OTFS vs. OFDM in the Presence of Sparsity: A Fair Comparison

Lorenzo Gaudio, *Graduate Student Member, IEEE*, Giulio Colavolpe, *Senior Member, IEEE*, and Giuseppe Caire, *Fellow, IEEE*

***Abstract*— Many recent works in the literature declare that Orthogonal Time-Frequency-Space (OTFS) modulation is a promising candidate technology for high mobility communication scenarios. However, a truly fair comparison with its direct concurrent and widely used Orthogonal Frequency-Division Multiplexing (OFDM) modulation has not yet been provided. In this paper, we present such a fair comparison between the two digital modulation formats in terms of achievable communication rate. In this context, we explicitly address the problem of channel estimation by considering, for each modulation, a pilot scheme and the associated channel estimation algorithm specifically adapted to sparse channels in the Doppler-delay domain, targeting the optimization of the pilot overhead to maximize the overall achievable rate. In our achievable rate analysis we consider also the presence of a guard interval or cyclic prefix. The results are supported by numerical simulations, for different time-frequency selective channels including multiple scattering components and under non-perfect channel state information resulting from the considered pilot schemes. This work does not claim to establish in a fully definitive way which is the best modulation format, since such choice depends on many other features which are outside the scope of this work (e.g., legacy, intellectual property, ease and know-how for implementation, and many other criteria). Nevertheless, we provide the foundations to properly compare multi-carrier communication systems in terms of their information theoretic achievable rate potential, within meaningful and sensible assumptions on the channel models and on the receiver complexity (both in terms of channel estimation and in terms of soft-output symbol detection).**

***Index Terms*— OFDM, OTFS, compressed sensing, Doppler-delay domain, sparsity, pragmatic capacity, channel state information.**

Manuscript received November 1, 2020; revised April 8, 2021 and August 26, 2021; accepted November 18, 2021. This work was supported by MIUR under the PRIN Liquid_Edge Contract. An earlier version of this paper titled "On achievable rate of OFDM and OTFS in the presence of sparsity" was presented in part at the IEEE International Conference on Communications 2021, Montreal, Canada [DOI: 10.1109/ICCWorkshops50388.2021.9473825]. The associate editor coordinating the review of this article and approving it for publication was C. Zhong. *(Corresponding author: Giulio Colavolpe.)*

Lorenzo Gaudio was with the Department of Engineering and Architecture, University of Parma, 43124 Parma, Italy (e-mail: lorenzo.gaudio@studenti.unipr.it).

Giulio Colavolpe is with the Department of Engineering and Architecture, University of Parma, 43124 Parma, Italy, and also with the CNIT Research Unit, 43124 Parma, Italy (e-mail: giulio.colavolpe@unipr.it).

Giuseppe Caire is with the Department of Electrical Engineering and Computer Science, Technical University of Berlin, 10623 Berlin, Germany (e-mail: caire@tu-berlin.de).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3129975.

Digital Object Identifier 10.1109/TWC.2021.3129975

I. Introduction

In ANY communication scenario, the channel state information (CSI), i.e., the knowledge of the communication channel, is required at the receiver in order to perform coherent detection [1]. The most common approach to acquire CSI is through the transmission of known pilot symbols [1]. Generally, these pilots are arranged within the block of information symbols following a chosen fixed pattern known to both transmitter and receiver (see e.g, [2]–[4]).[1] It is also well-known that, subject to the meaningful and widely used assumption of *block fading* (i.e., the propagation channel remains constant over blocks of consecutive time-domain symbols, and may change independently from block to block), pilot-aided schemes are indeed nearly information-theoretically optimal in terms of the capacity scaling in the high spectral efficiency/high signal-to-noise ratio (SNR) regime (see, e.g., [5]–[8]). On the other hand, both the pilot pattern and the channel estimation algorithm should be optimized for the particular communication scenario. By considering a noisy channel, the CSI is inevitably affected by an estimation error, whose magnitude depends on the channel SNR and number of pilots per block. The estimated channel coefficients are then used to perform coherent detection at the receiver side. Taking as the meaningful and most relevant performance metric the achievable communication rate, i.e., the amount of useful information sent in a block of symbols, there is a tension between the number of pilot symbols per block dedicated to CSI estimation and the number of information-bearing symbols. The optimization of this tradeoff is generally not trivial and depends on the modulation format and on the channel propagation characteristics.

In this work, we compare two digital modulation formats designed to handle *time-frequency selective* channels, namely orthogonal time frequency space (OTFS) [9] and orthogonal frequency division multiplexing (OFDM), in terms of pragmatic capacity, i.e., the achievable rate of the channel induced by the signal constellation and the detector

[1] In some communication scenarios the transmitter could send an entire block of pilot symbols, i.e., without any useful information, followed by blocks with no pilots. Thus, the symbol detection is based on the channel estimation made from the first block. In this work we are not considering such a case, by focusing on a *per block* channel estimation, with associated benefits and losses.

soft-output [10], [11].[2] In general, a soft-output detector at the receiver side produces an estimate of the posterior probability of the transmitted symbols given the received signal block (pilots and data). This estimated posterior probability (e.g., in the form of log-likelihood ratios) is then passed to a decoder, that treats the sequence of soft-output symbols as the output of a *virtual channel*. The pragmatic capacity is the capacity of such virtual channel, with discrete input represented by the modulation symbols, and soft-output generated by the detector. Hence, the pragmatic capacity is representative of the achievable rate under the assumption of *separate detection and decoding*, i.e., without "turbo" reprocessing of the decoder output (see, e.g., [13], [14] and references therein). In practice, iterative "turbo" detection is very hard to implement since often the detector is implemented in hardware (e.g., in an integrated circuit) and the decoder is implemented in software, and maybe even in a different location (as for example in the so-called 7.2 split between hardware and software, enabling cloud-based processing of the signals from remote radio heads [15]). For this reason, we believe that the pragmatic capacity for separated detection and decoding is a very meaningful performance metric to compare different modulation formats and the associated pilot schemes and soft-output detectors.

In order to make a fair comparison between the two modulation formats, we take into account the pilot overhead resulting from the optimization of the above mentioned trade-off. Since the loss associated to the presence of pilots cannot be neglected, the achievable communication rate inevitably deviates from the upper limit of additive white Gaussian noise (AWGN) capacity for the same constellation format. Moreover, we also consider the presence of a guard interval (GI) or cyclic prefix (CP), which additionally reduces the achievable rate. One GI for each block is used in OTFS to avoid inter-block interference (IBI) [16], while a CP is used for every OFDM symbol to avoid inter-symbol interference (ISI) and to make symbols orthogonal in the time-frequency domain. Here a first significant difference between the two modulation formats evidently appears. In fact, in order to accommodate typical channel delay spreads, the CP length in several communication standards based on OFDM may be a large fraction (up to 25%) of the symbol time (see, e.g., [17] and references therein), leading to a remarkable loss in terms of capacity. On the other hand, OTFS does not need a per symbol separation, but this comes at the cost of a non-negligible increase in signal processing complexity [18].[3]

The time-frequency selective channels of interest mainly target outdoor scenarios with high mobility, since OTFS has been specifically proposed for these cases, where the Doppler spreads are significant and where there are few reflectors (or group of reflectors with similar properties), and thus a small number of multipath components [9]. Since the properties of the communication channel represented in the Doppler-delay domain depend on the physical geometry of the environment, the scattering components are thus sparse in the Doppler-delay plane.[4] In this context, exploiting the channel sparsity plays a fundamental role since estimation algorithms built over this concept exhibit very good tradeoffs between pilot overhead, complexity, and estimation error.

The idea of exploiting the channel sparsity in multi-carrier systems is well established and shared by other estimation techniques in the current literature, which generally applies concepts from compressed sensing (CS) (e.g., see [19]–[26]), for both OFDM and OTFS modulations. However, it should be noticed that CS is meaningful when the measurements (i.e., the pilot symbols) are obtained in the dual domain with respect to the domain in which the channel is sparse. Thus, in our case, the channel is sparse in the Doppler-delay domain, for which the dual domain (related by a two-dimensional Fourier transform) is the time-frequency domain. In the considered pilot scheme for OTFS, the pilots are directly placed in the sparsity domain (Doppler-delay). Hence, direct estimation via the maximum likelihood (ML) approach is both efficient and computationally feasible. In contrast, the OFDM modulation format places the pilot symbols in the time-frequency domain. Therefore, in this case it makes sense to consider a CS-based channel estimation approach.

Based on the aforementioned discussion, the proposed channel estimation algorithm for OTFS takes into account a pilot scheme similar to the one proposed in [4] (adopted in some other works, see, e.g., [27]), which considers a high energy center pilot (or a cluster of pilots [19], to contrast eventual destructive non-linear amplification effects over the single pilot) surrounded by zeros in the transmitted two-dimensional (Doppler-delay domain) block of symbols. This configuration of pilots and information symbols is a natural consequence of the input-output relation of OTFS, where the effect of the channel cause a cyclic shift of the transmitted symbol of a quantity proportional to the delay and Doppler shifts associated to each channel path. The estimation algorithm for OTFS is based on the ML approach of [18], with the introduction of a low-complexity pilot-based mechanism to detect channel multipath components and achieve a first coarse estimation of relevant parameters. On the other hand, for OFDM we adopt a well-known CS-based estimation algorithm based on least-absolute shrinkage and selection operator (LASSO) [28] (i.e., l_1-norm regularized least squares minimization). In this case, the pilots placed in the time-frequency domain define the sensing matrix, and the proposed estimation algorithm makes use of the soft-thresholding iterative algorithm [29], optimized to efficiently work in our system setup [30]–[32]. The pilot configuration for OFDM is discussed in Sec. II.

The paper is organized as follows. In Sec. II and in Sec. III we present the input-output relation, the used pilot scheme, and the proposed channel estimation algorithm for OFDM

[2]Indeed, the analysis can be extended to the many multicarrier modulation formats proposed so far in the literature (e.g., see [12]). However, we consider here OFDM only since by far the most widely used scheme in modern standards.

[3]There exist OFDM-based OTFS systems, e.g., [19], which consider "OTFS-like" operations as precoding and equalization of an inner OFDM system. We will not consider such a type of systems, which deviate from the scope of this work.

[4]A richer scattering environment such as indoor scenarios, could break the sparsity constraint with subsequent problems for the proposed estimation algorithms. Thus, the proposed methods mainly focus outdoor scenarios with channels represented by few sparse components.

and OTFS modulation, respectively. Sec. IV discusses the numerical results, while Sec. V concludes the paper.

II. OFDM Modulation and the CS Algorithm

We consider OFDM with per-symbol CP transmitted over a time-frequency selective channel, assuming perfect symbol orthogonality and absence of inter-carrier interference (ICI). The channel impulse response (CIR) in time-frequency domain is given by [33]

$$H(t,f) = \sum_{p=0}^{P-1} h_p e^{j2\pi\nu_p t} e^{-j2\pi\tau_p f}, \tag{1}$$

where P is the number of multipath scattering components and h_p, ν_p, and τ_p are the complex channel gain including the pathloss, the Doppler shift, and the delay, associated to the p-th scattering component, respectively. Note that the maximum channel delay and Doppler shift are assumed to satisfy

$$\tau_{\max} < T, \quad \nu_{\max} < \Delta f, \tag{2}$$

where T is the symbol time and Δf is the subcarrier spacing. By discretizing the time axis at steps nT, for $n = 0, \ldots, N-1$, and the frequency axis at steps $m\Delta f$, for $m = 0, \ldots, M-1$, we obtain the discrete time-frequency channel matrix

$$\mathbf{H} = \sum_{p=0}^{P-1} h_p \mathbf{a}(\nu_p)\, \mathbf{b}^{\mathsf{H}}(\tau_p), \tag{3}$$

where

$$\mathbf{a}(\nu_p) = \left[1, e^{j2\pi\tau_p T}, \ldots, e^{j2\pi\tau_p (N-1)T}\right]^{\mathsf{T}}, \tag{4}$$

$$\mathbf{b}(\tau_p) = \left[1, e^{j2\pi\nu_p \Delta f}, \ldots, e^{j2\pi\nu_p (M-1)\Delta f}\right]^{\mathsf{T}}, \tag{5}$$

where $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ denote the transpose and the conjugate transpose (Hermitian) operation, respectively. By gathering the information symbols $\{x_{n,m}\}$ in a $N \times M$ matrix $\mathbf{X}$, the expression of the received samples after transmission over the channel in (3) is

$$\mathbf{Y} = \mathbf{X} \odot \mathbf{H} + \mathbf{Z}, \tag{6}$$

where $\odot$ denotes element-wise multiplication and $\mathbf{Z}$ is the AWGN with zero mean and covariance matrix $\sigma^2 \mathbf{I}_{NM}$. Information symbols may belong to any suitable complex modulation alphabet.

At this point, in order to help understanding the rationale behind the adopted channel estimation algorithm, let us represent the channel in a different form. Suppose to define a Doppler-delay grid Γ, with some grid steps (both in the Doppler and delay domain) and total dimension G, given by the product between Doppler and delay axis dimensions. For each grid point $\gamma_i \in \Gamma$, for $i = 0, \ldots, G-1$, the corresponding rank-1 channel component can be expressed, similarly to (3), as

$$\tilde{\mathbf{H}}(\gamma_i) = \mathbf{a}(\tau(\gamma_i))\, \mathbf{b}^{\mathsf{H}}(\nu(\gamma_i)), \tag{7}$$

in which $\tau(\gamma_i)$ and $\nu(\gamma_i)$ are two fixed values of delay and Doppler depending on the discretization point γ_i. By stacking the $N \times M$ matrices $\tilde{\mathbf{H}}(\gamma_i)$ to column vectors for all points $\gamma_i \in \Gamma$ (vec$(\cdot)$ operator) and concatenating the obtained vectors, we create a generally overcomplete dictionary matrix

$$\mathbf{D} = \left[\mathrm{vec}\left(\tilde{\mathbf{H}}(\gamma_0)\right), \ldots, \mathrm{vec}\left(\tilde{\mathbf{H}}(\gamma_{G-1})\right)\right], \tag{8}$$

of dimension $NM \times G$. We also define a $G \times 1$ vector $\mathbf{h}_{\mathrm{sp}}$ (read: "*h-sparse*"), representing the channel gains corresponding to each discrete Doppler and delay component $\gamma_i \in \Gamma$. Thus, the vectorized channel matrix $\bar{\mathbf{H}}$, which is an approximation of the true channel matrix $\mathbf{H}$ in (3), takes on the form

$$\mathrm{vec}(\bar{\mathbf{H}}) = \mathbf{D}\mathbf{h}_{\mathrm{sp}}. \tag{9}$$

Since the true channel contains only a small number $P \ll G$ of Doppler-delay components, the grid coefficient vector $\mathbf{h}_{\mathrm{sp}}$ is a sparse, in the form

$$\mathbf{h}_{\mathrm{sp}} = \left[0, \ldots, 0, \bar{h}_0, 0, \ldots, 0, \bar{h}_1, 0, \ldots, 0, \bar{h}_2, 0, \ldots\right]^{\mathsf{T}}, \tag{10}$$

where the positions of the approximated (to the nearest grid step) channel coefficients $\bar{h}_i$ select the columns of $\mathbf{D}$ with the pair of channel coefficients $(\tau(\gamma_i), \nu(\gamma_i))$, to overall represent the triplet $\left(\tau(\gamma_i), \nu(\gamma_i), \bar{h}_p\right)$ emulating the true channel parameters (τ_p, ν_p, h_p). Thus, the approximated received samples expression in vectorized form can be written as

$$\bar{\mathbf{y}} = \mathbf{x} \odot (\mathbf{D}\mathbf{h}_{\mathrm{sp}}) + \mathbf{z}, \tag{11}$$

where $\mathbf{x} = \mathrm{vec}(\mathbf{X})$ and $\mathbf{z} = \mathrm{vec}(\mathbf{Z})$. Moreover, by defining a selection matrix $\mathbf{S}$, of dimension $|\mathcal{P}| \times NM$, to choose $|\mathcal{P}|$ symbols (pilots) among the total NM ($\mathcal{P}$ is the set of pilots and $|\cdot|$ indicates its cardinality), the transmitted vector of pilots $\mathbf{x}_{\mathrm{pl,OFDM}}$ (read: "*x-pilot*") takes the form

$$\mathbf{x}_{\mathrm{pl,OFDM}} = \mathbf{S}\mathbf{x}, \tag{12}$$

and

$$\bar{\mathbf{y}}_{\mathrm{pl}} = \mathbf{x}_{\mathrm{pl,OFDM}} \odot (\mathbf{S}\mathbf{D}\mathbf{h}_{\mathrm{sp}}) + \mathbf{z}. \tag{13}$$

The estimation of the channel coefficients can be carried on by solving the problem known as LASSO [28], i.e.,

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \left\|\mathbf{y}_{\mathrm{pl}} - \mathbf{x}_{\mathrm{pl,OFDM}} \odot (\mathbf{S}\mathbf{D}\mathbf{h})\right\|_2^2 + \lambda \left\|\mathbf{h}\right\|_1, \tag{14}$$

where $\mathbf{S}\mathbf{D}$, under this configuration, takes the role of the sensing matrix of the CS configuration, and λ is the LASSO regularizer (see details in Appendix A). Notice that $\mathbf{y}_{\mathrm{pl}}$ is obtained through the transmission over the *actual* channel, and thus differs from $\bar{\mathbf{y}}_{\mathrm{pl}}$, obtained through the process of approximation of the channel matrix described before. The incurred approximation error should be kept small, by choosing appropriately the grid Γ with sufficiently fine discretization. In any case, the residual approximation error between $\mathbf{y}_{\mathrm{pl}}$ and $\bar{\mathbf{y}}_{\mathrm{pl}}$ is automatically included in the minimization or the overall quadratic error term $\left\|\mathbf{y}_{\mathrm{pl}} - \mathbf{x}_{\mathrm{pl,OFDM}} \odot \mathbf{S}\mathbf{D}\mathbf{h}\right\|_2^2$ in (14).

The LASSO minimization problem has been extensively studied in literature. It can be solved using many different algorithms [29], [30], [32], and it was also adopted for the specific case of channel estimation [20]. As a final outcome, the estimated channel matrix resulting from the minimization

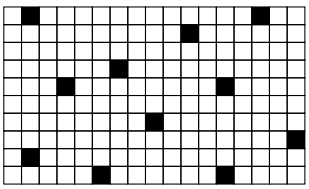


Fig. 1. Example of a random pilot scheme for OFDM modulation.

of (14) and subsequently passed to the symbol detector is given by

$$\hat{\mathbf{H}} = \mathbf{D}\hat{\mathbf{h}}. \tag{15}$$

For completeness and for the sake of reproducibility of our results, the details of the used LASSO solver, together with an analysis on its complexity, are described in Appendix A.

A. Pilot Scheme

The optimization of a deterministic sensing matrix for CS problems such as LASSO is up to now one of the most studied open problems in CS theory. In fact, the typical performance guarantees of CS require properties such as the restricted isometry property [20], [34], for which explicit constructions are not available and even checking the property for a given randomly generated matrix is exponentially complex [35]. On the other hand, ensembles of randomly generated matrices have the property of satisfying these properties with high probability [20]. Hence, here we resort to a pseudo-random pilot placement on the 2-dimensional time-frequency grid. We have verified by simulation that such random placement achieves with high probability the best performance with respect to regular "lattice" placements (e.g., equally spaced combs of subcarriers) usually specified in wireless standards [3]. An example of a random pilot scheme is depicted in Fig. 1. Generally, almost every configuration of a fixed number of pilots randomly placed within the 2-dimensional grid provides similar performance in terms of channel estimation. If pilots are not placed randomly but follow some periodic pattern, the algorithm for solving the LASSO produces far inferior results. This behavior is caused by the periodic sampling of a random Fourier matrix (i.e., $\mathbf{H}$ or $\mathbf{D}\mathbf{h}_{\mathrm{sp}}$). This is the reason why commonly used pilot schemes (see, e.g., [3] and references therein), generally structured or periodic, are not suitable for the CS-based estimation of OFDM systems (assuming that the OFDM channel is represented by a Fourier matrix). In fact, it is well-known that deterministic contribution of sensing matrices with small restricted isometry property (RIP) constant is a very hard problem (see [36] and references therein). In contrast, the RIP constant and therefore CS reconstruction guarantees of random sensing matrices have been extensively analyzed in the theoretical CS literature. In particular, it is well-known that "random DFT" matrices (i.e., matrices obtained by a random selection of the rows of a unitary DFT matrix) have very good properties, analyzed for example in the classical papers [37] and [38]. While there is no result saying that a pseudo-random selection of the pilot symbols position is the best possible (in fact, by definition, there must be a deterministic choice that performs better than average, by the usual random coding argument), an explicit algorithmic construction that performs better than random selection has not yet been found. This is a well-known open problem in compressed sensing. Since the focus of this paper is not to exhibit the best possible scheme, and random selection has theoretical guarantees of good performance with high probability, we have resorted to random selection. Moreover, with CS, the number of measurements, i.e., pilot symbols, is much less than the dimension of the target signal that we want to estimate. In this regime, standard LS-type channel estimation would fail since the underlying regression problem is under-determined. This is precisely the regime where novel compressed sensing schemes as the LASSO scheme used in our paper (which is pretty much the state of the art for noisy measurements as in our case) allow good estimation with a much reduced number of pilot symbols.

We aim at maximizing the overall achievable rate under random pilot placement. Hence, we can optimize the number of pilots per block to seek the optimal tradeoff between CSI estimation quality and pilot overhead (see (40) in the following and numerical results in Sec. IV).

B. Received Samples Expression — Real and Approximated Channel Conditions

Without entering into details of the complete input-output derivation of a CP OFDM system which can be found, e.g., in [39], we only provide the received sample expression. By considering real and approximated channel conditions, the received samples at time instant n and subcarrier m are respectively given by (16) and (17), shown at the bottom of the page, in which the ICI-free approximation follows the assumption $\nu_{\max}/\Delta f \ll 1$, and the last equality follows by using the orthogonality property. Note that (17) is equivalent

$$y[n,m] = \frac{1}{M}\sum_{p=0}^{P-1} h_p e^{j2\pi\nu_p nT} \sum_{m'=0}^{M-1} x[n,m'] e^{-j2\pi m'\Delta f\tau_p} \sum_{i=0}^{M-1} e^{j2\pi\frac{i}{M}\frac{\nu_p}{\Delta f}} e^{j2\pi\frac{i(m'-m)}{M}} \tag{16}$$

$$\approx \frac{1}{M}\sum_{p=0}^{P-1} h_p e^{j2\pi\nu_p nT} \sum_{m'=0}^{M-1} x[n,m'] e^{-j2\pi m'\Delta f\tau_p} \sum_{i=0}^{M-1} e^{j2\pi\frac{i(m'-m)}{M}}$$

$$= \sum_{p=0}^{P-1} h_p e^{j2\pi\nu_p nT} e^{-j2\pi\tau_p m\Delta f} x[n,m] \tag{17}$$

to (6). For numerical simulations we generally consider the approximated channel model (17) in which the assumption on the absence of ICI holds. However, we also show the results with a comparison of OFDM under real channel conditions, i.e., considering (16), showing the performance degradation in the case where the ICI-free assumption is not well satisfied.

III. OTFS Modulation and the Proposed Estimation Algorithm

A. OTFS Input-Output Relation

In this section we concisely review the derivation of the input-output relation of OTFS modulation (see, e.g., [18], [40]) and cast it in the notation of this paper.

Data symbols $\{x_{k,l}\}$, for $k = 0, \ldots, N-1$ and $l = 0, \ldots, M-1$ belonging to any modulation format (e.g., some quadrature amplitude modulation (QAM) constellation), are arranged in an $N \times M$ two-dimensional grid referred to as the Doppler-delay domain, i.e., $\Gamma = \{(k/NT, l/M\Delta f)\}$ for $k = 0, \ldots, N-1$ and $l = 0, \ldots, M-1$ (see (6) for the analogy with OFDM). The transmitter first applies the inverse symplectic finite Fourier transform (ISFFT) to convert data symbols $\{x_{k,l}\}$ into a block of samples $\{X[n,m]\}$ in the (dual) time-frequency domain, thus

$$X[n,m] = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x_{k,l} e^{j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)}, \tag{18}$$

for $n = 0, \ldots, N-1$ and $m = 0, \ldots, M-1$. Then, it generates the continuous-time signal

$$s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X[n,m] g_{\mathrm{tx}}(t - nT) e^{j2\pi m\Delta f(t-nT)}, \tag{19}$$

where $X[n,m]$ denotes the sample sent at time n and over subcarrier m, and $g_{\mathrm{tx}}(t)$ is a generic transmit shaping pulse. After transmission over the channel defined in (1), the continuous received signal without noise is

$$r(t) = \sum_{p=0}^{P-1} h_p s(t - \tau_p) e^{j2\pi\nu_p t}, \tag{20}$$

and the output of the receiver filter-bank adopting a generic receive shaping pulse $g_{\mathrm{rx}}(t)$ is given in (21), shown at the bottom of the page. By sampling at $t = nT$ and $f = m\Delta f$, we obtain (22), shown at the bottom of the page, where the time-frequency domain channel $h_{n,m}[n', m']$ is given in (23), shown at the bottom of the page, having defined the cross ambiguity function $C_{u,v}(\tau,\nu) \triangleq \int_{-\infty}^{\infty} u(s) v^*(s-\tau) e^{-j2\pi\nu s} ds$ as in [41], set $h'_p = h_p e^{j2\pi\nu_p\tau_p}$, and imposed the term $e^{-j2\pi mn'\Delta fT} = 1$, $\forall n', m$, which is always true under the assumption $T = 1/\Delta f$. Since $X[n,m]$ is generated via ISFFT, the received signal in the Doppler-delay domain is obtained by the application of the symplectic finite Fourier transform (SFFT)

$$y[k,l] = \sum_{n,m} \frac{y[n,m]}{NM} e^{j2\pi\left(\frac{ml}{M} - \frac{nk}{N}\right)} = \sum_{k',l'} x_{k',l'} g_{k,k'}[l,l'], \tag{24}$$

where the ISI coefficient of the Doppler-delay pair $[k', l']$ seen by sample $[k,l]$ is given by

$$g_{k,k'}[l,l'] = \sum_p h'_p \Psi^p_{k,k'}[l,l'], \tag{25}$$

with $\Psi^p_{k,k'}[l,l']$ defined as in (26), shown at the bottom of the page.

Using rectangular shaping pulses and after a suitable approximation of the cross ambiguity function as illustrated in the derivation in [18] and omitted here for the sake of brevity, the simplified version of $\Psi^p_{k,k'}[l,l']$ is given by (27), shown at the bottom of the page.

$$\begin{aligned} y(t,f) &= \int r(t') g^*_{\mathrm{rx}}(t'-t) e^{-j2\pi ft'} dt' = \int_{t'} g^*_{\mathrm{rx}}(t'-t) \sum_{p=0}^{P-1} h_p s(t'-\tau_p) e^{j2\pi\nu_p t'} e^{-j2\pi ft'} dt' \\ &= \sum_{p,n',m'} h_p X[n',m'] \int_{t'} g^*_{\mathrm{rx}}(t'-t) g_{\mathrm{tx}}(t'-\tau_p - n'T) e^{j2\pi m'\Delta f(t'-\tau_p-n'T)} e^{j2\pi(\nu_p - f)t'} dt' \end{aligned} \tag{21}$$

$$y[n,m] = y(t,f)|_{t=nT, f=m\Delta f} = \sum_{n'=0}^{N-1} \sum_{m'=0}^{M-1} X[n',m'] h_{n,m}[n',m'] \tag{22}$$

$$h_{n,m}[n',m'] = \sum_{p=0}^{P-1} h'_p C_{g_{\mathrm{tx}},g_{\mathrm{rx}}}((n-n')T - \tau_p, (m-m')\Delta f - \nu_p) e^{j2\pi n'T\nu_p} e^{-j2\pi m\Delta f\tau_p} \tag{23}$$

$$\Psi^p_{k,k'}[l,l'] = \sum_{n,n',m,m'} \frac{C_{g_{\mathrm{rx}},g_{\mathrm{tx}}}((n-n')T-\tau_p,(m-m')\Delta f-\nu_p)}{NM} e^{j2\pi n'T\nu_p} e^{-j2\pi m\Delta f\tau_p} e^{j2\pi\left(\frac{n'k'}{N} - \frac{m'l'}{M}\right)} e^{-j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)} \tag{26}$$

$$\begin{aligned} \Psi^p_{k,k'}[l,l'] \simeq \frac{1}{NM} & \frac{1 - e^{j2\pi\left(k'-k+\nu_p NT\right)}}{1 - e^{j2\pi\frac{\left(k'-k+\nu_p NT\right)}{N}}} \frac{1 - e^{j2\pi\left(l'-l+\tau_p M\Delta f\right)}}{1 - e^{j2\pi\frac{\left(l'-l+\tau_p M\Delta f\right)}{M}}} e^{-j2\pi\nu_p\frac{l'}{M\Delta f}} \\ & \times \begin{cases} 1 & \text{if } \; l' \in [0, M-1-\lceil \tau_p/(T/M) \rceil] \\ e^{-j2\pi\left(\frac{k'}{N} + \nu_p T\right)} & \text{if } \; l' \in [M - \lceil \tau_p/(T/M) \rceil, M-1] \end{cases} \end{aligned} \tag{27}$$

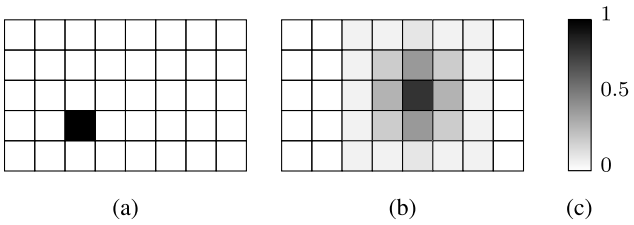


Fig. 2. Symbol (magnitude) translation in the Doppler-delay domain. (a) Transmitted symbol. (b) Received samples. (c) Heatmap (magnitude).

B. Pilot Scheme

In this section we describe the pilot scheme inspired by [4]. Using (24) expressed in matrix form, the OTFS input-output relation is given by

$$\mathbf{y} = \left(\sum_{p=0}^{P-1} h_p \mathbf{\Psi}_p \right) \mathbf{x} + \mathbf{z}, \tag{28}$$

where $\mathbf{z}$ denotes the AWGN with zero mean and covariance matrix $\sigma^2 \mathbf{I}_{NM}$. Note that $\mathbf{\Psi}_p$ implicitly takes into account a Doppler-delay pair (τ_p, ν_p), i.e., $\mathbf{\Psi}_p \triangleq \mathbf{\Psi}_p(\tau_p, \nu_p)$. Without entering into mathematical details, the effect of the channel to symbols arranged in blocks is described in the following.

Consider a block composed by all zero symbols but one non-zero with enough energy to be well distinguishable and positioned anywhere within the block (the position of the symbol does not influence the result, since the channel shift effect is circular within the block), transmitted over the time-frequency selective channel in (3). At the receiver, most of the energy concentrates at positions on the 2-dimensional block (one per multipath component), with some diffusion to the surrounding positions according to the Dirichlet kernel functions appearing in the OTFS channel matrix expression in (27) (see [18] and references therein for more details). Examples of blocks of transmitted symbols and received samples in the case of a single multipath component are depicted in Fig. 2 (a) and Fig. 2 (b), respectively. Since the channel is composed of P components and is linear, in general the resulting received signal is formed by the superposition of the effects of P multipath components, i.e., a single transmit symbol will be shifted in P different positions, each of which has some surrounding diffusion as qualitatively shown in Fig. 2. Intuitively, the estimation of the pairs (τ_p, ν_p) follows by searching the peaks of the magnitude of the received samples grid (as suggested in [4]). This intuitive estimation procedure is, however, only able to provide the integer parts of the Doppler and delay shifts, associated to the Doppler-delay grid point collecting the maximum energy. The fractional parts are linked to the dissipation of the energy around the peak points and must be treated separately, as we will do via Algorithm 1 and Algorithm 2 in the following.

The approximation of the channel behavior to integer Doppler and delay shifts, as it was done in [4], yields accurate results only under the non-realistic conditions of integer Doppler and delay shifts. Thus, based on our previously proposed ML estimator in [18], we will extend the idea of [4] providing a reliable estimation algorithm for general sparse channels with non-integer shifts.

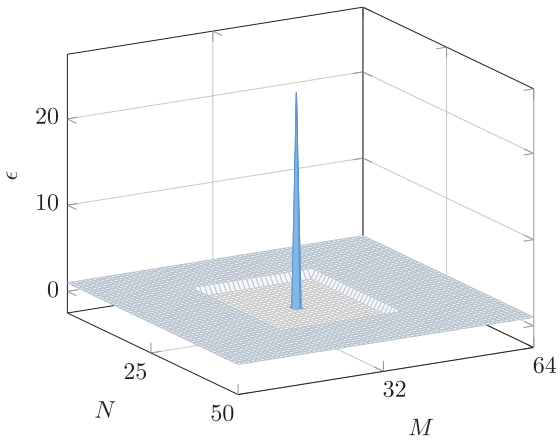


Fig. 3. Example of a pilot scheme for OTFS modulation. Within the block of dimension $N \times M$ it is well distinguishable the centered pilot with high energy ϵ, surrounded first by zero pilots (the hollow zone) and after by information symbols (here, for convenience, with unit energy).

A block of dimension $N \times M$ of transmitted symbols contains both information bearing symbols and pilot symbols. The arrangement of pilot symbols consists of a rectangular region placed in the block containing two types of symbols (see Fig. 3):

- *Zero Pilots*: Placed between information symbols and non-zero pilots to guarantee no significant interference between them. The Dirichlet kernel functions appearing in the OTFS input-output relation (27) are rapidly decreasing functions, hence, perfect orthogonality between information symbols and pilots cannot be achieved, but, at least, the Doppler-delay ISI can be reduced.
- *Peak Pilot*: A pilot symbol with high energy, collecting the energy of the whole pilot field, is placed at the grid center. Its shifts in the Doppler-delay grid are used to provide the initial coarse estimation of the Doppler-delay pairs, which results to be fast and simple.

Given this pilot arrangement, the number of pilot symbols has to be optimized to match the optimal performance-overhead tradeoff, while keeping constant the total block energy.

Note that in OFDM, symbols are independent, i.e., ISI and ICI free for modulation definition, and the pilot vector $\mathbf{x}_{\mathrm{pl,OFDM}}$, defined in (12) through a selection matrix $\mathbf{S}$, results to be a subset of symbols $\mathbf{x}$. Differently, in OTFS, since the channel, as depicted in Fig. 2, behaves *per block* and not *per symbol* as in OFDM, the processing at the receiver could not be based on a subset of samples but must take into account the entire block. As a result, the vector of pilots $\mathbf{x}_{\mathrm{pl,OTFS}}$ has dimension $NM \times 1$ and is composed of all zero entries (the positions of unknown data symbols are set to zero) but one, i.e., the peak pilot (Fig. 3).

C. Channel Estimation

The proposed channel estimation scheme is based on the ML approach of [18],[5] providing a parameter estimation of Doppler, delay, and complex channel gain associated to each multipath component. In the following, we provide a concise description of the scheme, since it plays a central role within the channel estimation algorithm for OTFS.

The objective is to estimate the set of parameters $\boldsymbol{\theta} = \{h'_p, \tau_p, \nu_p\} \in \mathcal{T}^P$, with $\mathcal{T} = \mathbb{C} \times \mathbb{R} \times \mathbb{R}$. By defining the ML function as

$$l(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x}_{\mathrm{pl,OTFS}}) = \left|\mathbf{y} - \textstyle\sum_p h'_p \boldsymbol{\Psi}_p \mathbf{x}_{\mathrm{pl,OTFS}}\right|^2, \quad (29)$$

the ML solution becomes

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathcal{T}^P} l(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x}_{\mathrm{pl,OTFS}}). \quad (30)$$

For a fixed set of $\{\tau_p, \nu_p\}$, the ML estimator of $\{h'_p\}$ is given by solving the following set of equations

$$\mathbf{x}^{\mathsf{H}}_{\mathrm{pl,OTFS}} \boldsymbol{\Psi}^{\mathsf{H}}_p \left(\sum_{q=0}^{P-1} h'_q \boldsymbol{\Psi}_q \right) \mathbf{x}_{\mathrm{pl,OTFS}} = \mathbf{x}^{\mathsf{H}}_{\mathrm{pl,OTFS}} \boldsymbol{\Psi}^{\mathsf{H}}_p \mathbf{y}. \quad (31)$$

By expanding (29), solving the system of equations in (31) to find the complex channel gains h'_p for every multipath component, and substituting these results in (29), after some long but relatively simple algebra (not given explicitly for the sake of brevity), we find that the minimization w.r.t. $\boldsymbol{\theta}$ reduces to maximizing the function

$$\begin{aligned} l_2(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x}_{\mathrm{pl,OTFS}}) &= \sum_p h'_p \mathbf{y}^{\mathsf{H}} \boldsymbol{\Psi}_p \mathbf{x}_{\mathrm{pl,OTFS}} \\ &= \sum_p S(\tau_p, \nu_p) - I_p(\{h'_q\}_{q \neq p}, \boldsymbol{\theta}), \end{aligned} \quad (32)$$

where $S(\tau_p, \nu_p, \phi_p)$ and $I_p(\{h'_q\}_{q \neq p}, \boldsymbol{\theta})$ (S_p and I_p in short hand notation) denote the useful signal and the interference term for the multipath component p, given respectively by

$$S_p = \frac{|\mathbf{y}^{\mathsf{H}} \boldsymbol{\Psi}_p \mathbf{x}_{\mathrm{pl,OTFS}}|^2}{|\boldsymbol{\Psi}_p \mathbf{x}_{\mathrm{pl,OTFS}}|^2}, \quad (33)$$

$$I_p = \frac{(\mathbf{y}^{\mathsf{H}} \boldsymbol{\Psi}_p \mathbf{x}_{\mathrm{pl,OTFS}})\, \mathbf{x}^{\mathsf{H}}_{\mathrm{pl,OTFS}} \left(\boldsymbol{\Psi}^{\mathsf{H}}_p \sum_{q \neq p} h'_q \boldsymbol{\Psi}_q \right) \mathbf{x}_{\mathrm{pl,OTFS}}}{|\boldsymbol{\Psi}_p \mathbf{x}_{\mathrm{pl,OTFS}}|^2}. \quad (34)$$

The algorithm to obtain the estimation of Doppler, delay, and complex channel coefficient of each multipath component is described in the following.

After the definition of the ML approach, Algorithm 2 describes the actual CSI estimation, i.e., the estimation of the unknown parameters of each multipath component. Note that, as for OFDM, if the number of multipath components P is not available at the receiver, we instead select all local maxima or peaks (of the groups of estimates) whose magnitude is above a certain threshold (to be defined).

[5]In [18] we proposed the ML method to estimate the Doppler shift and delay of the main path assuming LoS in the backscattered wave for a joint radar and communication application with OTFS modulation format. Since in [18] the estimation of the radar parameters is performed at the transmitter side, all modulation symbols in the block are known as they are generated by the transmitter itself. Therefore, they can be all treated as pilot symbols. Here we use the same ML approach, but applied to the specific pilot pattern considered in this paper.

Algorithm 1 Multipath Parameters Estimation

Result: The set $(\hat{h}'_p, \hat{\tau}_p, \hat{\nu}_p)$, for $p = 0, \ldots, P-1$.
It: Let $i = 0, 1, 2, \ldots$ be the iteration number;
Initialization: For $i = 0$, initialize $\hat{h}'_p[0] = 0$;
for $i = 1, 2, \ldots$ **do**
- ***Delay and Doppler update***: For each $p = 1, \ldots, P$, find the estimates $\hat{\tau}_p[i], \hat{\nu}_p[i]$ by solving the two-dimensional maximization

$$(\hat{\tau}_p[i], \hat{\nu}_p[i]) = \arg\max_{(\tau,\nu)} \Big\{ S_p - I_p \Big\}, \quad (35)$$

with S_p and I_p calculated for $(\hat{h}'_p[i], \tau, \nu, \hat{\phi}_p[i])$;
- ***Complex channel coefficients update***: Solve the linear system (31) with channel matrices $\boldsymbol{\Psi}_p$ with parameters $(\hat{h}'_p[i], \hat{\tau}_p[i], \hat{\nu}_p[i]$, and let the solution be denoted by $\hat{h}'_p[i]$.

end

Algorithm 2 CSI Estimation

Result: The set $(\hat{h}'_p, \hat{\tau}_p, \hat{\nu}_p)$, for $p = 0, \ldots, P-1$.
Coarse Estimation: By analyzing on-grid Doppler and delay shifts, get the first coarse estimation of the pairs $(\hat{\tau}_p, \hat{\nu}_p)$ through the shifts of the peak pilot w.r.t. the Doppler-delay grid, by selecting the first P local maxima;
It: Let $i = 0, 1, 2, \ldots$ be the iteration number;
for $i = 1, 2, \ldots$ **do**
- ***Grid step and interval refinement:*** Refine the granularity of the step around the estimated values within a refinement interval. The finer the step size and the larger the interval, the greater the computational complexity;
- Use the ML approach described in Alg. 1 to get a finer estimation of the unknown parameters;
- Select the first P local maxima.

end

The iterative process allows to refine the estimation through iterations while keeping the computational cost limited and speed-up the minimization of the estimation error. As indicated in [18], few number of iterations, e.g., less than 5, is enough to reach the algorithm convergence. The definition of the refinement interval and grid step is not mandatory and depends on the hardware capabilities of the system. Note that other iterative schemes based on coarse and refined estimation over discretized Doppler-delay-angle grids can be found in literature (see, e.g., [26] and references therein). However, our scheme relies on a rigorous derivation of OTFS with rectangular pulses and fractional Doppler and delay shifts, thus differs significantly from the approach proposed in [26], where OTFS is seen as a pre- and post-processing for an inner OFDM modulation, for which CS algorithms can be applied (due to the orthogonality of received samples, as discussed in Section II).

Remark 1: Note that, by considering the pilot scheme of Fig. 3 (and also the definition of $\mathbf{x}_{\mathrm{pl,OTFS}}$), the multiplication between the channel matrix $\mathbf{\Psi}_p$ and the pilot vector $\mathbf{x}_{\mathrm{pl,OTFS}}$ is just a column selection of matrix $\mathbf{\Psi}_p$, which significantly simplifies all the equations involved. We did not explicitly take advantage of this aspect, by keeping the treatment as general as possible (one can think to modify the pilot pattern configuration of Fig. 3), but note that the reduction in terms of computational complexity could be remarkable.

IV. Comparison in Terms of Pragramatic Capacity

Before proceeding to the numerical results, we introduce the simulation setting in terms of the pilot schemes and their overhead, further details on the channel estimation algorithms, the performance metric adopted for the comparison, and the soft-output symbol detection algorithms for OFDM and OTFS. All these details are listed in the following with the aim of providing a clear definition of the comparison of these two modulation formats.

- In general, we denote by $\mathcal{D}$ and by $\mathcal{P}$ the set of data and pilot symbols, respectively, in a NM frame of length NM for both OFDM and OTFS. An important aspect in the system optimization is the number of pilots $|\mathcal{P}|$ ($|\cdot|$ denoting the cardinality of a set). Consider first the OFDM modulation. It is well known in the CS literature that the minimum number of pilots (or measurements, from CS literature) to recover a sparse signal is given by the logarithmic scaling factor [42]

$$|\mathcal{P}| \geq P \log \frac{G}{P}, \tag{36}$$

  where P and G are the number of non-zero components and the dimension of the vector to be estimated, respectively. Thus, given a multipath channel with P paths and a sensing matrix of dimension $|\mathcal{P}| \times G$ (defined in (7)), the (asymptotically) minimum number of pilot symbols necessary to solve the minimization problem in (14) is given by (36). As a consequence, the required $|\mathcal{P}|$ slowly increases with G, i.e., with the resolution of the Doppler-delay grid (see also the Appendix A). Moreover, (36) provides only the scaling law of the number of measurements (up to some constant factor larger than 1) and the actual optimal system performance might be achieved for a number of pilots larger than the bound in (36). Note that this is not a precise quantitative analysis but it just gives a qualitative idea or intuition on how large the number of pilots per block should be, knowing that the aforementioned CS-based conditions are always given up to constant factors that depend on the specific problem, SNR, shape of the sensing matrix, and other variables. In other words, the analytical optimization of the number of pilots is an intractable problem. Thus, we based our comparison on a brute-force search over a suitable set of overhead values, identifying the best tradeoff between number of pilots and channel estimation performance for the given scenarios.
  For OTFS modulation, the pilot scheme presented in Section III-B and its estimation algorithm are independent of the block dimension, and depend only on the maximum Doppler shift and delay of the channel. Moreover, if the dimension of the block increases, one can set more pilots to zero to raise the power of the peak pilot, allowing the detection of low power scattering components when the number of paths P is not known a-priori, using the threshold-based approach explained in the following. However, limits on the block dimension come, in first place, from important restrictions on OTFS detection computational complexity [18], and then from the assumption that the channel parameters are time-invariant, which breaks down if the block becomes too large. Thus, realistic block sizes have to be considered for both OFDM and OTFS.
- For both OFDM and OTFS, during the channel estimation process, a "peak selection" has to be performed. In a genie-aided scenario where the number of propagation paths P is a-priori known, this results in the search of the P local maxima of the objective function. However, this information may not be available at the receiver and, in such a case, the algorithms should select all local maxima above a certain threshold (to be defined). By considering scattering components with decreasing power, which is generally the case, once the pathloss brings the power below the threshold, the corresponding component is neglected in the construction of the (estimated) channel matrix. Clearly, this results in a less accurate CSI, but, at the same time, the contribution of low- or very-low-energy paths has a minor impact. In our simulation results we consider the genie-aided case where P is known, while the threshold-based scheme is a conceptually straightforward extension.
- In the numerical results, we assess the performance of both schemes in terms of *pragmatic capacity*, i.e., the mutual information of the virtual channel having at its input the constellation symbols and at its output the detector soft-outputs [10], [11]. The pragmatic capacity is representative of the achievable rate under the assumption of separate detection and decoding, i.e., without "turbo" reprocessing of the decoder output. By considering a sequence of NM symbols $\{x_k\}$ belonging to any constellation $\mathcal{C}$, let $\{\hat{x}_k\}$ be the noisy estimates of the transmitted symbol. The pragmatic capacity is simply defined as the symbol-by-symbol mutual information (see [10], [11], [18] for more details), which can be easily computed by Monte Carlo simulations as

$$I_{\mathrm{PC}} = \log_2 |\mathcal{C}| - \frac{1}{NM} \sum_{k \in \mathcal{D}} P(x_k|\hat{x}_k) \log_2 \frac{1}{P(x_k|\hat{x}_k)}, \tag{37}$$

  where $P(x_k|\hat{x}_k)$ is the *a-posteriori* probability mass function of symbols $x_k \in \mathcal{C}$ given the detector soft-output $\hat{x}_k$, while $\mathcal{D}$ is the set of information symbols, i.e., excluding the pilots. Note that, since the numerator sums $|\mathcal{D}| \leq NM$ terms, while the denominator is the block size NM, the pilot overhead emerges naturally. An indication of the minimum length of the sequence to obtain reliable pragmatic capacity results is given in [43].

- For OFDM, we adopt the linear minimum mean square error (LMMSE) detector, whose soft-output under non-perfect CSI, i.e., employing the estimated channel matrix $\hat{\mathbf{H}}$,[6] is given by

$$\hat{\mathbf{x}} = \hat{\mathbf{H}}^{\mathrm{H}} \left(\hat{\mathbf{H}} \hat{\mathbf{H}}^{\mathrm{H}} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{y}. \tag{38}$$

Note that for OFDM, whose (estimated) channel matrix is diagonal, the LMMSE detector reduces to the symbol-by-symbol detection given by

$$\hat{x}_k = \frac{H_{k,k}^*}{|H_{k,k}|^2 + \sigma^2} y_k, \tag{39}$$

which significantly simplifies the computational complexity at the detector. On the other hand, the same detector cannot be adopted for OTFS modulation, because of the costly matrix inversion of a non-diagonal matrix in (38). Hence, different solutions have been proposed in the literature [18], [44]–[46]. However, some of these approaches rely on non-realistic model or channel assumptions (e.g., Doppler shifts and delays integer multiples of the symbol grid) and therefore their performance degrades significantly when applied to realistic channel conditions [18]. For this reason, in place of the very-high-complexity block-based LMMSE detector (38) for OTFS, we consider the low-complexity message-passing (MP) soft-output algorithm proposed in [18], which achieves linear complexity per block (i.e., constant complexity per symbol, comparable with the symbol-by-symbol LMMSE detector for OFDM).

As a reference benchmark, the mutual information for the considered input constellation transmitted over an AWGN channel (thus in the absence of fading), denoted to as AWGN (symmetric) capacity $C_{\mathrm{AWGN}}^{\mathrm{sym}}$, [47], will be also computed. At high SNR, where the estimation error is supposed to be small, the gap between pragmatic capacity curves and this benchmark is only due to the presence of the overhead of pilot symbols within the transmitted block. Asymptotically, the achievable rate loss $\mathcal{R}^{\ell}\mathrm{x}$ is given by

$$\mathcal{R}^{\ell} = \frac{\mathcal{P}}{NM} \times 100 \ [\%]. \tag{40}$$

and the rate simply becomes the AWGN capacity multiplied by the fraction of data symbols per block, i.e., $|\mathcal{D}|/NM = 1 - |\mathcal{P}|/NM$. Moreover, as already said, in order to make a fair comparison, the overhead introduced by a CP for OFDM and by a generic GI (between blocks) for OTFS must be taken into account. While a GI interposed between two OTFS blocks, to avoid IBI, introduces a small-to-negligible overhead (especially when the dimension of the block increases), the CP overhead of OFDM is kept constant within the entire block (whatever its dimension), introducing a considerable loss in terms of pragmatic capacity. For instance, by considering a CP of length $T/4$, being T the symbol time, the loss is

$$\frac{T}{T + T/4} = \frac{T}{5T/4} = \frac{4}{5} = 0.8 = 20\%. \tag{41}$$

This means that, with a modulation constellation of size $|\mathcal{C}|$, OFDM saturates at $0.8 \cdot \log_2 |\mathcal{C}|$ bits/symbol. The overall loss takes into account both the pilot overhead and the presence of a CP and/or GI (also of length $T/4$, for consistency).

- As anticipated at the beginning of this paper (Section II), in order to restrict to the classical low-complexity symbol-by-symbol minimum mean square error (MMSE) estimation for OFDM we have neglected the ICI. As already seen in (16), the ICI depends on the ratio between the subcarrier spacing Δf and the maximum Doppler shift introduced by the channel. In order to have negligible ICI, the necessary condition is $\Delta f \gg \nu_{\max}$, or, equivalently, $\nu_{\max}/\Delta f \ll 1$. Since $\Delta f = B/M$, with B total bandwidth, the condition may not be satisfied when the number of subcarriers M becomes too large, even for moderate Doppler. In this paper we insist on neglecting ICI and consider the range of system parameters for which this assumption is indeed virtually exact. Furthermore, we notice that while OFDM incurs in this additional limitation, OTFS remains not sensitive to the Doppler shift.

[6]Note that, always for fairness, both modulations use the same estimation grid granularity.

TABLE I
SYSTEM PARAMETERS

$f_c = 5.89$ [GHz]	$M = 64$
$B = 10$ [MHz]	$N = 50$
$\Delta f = B/M = 156.25$ [kHz]	$T = 1/\Delta f = 6.4$ [μs]

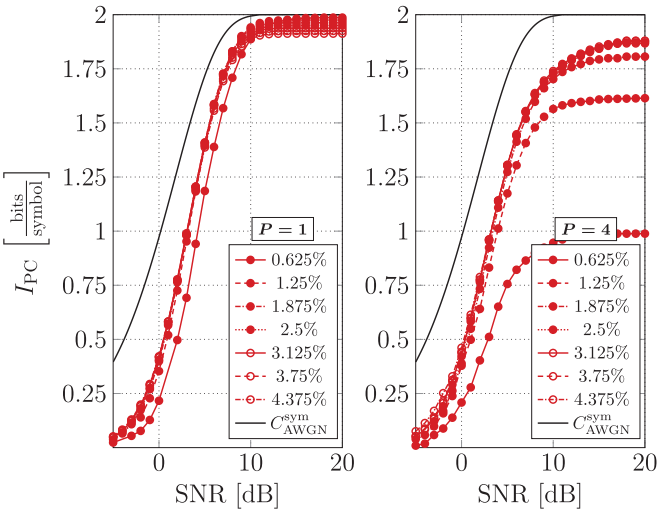


Fig. 4. Pragmatic Capacity vs. SNR for OFDM modulation with $P = 1$ and $P = 4$ and different pilot overhead (whose percentage is indicated in the legend).

A. Simulation Results

We present results in terms of pragmatic capacity vs. SNR for OTFS and OFDM with quadrature phase-shift keying (QPSK) modulated symbols, for a time-frequency multipath

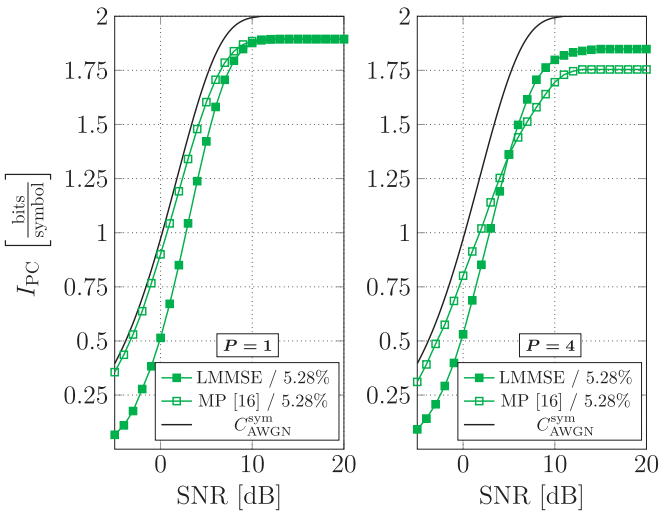


Fig. 5. Pragmatic Capacity vs. SNR for OTFS modulation with $P = 1$ and $P = 4$, different detection algorithms, i.e., LMMSE and MP approach of [18], for a fixed pilot overhead, in percentage 5.28%.

channel with P components and affected by AWGN, under the CSI estimation schemes and pilot schemes of Section II for OFDM and Section III for OTFS, respectively. As said, as a reference benchmark, we plot the mutual information for a QPSK input constellation transmitted over an AWGN channel (thus in the absence of fading) $C^{\mathrm{sym}}_{\mathrm{AWGN}}$ and without any pilot overhead. The system parameters are listed in Table I.

Fig. 4 shows the performance of OFDM for a different number of pilot symbols. For the case $P = 1$, it easy to note that the performance slightly changes for different pilot overheads, whose percentage is indicated in the legend. On the other hand, as suggested by (36), if the number of non-zero components to be estimated increases, i.e., with $P = 4$, the channel estimation algorithm needs more pilots to work efficiently. Given these results, from now on, we will consider a pilot overhead of 3.125%, achieving, in our setup, the best tradeoff between estimation accuracy and achievable pragmatic capacity (for any number of scattering components).

Fig. 5 shows the performance of OTFS with different detection algorithms. The MP soft-output detection approach of [18] is able to almost achieve the AWGN capacity under non-perfect CSI for a low number of scattering components, i.e., $P = 1$, together with a remarkable reduction of the computational complexity [18]. However, in line with the results of [18], the detector performance slightly decreases with increasing multipath. Note that the small loss w.r.t. $C^{\mathrm{sym}}_{\mathrm{AWGN}}$ under non-perfect CSI is an indicator of the performance of the channel estimation algorithm, which results to be very accurate (otherwise the curve would have deviated from the benchmark). In light of these results, we see that there is no reason to adopt the LMMSE estimator for OTFS, which results in high complexity and worse performance (see [18] for a more detailed analysis). Hence, from now on, for the comparison with OFDM modulation, we consider the MP approach of [18].

In Fig. 6, we plot the pragmatic capacity vs. SNR for OFDM and OTFS under the configurations mentioned above. First of all, it is possible to note that the performance decreases only

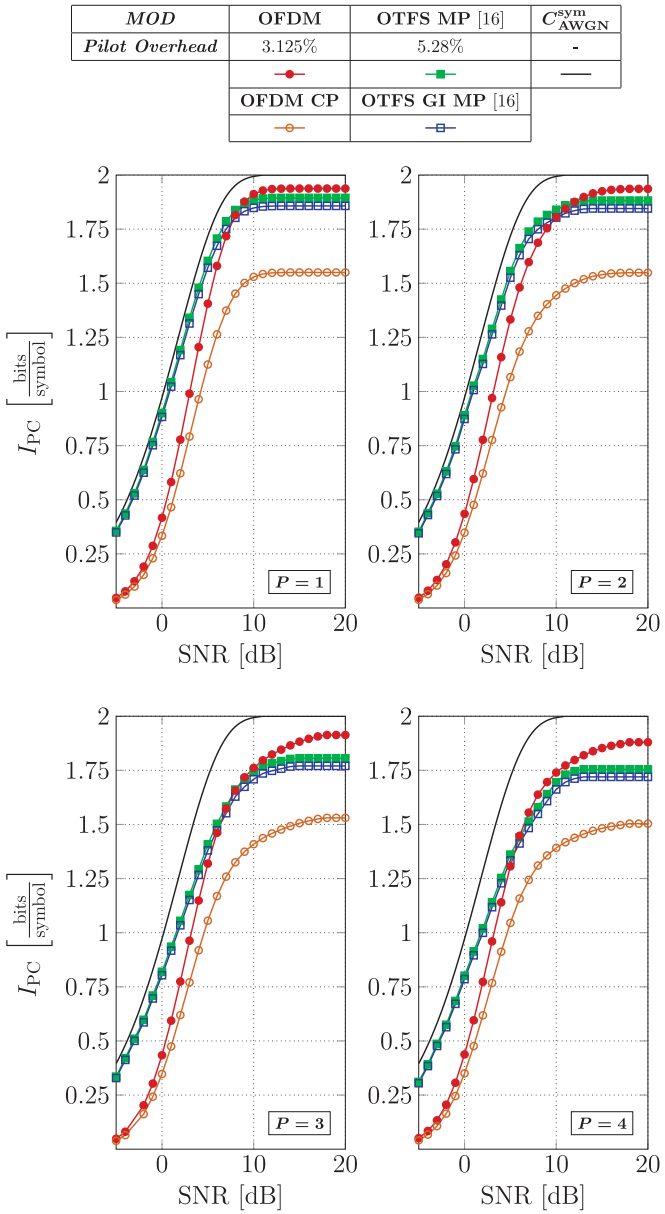


Fig. 6. The curves take into account the loss related to the presence of pilots in the block of $N \times M$ symbols. The OFDM CP curve includes the CP overhead which is 0.25 of the symbol time, while the OTFS GI curve includes a GI for the entire block. A QPSK modulation is used. The legend indicates the percentage of pilot symbols.

slightly while increasing the number of multipath components, proving the robustness of the pilot schemes and the algorithms proposed. The pilot overhead has been chosen *ad hoc* for both modulations. For OTFS, as said in Sec. III, the peak centered pilot collects all the energy of surrounding zero pilots, and the percentage of overhead (indicated in the figures) is an indicator of the peak pilot energy. For OFDM, as pointed out before, the optimal tradeoff between estimation performance and pilot overhead has been found by exhaustive searching over a suitable set of values (some of them are shown in Fig. 4). While the performance of the two modulations is similar, the presence of a *per symbol* CP for OFDM remarkably deteriorates the pragmatic capacity, while a *per block* GI for OTFS introduces a negligible loss.

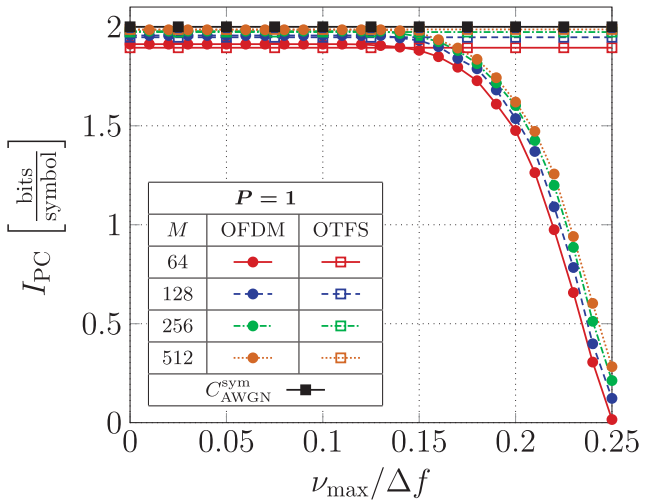


Fig. 7. Pragmatic Capacity vs. $\nu_{\max}/\Delta f$ ratio for OFDM modulation with $P = 1$, at SNR $= 18$ dB.

TABLE II

MAXIMUM SUPPORTABLE VELOCITY W.R.T. THE RATIO $\nu_{\max}/\Delta f$ FOR OFDM MODULATION WITH SUBCARRIER SPACING $\Delta f = B/M$ AND CARRIER FREQUENCY $f_c = 5.89$ GHz. THE VELOCITY IS GIVEN BY $v = c \cdot \nu_{\max}/(2f_c)$

		$\nu_{\max}/\Delta f$		
		0.1	0.15	0.2
M	64	1431 [km/h]	2147 [km/h]	2863 [km/h]
	128	715 [km/h]	1073 [km/h]	1431 [km/h]
	256	357 [km/h]	536 [km/h]	715 [km/h]
	512	178 [km/h]	268 [km/h]	357 [km/h]

In Fig. 7, we plot the pragmatic capacity of OFDM and OTFS for a fixed value of SNR, i.e., 18 dB, while changing the ratio between the maximum Doppler shift and the subcarrier spacing, i.e., $\nu_{\max}/\Delta f$, for different number of subcarriers M ($N = 50$ for all cases). In this case, in particular for OFDM, the received samples are obtained by considering a real channel taking into account the ICI, i.e., (16), while the channel estimation works under the hypothesis of an ideal interference-free channel. By first taking into account OFDM, intuitively, the performance degrades when the ICI becomes significant. Note that the estimation performance of the LASSO solver is independent of the number of subcarrier M and, for this reason, whatever the choice of $\nu_{\max}$ and M, the performance of OFDM depends only on their ratio. Fig. 7 shows that the PC performance starts decreasing significantly for $\nu_{\max}/\Delta f \simeq 0.15$. Almost the same behavior is shown for different number of subcarriers M (not reported here for the sake of space limitation), except for the percentage of pilot loss due to different block dimensions, supporting what stated above. However, as pointed out in Table II, while the performance is almost constant, the maximum supportable Doppler (or velocity), inversely proportional to M, is not. For these reasons, as expected, OFDM is not independent of the block dimension and the system has to be defined properly to operate in the range where the ICI is negligible.

Also in Fig. 7 we report the performance of OTFS in the same conditions, from which it is evident that OTFS is insensitive to the Doppler effects. This means also that the simulation results of OTFS depicted in Fig. 6 are valid for any Doppler spread.

V. CONCLUSION

In this paper we carried out a fair comparison between OFDM and OTFS modulation formats in terms of maximum achievable rate for practical separated detection and decoding, quantified by the Pragmatic Capacity measured at the soft-detector output.

We considered two pilot schemes and channel estimation algorithms each one specifically suited for the given modulation scheme. Both pilot and CSI estimation schemes are able to achieve very good performance (near genie-aided) under time-varying communication channel in the sparsity regime of a small number of number of multipath components. This conclusion is fully supported by numerical results, where simulation curves achieve the theoretical benchmark under non-perfect CSI, proving the quality of the proposed approaches.

OTFS achieves a better communication rate mainly because of the presence of a *per symbol* Guard Interval rather than a per-symbol Cyclic Prefix as in OFDM. This of course comes at the cost of a more complex channel estimation scheme, working on large block-wise operations.

In terms of soft-output data detection, the use of our Message-Passing soft-output detector, previously proposed in [18], yields constant per-symbol complexity for OTFS, which is the same scaling law of symbol-by-symbol MMSE detection for OFDM. Although we do not claim that the complexity of the two detectors is identical, in fact the actual complexity differ for some implementation-based constant.

Finally, we can observe that OTFS is indeed insensitive to the magnitude of the Doppler shifts, while the performance of OFDM degrades significantly even under small-to-moderate Doppler values if the number of subcarriers increases. Therefore, OTFS is effectively a good candidate for high-mobility systems in rural environments (e.g., high speed trains [48]) or aerial environments (e.g., UAVs [49]), where Doppler shifts may be large, and the propagation channel contains typically the line-of-sight and a few other reflection components (e.g., ground reflection, hills, large buildings), and it is therefore sparse in the Doppler-delay domain.

APPENDIX A
THE LASSO SOLVER

For the sake of completeness and reproducibility of the results, in this Appendix we give the details of the algorithm used to solve the LASSO minimization problem in (14).

0) *Initialization*: By defining the known support matrix $\mathbf{A} \triangleq \mathbf{X}_{\mathrm{pl}} \odot \mathbf{SD}$, in which $\mathbf{X}_{\mathrm{pl}}$ is a matrix of dimension $|\mathcal{P}| \times G$ composed of G equal vectors $\mathbf{x}_{\mathrm{pl}}$, i.e., $\mathbf{X}_{\mathrm{pl}} \triangleq [\mathbf{x}_{\mathrm{pl}}, \ldots, \mathbf{x}_{\mathrm{pl}}]$, let $\mathbf{\Lambda} \triangleq \mathbf{A}^{\mathrm{H}}\mathbf{A}$ and initialize the step size ϵ as [29]

$$\epsilon \triangleq \frac{1}{\|\mathbf{\Lambda}^{\mathrm{H}}\mathbf{\Lambda}\|_{\mathrm{F}}} = \frac{1}{\sqrt{\mathrm{trace}\,(\mathbf{\Lambda}^{\mathrm{H}}\mathbf{\Lambda})}}, \tag{42}$$

in which $\|\cdot\|_{\mathrm{F}}$ indicates the Frobenius norm and the $\mathrm{trace}(\cdot)$ operation takes the sum of the matrix main diagonal elements. The threshold t is set as $t = \lambda\epsilon$, where λ is the LASSO regularizer appearing in (14). The vector of estimated values $\hat{\mathbf{h}}$ is initialized to all zeros.

1) *Iterations* $i = 1, 2, \ldots$
   a) *Soft Thresholding*: with $\psi_{\mathrm{st}}(\cdot, t)$ soft thresholding operator with threshold t (see [29], [50]), compute
   $$\boldsymbol{\beta}^{i+1} = \psi_{\mathrm{st}}\left(\hat{\mathbf{h}}^{i} + \epsilon\mathbf{A}^{\mathsf{H}}\left(\mathbf{y} - \mathbf{A}\hat{\mathbf{h}}^{i}\right), t\right). \quad (43)$$
   b) *Nesterov's Acceleration Factor (Optional)* [29]: Introduce a tuning coefficient $\alpha_i \in [0, 1]$, which can be fixed or variable in t, and compute
   $$\hat{\mathbf{h}}^{i+1} = \boldsymbol{\beta}^{i+1} + \alpha_i\left(\boldsymbol{\beta}^{i+1} - \hat{\mathbf{h}}^{i}\right), \quad (44)$$
   with α_i defined, e.g., in [29]–[31] and reviewed for completeness in Appendix B.
   c) *Shrink*: Remove the entries of $\mathbf{y}$ and $\boldsymbol{\beta}$, the columns of $\mathbf{A}$, and the entries of $\hat{\mathbf{h}}$, corresponding to the zero entries of $\hat{\mathbf{h}}$.
2) *Restoring*: Restore the estimated vector $\hat{\mathbf{h}}$ to its full dimension (this operation is necessary after the shrink of the vectors during the iterations).

We used as stopping criterion the maximum number of iterations. Note that the shrinking operation is allowed because zero entries of vector $\hat{\mathbf{h}}$ at iteration i cannot assume a value $\neq 0$ at iteration $i' > i$ [32]. From a complexity point of view, the first iterations are the most costly, while the algorithm can run $> 10^6$ times keeping the complexity almost constant and the computational time linear (when the number of iterations is large enough, i.e., far away from starting costly ones).

A. Complexity of the LASSO Solver and Step Size Refinement

The sensing matrix $\mathbf{D}$ is composed of G columns of length NM. While the dimensions N and M depends on system settings and can be somehow controlled or tuned, the dimension G takes into account the estimation precision, or granularity, of the searching grid Γ. Hence, the larger the dimension G, the more reliable the result. Using some examples:

- If Γ is equivalent to the Doppler-delay grid (delay and Doppler shifts integer multiple of the grid), $G = NM$ and $\mathbf{D}$ is a $NM \times NM$ matrix. Blockwise operations adopted by any LASSO solver are feasible in this framework.
- If the step size for both Doppler and delay axis is a fraction $1/\rho$ of the Doppler-delay grid step, $G \simeq O\left(NM \cdot \rho^2\right)$ and $\mathbf{D}$ is approximately a $NM \times \left(NM \cdot \rho^2\right)$ matrix. Clearly, increasing the granularity of the grid quickly increases the complexity.

In order to overcome the complexity induced by searching grid with fine granularity, it is possible to refine the step size in successive phases, rather than directly defining a low fractional value for the entire grid. The proposed refinement scheme is illustrated in Algorithm 3. During the *Peak Selection* step, if the number of multipath components P is not available at the receiver, instead select all local maxima or peaks (of the groups of estimates) whose magnitude is above a certain threshold (to be defined).

Algorithm 3 Refinement of the Granularity

Result: Fine estimation $\hat{\mathbf{h}}$ for LASSO problem (14).
Coarse Estimation: For any LASSO solver, get a first coarse estimation $\hat{\mathbf{h}}$ such that the searching grid Γ is equivalent to the Doppler-delay grid (i.e., delay and Doppler shifts integer multiple of the grid). In this case $G = NM$ and $\mathbf{D}$ is a $NM \times NM$ matrix;
for *Iteration* $i = 1, 2, \ldots$ **do**
- ***Peak Selection***: Select the first P local maxima of $\hat{\mathbf{h}}$;
- ***Step Refinement Around Maxima***: Build a new sensing matrix based on an extension of matrix $\mathbf{D}$ such that the step size around the peaks is decreased (i.e., the granularity and the precision are increased);
- ***Finer Estimation***: For any LASSO solver, get a finer estimation $\hat{\mathbf{h}}$.

end

Note that during the *Coarse Estimation* step, i.e., when $\mathbf{D}$ is an $NM \times NM$ matrix, it is possible to adopt the approach proposed in [51] to solve the LASSO minimization in (14). The algorithm of [51], benefiting of the hierarchical structure of vector $\mathbf{h}$, is able to provide a first coarse and reliable estimation optimizing the computational complexity. However, if $\mathbf{h}$ takes off-grid values, the approach of [51] becomes inappropriate, as confirmed by the presented simulation results. For this reason, after a *Coarse Estimation*, i.e., within the *Iteration* step, another LASSO solver must be chosen to obtain the best performance in terms of channel estimation.

APPENDIX B
NESTEROV'S ACCELERATION FACTOR

The Nesterov's acceleration factor governs the dependency between two successive estimations, while remarkably reducing the convergence time of the algorithm [30]. The choice of the optimization coefficient α_i is not mandatory. By following the pioneering work [31], which inspired many other works, e.g., [29], α_i can be recursively defined as

$$\xi_{i+i} = \frac{1 + \sqrt{4\xi_i^2 + 1}}{2}, \quad (45)$$

$$\alpha_i = \frac{\xi_i - 1}{\xi_{i+i}}, \quad (46)$$

with $\xi_0 = 1$. Another choice simply based on the iteration index i is $\alpha_i = (i-1)/(i+2)$ [30], [31]. Both solutions describe a curve growing from an initial value ("far" from 1) up to 1. The associated plots, with their similar behaviors, can be seen in Fig. 8. The most conservative choice is $\alpha_i = 1$, for which the dependency from the previous estimated values is maximized, while the less conservative choice is $\alpha_i = 0$,

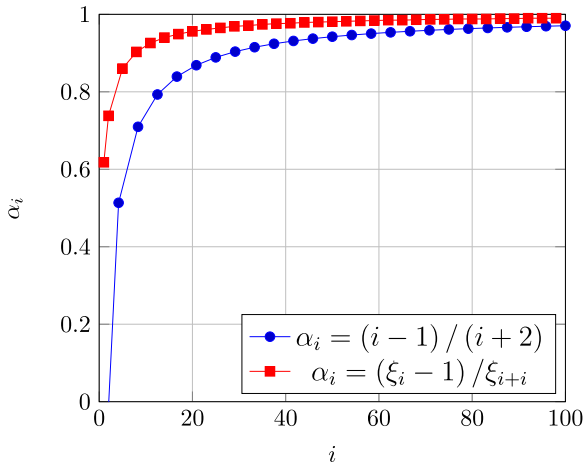


Fig. 8. The evolution of α_i over iterations i for different approaches.

completely forgetting the previous estimated values. Intuitively, a small value of α_i is preferable for the first noisy iterations, while a parameter α_i near to one should be chosen when the reliability of the estimation increases.

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[2] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul. 2000.

[3] J. Choi and Y.-H. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2083–2088, Sep. 2005.

[4] 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.

[5] A. Lancho, T. Koch, and G. Durisi, "On single-antenna Rayleigh block-fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 496–519, Jan. 2020.

[6] A. Lancho, G. Durisi, T. Koch, and G. Vazquez-Vilar, "Saddlepoint approximations for noncoherent single-antenna Rayleigh block-fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2019, pp. 612–616.

[7] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859–1874, Sep. 2004.

[8] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[9] R. Hadani *et al.*, "Orthogonal time frequency space (OTFS) modulation for millimeter-wave communications systems," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2017, pp. 681–683.

[10] A. Kavcic, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution, and code performance bounds," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1636–1652, Jul. 2003.

[11] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, "Determining and approaching achievable rates of binary intersymbol interference channels using multistage decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1416–1429, Apr. 2007.

[12] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?" *IEEE Commun. Mag.*, vol. 31, no. 6, pp. 80–93, Nov. 2014.

[13] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.

[14] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 944–957, May 2001.

[15] V. Q. Rodriguez, F. Guillemin, A. Ferrieux, and L. Thomas, "Cloud-RAN functional split for an efficient fronthaul network," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, 2020, pp. 245–250.

[16] 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.

[17] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[18] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "On the effectiveness of OTFS for joint radar parameter estimation and communication," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5951–5965, Sep. 2020.

[19] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, Jr., "Channel estimation for orthogonal time frequency space (OTFS) massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4204–4217, Aug. 2019.

[20] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[21] W. Dongming, H. Bing, Z. Junhui, G. Xiqi, and Y. Xiaohu, "Channel estimation algorithms for broadband MIMO-OFDM sparse channel," in *Proc. 14th IEEE Proc. Pers., Indoor Mobile Radio Commun.*, vol. 2, Sep. 2003, pp. 1929–1933.

[22] V. Rajan, A. A. Balakrishnan, and K. E. Nissar, "OFDM channel estimation using compressed sensing L1-regularized least square problem solver," in *Proc. 3rd Int. Conf. Adv. Comput. Commun.*, Aug. 2013, pp. 94–97.

[23] G. Tauböck, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, Apr. 2010.

[24] C.-J. Wu and D. W. Lin, "A group matching pursuit algorithm for sparse channel estimation for OFDM transmission," in *Proc. IEEE Int. Conf. Acoust. Speed Signal Process.*, May 2006, pp. 1–5.

[25] M. Zhang, F. Wang, X. Yuan, and L. Chen, "2D structured turbo compressed sensing for channel estimation in OTFS systems," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Dec. 2018, pp. 45–49.

[26] M. Li, S. Zhang, F. Gao, P. Fan, and O. A. Dobre, "A new path division multiple access for the massive MIMO-OTFS networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 903–918, Apr. 2020.

[27] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, "Transmitter and receiver window designs for orthogonal time-frequency space modulation," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2207–2223, Apr. 2021.

[28] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. SoC., B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[29] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[30] Y. Nesterov, "Gradient methods for minimizing composite functions," *Math. Program.*, vol. 140, no. 1, pp. 125–161, Aug. 2013.

[31] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate O($1/k^2$)," *Dokl. Akad. Nauk*, vol. 269, pp. 543–547, Dec. 1983. [Online]. Available: https://ci.nii.ac.jp/naid/10029946121/en/

[32] M. Schmidt, "Least squares optimization with L1-norm regularization," Univ. British Columbia, Vancouver, BC, Canada, Project Rep. CS542B, 2005, vol. 504, pp. 195–221.

[33] G. A. Vitetta, D. P. Taylor, G. Colavolpe, F. Pancaldi, and P. A. Martin, *Wireless Communications: Algorithmic Techniques*. Hoboken, NJ, USA: Wiley, 2013.

[34] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comput. Rendus Math.*, vol. 346, nos. 9–10, pp. 589–592, May 2008.

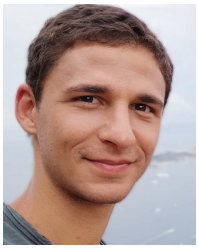
[35] P. Koiran and A. Zouzias, "Hidden cliques and the certification of the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4999–5006, Aug. 2014.

[36] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[37] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[38] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proc. 40th Annu. Conf. Inf. Sci. Syst.*, Mar. 2006, pp. 207–212.

[39] L. Gaudio, M. Kobayashi, B. Bissinger, and G. Caire, "Performance analysis of joint radar and communication using OFDM and OTFS," in *Proc. IEEE Int. Conf. Commun. Workshops*, May 2019, pp. 1–6.
[40] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Dec. 2017, pp. 1–6.
[41] G. Matz, H. Bolcskei, and F. Hlawatsch, "Time-frequency foundations of communications: Concepts and tools," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 87–96, Jun. 2013.
[42] B. Adcock, S. Brugiapaglia, and M. King-Roskamp, "Do log factors matter? On optimal wavelet approximation and the foundations of compressed sensing," 2019, *arXiv:1905.10028*.
[43] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.
[44] 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.
[45] W. Xu, T. Zou, H. Gao, Z. Bie, Z. Feng, and Z. Ding, "Low-complexity linear equalization for OTFS systems with rectangular waveforms," 2019, *arXiv:1911.08133*.
[46] J. Cheng, H. Gao, W. Xu, Z. Bie, and Y. Lu, "Low-complexity linear equalizers for OTFS exploiting two-dimensional fast Fourier transform," 2019, *arXiv:1909.00524*.
[47] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA, USA: Addison-Wesley, 1987.
[48] B. Ai *et al.*, "Challenges toward wireless communications for high-speed railway," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2143–2158, Oct. 2014.
[49] E. Haas, "Aeronautical channel modeling," *IEEE Trans. Veh. Technol.*, vol. 51, no. 2, pp. 254–264, Mar. 2002.
[50] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the Lasso," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 8, pp. 1902–1912, Nov. 2010.
[51] G. Wunder, S. Stefanatos, A. Flinth, I. Roth, and G. Caire, "Low-overhead hierarchically-sparse channel estimation for multiuser wideband massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2186–2199, Apr. 2019.



Lorenzo Gaudio (Graduate Student Member, IEEE) received the M.Sc. degree *(cum laude)* in communication engineering from the University of Parma, Parma, Italy, in 2017, and the Ph.D. degree in information technology with the Department of Engineering and Architecture, University of Parma in 2021. His main research interests include digital communications systems, channel coding and information theory, and signal processing.

Giulio Colavolpe (Senior Member, IEEE) received the Dr.Ing. degree *(cum laude)* in telecommunications engineering from the University of Pisa, Italy, in 1994, and the Ph.D. degree in information technologies from the University of Parma, Italy, in 1998. Since 1997, he has been at the University of Parma, where he is currently a Professor of telecommunications at the Dipartimento di Ingegneria e Architettura (DIA). In 2000, he was a Visiting Scientist at the Institut Eurécom, Valbonne, France. In 2013, he was a Visiting Scientist at the European Space Agency (ESTEC, Noordwijk, The Netherlands). His research interests include the design of digital communication systems, adaptive signal processing (with particular emphasis on iterative detection techniques for channels with memory), and channel coding and information theory. His research activity has led to more than 200 papers in refereed journals and in leading international conferences, and 18 industrial patents. He received the Best Paper Award at the 13th International Conference on Software, Telecommunications and Computer Networks (SoftCOM'05), Split, Croatia, September 2005, the Best Paper Award for Optical Networks and Systems at the IEEE International Conference on Communications (ICC 2008), Beijing, China, in May 2008, and the Best Paper Award at the 5th Advanced Satellite Mobile Systems Conference and 11th International Workshop on Signal Processing for Space Communications (ASMS&SPSC 2010), Cagliari, Italy. He served as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, and IEEE WIRELESS COMMUNICATIONS LETTERS, and an Executive Editor for *Transactions on Emerging Telecommunications Technologies* (ETT).

Giuseppe Caire (Fellow, IEEE) was born in Torino, in 1965. He received the B.Sc. degree in electrical engineering from the Politecnico di Torino in 1990, the M.Sc. degree in electrical engineering from Princeton University in 1992, and the Ph.D. degree from the Politecnico di Torino in 1994.

He was a Post-Doctoral Research Fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) from 1994 to 1995, an Assistant Professor in telecommunications with the Politecnico di Torino, an Associate Professor with the University of Parma, Italy, a Professor with the Department of Mobile Communications, Eurecom Institute, Sophia-Antipolis, France, a Professor of electrical engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, and is currently an Alexander von Humboldt Professor with the Faculty of Electrical Engineering and Computer Science, Technical University of Berlin, Germany. His main research interests are in the field of communications theory, information theory, and channel and source coding with particular focus on wireless communications. He has served in the Board of Governors for the IEEE Information Theory Society from 2004 to 2007, and as an Officer from 2008 to 2013. He was the President of the IEEE Information Theory Society in 2011. He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, the IEEE Communications Society and Information Theory Society Joint Paper Award in 2004 and in 2011, the Okawa Research Award in 2006, the Alexander von Humboldt Professorship in 2014, the Vodafone Innovation Prize in 2015, an ERC Advanced Grant in 2018, the Leonard G. Abraham Prize for best IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS paper in 2019, and the IEEE Communications Society Edwin Howard Armstrong Achievement Award in 2020, and was a recipient of the 2021 Leibinz Prize of the German National Science Foundation (DFG).