)(Q+1)$. (b) Decaying channel tap power, $\sigma^2_{c_{l,q}} = 1/(4^{|q|}3^{l+1})$.

Fig. 10. Capacity optimal power distribution (i.e., (32)), versus BER optimal power distribution.

Table 3
Results of the comparison with [13] for a simulated channel with $Q = 2$, $L = 6$. The SNR and capacity values are listed in dB and bit/s/Hz, respectively.

Island case $M = 128$, $N = 16$									Doppler slab $M = 686$, $N = 3$		Delay slab $M = 7$, $N = 294$	
Parameters from [13], cf Fig. 14				Resulting parameters [13]			Sub-optimal choice		Sub-optimal choice:		Optimal choice:	
SNR_p	SNR_c	σ^2	BER	SNR_{tx}	α	$\underline{C}(\alpha)$	α^*	$\underline{C}(\alpha^*)$	α^*	$\underline{C}(\alpha^*)$	α^*	$\underline{C}(\alpha^*)$
50	20	1	$\approx 1 \cdot 10^{-2}$	21.6	0.66	3.92	0.91	4.14	0.91	4.17	0.91	4.18
60	20	1	$\approx 8.5 \cdot 10^{-3}$	27.7	0.17	4.01	0.91	5.50	0.91	5.55	0.91	5.56
50	25	1	$\approx 2 \cdot 10^{-3}$	25.5	0.86	5.04	0.91	5.05	0.91	5.09	0.91	5.10
60	25	1	$\approx 1.5 \cdot 10^{-3}$	29.0	0.39	5.09	0.91	5.75	0.91	5.81	0.91	5.81

7. Conclusions

In this paper, we have aimed to contribute to the understanding of how to design pilot signals for OTFS. We conducted an investigation into the literature on pilot design and established connections between the work on LTI and LTV channels, SCM, OFDM and OTFS modulation.

We have proposed two new pilot allocations for OTFS, that adjust M or N according to the channel parameters L and Q. Our pilot allocations save approximately 50% on pilot overhead compared to the conventional allocation, while still achieving the MMSE for channel estimation. We have also addressed the aspect of optimizing the power distribution between the pilot and communication signal with respect to the channel capacity. Our results indicate that selecting an appropriate power distribution significantly enhances a (lower bound on the) channel capacity of the communication system. Moreover, we empirically show that, across different channel settings, our proposed power distribution, although optimized for capacity, also minimizes the BER.

In summary, our research demonstrates that the careful selection of OTFS parameters, together with pilot design (including the allocation and power distribution) can lead to a significant improvement in the performance of an OTFS based communication system.

CRediT authorship contribution statement

Ids van der Werf: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Richard Heusdens:** Writing – review & editing, Validation, Supervision. **Richard C. Hendriks:** Writing – review & editing, Validation, Supervision. **Geert Leus:** Writing – review & editing, Supervision, Formal analysis.

Acknowledgments

This work was partly funded by the Netherlands Organisation for Applied Scientific Research (TNO) and the Netherlands Defence Academy (NLDA), reference no. TNO-10026587.

Appendix A. Derivation 1

We derive an upper bound in Eq. (33) (which is given in Box I).

$$
\begin{aligned}
&\mathbb{E}[(\tilde{\mathbf{H}}_{\mathrm{c}}-\hat{\tilde{\mathbf{H}}}_{\mathrm{c}})(\tilde{\mathbf{H}}_{\mathrm{c}}-\hat{\tilde{\mathbf{H}}}_{\mathrm{c}})^{H}]\\
&=\mathbb{E}\left[\left(\boldsymbol{\Psi}_{\mathrm{c}}^{H}\left(\mathbf{H}_{\mathrm{DD}}-\hat{\mathbf{H}}_{\mathrm{DD}}\right)\boldsymbol{\Phi}_{\mathrm{c}}\right)\left(\boldsymbol{\Psi}_{\mathrm{c}}^{H}\left(\mathbf{H}_{\mathrm{DD}}-\hat{\mathbf{H}}_{\mathrm{DD}}\right)\boldsymbol{\Phi}_{\mathrm{c}}\right)^{H}\right]\\
&=\mathbb{E}\left[\left(\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}(c_{l,q}-\hat{c}_{l,q})\boldsymbol{\Psi}_{\mathrm{c}}^{H}(\mathbf{F}_N\otimes\mathbf{I}_M)(\boldsymbol{\Lambda}_q\mathbf{P}^l)(\mathbf{F}_N^H\otimes\mathbf{I}_M)\boldsymbol{\Phi}_{\mathrm{c}}\right)\left(\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}(c_{l,q}-\hat{c}_{l,q})\boldsymbol{\Psi}_{\mathrm{c}}^{H}(\mathbf{F}_N\otimes\mathbf{I}_M)(\boldsymbol{\Lambda}_q\mathbf{P}^l)(\mathbf{F}_N^H\otimes\mathbf{I}_M)\boldsymbol{\Phi}_{\mathrm{c}}\right)^{H}\right]\\
&=\mathbb{E}\left[\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}|(c_{l,q}-\hat{c}_{l,q})|^2\underbrace{\boldsymbol{\Psi}_{\mathrm{c}}^{H}(\mathbf{F}_N\otimes\mathbf{I}_M)(\boldsymbol{\Lambda}_q\mathbf{P}^l)(\mathbf{F}_N^H\otimes\mathbf{I}_M)\boldsymbol{\Phi}_{\mathrm{c}}\left(\boldsymbol{\Psi}_{\mathrm{c}}^{H}(\mathbf{F}_N\otimes\mathbf{I}_M)(\boldsymbol{\Lambda}_q\mathbf{P}^l)(\mathbf{F}_N^H\otimes\mathbf{I}_M)\boldsymbol{\Phi}_{\mathrm{c}}\right)^{H}}_{\text{Diagonal matrix with } K_{\mathrm{c}} \text{ ones and } (R_{\mathrm{c}}-K_{\mathrm{c}}) \text{ zeros.}}\right]\\
&\preceq\mathbb{E}\left[\mathrm{tr}\left((\mathbf{c}-\hat{\mathbf{c}})(\mathbf{c}-\hat{\mathbf{c}})^{H}\right)\right]\mathbf{I}_{R_{\mathrm{c}}}
\end{aligned}
\tag{33}
$$

Box I.

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\tilde{\mathbf{H}}}_{\mathrm{c}}\hat{\tilde{\mathbf{H}}}_{\mathrm{c}}^{H}\right]=\mathbb{E}\left[\left(\boldsymbol{\Psi}_{\mathrm{c}}^{H}\hat{\mathbf{H}}_{\mathrm{DD}}\boldsymbol{\Phi}_{\mathrm{c}}\right)\left(\boldsymbol{\Psi}_{\mathrm{c}}^{H}\hat{\mathbf{H}}_{\mathrm{DD}}\boldsymbol{\Phi}_{\mathrm{c}}\right)^{H}\right]\\
&=\mathbb{E}\left[\left(\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\hat{c}_{l,q}\boldsymbol{\Psi}_{\mathrm{c}}^{H}(\mathbf{F}_N\otimes\mathbf{I}_M)(\boldsymbol{\Lambda}_q\mathbf{P}^l)(\mathbf{F}_N^H\otimes\mathbf{I}_M)\boldsymbol{\Phi}_{\mathrm{c}}\right)\left(\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\hat{c}_{l,q}\boldsymbol{\Psi}_{\mathrm{c}}^{H}(\mathbf{F}_N\otimes\mathbf{I}_M)(\boldsymbol{\Lambda}_q\mathbf{P}^l)(\mathbf{F}_N^H\otimes\mathbf{I}_M)\boldsymbol{\Phi}_{\mathrm{c}}\right)^{H}\right]\\
&=\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\sigma_{\hat{c}_{l,q}}^2\underbrace{\left(\boldsymbol{\Psi}_{\mathrm{c}}^{H}(\mathbf{F}_N\otimes\mathbf{I}_M)(\boldsymbol{\Lambda}_q\mathbf{P}^l)(\mathbf{F}_N^H\otimes\mathbf{I}_M)\boldsymbol{\Phi}_{\mathrm{c}}\right)\left(\boldsymbol{\Psi}_{\mathrm{c}}^{H}(\mathbf{F}_N\otimes\mathbf{I}_M)(\boldsymbol{\Lambda}_q\mathbf{P}^l)(\mathbf{F}_N^H\otimes\mathbf{I}_M)\boldsymbol{\Phi}_{\mathrm{c}}\right)^{H}}_{\text{Diagonal matrix with } K_{\mathrm{c}} \text{ ones and } (R_{\mathrm{c}}-K_{\mathrm{c}}) \text{ zeros.}}
\end{aligned}
\tag{34}
$$

Box II.

Appendix B. Derivation 2

We can write Eq. (35) (which is given in Box II). Moreover, due to the K_{c} ones on the diagonal, we have

$$\mathrm{tr}\left(\mathbb{E}\left[\hat{\tilde{\mathbf{H}}}_{\mathrm{c}}\hat{\tilde{\mathbf{H}}}_{\mathrm{c}}^{H}\right]\right)=K_{\mathrm{c}}\sum_{q=-Q/2}^{Q/2}\sum_{l=0}^{L}\sigma_{\hat{c}_{l,q}}^2.$$

Note that $\sigma_{\hat{c}_{l,q}}^2$ is defined as in (35).

$$
\begin{aligned}
\mathbb{E}[\hat{\mathbf{c}}\hat{\mathbf{c}}^{H}]&=\mathbb{E}\Big[\left(\sigma_{\mathbf{n}}^2\mathbf{R}_{\mathbf{c}}^{-1}+\mathbf{Z}^{H}\mathbf{Z}\right)^{-1}\mathbf{Z}^{H}\left(\mathbf{Z}\mathbf{c}+\mathbf{w}_{\mathrm{p}}\right)\left(\mathbf{Z}\mathbf{c}+\mathbf{w}_{\mathrm{p}}\right)^{H}\\
&\quad\times\mathbf{Z}\left(\sigma_{\mathbf{n}}^2\mathbf{R}_{\mathbf{c}}^{-1}+\mathbf{Z}^{H}\mathbf{Z}\right)^{-1}\Big]\\
&=\left(\sigma_{\mathbf{n}}^2\mathbf{R}_{\mathbf{c}}^{-1}+\mathbf{Z}^{H}\mathbf{Z}\right)^{-1}\mathbb{E}\left[\mathbf{Z}^{H}\left(\mathbf{Z}\mathbf{c}\mathbf{c}^{H}\mathbf{Z}^{H}+\mathbf{w}_{\mathrm{p}}\mathbf{w}_{\mathrm{p}}^{H}\right)\mathbf{Z}\right]\\
&\quad\times\left(\sigma_{\mathbf{n}}^2\mathbf{R}_{\mathbf{c}}^{-1}+\mathbf{Z}^{H}\mathbf{Z}\right)^{-1}\\
&=\left(\sigma_{\mathbf{n}}^2\mathbf{R}_{\mathbf{c}}^{-1}+\mathbf{Z}^{H}\mathbf{Z}\right)^{-1}\left[\mathbf{Z}^{H}\mathbf{Z}\mathbf{R}_{\mathbf{c}}\mathbf{Z}^{H}\mathbf{Z}+\mathbf{Z}^{H}\mathbf{R}_{\mathbf{w}_{\mathrm{p}}}\mathbf{Z}\right]\left(\sigma_{\mathbf{n}}^2\mathbf{R}_{\mathbf{c}}^{-1}+\mathbf{Z}^{H}\mathbf{Z}\right)^{-1}\\
&=\frac{P_{\mathrm{p}}^2\mathrm{diag}\left([\sigma_{c_{0,0}}^2,\ldots,\sigma_{c_{L,Q}}^2]\right)+P_{\mathrm{p}}\sigma_{\mathbf{n}}^2\mathbf{I}_{(L+1)(Q+1)\times(L+1)(Q+1)}}{\left(\sigma_{\mathbf{n}}^2\mathrm{diag}\left([\sigma_{c_{0,0}}^2,\ldots,\sigma_{c_{L,Q}}^2]\right)^{-1}+P_{\mathrm{p}}\mathbf{I}_{(L+1)(Q+1)\times(L+1)(Q+1)}\right)^{2}}\\
&=\mathrm{diag}\left(\left[\frac{P_{\mathrm{p}}\sigma_{c_{0,0}}^4}{\sigma_{\mathbf{n}}^2+\sigma_{c_{0,0}}^2P_{\mathrm{p}}},\ldots,\frac{P_{\mathrm{p}}\sigma_{c_{L,Q}}^4}{\sigma_{\mathbf{n}}^2+\sigma_{c_{L,Q}}^2P_{\mathrm{p}}}\right]\right)\\
&=\mathrm{diag}\left(\left[\sigma_{\hat{c}_{0,0}}^2,\ldots,\sigma_{\hat{c}_{L,Q}}^2\right]\right)
\end{aligned}
\tag{35}
$$

Data availability

No data was used for the research described in the article.

References

[1] Z. Wang, G. Giannakis, Wireless multicarrier communications, IEEE Signal Process. Mag. 17 (3) (2000) 29–48.

[2] R. Hadani, S.S. Rakib, OTFS methods of data channel characterization and uses thereof, 2016, US Patent 9 444 514 B2. [Online]. Available: https://patents.google.com/patent/US9444514B2.

[3] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A.J. Goldsmith, A.F. Molisch, R. Calderbank, Orthogonal time frequency space modulation, in: 2017 IEEE Wireless Commun. Networking Conf., WCNC, 2017, pp. 1–6.

[4] R. Hadani, S. Rakib, S. Kons, M. Tsatsanis, A. Monk, C. Ibars, J. Delfeld, Y. Hebron, A.J. Goldsmith, A.F. Molisch, R. Calderbank, Orthogonal time frequency space modulation, 2018, [Online]. Available: arXiv:1808.00519.

[5] T. Ebihara, K. Mizutani, Underwater acoustic communication with an orthogonal signal division multiplexing scheme in doubly spread channels, IEEE J. Ocean. Eng. 39 (1) (2014) 47–58.

[6] P. Raviteja, K.T. Phan, Y. Hong, E. Viterbo, Interference cancellation and iterative detection for orthogonal time frequency space modulation, IEEE Trans. Wirel. Commun. 17 (10) (2018) 6501–6515.

[7] X.-G. Xia, Precoded and vector OFDM robust to channel spectral nulls and with reduced cyclic prefix length in single transmit antenna systems, IEEE Trans. Commun. 49 (8) (2001) 1363–1374.

[8] J. Zhang, A.D.S. Jayalath, Y. Chen, Asymmetric OFDM systems based on layered FFT structure, IEEE Signal Process. Lett. 14 (11) (2007) 812–815.

[9] P. Raviteja, E. Viterbo, Y. Hong, OTFS performance on static multipath channels, IEEE Wirel. Commun. Lett. 8 (3) (2019) 745–748.

[10] X.-G. Xia, Comments on "the transmitted signals of OTFS and VOFDM are the same", IEEE Trans. Wirel. Commun. 21 (12) (2022) 11252–11252.

[11] I. van der Werf, H. Dol, K. Blom, R. Heusdens, R.C. Hendriks, G. Leus, On the equivalence of OSDM and OTFS, Signal Process. 214 (2024) 109254.

[12] P. Raviteja, K.T. Phan, Y. Hong, E. Viterbo, Embedded delay-Doppler channel estimation for orthogonal time frequency space modulation, in: 2018 IEEE 88th Veh. Technol. Conf., VTC-Fall, 2018, pp. 1–5.

[13] P. Raviteja, K.T. Phan, Y. Hong, Embedded pilot-aided channel estimation for OTFS in delay–Doppler channels, IEEE Trans. Veh. Technol. 68 (5) (2019) 4906–4917.

[14] H. Qu, G. Liu, L. Zhang, M.A. Imran, S. Wen, Low-dimensional subspace estimation of continuous-Doppler-spread channel in OTFS systems, IEEE Trans. Commun. 69 (7) (2021) 4717–4731.

[15] S. P. S., A. Farhang, A practical pilot for channel estimation of OTFS, in: ICC 2023 - IEEE Int. Conf. Commun., 2023, pp. 1319–1325.

[16] P. Raviteja, Y. Hong, E. Viterbo, E. Biglieri, Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS, IEEE Trans. Veh. Technol. 68 (1) (2019) 957–961.
[17] M.K. Tsatsanis, G.B. Giannakis, Modelling and equalization of rapidly fading channels, Int. J. Adapt. Control. Signal Process. 10 (2–3) (1996) 159–176.
[18] G. Leus, On the estimation of rapidly time-varying channels, in: 2004 12th European Signal Process. Conf., 2004, pp. 2227–2230.
[19] D. Borah, B. Hart, Frequency-selective fading channel estimation with a polynomial time-varying channel model, IEEE Trans. Commun. 47 (6) (1999) 862–873.
[20] K. Teo, S. Ohno, Optimal MMSE finite parameter model for doubly-selective channels, in: GLOBECOM '05. IEEE Global Telecommun. Conf., 2005, vol. 6, 2005, p. 5, pp.–3507.
[21] T. Zemen, C. Mecklenbrauker, Time-variant channel estimation using discrete prolate spheroidal sequences, IEEE Trans. Signal Process. 53 (9) (2005) 3597–3607.
[22] K.R. Murali, A. Chockalingam, On OTFS modulation for high-Doppler fading channels, in: 2018 Inf. Theory and Applicat. Workshop, ITA, 2018, pp. 1–10.
[23] L. Jing, H. Wang, C. He, Y. Zhang, H. Yin, Two dimensional adaptive multichannel decision feedback equalization for OTFS system, IEEE Commun. Lett. 25 (3) (2021) 840–844.
[24] Y. Liu, Y.L. Guan, D. González G., Near-optimal BEM OTFS receiver with low pilot overhead for high-mobility communications, IEEE Trans. Commun. 70 (5) (2022) 3392–3406.
[25] H.B. Mishra, P. Singh, A.K. Prasad, R. Budhiraja, OTFS channel estimation and data detection designs with superimposed pilots, IEEE Trans. Wirel. Commun. 21 (4) (2022) 2258–2274.
[26] S. Adireddy, L. Tong, H. Viswanathan, Optimal placement of training for frequency-selective block-fading channels, IEEE Trans. Inform. Theory 48 (8) (2002) 2338–2353.
[27] M. Dong, L. Tong, Optimal design and placement of pilot symbols for channel estimation, IEEE Trans. Signal Process. 50 (12) (2002) 3055–3069.
[28] R. Negi, J. Cioffi, Pilot tone selection for channel estimation in a mobile OFDM system, IEEE Trans. Consum. Electron. 44 (3) (1998) 1122–1128.
[29] S. Ohno, G. Giannakis, Optimal training and redundant precoding for block transmissions with application to wireless OFDM, IEEE Trans. Commun. 50 (12) (2002) 2113–2123.
[30] X. Ma, G. Giannakis, S. Ohno, Optimal training for block transmissions over doubly selective wireless fading channels, IEEE Trans. Signal Process. 51 (5) (2003) 1351–1366.
[31] A. Kannu, P. Schniter, MSE-optimal training for linear time-varying channels, in: Proceedings. (ICASSP '05). IEEE Int. Conf. Acoust., Speech, and Signal Process., 2005, vol. 3, 2005, pp. 789–792.
[32] A.P. Kannu, P. Schniter, Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels, IEEE Trans. Signal Process. 56 (3) (2008) 1148–1160.
[33] T. Ebihara, G. Leus, Doppler-resilient orthogonal signal-division multiplexing for underwater acoustic communication, IEEE J. Ocean. Eng. 41 (2) (2016) 408–427.
[34] X. Li, P. Fan, Modified clustered comb pilot-aided fast time-varying channel estimation for OFDM system, J. Mod. Transp. 20 (2012) 220–226.
[35] K.M.Z. Islam, T.Y. Al-Naffouri, N. Al-Dhahir, On optimum pilot design for comb-type OFDM transmission over doubly-selective channels, IEEE Trans. Commun. 59 (4) (2011) 930–935.
[36] S. Ohno, G. Giannakis, Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels, IEEE Trans. Inform. Theory 50 (9) (2004) 2138–2145.
[37] R. Marsalek, J. Blumenstein, A. Prokes, T. Gotthans, Orthogonal time frequency space modulation: Pilot power allocation and nonlinear power amplifiers, in: 2019 IEEE Int. Symp. Signal Process. Inf. Technol., ISSPIT, 2019, pp. 1–4.
[38] E. Telatar, Capacity of multi-antenna Gaussian channels, Eur. Trans. Telecommun. 10 (6) (1999) 585–595.