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.

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 channel estimation (CSI), i.e., the knowledge of the communication channel, which is generally not trivial and depends on the 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).

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]). 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 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 soft-output [10], [11]. 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

1In 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.

2Indeed, 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.
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 tradeoff. 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].

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. 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 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]

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

where $P$ is the number of multipath scattering components and $h_p$, $\nu_p$, and $\tau_p$ are the complex channel gain including the
pathloss, the Doppler shift, and the delay, associated to the $\nu$-th scattering component, respectively. Note that the maximum channel delay and Doppler shift are assumed to satisfy
\[
\tau_{\text{max}} < T, \quad \nu_{\text{max}} < \Delta f,
\] where $T$ is the symbol time and $\Delta f$ 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
\[
H = \sum_{p=0}^{P-1} h_p \mathbf{a}(\nu_p) \mathbf{b}^\dagger(\tau_p), \tag{3}
\]
where
\[
\mathbf{a}(\nu_p) = [1, e^{j2\pi \nu_p T}, \ldots, e^{j2\pi \nu_p(N-1)T}]^\dagger, \tag{4}
\]
\[
\mathbf{b}(\tau_p) = [1, e^{j2\pi \nu_p \Delta f}, \ldots, e^{j2\pi \nu_p(M-1)\Delta f}]^\dagger. \tag{5}
\]
where $(\cdot)^\dagger$ and $(\cdot)^\dagger$ denote the transpose and the conjugate transpose (Hermitian) operation, respectively. By gathering the information symbols $\{x_{n,m}\}$ in a $N \times M$ matrix $X$, the expression of the received samples after transmission over the channel in (3) is
\[
Y = X \odot H + Z, \tag{6}
\]
and where $\odot$ denotes element-wise multiplication and $Z$ is the AWGN with zero mean and covariance matrix $\sigma^2 I_{NM}$. 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 component, the grid coefficient vector $\hat{h}$, select the columns of $D$ with the pair of channel coefficients $(\tau(\gamma_i), \nu(\gamma_i))$, to overall represent the triplet $(\tau(\gamma_i), \nu(\gamma_i), \hat{h}_p)$, emulating the true channel parameters $(\tau_p, \nu_p, \hat{h}_p)$. Thus, the approximated received samples expression in vectorized form can be written as
\[
y = x \odot (Dh_{\text{sp}}) + z, \tag{11}
\]
where $x = \text{vec}(X)$ and $z = \text{vec}(Z)$. Moreover, by defining a selection matrix $S$, of dimension $|P| \times NM$, to choose $|P|$ symbols (pilots) among the total $NM$ ($P$ is the set of pilots and $|$ indicates its cardinality), the transmitted vector of pilots $x_{\text{pl,OFDM}}$ (read: “x-pilot”) takes the form
\[
x_{\text{pl,OFDM}} = Sx, \tag{12}
\]
and
\[
y_{\text{pl}} = x_{\text{pl,OFDM}} \odot (SDh_{\text{sp}}) + z. \tag{13}
\]
The estimation of the channel coefficients can be carried on by solving the problem known as LASSO [28], i.e.,
\[
\hat{h} = \arg \min_h \|y_{\text{pl}} - x_{\text{pl,OFDM}} \odot (SDh)\|_2^2 + \lambda \|h\|_1, \tag{14}
\]
where $SD$, under this configuration, takes the role of the sensing matrix of the CS configuration, and $\lambda$ is the LASSO regularizer (see details in Appendix A). Notice that $y_{\text{pl}}$ is obtained through the transmission over the actual channel, and thus differs from $y_{\text{pl}}$, 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 $\Gamma$ with sufficiently fine discretization. In any case, the residual approximation error between $y_{\text{pl}}$ and $\hat{y}_{\text{pl}}$ is automatically included in the minimization or the overall quadratic error term $\|y_{\text{pl}} - x_{\text{pl,OFDM}} \odot SDh\|_2^2$ 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 of (14) and subsequently passed to the symbol detector is given by
\[
\hat{H} = D\hat{h}. \tag{15}
\]
For completeness and for the sake of reproductibility 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., $H$ or $D_{\text{sp}}$). 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) at the top of next page, in which the ICI-free approximation follows the assumption $\nu_{\text{max}}/\Delta f \ll 1$, and the last equality follows by using the orthogonality property. Note that (17) is equivalent 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

$$X[n, m] = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x_{k,l} e^{i2\pi(\frac{nk}{NT} - \frac{lm}{M\Delta f})}, \quad (18)$$

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

$$s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X[n, m] g_{tx}(t-nT)e^{i2\pi m\Delta f(t-nT)}, \quad (19)$$

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

$$r(t) = \sum_{p=0}^{P-1} h_p s(t-\tau_p) e^{i2\pi f_p t}, \quad (20)$$

and the output of the receiver filter-bank adopting a generic receive shaping pulse $g_{rx}(t)$ is given in (21). By sampling
where the ISI coefficient of the Doppler-delay pair is given by (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 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 counting 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.

A block of dimension N × 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
\[
y(t, f) = \int r(t')g^*_{tx}(t' - t)e^{-j2\pi ft'}dt' = \int^P_{t'=0} g^*_{tx}(t' - t)\sum_{p=0}^{P-1} h_p s(t' - \tau_p)e^{-j2\pi nv_p f'}e^{-j2\pi ft'}dt'
\]
\[
= \sum_{p,n',m'} h_p X[n', m'] \int^P_{t'=0} g^*_{tx}(t' - t)g_{tx}(t' - \tau_p - n'T)e^{j2\pi m'\tau f(t'-\tau_p-n'T)}e^{j2\pi (\nu_p-f)t'}dt'
\]
\[
y[n, m] = y(t, f) |_{t=nT, f=m\Delta f} = \sum_{n'=0}^{N-1} \sum_{m'=0}^{M-1} X[n', m']h_{n,m}[n', m']
\]
\[
h_{n,m}[n', m'] = \sum_{p=0}^{P-1} h_p^* C_{g_{tx},g_{rx}}((n - n')T - \tau_p, (m - m')\Delta f - \nu_p)e^{j2\pi n'T\nu_p}e^{-j2\pi m\Delta f}\tau_p
\]
\[
\Psi^p_{k,k}[l, l'] = \sum_{n,n',m,m'} \frac{C_{g_{tx},g_{rx}}((n - n')T - \tau_p, (m - m')\Delta f - \nu_p)e^{j2\pi n'T\nu_p}e^{-j2\pi m\Delta f}\tau_p}{NM}e^{-j2\pi \left( \frac{n'}{N} - \frac{n}{N} \right)}e^{-j2\pi \left( \frac{m'}{M} - \frac{m}{M} \right)}
\]

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 \(x_{pl,OFDM}\), defined in (12) through a selection matrix \(S\), results to be a subset of symbols \(x\). 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 \(x_{pl,OTFS}\) 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].\(^3\) 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 \(\theta = \{h'_p, \tau_p, \nu_p\} \in \mathcal{T}^P\), with \(\mathcal{T} = \mathbb{C} \times \mathbb{R} \times \mathbb{R}\). By defining the ML function as

\[
l(y|\theta, x_{pl,OTFS}) = \left| y - \sum_{p} h_p^* \Psi_p x_{pl,OTFS} \right|^2,
\]

the ML solution becomes

\[
\hat{\theta} = \arg \min_{\theta \in \mathcal{T}^P} l(y|\theta, x_{pl,OTFS}).
\]

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

\[
x_{pl,OTFS} \Psi_{p} \left( \sum_{q=0}^{P-1} h'_q \Psi_q \right) x_{pl,OTFS} = x_{pl,OTFS} \Psi_{p} \Psi_{p} x_{pl,OTFS}.
\]

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. \(\theta\) reduces to maximizing the function

\[
l_2(y|\theta, x_{pl,OTFS}) = \sum_{p} h'_p y^n x_{pl,OTFS} \Psi_p x_{pl,OTFS}
\]

\[
= \sum_{p} S(\tau_p, \nu_p) - I_p(\{h'_q\}_{q\neq p}, \theta),
\]

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

\[
S_p = \frac{|y^n \Psi_p x_{pl,OTFS}|^2}{|\Psi_p x_{pl,OTFS}|^2},
\]

\[
I_p(\{h'_q\}_{q\neq p}, \theta) = \sum_{q=0}^{P-1} h'_q \Psi_q x_{pl,OTFS} \Psi_{p} x_{pl,OTFS}.
\]
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).

**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 $x_{pl, OTFS}$), the multiplication between the channel matrix $\Psi_p$ and the pilot vector $x_{pl, OTFS}$ is just a column selection of matrix $\Psi_p$, 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

After the definition of the ML approach, Algorithm 2 describes the actual CSI estimation, i.e., the estimation of the
configuration of Fig. 3), but note that the reduction in terms of computational complexity could be remarkable.

IV. COMPARISON IN TERMS OF PRAGMATIC 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 \( D \) 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 base 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

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

where \( P(x_k|\hat{x}_k) \) is the a-posteriori probability mass function of symbols \( x_k \) belonging to \( \mathcal{C} \) 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
\( \hat{\mathbf{H}} \) is given by
\[
\hat{x} = \mathbf{H}^* (\mathbf{H} \mathbf{H}^* + \sigma^2 \mathbf{I})^{-1} \mathbf{y} .
\] (38)

Note that for OFDM, whose (estimated) channel matrix is diagonal, the LMMSE detector reduces to the symbol-by-symbol detection given by
\[
\hat{x}_k = \frac{H^*_{k,k}}{|H_{k,k}|^2 + \sigma_k^2} y_k ,
\] (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 as AWGN (symmetric) capacity \( C_{\text{sym}}^{\text{AWGN}} \), [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 \( R^\ell \) is given by
\[
R^\ell = \frac{P}{NM} \times 100 \, \% .
\] (40)

and the rate simply becomes the AWGN capacity multiplied by the fraction of data symbols per block, i.e., \( |D|/NM = 1 - |P|/NM \). 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
\[
\frac{T}{T + T/4} = \frac{T}{5T/4} = \frac{4}{5} = 0.8 = 20\% .
\] (41)

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

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 \( \Delta f \) and the maximum Doppler shift introduced by the channel. In order to have negligible ICI, the necessary condition is \( \Delta f \gg \nu_{\text{max}} \), or, equivalently, \( \nu_{\text{max}}/\Delta f \ll 1 \). 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.

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 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) $C_{AWGN}$ 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 is 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. $C_{AWGN}$ 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 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.

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., \( \nu_{\text{max}}/\Delta f \), 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 \( \nu_{\text{max}} \) and \( M \), the performance of OFDM depends only on their ratio. Fig. 7 shows that the PC performance starts decreasing significantly for \( \nu_{\text{max}}/\Delta f \approx 0.15 \). 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. Conclusions

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.

### 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 \( A \triangleq X_{\text{pl}} \odot \Sigma D \), in which \( X_{\text{pl}} \) is a matrix of dimension \( |\mathcal{P}| \times G \) composed of \( G \) equal vectors \( x_{\text{pl}} \), i.e.,

\[
\begin{array}{cccc}
\nu_{\text{max}}/\Delta f & 0.1 & 0.15 & 0.2 \\
64 & 1431 [\text{km/h}] & 2147 [\text{km/h}] & 2863 [\text{km/h}] \\
128 & 715 [\text{km/h}] & 1073 [\text{km/h}] & 1431 [\text{km/h}] \\
256 & 357 [\text{km/h}] & 536 [\text{km/h}] & 715 [\text{km/h}] \\
512 & 178 [\text{km/h}] & 268 [\text{km/h}] & 357 [\text{km/h}] \\
\end{array}
\]
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 define 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 \( \Gamma \) 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 \( \hat{h} \);
    • 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 \( \hat{\mathbf{h}} \), is able to provide a first coarse and reliable estimation optimizing the computational complexity. However, if \( \mathbf{D} \) 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 \( \alpha_i \) is not mandatory. By following the pioneering work [31], which inspired many other works, e.g., [29], \( \alpha_i \) can be recursively defined as

\[
\alpha_i = \frac{\xi_i - 1}{\xi_{i+i}},
\]

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
Fig. 8: The evolution of $\alpha_i$ over iterations $i$ for different approaches.

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 $\alpha_i = 0$, completely forgetting the previous estimated values. Intuitively, a small value of $\alpha_i$ is preferable for the first noisy iterations, while a parameter $\alpha_i$ near to one should be chosen when the reliability of the estimation increases.

REFERENCES

[1] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[2] Ye Li, “Pilot-symbol-aided channel estimation for OFDM in wireless systems,” IEEE Trans. Veh. Technol., vol. 49, no. 4, pp. 1207–1215, July 2000.

[3] Ji-Woong Choi and Yong-Hwan 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, 2020.

[6] A. Lancho, J. Ostman, 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, 2019, pp. 612–616.

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

[8] Lizhong 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, 2003.

[9] R. Hadani, S. Rakih, A. F. Molisch, C. Ibars, A. Monk, M. Tsatsanis, J. Delfeld, A. Goldsmith, and R. Calderbank, “Orthogonal Time Frequency Space (OTFS) modulation for millimeter-wave communications systems,” in 2017 IEEE MTTS Int. Microw. Symp. (IMS), June 2017, pp. 681–683.

[10] A. Kavcic, Xiao 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, July 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, April 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?” 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, 2002.

[14] A. Chindapol and J. A. Rice, “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, 2001.

[15] V. Q. Rodriguez, F. Guillemin, A. Ferriex, and L. Thomas, “Cloud-RAN functional split for an efficient fronthaul network,” in 2020 International Wireless Communications and Mobile Computing (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. Technol., 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 Communications Surveys Tutorials, vol. 18, no. 3, pp. 1617–1655, 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, 2020.

[19] W. Shen, L. Dai, J.-p. An, P. Fan, and R. W. Heath, “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, June 2010.

[21] Wang Dongming, Han Bing, Zhao Junhui, Gao Xiqi, and You Xiaohu, “Channel estimation algorithms for broadband MIMO-OFDM sparse channel,” in 14th IEEE Proc. on Personal, Indoor and Mobile Radio Comm., 2003. PIMRC 2003., vol. 2, Sep. 2003, pp. 1929–1933 vol.2.

[22] V. Rajan, A. A. Balakrishnan, and K. E. Nissar, “OFDM channel estimation using compressed sensing II-regularized least square problem solver,” in 2013 Third International Conference on Advances in Computing and Communications, Aug 2013, pp. 94–97.

[23] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Raubath, “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, April 2010.

[24] Chun-Jung Wu and D. W. Lin, “A group matching pursuit algorithm for sparse channel estimation for OFDM transmission,” in 2006 IEEE Int. Conf. on Acoust. Speech and Signal Process, Proc., vol. 4, May 2006, pp. IV–IV.

[25] M. Zhang, F. Wang, X. Yuan, and L. Chen, “2D structured turbo compressed sensing for channel estimation in OTFS systems,” in 2018 IEEE Int. Conf. on 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., pp. 1–1, 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. Wireless Commun., pp. 1–1, 2021.

[28] R. Tibshirani, “Regression shrinkage and selection via the lasso,” Journal of the Royal Statistical Society: Series B (Methodological), 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 Journal on Imaging Sciences, vol. 2, no. 1, pp. 183–202, 2009.

[30] Y. Nesterov, “Gradient methods for minimizing composite functions,” Mathematical Programming, vol. 140, no. 1, pp. 125–161, 2013.

[31] R. Tibshirani, “The lasso method for variable selection in the cox proportional hazards model,” in International Conference on Machine Learning, 2000, p. 911.

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

[33] M. Schmidt, “Least squares optimization with L1-norm regularization,” Mathematical Programming, vol. 346, no. 1, pp. 251–267, 2008.

[34] G. Taubock, D. Tse, S. Buzzi, G. Colavolpe, and P. A. Martin, Wireless communications: algorithmic techniques. John Wiley & Sons, 2013.

[35] E. J. Candes et al., “The restricted isometry property and its implications for compressed sensing,” Comptes rendus mathematique, vol. 346, no. 9-10, pp. 589–592, 2008.

[36] P. Koiran and A. Zouzias, “Hidden cliques and the certification of the restricted isometry property,” IEEE transactions on information theory, vol. 60, no. 8, pp. 4999–5006, 2014.
Lorenzo Gaudio (Graduate Student Member, IEEE) received the Master of Science degree in Communication Engineering (cum laude) from the University of Parma, Parma, Italy, in 2017, and the Ph.D. degree in Information Technology with the Department of Engineering and Architecture at the University of Parma, Parma, Italy, in 2021. His main research interests include digital communications systems, channel coding and information theory, and signal processing.

Giuseppe Caire (S’92 – M’94 – SM’03 – F’05) was born in Torino in 1965. He received the B.Sc. in Electrical Engineering from Politecnico di Torino in 1990, the M.Sc. in Electrical Engineering from Princeton University in 1992, and the Ph.D. from Politecnico di Torino in 1994. He has been a post-doctoral research fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) for more than 20 years in the field of communications theory, information theory, channel and source coding with particular focus on wireless communications.