Compressive Estimation of Doubly Selective Channels in Multicarrier Systems: Leakage Effects and Sparsity-Enhancing Processing

Georg Tauböck, *Member, IEEE*, Franz Hlawatsch, *Senior Member, IEEE*, Daniel Eiwen, *Student Member, IEEE*, and Holger Rauhut

Abstract—We consider the application of *compressed sensing* (CS) to the estimation of doubly selective channels within pulse-shaping multicarrier systems (which include orthogonal frequency-division multiplexing (OFDM) systems as a special case). By exploiting sparsity in the delay-Doppler domain, CS-based channel estimation allows for an increase in spectral efficiency through a reduction of the number of pilot symbols. For combating leakage effects that limit the delay-Doppler sparsity, we propose a sparsity-enhancing basis expansion and a method for optimizing the basis with or without prior statistical information about the channel. We also present an alternative CS-based channel estimator for (potentially) strongly time–frequency dispersive channels, which is capable of estimating the "off-diagonal" channel coefficients characterizing intersymbol and intercarrier interference (ISI/ICI). For this estimator, we propose a basis construction combining Fourier (exponential) and prolate spheroidal sequences. Simulation results assess the performance gains achieved by the proposed sparsity-enhancing processing techniques and by explicit estimation of ISI/ICI channel coefficients.

Index Terms—channel estimation, compressed sensing, CoSaMP, dictionary learning, doubly selective channel, intercarrier interference, intersymbol interference, Lasso, multicarrier modulation, orthogonal frequency-division multiplexing (OFDM), orthogonal matching pursuit (OMP), sparse reconstruction.

I. INTRODUCTION

THE recently introduced principle and methodology of *compressed sensing* (CS) allows the efficient reconstruction of sparse signals from a very limited number of measurements (samples) [1], [2]. CS has gained a fast-growing interest in applied mathematics and signal processing [3]. In this paper, we apply CS to the estimation of doubly selective (doubly dispersive, doubly spread) channels. We consider

pulse-shaping multicarrier (MC) systems, which include orthogonal frequency-division multiplexing (OFDM) as a special case [4], [5]. OFDM is part of, or proposed for, numerous wireless standards like WLANs (IEEE 802.11a,g,n, Hiperlan/2), fixed broadband wireless access (IEEE 802.16), wireless personal area networks (IEEE 802.15), digital audio and video broadcasting (DAB, DRM, DVB), and future cellular communication systems (3GPP LTE) [6]–[11].

Coherent detection in such systems requires channel state information (CSI) at the receiver. Usually, CSI is obtained by embedding pilot symbols in the transmit signal and using a least-squares (LS) [12] or minimum mean-square error (MMSE) [13] channel estimator. More advanced channel estimators for MC transmissions include estimators employing 1-D, double 1-D, or 2-D MMSE filtering algorithms [14]–[16]; 2-D irregular sampling techniques [17]; or basis expansion models [18]–[20]. The CS-based ("compressive") channel estimation methodology proposed in this paper exploits the fact that doubly selective multipath channels tend to be dominated by a relatively small number of clusters of significant paths, especially for large signaling bandwidths and durations [21]. Conventional methods for channel estimation do not take advantage of this *inherent sparsity* of the channel. In [22] and [23], we proposed CS-based techniques for estimating doubly selective channels within MC systems. We demonstrated that CS provides a way to exploit channel sparsity in the sense that the number of pilot symbols that have to be transmitted for accurate channel estimation can be reduced. Transmitting fewer pilots leaves more symbols for transmitting data, which yields an increase in spectral efficiency.

For sparse channel estimation, several other authors have independently proposed the application of CS methods or methods inspired by the literature on sparse signal representations [21], [24]–[31]. Both [24] and [26] considered single-carrier signaling and proposed variants of the matching pursuit algorithm [32] for channel estimation. The results were primarily based on simulation and experimental implementations, without a CS theoretical background. The channel estimation techniques presented in [24], [27], and [28] limited themselves to sparsity in the delay domain, i.e., they did not exploit Doppler sparsity. The recent work in [29] and its extension to multiple-input/multiple-output (MIMO) channels [30], on the other hand, considered both MC signaling (besides single-carrier signaling) and sparsity in the delay-Doppler domain, somewhat similar to [22]; however, a different CS recovery technique was used. In [33], it is shown experimentally

for MC communications over underwater acoustic channels that compressive channel estimation outperforms traditional subspace algorithms (root-MUSIC and ESPRIT).

In this paper, extending our work in [22] and [23], we present CS-based techniques for estimating doubly selective channels that are potentially strongly time- and/or frequency-dispersive. In MC systems, strong channel dispersion may cause inter-symbol interference (ISI) and/or intercarrier interference (ICI) [4]. One of the proposed techniques enables the estimation of ISI/ICI channel coefficients. We first present a basic compressive estimator for mildly dispersive channels that yields estimates of the "diagonal" channel coefficients. Our focus is on *leakage effects* that limit the delay-Doppler sparsity, and which have not been considered in [21], [24]–[31]. For combating leakage effects and, hence, enhancing sparsity, we then replace the discrete Fourier transform (DFT) used in conventional compressive channel estimation by a more suitable basis expansion. We also develop an iterative basis-optimization procedure that is similar in spirit—but not algorithmically—to dictionary learning techniques recently proposed in [34]–[36]. This procedure is able to take into account prior statistical information about the channel. Next, we present an alternative compressive method for estimating also the "off-diagonal" ISI/ICI channel coefficients of potentially strongly dispersive channels (e.g., highly mobile wireless channels or underwater acoustic channels [26], [33]). Here, motivated by [20], [37], we propose a sparsity-enhancing basis expansion that combines Fourier (exponential) and prolate spheroidal sequences.

This paper is organized as follows. In Section II, we describe the MC system model. In Section III, we present the basic compressive estimator for mildly dispersive channels. An analysis of delay-Doppler leakage and its effect on delay-Doppler sparsity is performed in Section IV. A sparsity-enhancing basis expansion and a framework and iterative algorithm for optimizing the basis (with or without prior statistical information about the channel) are developed in Sections V and VI, respectively. In Section VII, we propose a compressive estimator and a basis expansion for (potentially) strongly dispersive channels. Finally, simulation results presented in Section VIII assess the performance gains achieved by the proposed sparsity-enhancing basis expansions and by the estimation of ISI/ICI channel coefficients.

II. MULTICARRIER SYSTEM MODEL

We assume a *pulse-shaping* MC system for the sake of generality and because of its advantages over conventional cyclic-prefix (CP) OFDM [4], [38]–[41]. This framework includes CP-OFDM as a special case. The complex baseband domain is considered throughout.

A. Modulator, Channel, Demodulator

The MC modulator generates the discrete-time transmit signal [4]

$$s[n] = \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} a_{l,k}\, g_{l,k}[n] \qquad (1)$$

where L and K denote the numbers of transmitted MC symbols and subcarriers, respectively; $a_{l,k} \in \mathcal{A}$ ($l = 0, \ldots, L-1; k = 0, \ldots, K-1$) denotes the complex data symbols, drawn from a finite symbol alphabet $\mathcal{A}$; and $g_{l,k}[n] \triangleq g[n-lN]e^{j2\pi k(n-lN)/K}$ is a time–frequency shift of a transmit pulse $g[n]$ ($N \geq K$ is the symbol duration). Using an interpolation filter with impulse response $f_1(t)$, $s[n]$ is converted into the continuous-time transmit signal

$$s(t) = \sum_{n=-\infty}^{\infty} s[n] f_1(t - nT_s) \qquad (2)$$

where T_s is the sampling period. This signal is transmitted over a noisy, doubly selective channel, at whose output the receive signal

$$r(t) = \int_{-\infty}^{\infty} h(t,\tau) s(t-\tau)\, d\tau + z(t) \qquad (3)$$

is obtained. Here, $h(t,\tau)$ is the channel's time-varying impulse response and $z(t)$ is complex noise. At the receiver, $r(t)$ is converted into the discrete-time receive signal

$$r[n] = \int_{-\infty}^{\infty} r(t) f_2(nT_s - t)\, dt \qquad (4)$$

where $f_2(t)$ is the impulse response of an anti-aliasing filter. Subsequently, the MC demodulator calculates the "demodulated symbols"

$$r_{l,k} = \langle r, \gamma_{l,k} \rangle = \sum_{n=-\infty}^{\infty} r[n] \gamma_{l,k}^*[n],$$
$$l = 0, \ldots, L-1, \quad k = 0, \ldots, K-1. \qquad (5)$$

Here, $\gamma_{l,k}[n] \triangleq \gamma[n-lN]e^{j2\pi k(n-lN)/K}$ with a receive pulse $\gamma[n]$. Finally, the demodulated symbols $r_{l,k}$ are equalized and quantized according to the data symbol alphabet $\mathcal{A}$.

Combining (2)–(4), we obtain an equivalent discrete-time channel that is described by the following relation between the discrete-time signals $s[n]$ and $r[n]$:

$$r[n] = \sum_{m=-\infty}^{\infty} h[n,m] s[n-m] + z[n] \qquad (6)$$

with the discrete-time time-varying impulse response $h[n,m] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t+nT_s,\tau) f_1(t-\tau+mT_s) f_2(-t)\, dt\, d\tau$ and the discrete-time noise $z[n] = \int_{-\infty}^{\infty} z(t) f_2(nT_s - t)\, dt$.

CP-OFDM is a simple special case of the pulse-shaping MC framework; it is obtained for a rectangular transmit pulse $g[n]$ that is 1 for $n = 0, \ldots, N-1$ and 0 otherwise, and a rectangular receive pulse $\gamma[n]$ that is 1 for $n = N-K, \ldots, N-1$ and 0 otherwise ($N-K \geq 0$ is the CP length).

B. System Channel

Next, we consider the equivalent *system channel* that subsumes the MC modulator, interpolation filter, physical channel,

anti-aliasing filter, and MC demodulator. Combining (5), (6), and (1), we obtain

$$r_{l,k} = \sum_{l'=0}^{L-1} \sum_{k'=0}^{K-1} H_{l,k;l',k'} a_{l',k'} + z_{l,k},$$

$$l = 0,\ldots,L-1, \quad k = 0,\ldots,K-1 \quad (7)$$

with $z_{l,k} = \langle z, \gamma_{l,k} \rangle = \sum_{n=-\infty}^{\infty} z[n]\gamma_{l,k}^*[n]$. The system channel coefficients $H_{l,k;l',k'}$ describe ICI for $k \neq k'$ and $l = l'$ and ISI for $l \neq l'$; they can be expressed in terms of $h[n,m], g[n]$, and $\gamma[n]$ [4].

Let $\gamma[n]$ be zero outside $\{0,\ldots,L_\gamma\}$. To compute $r_{l,k}$ in (5) for $l = 0,\ldots,L-1$, we need to know $r[n]$ for $n = 0,\ldots,N_r-1$, where $N_r \triangleq (L-1)N + L_\gamma + 1$. In this interval, we can rewrite (6) as

$$r[n] = \sum_{m=-\infty}^{\infty} \sum_{i=0}^{N_r-1} S_h[m,i]s[n-m]e^{j2\pi\frac{ni}{N_r}} + z[n] \quad (8)$$

with the *discrete-delay-Doppler spreading function* [42]

$$S_h[m,i] \triangleq \frac{1}{N_r} \sum_{n=0}^{N_r-1} h[n,m]e^{-j2\pi\frac{in}{N_r}}, \quad m,i \in \mathbb{Z} \quad (9)$$

which represents the channel in terms of discrete delay (time shift) m and discrete Doppler frequency shift i. Combining (5), (8), and (1), and assuming that $h[n,m]$ is causal with maximum delay at most $K-1$, i.e., $h[n,m] = 0$ for $m \notin \{0,\ldots,K-1\}$, we reobtain the system channel relation (7), however with the system channel coefficients $H_{l,k;l',k'}$ now expressed in terms of the delay-Dopler representation $S_h[m,i]$. Specializing this expression to $(l',k') = (l,k)$ and using the approximation $N_r \approx LN$ (which is exact for CP-OFDM) yields the following expression for the *diagonal* channel coefficients $H_{l,k} \triangleq H_{l,k;l,k}$ (L is assumed even for mathematical convenience)

$$H_{l,k} = \sum_{m=0}^{K-1} \sum_{i=-L/2}^{L/2-1} F[m,i]e^{-j2\pi(\frac{km}{K} - \frac{li}{L})},$$

$$l = 0,\ldots,L-1, \quad k = 0,\ldots,K-1 \quad (10)$$

with

$$F[m,i] \triangleq \sum_{q=0}^{N-1} S_h[m,i+qL]A_{\gamma,g}^*\left(m, \frac{i+qL}{N_r}\right), \quad i \in \mathbb{Z}. \quad (11)$$

Here, $A_{\gamma,g}(m,\xi) \triangleq \sum_{n=-\infty}^{\infty} \gamma[n]g^*[n-m]e^{-j2\pi\xi n}$ is the *cross-ambiguity function* [43] of $\gamma[n]$ and $g[n]$.

III. COMPRESSIVE CHANNEL ESTIMATION

We now present the basic compressive channel estimation method [22], [29]. This method enables estimation of the diagonal channel coefficients $H_{l,k} = H_{l,k;l,k}$, which is sufficient for mildly dispersive channels.

A. Pilot-Assisted Channel Estimation

Our goal is to estimate the system channel coefficients $H_{l,k} = H_{l,k;l,k}$ from the system channel output $r_{l,k}$, aided by some known pilot symbols. For practical (underspread [42]) wireless channels and practical transmit and receive pulses, $F[m,i]$ in (11) is effectively supported in a subregion of the delay-Doppler plane. Thus, hereafter we assume that the support of $F[m,i]$ (within the fundamental i period $\{-L/2,\ldots,L/2-1\}$; note that $F[m,i]$ is L-periodic in i) is contained in $\{0,\ldots,D-1\} \times \{-J/2,\ldots,J/2-1\}$, where $D \leq K$ and $J \leq L$. Here, J is chosen even, and D and J are such that $\Delta K \triangleq K/D$ and $\Delta L \triangleq L/J$ are integers. Note that we also allow the limiting case of full support in either or both dimensions, that is, $D = K$ (i.e., $\Delta K = 1$) and/or $J = L$ (i.e., $\Delta L = 1$). Because of (10), the $H_{l,k}$ are then uniquely specified by their values on the *subsampled time–frequency grid*

$$\mathcal{G} \triangleq \{(l,k) = (\lambda\Delta L, \kappa\Delta K) : \lambda = 0,\ldots,J-1,$$
$$\kappa = 0,\ldots,D-1\}.$$

These subsampled values are given by

$$H_{\lambda\Delta L, \kappa\Delta K} = \sum_{m=0}^{D-1} \sum_{i=-J/2}^{J/2-1} F[m,i]e^{-j2\pi(\frac{\kappa m}{D} - \frac{\lambda i}{J})},$$

$$\lambda = 0,\ldots,J-1, \quad \kappa = 0,\ldots,D-1. \quad (12)$$

The time–frequency subsampling is desirable because it reduces the dimensionality of the estimation problem, and thus tends to result in better estimation performance.

Suppose now that pilot symbols $a_{l,k} = p_{l,k}$ are transmitted at time–frequency positions $(l,k) \in \mathcal{P}$, where $\mathcal{P} \subset \mathcal{G}$, i.e., the *pilot position set* $\mathcal{P}$ is a subset of the subsampled time–frequency grid $\mathcal{G}$. For mildly dispersive channels, the ISI and ICI are small. Then, at the pilot positions $(l,k) \in \mathcal{P}$, it is convenient to rewrite the system channel relation (7) as $r_{l,k} = H_{l,k}p_{l,k} + \tilde{z}_{l,k}$, where all ISI and ICI are now subsumed by the noise/interference term $\tilde{z}_{l,k}$. Based on this relation and the known $p_{l,k}$, the receiver calculates channel coefficient estimates $\hat{H}_{l,k}$ at the pilot positions according to

$$\hat{H}_{l,k} \triangleq \frac{r_{l,k}}{p_{l,k}} = H_{l,k} + \frac{\tilde{z}_{l,k}}{p_{l,k}}, \qquad (l,k) \in \mathcal{P}. \quad (13)$$

The last expression shows that the $H_{l,k}$ for $(l,k) \in \mathcal{P}$ are known up to additive noise/interference terms $\tilde{z}_{l,k}/p_{l,k}$. A conventional channel estimator then uses some interpolation technique to calculate channel estimates $\hat{H}_{l,k}$ for all (l,k) from the $\hat{H}_{l,k}$ for $(l,k) \in \mathcal{P}$ (e.g., [12]–[17]). In contrast, the proposed compressive channel estimator uses a CS recovery technique to obtain an estimate of $F[m,i]$ and, in turn, of the $H_{l,k}$.

B. Some CS Fundamentals

Before presenting the CS-based channel estimator, we need to review some CS fundamentals [1], [2]. CS considers the *sparse reconstruction problem* of estimating an (approximately) sparse

vector $\mathbf{x} \in \mathbb{C}^M$ from an observed vector of measurements $\mathbf{y} \in \mathbb{C}^Q$ based on the linear model ("measurement equation")

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{z}. \qquad (14)$$

Here, $\boldsymbol{\Phi} \in \mathbb{C}^{Q \times M}$ is a known measurement matrix and $\mathbf{z} \in \mathbb{C}^Q$ is an unknown vector that accounts for measurement noise and modeling errors. The reconstruction is subject to the constraint that $\mathbf{x}$ is (approximately) S-sparse, i.e., at most S of its entries are not (approximately) zero. The positions (indices) of the significantly nonzero entries of $\mathbf{x}$ are unknown. Typically, the number of variables to be estimated is much larger than the number of measurements, i.e., $M \gg Q$. Thus, $\boldsymbol{\Phi}$ is a fat matrix.

We briefly review some CS recovery methods. *Basis pursuit* (BP) [44], [45], and *orthogonal matching pursuit* (OMP) [46] are probably the most popular ones. Whereas for BP theoretical performance guarantees are available, OMP lacks similar results. However, OMP allows a faster implementation, and simulation results even demonstrate a better performance. Low computational complexity is important since the channel has to be estimated in real time. *CoSaMP* [47] allows an even faster implementation than OMP. (Note that *subspace pursuit* [48] is a very similar method.) Using an efficient implementation of the pseudoinverse by means of the LSQR algorithm [49], we observed a run time that was only less than half that of OMP, and a performance that was only slightly poorer. An advantage of CoSaMP is the availability of performance bounds. Hence, CoSaMP offers a good compromise between low complexity, good practical performance, and provable performance guarantees.

The performance guarantees of BP and CoSaMP are phrased as an upper bound on the approximation error $\|\hat{\mathbf{x}} - \mathbf{x}\|_2$, where $\hat{\mathbf{x}}$ denotes the estimate of $\mathbf{x}$. This bound is valid if the measurement matrix $\boldsymbol{\Phi}$ satisfies $(1-\delta)\|\mathbf{x}\|_2^2 \leq \|\boldsymbol{\Phi}\mathbf{x}\|_2^2 \leq (1+\delta)\|\mathbf{x}\|_2^2$ for all S-sparse vectors $\mathbf{x} \in \mathbb{C}^M$, with some positive constant δ. This is known as the *restricted isometry property* (RIP), and the smallest δ is termed the *restricted isometry constant* δ_S. For a small bound on $\|\hat{\mathbf{x}} - \mathbf{x}\|_2$, δ_S should be small. It has been shown [1], [50], [51] that if $\boldsymbol{\Phi} \in \mathbb{C}^{Q \times M}$ is constructed by selecting uniformly at random Q rows[1] from a unitary $M \times M$ matrix $\mathbf{U}$ and normalizing the columns (so that they have unit ℓ_2 norms), a sufficient condition for $\boldsymbol{\Phi}$ to satisfy the RIP with a restricted isometry constant that is bounded as $\delta_S \leq \gamma$ with probability $1 - \eta$ is provided by the following lower bound on the number of observations:

$$Q \geq C\gamma^{-2}(\ln M)^4 \mu_{\mathbf{U}}^2 S \ln(1/\eta). \qquad (15)$$

Here, $\mu_{\mathbf{U}} \triangleq \sqrt{M}\max_{i,j} |U_{i,j}|$ (known as the *coherence* of $\mathbf{U}$) and C is a constant.

Further CS recovery methods include *thresholding* [52], the *stagewise OMP* [53], the *LARS method* [54], [55], the *Lasso* [56], [57] (equivalent to *BP denoising* [57]), and Bayesian methods [58], [59]. In [29] and [30], the *Dantzig selector* (DS) [60] was applied to sparse channel estimation. DS satisfies optimal asymptotic performance bounds when the noise vector

[1]That is, all possible choices of Q rows are equally likely.

$\mathbf{z}$ is modeled as random. However, for the practically relevant case of finite (moderate) Q and M, the performance of DS is not necessarily superior. In our experiments, we did not observe any performance or complexity advantages of DS over BP, OMP, and CoSaMP.

C. Basic Compressive Channel Estimator

We now combine pilot-assisted channel estimation with CS recovery. The central assumption of compressive channel estimation is that $S_h[m, i]$ is "compressible" [45] or approximately S-sparse, i.e., at most S values of $S_h[m, i]$ (in the fundamental i period $\{-L/2, \ldots, L/2 - 1\}$) are not approximately zero. This approximate "delay-Doppler sparsity" assumption will be further considered in Section IV. Note that it implies that also $F[m, i] = \sum_{q=0}^{N-1} S_h[m, i + qL]A_{\gamma,g}^*(m, (i+qL)/N_r)$ is approximately S-sparse.

Our starting-point is the 2-D DFT relation (12), which can be written as the 2-D expansion

$$H_{\lambda\Delta L, \kappa\Delta K} = \sum_{m=0}^{D-1} \sum_{i=-J/2}^{J/2-1} \alpha_{m,i} u_{m,i}[\lambda, \kappa] \qquad (16)$$

with $\alpha_{m,i} \triangleq \sqrt{JD}F[m, i]$ and $u_{m,i}[\lambda, \kappa] \triangleq (1/\sqrt{JD})e^{-j2\pi(\kappa m/D - \lambda i/J)}$. The functions $H_{\lambda\Delta L, \kappa\Delta K}$ and $u_{m,i}[\lambda, \kappa]$ are defined for $\lambda = 0, \ldots, J-1$ and $\kappa = 0, \ldots, D-1$ and may thus be considered as $J \times D$ matrices. Define the vectors $\mathbf{h} \triangleq \mathrm{vec}\{H_{\lambda\Delta L, \kappa\Delta K}\}$ and $\mathbf{u}_{m,i} \triangleq \mathrm{vec}\{u_{m,i}[\lambda, \kappa]\}$ of length JD by stacking all columns of these matrices (e.g., $\mathbf{h} = [h_1 \cdots h_{JD}]^T$ with $h_{\kappa J + \lambda + 1} = H_{\lambda\Delta L, \kappa\Delta K}$). We can then rewrite (16) as

$$\mathbf{h} = \sum_{m=0}^{D-1} \sum_{i=-J/2}^{J/2-1} \alpha_{m,i}\mathbf{u}_{m,i} = \mathbf{U}\boldsymbol{\alpha} \qquad (17)$$

where $\boldsymbol{\alpha} \triangleq \mathrm{vec}\{\alpha_{m,i}\}$ and $\mathbf{U}$ is the $JD \times JD$ matrix whose $((i + J/2)D + m + 1)$th column is given by the vector $\mathbf{u}_{m,i}$. Because the $\mathbf{u}_{m,i}$ are orthonormal, $\mathbf{U}$ is a unitary matrix.

According to Section III-A, there are $|\mathcal{P}|$ pilot symbols at time–frequency positions $(l, k) \in \mathcal{P}$. Thus, $|\mathcal{P}|$ of the JD entries of $\mathbf{h}$ are given by the channel coefficients $H_{l,k}$ at the pilot positions $(l, k) \in \mathcal{P}$. Let $\mathbf{h}^{(\mathrm{p})}$ denote the corresponding length-$|\mathcal{P}|$ subvector of $\mathbf{h}$, and let $\mathbf{U}^{(\mathrm{p})}$ denote the $|\mathcal{P}| \times JD$ submatrix of $\mathbf{U}$ constituted by the corresponding $|\mathcal{P}|$ rows of $\mathbf{U}$. Reducing (17) to the pilot positions, we obtain

$$\mathbf{h}^{(\mathrm{p})} = \mathbf{U}^{(\mathrm{p})}\boldsymbol{\alpha} = \boldsymbol{\Phi}\mathbf{x} \qquad (18)$$

with $\boldsymbol{\Phi} \triangleq \sqrt{JD/|\mathcal{P}|}\mathbf{U}^{(\mathrm{p})}$ and $\mathbf{x} \triangleq \sqrt{|\mathcal{P}|/(JD)}\boldsymbol{\alpha}$. Note that $\boldsymbol{\Phi}$ is normalized such that its columns have unit ℓ_2-norm, and that the length-JD vector $\mathbf{x}$ is, up to a constant factor, the vector form of $F[m, i]$.

Our task is to estimate $\mathbf{x}$ based on relation (18). The vector $\mathbf{h}^{(\mathrm{p})}$ is unknown, but we can approximate it by the corresponding vector of pilot-based channel coefficient estimates $\hat{H}_{l,k}\big|_{(l,k) \in \mathcal{P}}$ [see (13)]. For consistency with the notation used in Section III-B, this latter vector will be denoted as $\mathbf{y}$ (rather

than $\hat{\mathbf{h}}^{(\mathrm{p})}$). According to (13), $\mathbf{y} = \mathbf{h}^{(\mathrm{p})} + \mathbf{z}$, where $\mathbf{z}$ is the vector of noise/interference terms $\tilde{z}_{l,k}/p_{l,k}\big|_{(l,k)\in\mathcal{P}}$. Inserting (18), we finally obtain the *measurement equation*

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{z}. \qquad (19)$$

The vector $\mathbf{x}$ is approximately S-sparse because $S_h[m,i]$ was assumed approximately S-sparse. Thus, (19) is seen to be a sparse reconstruction problem of the form (14), with dimensions $M = \dim\{\mathbf{x}\} = JD$ and $Q = \dim\{\mathbf{y}\} = |\mathcal{P}|$ and sparsity S. We can hence use one of the CS recovery techniques reviewed in Section III-B to obtain an estimate of $\mathbf{x}$ or, equivalently, of $\boldsymbol{\alpha} = \sqrt{JD/|\mathcal{P}|}\mathbf{x}$ or of $F[m,i] = \alpha_{m,i}/\sqrt{JD}$. From the estimate $\hat{F}[m,i]$ of $F[m,i]$, estimates of all channel coefficients $H_{l,k}$ are finally obtained via (10).

According to its definition $\mathbf{\Phi} = \sqrt{JD/|\mathcal{P}|}\mathbf{U}^{(\mathrm{p})}$, the measurement matrix $\mathbf{\Phi}$ is constructed by selecting $|\mathcal{P}|$ rows of the unitary $JD \times JD$ matrix $\mathbf{U}$ and normalizing the resulting columns. This agrees with the construction of $\mathbf{\Phi}$ described in Section III-B in the context of BP and CoSaMP. To be fully consistent with that construction, we have to select the $|\mathcal{P}|$ rows of $\mathbf{U}$ uniformly at random. The indices of these rows equal the $|\mathcal{P}|$ indices within the index range $\{1,\dots,JD\}$ of the channel vector $\mathbf{h}$ that correspond to the set of pilot positions $\mathcal{P}$. We conclude that the pilot positions $(l,k) \in \mathcal{P}$ have to be selected uniformly at random within the subsampled time–frequency grid $\mathcal{G}$, in the sense that the $|\mathcal{P}|$ "pilot indices" within the index range $\{1,\dots,JD\}$ of $\mathbf{h}$ are selected uniformly at random.

For BP and CoSaMP, in order to achieve a small upper bound on the reconstruction error $\|\hat{\mathbf{x}} - \mathbf{x}\|_2$ as discussed in Section III-B, the number of pilots should satisfy condition (15). In our case, this (sufficient) condition becomes

$$|\mathcal{P}| \geq C\gamma^{-2}(\ln(JD))^4 S\ln(1/\eta)$$

with an appropriately chosen γ (note that $\mu_{\mathbf{U}} = 1$). This bound suggests that the required number of pilots scales at most linearly with the delay-Doppler sparsity parameter S and poly-logarithmically with the system design parameters J and D. Note that the pilot positions are randomly chosen (and communicated to the receiver) before the beginning of data transmission; they are fixed during data transmission.

IV. DELAY-DOPPLER SPARSITY AND LEAKAGE EFFECT

In this section, we analyze the sparsity of the channel's delay-Doppler representation for a simple time-varying multipath channel model comprising P specular (point) scatterers with fixed delays τ_p and Doppler frequency shifts ν_p for $p = 1,\dots,P$. This simple model is often a good approximation to real mobile radio channels [61], [62]. The channel impulse response thus has the form

$$h(t,\tau) = \sum_{p=1}^{P} \eta_p \delta(\tau - \tau_p)e^{j2\pi\nu_p t} \qquad (20)$$

where η_p characterizes the attenuation and initial phase of the pth propagation path and $\delta(\cdot)$ is the Dirac delta. The discrete-delay-Doppler spreading function (9) then becomes

$$\begin{aligned}
S_h[m,i] &= \frac{1}{N_r}\sum_{p=1}^{P}\eta_p\phi^{(\nu_p)}\left(m - \frac{\tau_p}{T_{\mathrm{s}}}\right)\sum_{n=0}^{N_r-1}e^{j2\pi(\nu_p T_{\mathrm{s}} - \frac{i}{N_r})n} \\
&= \sum_{p=1}^{P}\eta_p e^{j\pi(\nu_p T_{\mathrm{s}} - \frac{i}{N_r})(N_r-1)} \\
&\quad \times \Lambda^{(\nu_p)}\left(m - \frac{\tau_p}{T_{\mathrm{s}}}, i - \nu_p T_{\mathrm{s}} N_r\right)
\end{aligned} \qquad (21)$$

with

$$\Lambda^{(\nu)}(x,y) \triangleq \phi^{(\nu)}(x)\psi(y)$$

where

$$\phi^{(\nu)}(x) \triangleq \int_{-\infty}^{\infty}f_1(T_{\mathrm{s}}x - t)f_2(t)e^{-j2\pi\nu t}dt \qquad (22)$$

$$\begin{aligned}
\psi(y) &\triangleq \frac{1}{N_r}e^{j\pi\frac{y}{N_r}(N_r-1)}\sum_{n=0}^{N_r-1}e^{-j2\pi\frac{y}{N_r}n} \\
&= \frac{\sin(\pi y)}{N_r\sin(\pi y/N_r)}.
\end{aligned} \qquad (23)$$

It is seen from (21) that, although we assumed specular scattering, $S_h[m,i]$ does not consist of Dirac-like functions at the delay-Doppler points of the scatterers, $(\tau_p/T_{\mathrm{s}}, \nu_p T_{\mathrm{s}} N_r)$. Rather, there occurs a *leakage effect* which is characterized by the function $\Lambda^{(\nu)}(x,y) = \phi^{(\nu)}(x)\psi(y)$, and which is stronger for a broader $\Lambda^{(\nu)}(x,y)$. The leakage effect is due to the finite transmit bandwidth ($\approx 1/T_{\mathrm{s}}$) and the finite blocklength ($N_r \approx LN$). It is important for compressive channel estimation because it implies a poorer sparsity of $S_h[m,i]$. Note that whereas a large blocklength reduces the leakage effect, it also implies that the specular model with constant parameters (20) is a less accurate approximation and, thus, that the continuous-delay-Doppler spreading function [42] is less sparse. This motivates an extension of the compressive channel estimation method that is able to reduce the leakage effect (see Section V).

In view of (21), studying the sparsity of $S_h[m,i]$ essentially amounts to studying the sparsity of $\Lambda^{(\nu_p)}(m - \tau_p/T_{\mathrm{s}}, i - \nu_p T_{\mathrm{s}} N_r) = \phi^{(\nu_p)}(m - \tau_p/T_{\mathrm{s}})\psi(i - \nu_p T_{\mathrm{s}} N_r)$. To this end, we first consider the energy of those samples of $\phi^{(\nu_p)}(m - \tau_p/T_{\mathrm{s}})$ whose distance from τ_p/T_{s} is greater than $\Delta m \in \{1,2,\dots\}$, i.e., $|m - \tau_p/T_{\mathrm{s}}| > \Delta m$. We assume that $\phi^{(\nu)}(x)$ exhibits at least a polynomial decay, i.e., $|\phi^{(\nu)}(x)| \leq C(1 + |x/x_0|)^{-s}$ with $s \geq 1$, for some positive constants C and x_0. This includes the following important special cases: 1) the ideal low-pass filter, i.e., $f_1(t) = f_2(t) = \sqrt{1/T_{\mathrm{s}}}\,\mathrm{sinc}(t/T_{\mathrm{s}})$ with $\mathrm{sinc}(x) \triangleq (\sin(\pi x))/(\pi x)$, here $s = 1$; and 2) the family of root-raised-cosine filters: if both $f_1(t)$ and $f_2(t)$ are equal to the root-raised-cosine filter with roll-off factor ρ, then, for ν not too large, $\phi^{(\nu)}(x) \approx \mathrm{sinc}(x)\cos(\rho\pi x)/[1-(2\rho x)^2]$ and $s = 3$. Based on the polynomial-decay assumption, one can show the

following bound [23] on the energy of all $\phi^{(\nu_p)}(m - \tau_p/T_\mathrm{s})$ with $|m - \tau_p/T_\mathrm{s}| > \Delta m$:

$$\sum_{|m - \tau_p/T_\mathrm{s}| > \Delta m} \left| \phi^{(\nu_p)}\left(m - \frac{\tau_p}{T_\mathrm{s}}\right) \right|^2$$

$$\leq \frac{2C^2 x_0}{2s - 1}\left(1 + \frac{\Delta m - 1}{x_0}\right)^{-2s + 1}.$$

Hence, the energy of $\phi^{(\nu_p)}(m - \tau_p/T_\mathrm{s})$ outside the interval $[\lfloor \tau_p/T_\mathrm{s} - \Delta m \rfloor, \lceil \tau_p/T_\mathrm{s} + \Delta m \rceil]$ decays polynomially of order $2s - 1$ with respect to Δm.

In a similar manner, we consider the energy of those samples of $\psi(i - \nu_p T_\mathrm{s} N_r)$ whose distance (up to the modulo-N_r operation, see below) from $\nu_p T_\mathrm{s} N_r$ is greater than $\Delta i \in \{2, \ldots, \lfloor N_r/2 \rfloor\}$. Let $\mathcal{I}$ denote the set $\{0, \ldots, N_r - 1\}$ with the exception of all $i = i_\mathbb{Z} \bmod N_r$, where $i_\mathbb{Z}$ is any integer with $|i_\mathbb{Z} - \nu_p T_\mathrm{s} N_r| \leq \Delta i$. From (23), one can obtain the bound [22]

$$\sum_{i \in \mathcal{I}} |\psi(i - \nu_p T_\mathrm{s} N_r)|^2 \leq \frac{1}{\pi(\Delta i - 1)}$$

which shows that the energy of $\psi(i - \nu_p T_\mathrm{s} N_r)$ outside the interval $[\lfloor \nu_p T_\mathrm{s} N_r - \Delta i \rfloor, \lceil \nu_p T_\mathrm{s} N_r + \Delta i \rceil]$ (modulo N_r) decays linearly (polynomially of order 1) with respect to Δi.

From these decay results, it follows that $\Lambda^{(\nu_p)}(m - \tau_p/T_\mathrm{s}, i - \nu_p T_\mathrm{s} N_r) = \phi^{(\nu_p)}(m - \tau_p/T_\mathrm{s})\psi(i - \nu_p T_\mathrm{s} N_r)$ can be considered as an *approximately sparse* (or *compressible*, in CS terminology [45]) function. Thus, as an approximation, we can model $\Lambda^{(\nu_p)}(m - \tau_p/T_\mathrm{s}, i - \nu_p T_\mathrm{s} N_r)$ as N_Λ-sparse, with an appropriately chosen sparsity parameter N_Λ. It then follows from (21) that $S_h[m, i]$ is $P N_\Lambda$-sparse, and the same is true for $F[m, i]$ in (11). Unfortunately, N_Λ cannot be chosen extremely small because of the strong leakage that is due to the slowly (only linearly) decaying factor $\psi(i - \nu_p T_\mathrm{s} N_r)$. This limitation motivates the introduction of a sparsity-enhancing basis expansion in the next section.

V. Sparsity-Enhancing Basis Expansion

The 2-D DFT relation (12) underlying the basic compressive channel estimator is an expansion of the subsampled channel coefficients $H_{\lambda \Delta L, \kappa \Delta K}$ into the 2-D DFT basis $u_{m,i}[\lambda, \kappa] = (1/\sqrt{JD})e^{-j2\pi(\kappa m/D - \lambda i/J)}$ [see (16)]. The sparsity of the expansion coefficients $\alpha_{m,i} = \sqrt{JD} F[m, i]$ was shown above to be limited by the slowly (only linearly) decaying function $\psi(i - \nu_p T_\mathrm{s} N_r)$. In order to enhance the sparsity, we now introduce a generalized 2-D expansion of $H_{\lambda \Delta L, \kappa \Delta K}$ into orthonormal basis functions $v_{m,i}[\lambda, \kappa]$

$$H_{\lambda \Delta L, \kappa \Delta K} = \sum_{m=0}^{D-1} \sum_{i=-J/2}^{J/2-1} \beta_{m,i} v_{m,i}[\lambda, \kappa],$$

$$\lambda = 0, \ldots, J - 1, \ \ \kappa = 0, \ldots, D - 1. \quad (24)$$

Clearly, our previous 2-D DFT expansion (12), (16) is a special case of (24).

A. 1-D and 2-D Basis Expansions

We will choose a basis $\{v_{m,i}[\lambda, \kappa]\}$ that is adapted to the channel model (20) (but not to the specific channel parameters $P, \eta_p, \tau_p,$ and ν_p in (20)). Equation (20) suggests that the coefficients $\beta_{m,i}$ should be sparse for the elementary single-scatterer channel $h^{(\tau_1, \nu_1)}(t, \tau) \triangleq \delta(\tau - \tau_1)e^{j2\pi\nu_1 t}$, for all $\tau_1 \in [0, \tau_{\max}]$ and $\nu_1 \in [-\nu_{\max}, \nu_{\max}]$. Specializing (21) to $P = 1$ and $\eta_1 = 1$, and using (11), the 2-D DFT expansion (12) yields after a straightforward calculation

$$H_{\lambda \Delta L, \kappa \Delta K} = \sum_{m=0}^{D-1} \phi^{(\nu_1)}\left(m - \frac{\tau_1}{T_\mathrm{s}}\right) C^{(\nu_1)}[m, \lambda] e^{-j2\pi\frac{\kappa m}{D}}. \quad (25)$$

Here, we have set

$$C^{(\nu_1)}[m, \lambda] \triangleq \sum_{i=-J/2}^{J/2-1} \tilde\alpha_{m,i}^{(\nu_1)} \frac{1}{\sqrt{J}} e^{j2\pi\frac{\lambda i}{J}} \quad (26)$$

where

$$\tilde\alpha_{m,i}^{(\nu_1)} \triangleq \sqrt{J} \sum_{q=0}^{N-1} \psi^{(\nu_1)}[i + qL] A_{\gamma,g}^*\left(m, \frac{i + qL}{N_r}\right) \quad (27)$$

with $\psi^{(\nu_1)}[i] \triangleq e^{j\pi(\nu_1 T_\mathrm{s} - i/N_r)(N_r - 1)}\psi(i - \nu_1 T_\mathrm{s} N_r)$.

According to (27), the poor decay of $\psi(x)$ entails a poor decay of $\tilde\alpha_{m,i}^{(\nu_1)}$ with respect to i. To improve the decay, we replace the 1-D DFT (26) by a general 1-D basis expansion

$$C^{(\nu_1)}[m, \lambda] = \sum_{i=-J/2}^{J/2-1} \tilde\beta_{m,i}^{(\nu_1)} b_{m,i}[\lambda],$$

$$m = 0, \ldots, D - 1, \ \ \lambda = 0, \ldots, J - 1 \quad (28)$$

with a family of bases $\{b_{m,i}[\lambda]\}_{i=-J/2,\ldots,J/2-1}, m = 0, \ldots, D - 1$ that are orthonormal (i.e., $\sum_{\lambda=0}^{J-1} b_{m,i_1}[\lambda] b_{m,i_2}^*[\lambda] = \delta[i_1 - i_2]$ for all m) and do not depend on the value of ν_1 in $C^{(\nu_1)}[m, \lambda]$. The idea is to choose the 1-D bases $\{b_{m,i}[\lambda]\}_{i=-J/2,\ldots,J/2-1}$ such that the coefficient vector $[\tilde\beta_{m,-J/2}^{(\nu_1)} \cdots \tilde\beta_{m,J/2-1}^{(\nu_1)}]^T$ is sparse for all m and all $\nu_1 \in [-\nu_{\max}, \nu_{\max}]$. Substituting (28) back into (25), we obtain

$$H_{\lambda \Delta L, \kappa \Delta K} = \sum_{m=0}^{D-1} \sum_{i=-J/2}^{J/2-1} \phi^{(\nu_1)}\left(m - \frac{\tau_1}{T_\mathrm{s}}\right) \tilde\beta_{m,i}^{(\nu_1)}$$

$$\times b_{m,i}[\lambda] e^{-j2\pi\frac{\kappa m}{D}}.$$

This can now be identified with the 2-D basis expansion (24), with the orthonormal 2-D basis

$$v_{m,i}[\lambda, \kappa] \triangleq \frac{1}{\sqrt{D}} b_{m,i}[\lambda] e^{-j2\pi\frac{\kappa m}{D}} \quad (29)$$

and the 2-D coefficients $\beta_{m,i}^{(\tau_1, \nu_1)} \triangleq \sqrt{D}\phi^{(\nu_1)}(m - \tau_1/T_\mathrm{s})\tilde\beta_{m,i}^{(\nu_1)}$. The basis functions $v_{m,i}[\lambda, \kappa]$ are seen to agree with our previous 2-D DFT basis functions $u_{m,i}[\lambda, \kappa] =$

$(1/\sqrt{JD})e^{-j2\pi(\kappa m/D - \lambda i/J)}$ with respect to κ, but they are different with respect to λ because $(1/\sqrt{J})e^{j2\pi\lambda i/J}$ is replaced by $b_{m,i}[\lambda]$. Furthermore, the sparsity of $\beta_{m,i}^{(\tau_1,\nu_1)}$ in the i direction is governed by the new 1-D coefficients $\tilde{\beta}_{m,i}^{(\nu_1)}$, which are potentially sparser than the previous 1-D coefficients $\tilde{\alpha}_{m,i}^{(\nu_1)}$ in (26) that were based on the 1-D DFT basis $\{(1/\sqrt{J})e^{j2\pi\lambda i/J}\}$.

These considerations can be immediately extended to the multiple-scatterer case. When the channel comprises P scatterers as in (20), the coefficients are $\beta_{m,i} = \sum_{p=1}^{P} \eta_p \beta_{m,i}^{(\tau_p,\nu_p)}$. If each coefficient sequence $\beta_{m,i}^{(\tau_p,\nu_p)}$ is S-sparse, $\beta_{m,i}$ is PS-sparse. Note that, by construction, our basis $\{v_{m,i}[\lambda,\kappa]\}$ does not depend on the channel parameters P, η_p, τ_p, and ν_p, and its formulation is not explicitly based on the channel model (20). The use of the generalized 2-D basis $\{v_{m,i}[\lambda,\kappa]\}$ in (29) comes at the cost of an increased computational complexity, because efficient FFT algorithms can only be applied with respect to κ but not with respect to λ. However, if J is not too large, the additional complexity is small. Optimal designs of the 1-D bases $\{b_{m,i}[\lambda]\}_{i=-J/2,\dots,J/2-1}$ will be presented in Section VI.

B. Generalized Compressive Channel Estimator

A CS-based channel estimation scheme that uses the generalized basis expansion (24) can be developed similarly as in Section III-C. We can write (24) as [cf. (17)] $\mathbf{h} = \mathbf{V}\boldsymbol{\beta}$, with a unitary matrix $\mathbf{V}$. Here, $\boldsymbol{\beta}$ and $\mathbf{V}$ are defined in an analogous manner as, respectively, $\boldsymbol{\alpha}$ and $\mathbf{U}$ were defined in Section III-C. Reducing this relation to the pilot positions yields [cf. (18)] $\mathbf{h}^{(\mathrm{p})} = \mathbf{V}^{(\mathrm{p})}\boldsymbol{\beta} = \boldsymbol{\Phi}\mathbf{x}$, with $\boldsymbol{\Phi} \triangleq \mathbf{V}^{(\mathrm{p})}\mathbf{D}$ and $\mathbf{x} \triangleq \mathbf{D}^{-1}\boldsymbol{\beta}$, where the diagonal matrix $\mathbf{D}$ is chosen such that all columns of $\boldsymbol{\Phi}$ have unit ℓ_2-norm. Finally, we replace the unknown vector $\mathbf{h}^{(\mathrm{p})}$ by its pilot-based estimate, again denoted as $\mathbf{y}$. Using (13), we then obtain the *measurement equation* [cf. (19)] $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{z}$, where $\mathbf{z}$ is again the vector with entries $\tilde{z}_{l,k}/p_{l,k}|_{(l,k)\in\mathcal{P}}$. As in Section III-C, our task is to recover the length-JD vector $\mathbf{x}$ from the known length-$|\mathcal{P}|$ vector $\mathbf{y}$, based on the measurement equation. From the resulting estimate of $\mathbf{x}$, estimates of the channel coefficients $H_{l,k}$ on the subsampled grid $\mathcal{G}$ are obtained via (24) by means of the equivalence of $\beta_{m,i}$ and $\boldsymbol{\beta} = \mathbf{Dx}$. Inverting[2] (12) and applying (10) then yields estimates of all channel coefficients $H_{l,k}$. As discussed further above, we can expect $\boldsymbol{\beta}$ and, in turn, $\mathbf{x}$ to be approximately sparse provided the 1-D bases $\{b_{m,i}[\lambda]\}$ are chosen appropriately. Hence, our channel estimation problem is again recognized to be a sparse reconstruction problem of the form (14), with dimensions $M = \dim\{\mathbf{x}\} = JD$ and $Q = \dim\{\mathbf{y}\} = |\mathcal{P}|$. We can thus use a CS recovery technique to obtain an estimate of $\mathbf{x}$.

For consistency with the CS framework of Section III-C, we select the pilot positions uniformly at random within the subsampled time–frequency grid $\mathcal{G}$. For BP and CoSaMP, to

achieve a small upper bound on the reconstruction error, the number of pilots should satisfy condition (15), i.e.,

$$|\mathcal{P}| \geq C\gamma^{-2}(\ln(JD))^4\mu_{\mathbf{V}}^2 S\ln(1/\eta)$$

where S is the sparsity of $\mathbf{x}$ and $\mu_{\mathbf{V}}$ is the coherence of $\mathbf{V}$. Note that S depends on the chosen basis $\{v_{m,i}[\lambda,\kappa]\}$; furthermore, $\mu_{\mathbf{V}} \geq 1$ (for the DFT basis, we had $\mu_{\mathbf{U}} = 1$). Thus, the performance gain due to the better sparsity may be reduced to a certain extent because of the larger coherence.

VI. BASIS OPTIMIZATION

We now discuss the optimal design of the 1-D bases $\{b_{m,i}[\lambda]\}$.

A. Basis Optimization Framework

The orthonormal 1-D bases $\{b_{m,i}[\lambda]\}_{i=-J/2,\dots,J/2-1}, m = 0,\dots,D-1$ should be such that the coefficient vectors $[\tilde{\beta}_{m,-J/2}^{(\nu)} \cdots \tilde{\beta}_{m,J/2-1}^{(\nu)}]^T$ are sparse for all m and all $\nu \in [-\nu_{\max}, \nu_{\max}]$ (the maximum Doppler frequency shift $\nu_{\max}$ is assumed known). For our optimization, we slightly relax this requirement in that we only require a sparse coefficient vector for a finite number of uniformly spaced Doppler frequencies $\nu \in \mathcal{D}$, where $\mathcal{D} \triangleq \{\nu_{\Delta}d, d = -\lceil\nu_{\max}/\nu_{\Delta}\rceil,\dots,\lceil\nu_{\max}/\nu_{\Delta}\rceil\}$ with some Doppler frequency spacing ν_{Δ}.

Regarding the choice of ν_{Δ}, it is interesting to note that for the "canonical spacing" given by $\nu_{\Delta} = 1/(T_{\mathrm{s}}N_r)$, the coefficients $\tilde{\alpha}_{m,i}^{(\nu_{\Delta}d)}$ in the 1-D DFT expansion (26) are 1-sparse with respect to i. Indeed, $\psi^{(\nu_1)}[i] = e^{j\pi(\nu_1 T_{\mathrm{s}} - i/N_r)(N_r-1)}\psi(i - \nu_1 T_{\mathrm{s}}N_r)$ here simplifies to $\psi^{(\nu_{\Delta}d)}[i] = e^{j\pi(d-i)(N_r-1)/N_r}\psi(i - d) = \delta_{N_r}[i-d]$, where $\delta_{N_r}[i]$ is the N_r-periodic unit sample (i.e., $\delta_{N_r}[i]$ is 1 if i is a multiple of N_r and 0 otherwise). Expression (27) then reduces to

$$\tilde{\alpha}_{m,i}^{(\nu_{\Delta}d)} = \sqrt{J}\sum_{q=0}^{N-1}\delta_{N_r}[i - d + qL]A_{\gamma,g}^*\left(m, \frac{i+qL}{N_r}\right)$$

$$= \delta_{N_r}[i - \tilde{d}]A_{\gamma,g}^*\left(m, \frac{d}{N_r}\right)$$

where $\tilde{d}$ depends on d but not on i. Thus, for $\nu_{\Delta} = 1/(T_{\mathrm{s}}N_r)$, the coefficients obtained using the 1-D DFT basis $\{b_{m,i}[\lambda] = (1/\sqrt{J})e^{j2\pi\lambda i/J}\}$ are 1-sparse (no leakage effect). This means that the 1-D DFT basis would be optimal; no other basis could do better. We therefore choose a Doppler spacing that is twice as dense, i.e., $\nu_{\Delta} = 1/(2T_{\mathrm{s}}N_r)$. That is, we define $\mathcal{D}$ such that it includes also the Doppler frequencies located midway between any two adjacent canonical sampling points. For these frequencies—given by $\nu_{\Delta}d$ for odd d—the leakage (obtained with the DFT basis) is maximal.

Because the basis $\{b_{m,i}[\lambda]\}$ is orthonormal, the expansion coefficients $\tilde{\beta}_{m,i}^{(\nu)}$ defined by (28) can be calculated as the inner products $\tilde{\beta}_{m,i}^{(\nu)} = \sum_{\lambda=0}^{J-1}C^{(\nu)}[m,\lambda]b_{m,i}^*[\lambda], i = -J/2,\dots,J/2-1$. This can be rewritten as

$$\tilde{\boldsymbol{\beta}}_m^{(\nu)} = \mathbf{B}_m\mathbf{c}_m^{(\nu)}$$

---

[2]Note that the 1-D part of (24) corresponding to index m equals the respective 1-D part of (12) (1-D DFT), since $v_{m,i}[\lambda,\kappa] = (1/\sqrt{D})b_{m,i}[\lambda]e^{-j2\pi\kappa m/D}$. Hence, the transformation (24) and the inverted transformation (12) have to be applied only with respect to the index i.

with the length-J vectors $\tilde{\boldsymbol{\beta}}_m^{(\nu)} \triangleq [\tilde{\beta}_{m,-J/2}^{(\nu)} \cdots \tilde{\beta}_{m,J/2-1}^{(\nu)}]^T$ and $\mathbf{c}_m^{(\nu)} \triangleq [C^{(\nu)}[m,0] \cdots C^{(\nu)}[m,J-1]]^T$ and the unitary $J \times J$ matrix $\mathbf{B}_m$ with entries $(\mathbf{B}_m)_{i+1,\lambda+1} = b_{m,i-J/2}^*[\lambda]$. We can now state the basis optimization problem as follows. *For given vectors $\mathbf{c}_m^{(\nu)}$, $m = 0, \ldots, D-1$, with $\mathbf{c}_m^{(\nu)}$ defined as described above, find unitary $J \times J$ matrices $\mathbf{B}_m$ not dependent on ν such that the vectors $\tilde{\boldsymbol{\beta}}_m^{(\nu)} = \mathbf{B}_m \mathbf{c}_m^{(\nu)}$ are maximally sparse for all $\nu \in \mathcal{D}$.*

For the sake of algorithmic simplicity, we will measure the sparsity of $\tilde{\boldsymbol{\beta}}_m^{(\nu)}$ by the ℓ_1-norm or, more precisely, by the ℓ_1-norm averaged over all $\nu \in \mathcal{D}$, i.e., $(1/|\mathcal{D}|)\sum_{\nu \in \mathcal{D}}\|\tilde{\boldsymbol{\beta}}_m^{(\nu)}\|_1 = (1/|\mathcal{D}|)\sum_{\nu \in \mathcal{D}}\|\mathbf{B}_m \mathbf{c}_m^{(\nu)}\|_1$. Thus, our basis optimization problem is formulated as the D constrained minimization problems[3]

$$\hat{\mathbf{B}}_m = \arg \min_{\mathbf{B}_m \in \mathcal{U}} \sum_{\nu \in \mathcal{D}} \left\| \mathbf{B}_m \mathbf{c}_m^{(\nu)} \right\|_1, \quad m = 0, \ldots, D-1 \quad (30)$$

where $\mathcal{U}$ denotes the set of all unitary $J \times J$ matrices. Note that the vectors $\mathbf{c}_m^{(\nu)}$ are known because they follow from the function $C^{(\nu)}[m,\lambda]$, which is given by [see (26), (27)] $C^{(\nu)}[m,\lambda] = \sum_{i=-J/2}^{J/2-1} \sum_{q=0}^{N-1} \psi^{(\nu)}[i+qL] A_{\gamma,g}^*(m,(i+qL)/N_r)e^{j2\pi\lambda i/J}$. It is seen that the optimal bases characterized by the matrices $\hat{\mathbf{B}}_m$ depend on $N, L, J, g[n], \gamma[n]$, and (via the definition of $\mathcal{D}$) $\nu_{\max}$, but not on any other channel properties.

For classical CP-OFDM with CP length $N - K \geq D - 1$, we have $A_{\gamma,g}(m,\xi) = A_{\gamma,g}(0,\xi)$ for all $m = 1, \ldots, D-1$, so $C^{(\nu)}[m,\lambda] = C^{(\nu)}[0,\lambda]$ [see (26), (27)] and thus $\mathbf{c}_m^{(\nu)} = \mathbf{c}_0^{(\nu)}$. Because $\mathbf{c}_m^{(\nu)}$ no longer depends on m, only one basis $\mathbf{B}$ (instead of D different bases $\mathbf{B}_m$, $m = 0, \ldots, D-1$) has to be optimized.

B. Statistical Basis Optimization

The basis optimization framework presented above can be extended to take into account prior statistical information about the channel. Let us again consider the single-scatterer channel $h^{(\tau_1,\nu_1,\eta_1)}(t,\tau) = \eta_1 \delta(\tau - \tau_1)e^{j2\pi\nu_1 t}$, now including a path gain η_1. We assume that τ_1, ν_1, and η_1 are random, with (τ_1,ν_1) distributed according to a known probability density function (pdf) $p(\tau_1,\nu_1)$, and η_1 given (τ_1,ν_1) being zero-mean, circularly symmetric complex Gaussian with known variance $\sigma^2(\tau_1,\nu_1)$. As before, we consider a 2-D expansion of the subsampled channel coefficients $H_{\lambda\Delta L,\kappa\Delta K}$ into (deterministic) orthonormal basis functions $v_{m,i}[\lambda,\kappa]$, i.e., $H_{\lambda\Delta L,\kappa\Delta K} = \sum_{m=0}^{D-1}\sum_{i=-J/2}^{J/2-1} \beta_{m,i} v_{m,i}[\lambda,\kappa]$, $\lambda = 0, \ldots, J-1$, $\kappa = 0, \ldots, D-1$. Clearly, the vector $\boldsymbol{\beta}$ of expansion coefficients $\beta_{m,i}$ (which is defined as in Section V-B) now is a random vector. Our goal is to find basis functions $v_{m,i}[\lambda,\kappa]$ (or, equivalently, a unitary matrix $\mathbf{V}$, defined as in Section V-B) such that $\boldsymbol{\beta} = \boldsymbol{\beta}(\mathbf{V})$ is maximally sparse

on average. Measuring the sparsity of $\boldsymbol{\beta}$ by the ℓ_1-norm for convenience, we obtain the optimization problem

$$\hat{\mathbf{V}} = \arg \min_{\mathbf{V} \in \mathcal{U}'} \mathrm{E}\{\|\boldsymbol{\beta}(\mathbf{V})\|_1\} \quad (31)$$

where $\mathrm{E}\{\cdot\}$ denotes expectation and $\mathcal{U}'$ denotes the set of all unitary $JD \times JD$ matrices.

Again, we set $v_{m,i}[\lambda,\kappa] \triangleq (1/\sqrt{D})b_{m,i}[\lambda]e^{-j2\pi\kappa m/D}$ with a family of orthonormal 1-D bases $\{b_{m,i}[\lambda]\}$. Then, (31) reduces to the minimization of $\mathrm{E}\{\|\boldsymbol{\beta}\|_1\}$ with respect to $\{b_{m,i}[\lambda]\}$. For the single-scatterer channel, the ℓ_1-norm of $\boldsymbol{\beta} = \boldsymbol{\beta}^{(\tau_1,\nu_1,\eta_1)}$ can be shown to be

$$\left\| \boldsymbol{\beta}^{(\tau_1,\nu_1,\eta_1)} \right\|_1 = \sqrt{D}|\eta_1| \sum_{m=0}^{D-1} \left| \phi^{(\nu_1)}\left(m - \frac{\tau_1}{T_s} \right) \right|$$
$$\times \sum_{i=-J/2}^{J/2-1} \left| \sum_{\lambda=0}^{J-1} C^{(\nu_1)}[m,\lambda]b_{m,i}^*[\lambda] \right|$$

with $C^{(\nu_1)}[m,\lambda]$ as in (26), (27). We note that $|\eta_1|$ given (τ_1,ν_1) is Rayleigh distributed with mean $\sigma(\tau_1,\nu_1)\sqrt{\pi/2}$. Hence, $\mathrm{E}\{\|\boldsymbol{\beta}^{(\tau_1,\nu_1,\eta_1)}\|_1\}$ is given by (hereafter, we write τ, ν, η instead of τ_1, ν_1, η_1)

$$\mathrm{E}\left\{ \left\| \boldsymbol{\beta}^{(\tau,\nu,\eta)} \right\|_1 \right\} = \sqrt{\frac{D\pi}{2}} \int_{-\infty}^{\infty} \sum_{m=0}^{D-1} G^{(\nu)}[m]$$
$$\times \sum_{i=-J/2}^{J/2-1} \left| \sum_{\lambda=0}^{J-1} C^{(\nu)}[m,\lambda]b_{m,i}^*[\lambda] \right| d\nu \quad (32)$$

with

$$G^{(\nu)}[m] \triangleq \int_{-\infty}^{\infty} \sigma(\tau,\nu)\left| \phi^{(\nu)}\left(m - \frac{\tau}{T_s} \right) \right| p(\tau,\nu)d\tau > 0.$$

It follows that minimizing (32) with respect to $\{b_{m,i}[\lambda]\}$ amounts to minimizing

$$\int_{-\infty}^{\infty} \sum_{i=-J/2}^{J/2-1} \left| \sum_{\lambda=0}^{J-1} C^{(\nu)}[m,\lambda]b_{m,i}^*[\lambda] \right| G^{(\nu)}[m]d\nu \quad (33)$$

for all $m = 0, \ldots, D-1$. Note that $G^{(\nu)}[m]$ can be computed from the known statistics. In vector-matrix notation, with $\mathbf{c}_m^{(\nu)} \triangleq [C^{(\nu)}[m,0] \cdots C^{(\nu)}[m,J-1]]^T$ and the unitary $J \times J$ matrix $\mathbf{B}_m$ with entries $(\mathbf{B}_m)_{i+1,\lambda+1} \triangleq b_{m,i-J/2}^*[\lambda]$, minimization of (33) can be equivalently written as minimization of

$$\int_{-\infty}^{\infty} \left\| \mathbf{B}_m \mathbf{c}_m^{(\nu)} \right\|_1 G^{(\nu)}[m]d\nu \quad (34)$$

over the set $\mathcal{U}$ of all unitary $J \times J$ matrices $\mathbf{B}_m$, for $m = 0, \ldots, D-1$. Approximating this integral by its Riemannian sum[4] over the set $\mathcal{D} \triangleq \{\nu_\Delta d, d = -\lceil \nu_{\max}/\nu_\Delta \rceil,$

---

[3] We note that the optimization problem (30) is similar to *dictionary learning* problems that have recently been considered in [34]–[36]. In [36], conditions for the local identifiability of orthonormal bases by means of ℓ_1 minimization have been derived. An ℓ_0-norm based sparsity-enhancing basis design has been proposed in the MIMO context in [63]. Furthermore, basis adaptation and selection at the receiver has been considered in the ultrawideband context in [64].

[4] Alternatively, the integral can be interpreted as an expectation with respect to ν and computed by means of Monte Carlo techniques. This is especially advantageous if the maximum Doppler frequency is unknown.

$\ldots, \lceil \nu_{\max}/\nu_\Delta \rceil \}$ with $\nu_\Delta = 1/(2T_s N_r)$, for a given maximum Doppler frequency $\nu_{\max}$, the minimization problem can be finally stated as

$$\hat{\mathbf{B}}_m = \arg \min_{\mathbf{B}_m \in \mathcal{U}} \sum_{\nu \in \mathcal{D}} \left\| \mathbf{B}_m \tilde{\mathbf{c}}_m^{(\nu)} \right\|_1, \text{ with } \tilde{\mathbf{c}}_m^{(\nu)} \triangleq \mathbf{c}_m^{(\nu)} G^{(\nu)}[m]$$

(35)

for $m = 0, \ldots, D-1$. This is recognized to be of the same form as (30).

In practice, the channel statistics $p(\tau, \nu), \sigma^2(\tau, \nu)$ will deviate from the true statistics to some extent, so that the basis matrices $\hat{\mathbf{B}}_m$ obtained as described above will be different from the truly optimal ones. An interesting question is as to how this difference affects the average sparsity of the expansion coefficient vector $\boldsymbol{\beta}^{(\tau,\nu,\eta)}$. For simplicity, we measure the average sparsity by $\mathrm{E}\{\|\boldsymbol{\beta}\|_1\}$, and we assume that the optimization criterion is minimization of (34) [which, after all, is almost equivalent to (35)] and, further, that $\Delta L = 1$ or equivalently $J = L$ (i.e., no subsampling with respect to l). Let $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$ denote the expansion coefficient vectors obtained for the true and incorrect bases, respectively. Then, one can show the following bound on the normalized difference of the average sparsities of $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$:

$$\frac{\left| \mathrm{E}\{\|\tilde{\boldsymbol{\beta}}\|_1\} - \mathrm{E}\{\|\boldsymbol{\beta}\|_1\} \right|}{\mathrm{E}\{\|\boldsymbol{\beta}\|_1\}}$$
$$\leq 2\sqrt{L} \frac{\int_{-\infty}^{\infty} \sum_{m=0}^{D-1} \left| \tilde{G}^{(\nu)}[m] - G^{(\nu)}[m] \right| |A_{\gamma,g}(m, \nu T_s)| \, d\nu}{\int_{-\infty}^{\infty} \sum_{m=0}^{D-1} G^{(\nu)}[m] |A_{\gamma,g}(m, \nu T_s)| \, d\nu}$$

where $\tilde{G}^{(\nu)}[m]$ is defined analogously to $G^{(\nu)}[m]$ but with the incorrect statistics.

C. Basis Optimization Algorithm

Because the minimization problems (30) and (35) are non-convex (since $\mathcal{U}$ is not a convex set), standard convex optimization techniques cannot be used. We therefore propose an approximate iterative algorithm that relies on the following facts [65]. 1) Every unitary $J \times J$ matrix $\mathbf{B}$ can be represented in terms of a Hermitian $J \times J$ matrix $\mathbf{A}$ as $\mathbf{B} = e^{j\mathbf{A}}$. 2) The matrix exponential $\mathbf{B} = e^{j\mathbf{A}}$ can be approximated by its first-order Taylor expansion, i.e., $\mathbf{B} \approx \mathbf{I}_J + j\mathbf{A}$, where $\mathbf{I}_J$ is the $J \times J$ identity matrix. Even though $\mathbf{B}$ is unitary and $\mathbf{I}_J + j\mathbf{A}$ is not, this approximation will be good if $\|\mathbf{A}\|_\infty$ is small, where $\|\mathbf{A}\|_\infty$ denotes the largest modulus of all entries of $\mathbf{A}$. Because of this condition, we construct $\mathbf{B}_m$ iteratively: starting with the DFT basis, we perform a *small* update at each iteration, using the approximation $\mathbf{B} \approx \mathbf{I}_J + j\mathbf{A}$ in the optimization criterion *but not for actually updating* $\mathbf{B}_m$ (thus, the iterated $\mathbf{B}_m$ is always unitary). More specifically, at the rth iteration, we consider the following update of the unitary matrix $\mathbf{B}_m^{(r)}$:

$$\mathbf{B}_m^{(r+1)} = e^{j\mathbf{A}_m^{(r)}} \mathbf{B}_m^{(r)}$$

where $\mathbf{A}_m^{(r)}$ is a small Hermitian matrix that remains to be optimized. Note that $\mathbf{B}_m^{(r+1)}$ is again unitary because both $\mathbf{B}_m^{(r)}$ and $e^{j\mathbf{A}_m^{(r)}}$ are unitary.

Ideally, we would like to optimize $\mathbf{A}_m^{(r)}$ according to (30) [or (35)], i.e., by minimizing $\sum_{\nu \in \mathcal{D}} \|\mathbf{B}_m^{(r+1)} \mathbf{c}_m^{(\nu)}\|_1 = \sum_{\nu \in \mathcal{D}} \|e^{j\mathbf{A}_m^{(r)}} \mathbf{B}_m^{(r)} \mathbf{c}_m^{(\nu)}\|_1$. Since this problem is still non-convex, we use the approximation $e^{j\mathbf{A}} \approx \mathbf{I}_J + j\mathbf{A}$, and thus the final minimization problem at the rth iteration is

$$\hat{\mathbf{A}}_m^{(r)} = \arg \min_{\mathbf{A} \in \mathcal{A}_r} \sum_{\nu \in \mathcal{D}} \left\| (\mathbf{I}_J + j\mathbf{A}) \mathbf{B}_m^{(r)} \mathbf{c}_m^{(\nu)} \right\|_1. \quad (36)$$

Here, $\mathcal{A}_r$ is the set of all Hermitian $J \times J$ matrices $\mathbf{A}$ that are small in the sense that $\|\mathbf{A}\|_\infty \leq \rho_r$, where ρ_r is a positive constraint level (a small ρ_r ensures a good accuracy of our approximation $\mathbf{B} \approx \mathbf{I}_J + j\mathbf{A}$ and also that $e^{j\hat{\mathbf{A}}_m^{(r)}}$ is close to $\mathbf{I}_J$). The problem (36) is convex and thus can be solved by standard convex optimization techniques [66].

The next step at the rth iteration is to test whether the cost function is smaller for the new unitary matrix $e^{j\hat{\mathbf{A}}_m^{(r)}} \mathbf{B}_m^{(r)}$, i.e., whether $\sum_{\nu \in \mathcal{D}} \|e^{j\hat{\mathbf{A}}_m^{(r)}} \mathbf{B}_m^{(r)} \mathbf{c}_m^{(\nu)}\|_1 < \sum_{\nu \in \mathcal{D}} \|\mathbf{B}_m^{(r)} \mathbf{c}_m^{(\nu)}\|_1$. In the positive case, we actually perform the update of $\mathbf{B}_m^{(r)}$ and we retain the constraint level ρ_r for the next iteration, i.e.,

$$\mathbf{B}_m^{(r+1)} = e^{j\hat{\mathbf{A}}_m^{(r)}} \mathbf{B}_m^{(r)}, \qquad \rho_{r+1} = \rho_r.$$

Otherwise, we reject the update of $\mathbf{B}_m^{(r)}$ and reduce the constraint level ρ_r, i.e.,

$$\mathbf{B}_m^{(r+1)} = \mathbf{B}_m^{(r)}, \qquad \rho_{r+1} = \frac{\rho_r}{2}.$$

By this construction, the cost function sequence $\sum_{\nu \in \mathcal{D}} \|\mathbf{B}_m^{(r)} \mathbf{c}_m^{(\nu)}\|_1, r = 0, 1, \ldots$ is guaranteed to be monotonically decreasing.

The above iteration process is terminated if ρ_r falls below a prescribed threshold or if the number of iterations exceeds a certain value. The iteration process is initialized by the $J \times J$ DFT matrix $\mathbf{F}_J$, i.e., $\mathbf{B}_m^{(0)} = \mathbf{F}_J$, because the DFT basis was seen in Section IV to yield a relatively sparse coefficient vector. We note that efficient algorithms for computing the matrix exponentials $e^{j\hat{\mathbf{A}}_m^{(r)}}$ exist [65]. Since the bases $\{b_{m,i}[\lambda]\}$ (or, equivalently, the basis matrices $\mathbf{B}_m$) do not depend on the received signal, they have to be optimized only once before the actual channel estimation starts.

In Fig. 1, we compare the expansion coefficients $\alpha_{m,i}$ obtained with the DFT basis [see (16)] and $\beta_{m,i}$ obtained with the deterministically optimized basis [see (24), (29)] for one channel realization. The system parameters are as in Sections VIII-A and VIII-B (first scenario). For the minimization (36) (not m-dependent, since we consider a CP-OFDM system), we used the convex optimization package CVX [67]. It is seen that the basis optimization yields a significant enhancement of sparsity.

VII. CHANNEL ESTIMATION FOR STRONGLY DISPERSIVE CHANNELS

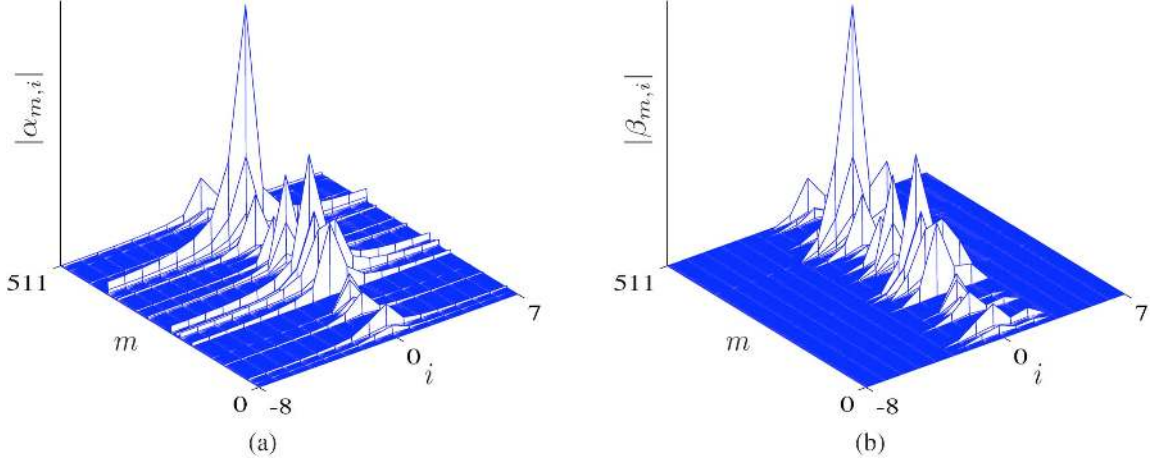For strongly dispersive channels, the off-diagonal system channel coefficients (ISI/ICI coefficients)

Fig. 1.   Sparsity enhancement obtained with the proposed iterative basis optimization algorithm: Modulus of the expansion coefficients for (a) the DFT basis and (b) the optimized basis.

$\{H_{l,k;l',k'}\}_{(l,k)\neq(l',k')}$ in (7) are no longer negligible. Therefore, we now present a compressive channel estimator that is able to produce reliable estimates of *all* channel coefficients $H_{l,k;l',k'}$.

A. Basis Expansion Model

The proposed channel estimator uses a basis expansion model [18]–[20] that is different from the basis expansion considered in Sections V and VI. The discrete-time channel impulse response $h[n,m]$ is expanded with respect to n into orthonormal basis functions $\psi_i[n], i=0,\ldots,N_r-1$, i.e.,

$$h[n,m] = \sum_{i=0}^{N_r-1} T_h[m,i]\psi_i[n], \quad n=0,\ldots,N_r-1 \quad (37)$$

with m-dependent expansion coefficients

$$T_h[m,i] \triangleq \sum_{n=0}^{N_r-1} h[n,m]\psi_i^*[n]. \quad (38)$$

The function $T_h[m,i]$ generalizes the discrete-delay-Doppler spreading function $S_h[m,i]$ in (9), which is reobtained for $\psi_i[n] = (1/\sqrt{N_r})e^{j2\pi in/N_r}$ (up to a constant factor). Similarly to (8), the discrete-time channel can now be rewritten as

$$r[n] = \sum_{m=-\infty}^{\infty} \sum_{i=0}^{N_r-1} T_h[m,i]s[n-m]\psi_i[n] + z[n],$$
$$n=0,\ldots,N_r-1. \quad (39)$$

We assume that the support of $T_h[m,i]$ is contained in $\{0,\ldots,D-1\} \times \{0,\ldots,J-1\}$ ($h[n,m]$ is assumed causal with maximum delay at most $D-1$). Combining (5), (39), and

(1), we then reobtain the system channel relation (7), with the channel coefficients $H_{l,k;l',k'}$ expressed as

$$H_{l,k;l',k'}$$
$$= e^{-j2\pi\frac{N}{K}k'(l'-l)} \sum_{m=0}^{D-1}\sum_{i=0}^{J-1} T_h[m,i]\left[\sum_{n=-\infty}^{\infty}\gamma^*[n]e^{j2\pi n\frac{k'-k}{K}}\right.$$
$$\left. \times g[n-m-(l'-l)N]\psi_i[n+lN]\right]e^{-j2\pi\frac{k'm}{K}}. \quad (40)$$

Note that the limiting cases $D = K$ and $J = N_r$ are also allowed.

B. Compressive Channel Estimator

The proposed compressive channel estimator operates in an iterative, decision-directed fashion. At the first iteration, it utilizes the knowledge of some pilots $p_{l,k} \in \mathcal{A}$ with $(l,k) \in \mathcal{P}$. The pilot position set $\mathcal{P}$ is selected uniformly at random within $\{0,\ldots,L-1\}\times\{0,\ldots,K-1\}$. At later iterations, the estimator additionally uses *virtual* pilots, which are based on the symbol decisions produced by a suitable ISI/ICI equalizer (e.g., [40] and [68]–[71]) followed by the quantizer. Typically, the equalizer will use the (estimated) channel coefficients $H_{l,k;l',k'}$ only within a certain "off-diagonal bandwidth," i.e., for $|l-l'| \le l_{\max}$ and $|k-k'| \le k_{\max}$ (modulo K).

At the rth iteration, let $p_{l,k}^{(r)}$ denote "extended pilots" (pilots augmented by virtual pilots) on an *extended pilot position set* $\mathcal{P}^{(r)}$. This set is defined as $\mathcal{P}^{(r)} \triangleq \mathcal{H}^{(r)} \oplus \mathcal{V} = \{(l,k) = (l_1 + l_2, (k_1 + k_2) \bmod K) : (l_1,k_1) \in \mathcal{H}^{(r)}, (l_2,k_2) \in \mathcal{V}\}$, where $\mathcal{V} \triangleq \{(l,k) : l = -l_{\max},\ldots,l_{\max}; k = -k_{\max},\ldots,k_{\max}\}$ and $\mathcal{H}^{(r)}$ will be specified later. Note that by this construction, for an extended pilot in $\mathcal{H}^{(r)}$, all neighboring symbols (which yield the largest interference) are also included in $\mathcal{P}^{(r)}$. Then, for $(l,k) \in \mathcal{H}^{(r)}$, relation (7) can be written as

$$r_{l,k} = \sum_{(l',k')\in\{(l,k)\}\oplus\mathcal{V}} H_{l,k;l',k'}p_{l',k'}^{(r)} + z_{l,k}^{(r)},$$
$$(l,k) \in \mathcal{H}^{(r)} \quad (41)$$

where the noise/interference term $z_{l,k}^{(r)}$ includes noise, ISI/ICI from *outside* the set $\{(l,k)\} \oplus \mathcal{V}$, and—possibly—some additional errors if $p_{l',k'}^{(r)} \neq a_{l',k'}$. If $\mathcal{V}$ is chosen sufficiently large, the ISI/ICI part in $z_{l,k}^{(r)}$ is negligible. Inserting (40) into (41) yields the noisy 2-D expansion

$$r_{l,k} = \sum_{m=0}^{D-1} \sum_{i=0}^{J-1} \theta_{m,i} w_{m,i}^{(r)}[l,k] + z_{l,k}^{(r)}, \quad (l,k) \in \mathcal{H}^{(r)} \quad (42)$$

with $\theta_{m,i} \triangleq T_h[m,i]$ and $w_{m,i}^{(r)}[l,k] \triangleq \sum_{(l',k') \in \{(l,k)\} \oplus \mathcal{V}} p_{l',k'}^{(r)}$ $\times e^{-j2\pi Nk'(l'-l)/K} [\sum_{n=-\infty}^{\infty} \gamma^*[n] e^{j2\pi n(k'-k)/K} g[n - m - (l'-l)N] \psi_i[n+lN]] e^{-j2\pi k'm/K}$. Differently from (16) and (24), this is an expansion of the demodulated symbols $r_{l,k}$ and not of the channel coefficients $H_{l,k}$. Note also that the basis functions $w_{m,i}^{(r)}[l,k]$ depend on the extended pilots $p_{l,k}^{(r)}, (l,k) \in \mathcal{P}^{(r)}$.

Using a stacking as in Section III-C, the expansion (42) can be expressed as $\mathbf{r}^{(r)} = \mathbf{W}^{(r)}\boldsymbol{\theta} + \mathbf{z}^{(r)}$, where the $|\mathcal{H}^{(r)}|$-dimensional vectors $\mathbf{r}^{(r)}$ and $\mathbf{z}^{(r)}$, the JD-dimensional vector $\boldsymbol{\theta}$, and the $|\mathcal{H}^{(r)}| \times JD$ matrix $\mathbf{W}^{(r)}$ are defined in an analogous manner as, respectively, $\mathbf{h}^{(\mathrm{p})}, \mathbf{z}, \boldsymbol{\alpha}$, and $\mathbf{U}^{(\mathrm{p})}$ in Section III-C. With $\mathbf{y}^{(r)} \triangleq \mathbf{r}^{(r)}, \boldsymbol{\Phi}^{(r)} \triangleq \mathbf{W}^{(r)}\mathbf{D}^{(r)}$, and $\mathbf{x}^{(r)} \triangleq (\mathbf{D}^{(r)})^{-1}\boldsymbol{\theta}$, where the diagonal matrix $\mathbf{D}^{(r)}$ is chosen such that all columns of $\boldsymbol{\Phi}^{(r)}$ have unit ℓ_2-norm, we obtain[5] the *measurement equation* [cf. (19)] $\mathbf{y}^{(r)} = \boldsymbol{\Phi}^{(r)}\mathbf{x}^{(r)} + \mathbf{z}^{(r)}$. As in Section III-C, we would like to recover the length-JD vector $\mathbf{x}^{(r)}$ from the known length-$|\mathcal{H}^{(r)}|$ vector $\mathbf{y}^{(r)}$. If the basis functions $\psi_i[n]$ in (37) and (38) are chosen such that $T_h[m,i]$ (or, equivalently, $\boldsymbol{\theta}$) is sparse, then also $\mathbf{x}^{(r)} = (\mathbf{D}^{(r)})^{-1}\boldsymbol{\theta}$ is sparse. Hence, our problem is again a sparse reconstruction problem of the form (14), with dimensions $M = \dim\{\mathbf{x}^{(r)}\} = JD$ and $Q = \dim\{\mathbf{y}^{(r)}\} = |\mathcal{H}^{(r)}|$. We can thus use a CS recovery technique[6] to obtain an estimate $\hat{\mathbf{x}}^{(r)}$ of $\mathbf{x}^{(r)}$ and, in turn, an estimate $\hat{\boldsymbol{\theta}}^{(r)} = \mathbf{D}^{(r)}\hat{\mathbf{x}}^{(r)}$ or, equivalently, $\hat{T}_h^{(r)}[m,i]$.

From $\hat{T}_h^{(r)}[m,i]$, estimates of the channel coefficients $H_{l,k;l',k'}$ for all $l, l' = 0, \ldots, L-1$ and $k, k' = 0, \ldots, K-1$ are obtained via (40). Then, an ISI/ICI equalizer yields symbol estimates $\tilde{a}_{l,k}^{(r)}$ and, subsequently, a quantizer produces detected symbols $\hat{a}_{l,k}^{(r)}, l = 0, \ldots, L-1, k = 0, \ldots, K-1$. On $\mathcal{P}$, these are replaced by the known pilots, i.e., we set $\hat{a}_{l,k}^{(r)} \triangleq p_{l,k}$ for $(l,k) \in \mathcal{P}$.

Next, we determine $\mathcal{H}^{(r+1)}$ as the largest subset of $\{0, \ldots, L-1\} \times \{0, \ldots, K-1\}$ such that the new extended pilot set $\mathcal{P}^{(r+1)} \triangleq \mathcal{H}^{(r+1)} \oplus \mathcal{V}$ contains only "reliable" detected symbols $\hat{a}_{l,k}^{(r)}$, and we define the new extended pilots as $p_{l,k}^{(r+1)} \triangleq \hat{a}_{l,k}^{(r)}$ for $(l,k) \in \mathcal{P}^{(r+1)}$. Here, following [71], a detected symbol $\hat{a}_{l,k}^{(r)}$ will be considered as "reliable" either if

---

[5]The computation of the measurement matrix essentially requires $L(2l_{\max} + 1)(2k_{\max} + 1)J$ FFTs of length K. Note that J is typically very small, cf. Section VII-C.

[6]Whether $\boldsymbol{\Phi}^{(r)}$ satisfies the RIP with a small restricted isometry constant depends on the basis functions $\psi_i[n]$ as well as on the extended pilot position set $\mathcal{P}^{(r)}$; hence, performance guarantees cannot be made in general.

$(l,k) \in \mathcal{P}$ or, for $(l,k) \notin \mathcal{P}$, if the corresponding symbol estimate $\tilde{a}_{l,k}^{(r)}$ (result of equalization, before quantization) is significantly closer to $\hat{a}_{l,k}^{(r)}$ than to any other symbol in $\mathcal{A}$. For example, for the QPSK alphabet $\mathcal{A} \triangleq \{1+j, 1-j, -1+j, -1-j\}$, $\hat{a}_{l,k}^{(r)}$ will be considered as reliable either if $(l,k) \in \mathcal{P}$ or if both $|\Re\{\tilde{a}_{l,k}^{(r)}\}| > \epsilon$ and $|\Im\{\tilde{a}_{l,k}^{(r)}\}| > \epsilon$ for a certain threshold $\epsilon > 0$.

Proceeding iteratively in this fashion, we successively construct extended pilots $p_{l,k}^{(r)}$, which are used to estimate $T_h[m,i]$ and, via (40), the channel coefficients $H_{l,k;l',k'}$. The reliability criterion ensures that most of the extended pilots equal the true transmitted symbols. Since the $p_{l,k}^{(r)}$ are improved with the iterations, we expect $|\mathcal{H}^{(r+1)}| > |\mathcal{H}^{(r)}|$ in general. The iterative algorithm is initialized with $p_{l,k}^{(0)} = p_{l,k}$ and $\mathcal{P}^{(0)} = \mathcal{H}^{(0)} = \mathcal{P}$ (for $r = 0, \mathcal{V} = \{0\}$, whereas later $\mathcal{V} = \{(l,k) : l = -l_{\max}, \ldots, l_{\max}; k = -k_{\max}, \ldots, k_{\max}\}$). Accordingly, we use the conventional one-tap equalizer (without ISI/ICI equalization) at the first iteration. The algorithm is terminated either if the difference between $\hat{H}_{l,k;l',k'}^{(r+1)}$ and $\hat{H}_{l,k;l',k'}^{(r)}$ (measured by a suitable norm) falls below a certain threshold or after a fixed number of iterations. While a proof of convergence for this iterative algorithm is not available, we always observed convergence for reasonably chosen $\psi_i[n]$ (see Section VII-C), $|\mathcal{P}|$, and ϵ.

The proposed algorithm is not limited to strongly dispersive channels. For weakly dispersive channels, we simply set $\mathcal{V} = \{0\}$ at all iterations and replace the ISI/ICI equalizer by the conventional one-tap equalizer. This effectively amounts to a decision-directed, iterative extension of the compressive channel estimator discussed in Sections III–VI. This extension can improve the estimation accuracy. Moreover, it can increase the spectral efficiency of the system even further, since the pilot set $\mathcal{P}$ can be chosen quite small due to the successive improvements achieved by the iterations. However, these gains come at the cost of some additional complexity.

C. Sparsity-Inducing Basis Functions

The basis functions $\psi_i[n], i = 0, \ldots, N_r - 1$ have to be chosen such that the generalized spreading function $T_h[m,i]$ in (38) is sparse. In particular, (20) suggests that $T_h[m,i]$ should be sparse for the single-scatterer channel $h^{(\tau_1, \nu_1)}(t, \tau) = \delta(\tau - \tau_1)e^{j2\pi\nu_1 t}$, for all $\tau_1 \in [0, \tau_{\max}]$ and $\nu_1 \in [-\nu_{\max}, \nu_{\max}]$. For this channel

$$T_h[m,i] = \phi^{(\nu_1)}\left(m - \frac{\tau_1}{T_s}\right)\vartheta^{(\nu_1)}[i],$$

$$\text{with} \quad \vartheta^{(\nu)}[i] \triangleq \sum_{n=0}^{N_r-1} e^{j2\pi\nu nT_s}\psi_i^*[n]. \quad (43)$$

The factor $\phi^{(\nu_1)}(m - \tau_1/T_s)$ [see (22)] is already sparse due to its fast decay as discussed in Section IV. Thus, we have to design the $\psi_i[n]$ such that the factor $\vartheta^{(\nu)}[i]$ is sparse for all $\nu \in [-\nu_{\max}, \nu_{\max}]$.

For this purpose, we can adapt the basis optimization of Section VI. Let $\mathcal{D} \triangleq \{\nu_\Delta d, d = -\lceil \nu_{\max}/\nu_\Delta \rceil, \ldots, \lceil \nu_{\max}/\nu_\Delta \rceil\}$ with $\nu_\Delta = 1/(2T_s N_r)$ and rewrite the second equation in (43) as $\boldsymbol{\vartheta}^{(\nu)} = \mathbf{P}\mathbf{e}^{(\nu)}$, with the length-$N_r$ vectors $\boldsymbol{\vartheta}^{(\nu)} \triangleq [\vartheta^{(\nu)}[0] \cdots \vartheta^{(\nu)}[N_r - 1]]^T$ and $\mathbf{e}^{(\nu)} \triangleq$

$[1 \; e^{j2\pi\nu T_{\mathrm{s}}} \cdots e^{j2\pi\nu(N_r-1)T_{\mathrm{s}}}]^T$ and the unitary $N_r \times N_r$ matrix $\mathbf{P}$ with entries $(\mathbf{P})_{i+1,n+1} = \psi_i^*[n]$. Optimal basis functions $\psi_i[n]$ are now defined as $\hat{\mathbf{P}} = \arg\min_{\mathbf{P}\in\mathcal{U}} \sum_{\nu\in\mathcal{D}} \|\mathbf{P}\mathbf{e}^{(\nu)}\|_1$, so that the iterative optimization algorithm of Section VI-C can be used. However, for large $N_r \approx NL$, the computational cost of this approach is quite high.

As a practical alternative, we propose a construction of the $\psi_i[n]$ that involves discrete prolate spheroidal sequences (DPSSs) [37]. Basis expansion models using DPSSs have been considered previously [20]. If their design parameters are chosen according to maximum Doppler frequency $\nu_{\max}$, sampling period T_{s}, and blocklength N_r, the corresponding functions $\vartheta_{\mathrm{p}}^{(\nu)}[i]$ in (43) will have an effective support $\{0,\ldots,J-1\}$ for all $\nu \in [-\nu_{\max},\nu_{\max}]$, where J is small compared with N_r. Unfortunately, within this support interval, the $\vartheta_{\mathrm{p}}^{(\nu)}[i]$ are not sparse in general.

We will therefore use a specific combination of DPSSs and DFT basis functions, which yields functions $\vartheta^{(\nu)}[i]$ that are still effectively zero outside $\{0,\ldots,J-1\}$ but, within that interval, preserve the sparsity obtained with the DFT basis. Let $\psi_i^{(\mathrm{p})}[n], n \in \mathbb{Z}, i = 0,\ldots,N_r - 1$ denote the DPSSs that are bandlimited to $[-\nu_{\max}T_{\mathrm{s}},\nu_{\max}T_{\mathrm{s}}]$ and have maximum energy concentration in $\{0,\ldots,N_r-1\}$ [37]. In what follows, the DPSSs $\psi_i^{(\mathrm{p})}[n]$ will be truncated to $\{0,\ldots,N_r-1\}$. Then, for large N_r, the support of $\vartheta_{\mathrm{p}}^{(\nu)}[i] \triangleq \sum_{n=0}^{N_r-1} e^{j2\pi\nu nT_{\mathrm{s}}}\psi_i^{(\mathrm{p})*}[n]$ is effectively contained in $\{0,\ldots,J-1\}$ for all $\nu \in [-\nu_{\max},\nu_{\max}]$, where $J \triangleq 2J_0 + J_1$ with $J_0 \triangleq \lfloor \nu_{\max}T_{\mathrm{s}}N_r \rfloor$ and $J_1 \geq 2$ a small integer. In addition, we consider the $2J_0 + 1$ orthonormal DFT basis functions $\psi_i^{(\mathrm{f})}[n] \triangleq (1/\sqrt{N_r})e^{j2\pi in/N_r}, n = 0,\ldots,N_r-1$, for $i = -J_0,\ldots,J_0$. For these i, $\nu[i] \triangleq i/(N_rT_{\mathrm{s}})$ is in $[-\nu_{\max},\nu_{\max}]$. We thus have for all $i_1 = -J_0,\ldots,J_0$ and $i_2 = J,\ldots,N_r-1$

$$\left\langle \psi_{i_1}^{(\mathrm{f})}, \psi_{i_2}^{(\mathrm{p})} \right\rangle = \frac{1}{\sqrt{N_r}} \sum_{n=0}^{N_r-1} e^{j2\pi\frac{i_1 n}{N_r}}\psi_{i_2}^{(\mathrm{p})*}[n]$$
$$= \frac{1}{\sqrt{N_r}}\vartheta_{\mathrm{p}}^{(\nu[i_1])}[i_2]$$
$$\approx 0 \qquad (44)$$

because $\nu[i_1] \in [-\nu_{\max},\nu_{\max}]$ but $i_2 \notin \{0,\ldots,J-1\}$. That is, $\psi_{i_1}^{(\mathrm{f})}$ and $\psi_{i_2}^{(\mathrm{p})}$ are effectively orthogonal for the specified ranges of i_1 and i_2. Let us now define the following ordered set of (in total N_r) DFT functions and (truncated) DPSSs:

$$\mathcal{M}' \triangleq \left\{\psi_{-J_0}^{(\mathrm{f})},\ldots,\psi_{J_0}^{(\mathrm{f})},\psi_{2J_0+1}^{(\mathrm{p})},\ldots,\psi_{N_r-1}^{(\mathrm{p})}\right\}.$$

Due to (44) and the orthonormality of the $\psi_i^{(\mathrm{p})}$ [37], all functions in $\mathcal{M}'$ are (effectively) mutually orthonormal with the exception of the DPSSs $\psi_i^{(\mathrm{p})}$ within the index range $i = 2J_0+1,\ldots,J-1$, which are not orthonormal to the DFT functions. Therefore, we derive the final set of basis functions $\mathcal{M} \triangleq \{\psi_0,\ldots,\psi_{N_r-1}\}$ by Gram–Schmidt orthonormalization [65] of $\mathcal{M}'$. This amounts to setting $\psi_i = \psi_{i-J_0}^{(\mathrm{f})}$ for $i = 0,\ldots,2J_0$ and $\psi_i = \sum_{n=0}^{2J_0} c_n \psi_{i-J_0}^{(\mathrm{f})} + \sum_{n=2J_0+1}^{i} c_n \psi_n^{(\mathrm{p})}$ for $i \geq 2J_0 + 1$, with suitable coefficients c_n. It follows that
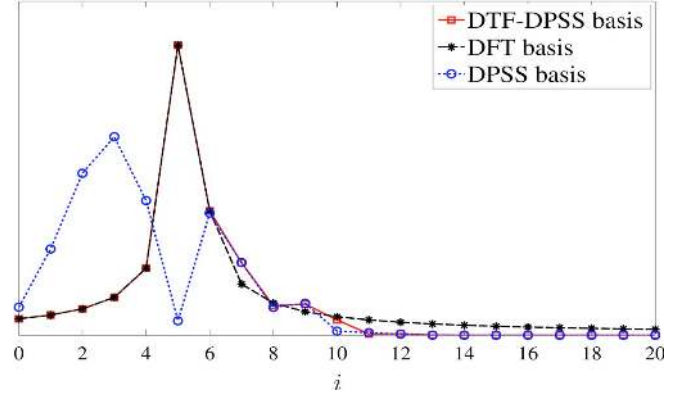

Fig. 2. Sparsity enhancement in $\vartheta^{(\nu)}[i]$ obtained with the proposed combined DFT-DPSS basis, relative to a pure DFT basis and a pure DPSS basis.

$\langle \psi_{i_1}, \psi_{i_2}^{(\mathrm{p})} \rangle \approx 0$ for all $i_1 = 0,\ldots,J-1$ and $i_2 = J,\ldots,N_r-1$. Hence, the Gram–Schmidt orthonormalization algorithm yields $\psi_i \approx \psi_i^{(\mathrm{p})}$ for all $i = J,\ldots,N_r - 1$, i.e., the last $N_r - J$ basis functions of $\mathcal{M}$ are effectively known *a priori*, and the algorithm can therefore be terminated after J steps. In fact, only $J_1 - 1$ steps are required, because the first $2J_0+1 = J - J_1+1$ (DFT) basis functions are also known.

With this construction of the $\psi_i[n]$, the support of $\vartheta^{(\nu)}[i] = \sum_{n=0}^{N_r-1} e^{j2\pi\nu nT_{\mathrm{s}}}\psi_i^*[n]$ is approximately contained in $\{0,\ldots,J-1\}$ for all $\nu \in [-\nu_{\max},\nu_{\max}]$. Furthermore, for $i = 0,\ldots,J-J_1$, the $\psi_i[n]$ are DFT basis functions, so that the sparsity of $\vartheta^{(\nu)}[i]$ corresponds to the sparsity given by the DFT basis for these indices i. For the $J_1 - 1$ remaining indices $i = J - J_1+1,\ldots,J-1$ within the support interval, we cannot expect any sparsity of $\vartheta^{(\nu)}[i]$. However, J_1 is quite small, so that the overall sparsity of $\vartheta^{(\nu)}[i]$ is not deteriorated significantly.

For $N_r = NL = (2048 + 512)16 = 40960$ and $\nu_{\max}T_{\mathrm{s}} = 0.2/K = 0.2/2048$ (corresponding to a maximum Doppler frequency of 20% of the subcarrier spacing), Fig. 2 depicts $|\vartheta^{(\nu)}[i]|, i = 0,\ldots,20$ for $\nu T_{\mathrm{s}} = 0.115/K = 0.115/2048$. For comparison, $|\vartheta_{\mathrm{f}}^{(\nu)}[i]|$ (obtained with a pure DFT basis) and $|\vartheta_{\mathrm{p}}^{(\nu)}[i]|$ (obtained with a pure DPSS basis) are also shown. We see that the proposed DFT-DPSS basis leads to the sparsest result: for the pure DPSS basis, there is no sparsity within the support interval, while for the pure DFT basis, the sparsity is impaired by a strong leakage effect.

VIII. SIMULATION RESULTS

Next, we demonstrate the performance gains that can be achieved with our sparsity-enhancing basis expansions and estimation of ISI/ICI channel coefficients, relative to the basic compressive estimator. We show results for three different recovery algorithms, namely, Lasso (equivalent to BP denoising), OMP, and CoSaMP.

A. Simulation Setup

MC system parameters. We simulated CP-OFDM systems with $K \in \{512, 1024, 2048\}$ subcarriers and CP length ratio $(N - K)/K = 1/4$. The systems employed 4-QAM symbols with Gray labeling, a rate-1/2 convolutional code, and 32×16
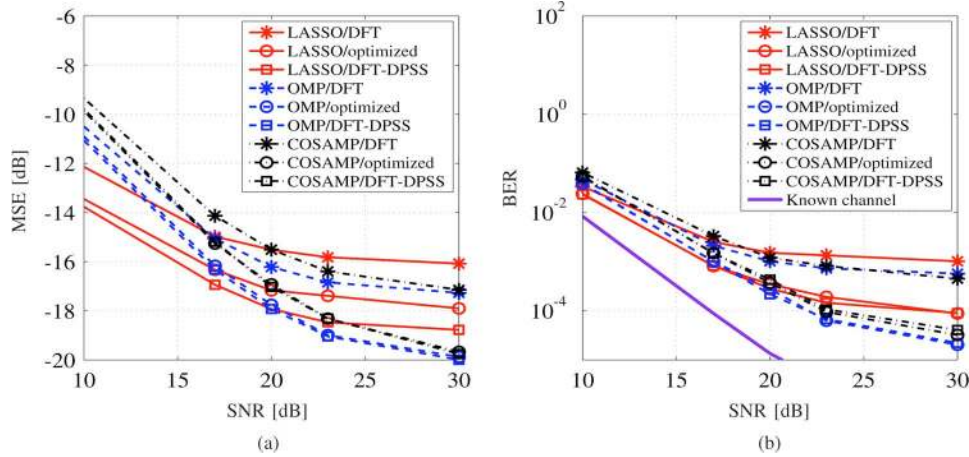
Fig. 3. Performance of compressive estimators versus the SNR. (a) MSE. (b) BER.

row-column interleaving. The interpolation/anti-aliasing filters $f_1(t) = f_2(t)$ were chosen as root-raised-cosine filters with roll-off factor $\rho = 1/4$.

Recovery method. For Lasso, we used the corresponding MATLAB function from the toolbox SPGL1 [72]. The required regularization parameters were found by trial and error. CoSaMP requires a prior estimate of the sparsity of $\mathbf{x}$. In all simulations of Section VIII-B, we used the fixed sparsity estimate $S = 262$, which was determined via the formula $S = \lceil Q/(2 \log M) \rceil$ suggested in [47], where we set $Q \equiv |\mathcal{P}| = 2048$. (Note that in most scenarios where CoSaMP was applied, we actually used 2048 pilots.) The number of CoSaMP iterations was 15. For OMP, we also used the sparsity estimate $S = 262$ (and, hence, 262 iterations), except for the strongly dispersive scenario of Section VIII-C. Therefore, in Section VIII-B, the vectors produced by OMP and CoSaMP were exactly S-sparse with $S = 262$.

Channel. We simulated and estimated the channel during blocks of L transmitted OFDM symbols (L will be specified in the individual subsections). For a more realistic simulation, the channel contained a diffuse part in addition to a sparse (specular) part, with 20 dB less total power than for the sparse part. The scattering function of the diffuse part was bricked-shaped within a rectangular delay-Doppler region $\{0, \ldots, K/4 - 1\} \times [-\nu_{\max}T_{\mathrm{s}}, \nu_{\max}T_{\mathrm{s}}]$. The discrete-delay-Doppler spreading function $S_h[m, i]$ of the sparse part was computed from (21). We always assumed $P = 20$ propagation paths with scatterer delay-Doppler positions $(\tau_p/T_{\mathrm{s}}, \nu_p T_{\mathrm{s}})$ chosen uniformly at random within (or within a subset of, cf. Section VIII-B) $\{0, \ldots, K/4 - 1\} \times [-\nu_{\max}T_{\mathrm{s}}, \nu_{\max}T_{\mathrm{s}}]$ for each block of L OFDM symbols. The scatterer amplitudes η_p were randomly drawn from zero-mean, complex Gaussian distributions with three different variances (three strong scatterers of equal mean power, seven medium scatterers with 10 dB less mean power, and ten weak scatterers with 20 dB less mean power). Furthermore, we added complex white Gaussian noise $z[n]$ whose variance was adjusted to achieve a prescribed receive signal-to-noise ratio (SNR) defined as (cf. (6)) $\sum_{n=0}^{N_r-1} \mathrm{E}\{|r[n] - z[n]|^2\} / \sum_{n=0}^{N_r-1} \mathrm{E}\{|z[n]|^2\}$.

Subsampling and pilots. All estimators employed a subsampled time–frequency grid with $\Delta K = 4$ and $\Delta L = 1$, on which the pilots were selected uniformly at random.

Performance measures. For all simulations, the performance is measured by the mean square error (MSE) normalized by the mean energy of the channel coefficients, as well as by the bit error rate (BER).

B. Performance Gains Through Basis Expansions

We first compare the performance of compressive channel estimation using the DFT basis (underlying the basic estimator of Section III), the optimized basis of Section VI (without knowledge of channel statistics), and the combined DFT-DPSS basis of Section VII. The number of subcarriers is $K = 2048$, the blocklength is $L = 16$, and the maximum Doppler frequency is $\nu_{\max}T_{\mathrm{s}} = 0.03/K$ (i.e., 3% of the subcarrier spacing). Here, the maximum Doppler frequency is quite small; accordingly, the estimator of Section VII-B only performs its initial iteration (where $\mathcal{V} = \{0\}$). All estimators use the same constellation of $|\mathcal{P}| = 2048$ pilots, corresponding to 6.25% of all symbols. Fig. 3 depicts the performance versus the SNR for the three recovery algorithms employed. The performance of the optimized basis and the combined DFT-DPSS basis is seen to be similar and clearly superior to that of the pure DFT basis, especially at high SNR. This performance gain is due to the better sparsity achieved, and it is obtained even though the coherence of the optimized basis ($\mu_{\mathbf{V}} = 2.237$) is greater than that of the DFT basis ($\mu_{\mathbf{U}} = 1$) and the measurement matrix for the combined DFT-DPSS basis is not constructed from an (ideally) unitary matrix. The larger gap to the known-channel BER performance observed in Fig. 3(b) at high SNR occurs because 1) the number of pilots is too small for the channel's sparsity, and 2) the OMP-based and CoSaMP-based estimators produce S-sparse signals with $S = 262$, which is too small for the channel's sparsity.

The number of pilots, $|\mathcal{P}|$, is an important design parameter because it equals the number of measurements available for sparse reconstruction. Fig. 4 depicts the performance versus $|\mathcal{P}| \in \{512, \ldots, 8192\}$ (corresponding to 1.5625% ... 25% of all symbols) at an SNR of 17 dB. As a reference, the known-
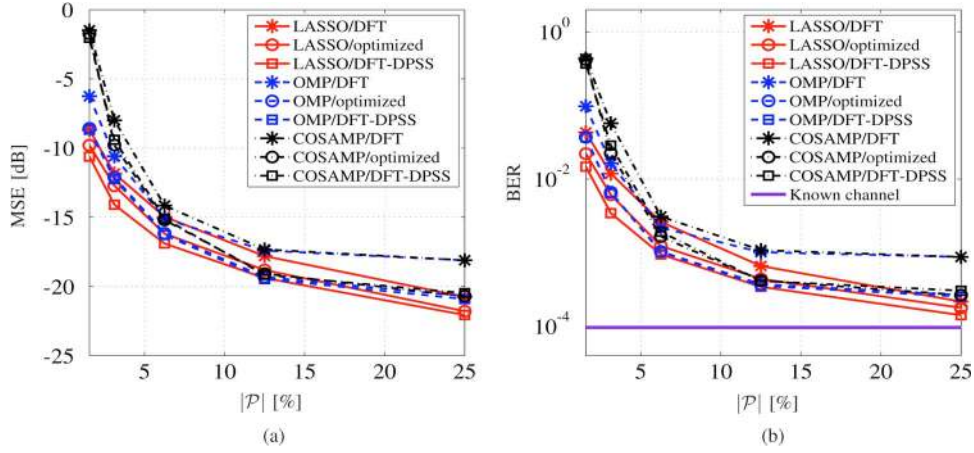
Fig. 4. Performance of compressive estimators versus the number of pilots. (a) MSE. (b) BER.
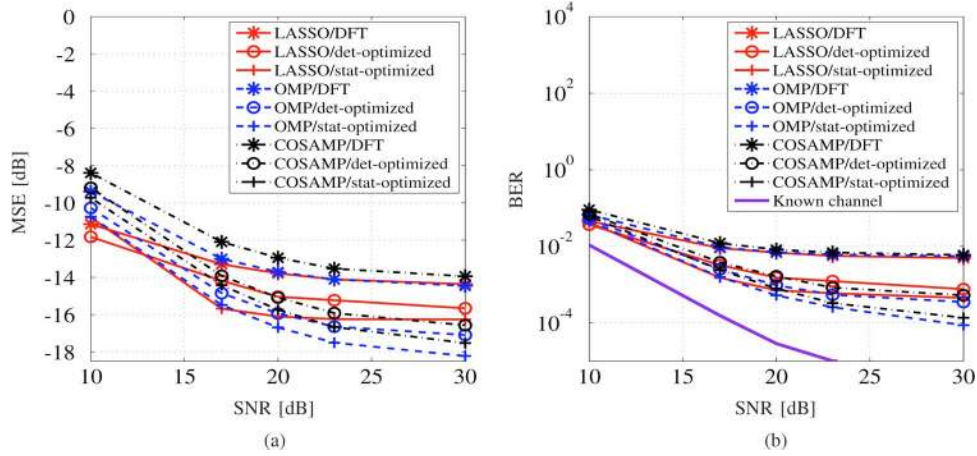

Fig. 5. Performance of DFT-based, deterministically optimized, and statistically optimized compressive estimators versus the SNR. (a) MSE. (b) BER.

channel BER is also plotted as a horizontal line. It is seen that, as expected, the performance of all estimators improves with growing $|\mathcal{P}|$. The optimized basis and the combined DFT-DPSS basis are again superior to the DFT basis.

Next, we demonstrate performance gains that can be achieved by the statistically optimized basis expansion of Section VI-B. The system and channel parameters are $K = 512, L = 64, \nu_{\max}T_{\mathrm{s}} = 0.05/K$ (5% of the subcarrier spacing), and $|\mathcal{P}| = 2048$ (6.25% of all symbols). For the sparse channel part, the 20 scatterer delay-Doppler positions $(\tau_p/T_{\mathrm{s}}, \nu_p T_{\mathrm{s}})$ now are chosen uniformly at random only within $\{0, \ldots, 127\} \times ([-0.05/K, -0.0375/K] \cup [0.0375/K, 0.05/K])$. This serves as a rough approximation to the Jakes Doppler spectrum [73], according to which the scatterers are stronger when they are closer to the maximum Doppler frequency. In order to optimize the basis expansion with this prior statistical knowledge, the pdf $p(\tau_1, \nu_1)$ (see Section VI-B) is set equal to a constant $c_1 > 0$ within $[0, 127T_{\mathrm{s}}] \times ([-0.05/(KT_{\mathrm{s}}), -0.0375/(KT_{\mathrm{s}})] \cup [0.0375/(KT_{\mathrm{s}}), 0.05/(KT_{\mathrm{s}})])$ and equal to zero outside. The variance of η_1 given (τ_1, ν_1) is assumed constant, i.e., $\sigma^2(\tau_1, \nu_1) = c_2 > 0$. Fig. 5 depicts the resulting performance versus the SNR. For comparison, we also

show the performance of the deterministically optimized basis expansion, which uses only knowledge of $\nu_{\max}$, as well as the performance of the DFT basis and the known-channel BER performance. The statistically optimized basis is seen to outperform the other bases. This can be explained by the fact that it reduces the leakage effects occurring within the Doppler interval $[-0.0375/(KT_{\mathrm{s}}), 0.0375/(KT_{\mathrm{s}})]$.

C. Performance Gains Through ISI/ICI Coefficient Estimation

Finally, we assess the performance of the compressive, iterative, decision-directed estimator of Section VII, which is able to estimate also off-diagonal (ISI/ICI) channel coefficients. We consider a wide range of maximum Doppler frequencies, corresponding also to strongly frequency-dispersive channels; more specifically, $\nu_{\max}T_{\mathrm{s}} \in [0.03/K, 0.25/K]$ or $3\% \ldots 25\%$ of the subcarrier spacing. The system parameters are $K = 1024, L = 4, \mathrm{SNR} = 17$ dB, and $|\mathcal{P}| = 128$ (i.e., only 3.125% of all symbols). There occurs no ISI, only ICI. The estimator uses $\mathcal{V} = \{(0, -3), \ldots, (0, 3)\}$ for all iterations $r \geq 1$, so that the ICI equalizer processes the main diagonal plus the first three upper and lower off-diagonals. The reliability threshold is $\epsilon = 0.2$. For ICI equalization, we use the LSQR equalizer proposed in [70], with a fixed number of 15 iterations. Furthermore, we
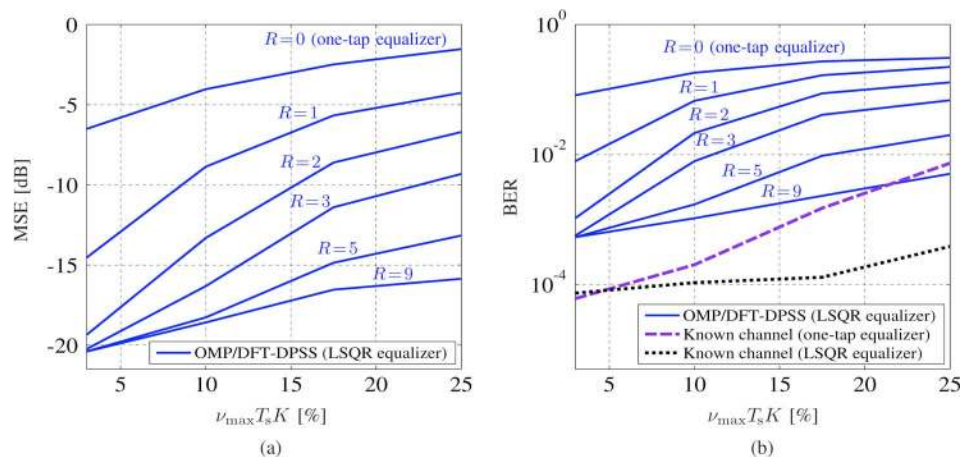
Fig. 6. Performance of the decision-directed compressive estimator versus the channel's maximum normalized Doppler frequency for different numbers of iterations R. (a) MSE. (b) BER.

use OMP with 90 iterations for CS recovery, and the combined DFT-DPSS basis of Section VII-C.

Fig. 6 depicts the performance of the estimator versus the maximum Doppler frequency for iterations up to $r = R$, with $R \in \{0, \dots, 9\}$. For comparison, the known-channel BER performance of conventional one-tap equalization and of LSQR-based ICI equalization is also shown. The MSE takes into account the estimated diagonal and first three upper and lower off-diagonal channel coefficients; it is normalized accordingly. For $R = 0$, where only the diagonal channel coefficients are estimated, the off-diagonal coefficients of the estimated channel are set to zero when calculating the MSE. It is seen from Fig. 6 that for $R = 0$, the performance is very poor even for small $\nu_{\max}$ (weakly dispersive channels). This is due to the small number of pilots used. However, the performance is improved with an increasing number R of iterations, thus demonstrating the benefits of off-diagonal coefficient estimation and the use of virtual pilots. The initial improvement is slower for larger $\nu_{\max}$, again because of the small number of pilots. It is furthermore seen that for $R = 9$ iterations, for large $\nu_{\max}$, the proposed compressive estimator is superior to the known-channel performance of one-tap equalization. Our results also show that the proposed decision-directed method is advantageous not only for coping with strongly dispersive channels; it is equally useful for further improving the spectral efficiency, even for mildly dispersive channels, because of the smaller number of pilots required.

IX. CONCLUSION

We considered the application of compressed sensing techniques to the estimation of doubly selective multipath channels within pulse-shaping multicarrier systems (which include OFDM systems as a special case). The channel coefficients on a subsampled time–frequency grid are estimated in a way that exploits the channel's sparsity in a dual delay-Doppler domain. We demonstrated that this delay-Doppler sparsity is limited by leakage effects. For combating leakage effects and, thus, enhancing sparsity, we proposed the use of an explicit basis expansion that replaces the Fourier transform used in the basic compressive channel estimation method. We also developed an iterative basis design algorithm, and we extended our basis design to the case where prior statistical information about the channel is available.

For strongly time–frequency dispersive channels, we then presented an alternative compressive channel estimator that is capable of estimating the "off-diagonal" channel coefficients characterizing intersymbol and intercarrier interference (ISI/ICI). Sparsity of the channel representation was here achieved by a basis expansion combining the advantages of Fourier (exponential) and prolate spheroidal sequences.

Simulation results demonstrated considerable performance gains achieved by the proposed sparsity-enhancing basis expansions and by explicit estimation of ISI/ICI channel coefficients. The additional computational complexity required by the basis expansions is moderate; in particular, the bases can be precomputed before the start of data transmission.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[3] Compressive Sensing Resources, Rice University. Houston, TX [Online]. Available: http://www.dsp.ece.rice.edu/cs/
[4] W. Kozek and A. F. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1579–1589, Oct. 1998.
[5] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.
[6] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11 [Online]. Available: http://grouper.ieee.org/groups/802/11/index.html
[7] *Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Standard 802.16 [Online]. Available: http://grouper.ieee.org/groups/802/16/index.html

[8] *Digital Video Broadcasting (DVB); Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television*, EN 300 744, V1.4.1, ETSI, 2001 [Online]. Available: http://www.etsi.org

[9] *Digital Audio Broadcasting (DAB) to Mobile, Portable and Fixed Receivers*, ETS 300 401, ETSI, 1995 [Online]. Available: http://www.etsi.org

[10] *Digital Radio Mondiale (DRM): System Specification*, EN 201 980, V2.1.1, ETSI, 2004 [Online]. Available: http://www.etsi.org

[11] "UTRA-UTRAN Long Term Evolution (LTE) and 3GPP System Architecture Evolution (SAE)," [Online]. Available: http://www.3gpp.org/article/lte

[12] E. G. Larsson, G. Liu, J. Li, and G. B. Giannakis, "Joint symbol timing and channel estimation for OFDM based WLANs," *IEEE Commun. Lett.*, vol. 5, no. 8, pp. 325–327, Aug. 2001.

[13] Y. Li, L. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.

[14] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[15] P. Hoeher, S. Kaiser, and P. Robertson, "Pilot-symbol-aided channel estimation in time and frequency," in *Proc. IEEE GLOBECOM-97*, Phoenix, AZ, Nov. 1997, pp. 90–96.

[16] 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.

[17] P. Fertl and G. Matz, "Efficient OFDM channel estimation in mobile environments based on irregular sampling," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct.–Nov. 2006, pp. 1777–1781.

[18] G. Leus, "On the estimation of rapidly time-varying channels," in *Proc. EUSIPCO'04*, Vienna, Austria, Sep. 2004, pp. 2227–2230.

[19] D. K. Borah and B. T. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, Jun. 1999.

[20] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sep. 2005.

[21] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 5, pp. 357–371, Oct. 2007.

[22] G. Tauböck and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. IEEE ICASSP'08*, Las Vegas, NV, Mar./Apr. 2008, pp. 2885–2888.

[23] G. Tauböck and F. Hlawatsch, "Compressed sensing based estimation of doubly selective channels using a sparsity-optimized basis expansion," in *Proc. EUSIPCO'08*, Lausanne, Switzerland, Aug. 2008.

[24] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.

[25] O. Rabaste and T. Chonavel, "Estimation of multipath channels with long impulse response at low SNR via an MCMC method," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1312–1325, Apr. 2007.

[26] W. Li and J. C. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Oceanic Eng.*, vol. 32, no. 4, pp. 927–939, Oct. 2007.

[27] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. 42nd Annu. Conf. Inf. Syst. (CISS'08)*, Princeton, NJ, Mar. 2008, pp. 5–10.

[28] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," in *Proc. IEEE ICASSP'08*, Las Vegas, NV, Apr. 2008, pp. 3469–3472.

[29] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proc. 46th Annu. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sep. 2008, pp. 575–582.

[30] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Compressed sensing of wireless channels in time, frequency, and space," in *Proc. 42nd Asilomar Conf. Sig., Syst., Comput.*, Pacific Grove, CA, Oct. 2008, pp. 2048–2052.

[31] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Sparse multipath channels: Modeling and estimation," in *Proc. 13th IEEE Digital Signal Process. Workshop*, Marco Island, FL, Jan. 2009, pp. 320–325.

[32] S. G. Mallat and Z. Zhang, "Matching pursuits and time–frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[33] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," in *Proc. IEEE OCEANS'09*, Bremen, Germany, May 2009, pp. 1–8.

[34] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 11, no. 54, pp. 4311–4322, 2006.

[35] K. Kreutz-Delgado, J. F. Murray, and B. D. Rao, "Dictionary learning algorithms for sparse representation," *Neural Comput.*, vol. 15, pp. 349–396, 2003.

[36] R. Gribonval and K. Schnass, "Dictionary identifiability from few training samples," in *Proc. EUSIPCO'08*, Lausanne, Switzerland, Aug. 2008.

[37] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1430, 1978.

[38] K. Liu, T. Kadous, and A. M. Sayeed, "Orthogonal time–frequency signaling over doubly dispersive channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2583–2603, Nov. 2004.

[39] P. Schniter, "On doubly dispersive channel estimation for pilot-aided pulse-shaped multicarrier modulation," in *Proc. 40th Annu. Conf. Inf. Sci. Syst. (CISS'06)*, Princeton, NJ, Mar. 2006, pp. 1296–1301.

[40] S. Das and P. Schniter, "Max-SINR ISI/ICI-shaping multicarrier communication over the doubly dispersive channel," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5782–5795, Dec. 2007.

[41] G. Matz, D. Schafhuber, K. Gröchenig, M. Hartmann, and F. Hlawatsch, "Analysis, optimization, and implementation of low-interference wireless multicarrier systems," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1921–1931, May 2007.

[42] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 4, pp. 360–393, Dec. 1963.

[43] P. Flandrin, *Time–Frequency/Time–Scale Analysis*. San Diego, CA: Academic, 1999.

[44] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[45] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223, Aug. 2006.

[46] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, Oct. 2004.

[47] J. A. Tropp and D. Needell, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301–321, May 2009.

[48] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[49] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Trans. Math. Software*, vol. 8, pp. 43–71, Mar. 1982.

[50] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proc. 40th Annu. Conf. Inf. Sci. Syst. (CISS'06)*, Princeton, NJ, Mar. 2006, pp. 207–212.

[51] H. Rauhut, "Stability results for random sampling of sparse trigonometric polynomials," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5661–5670, Dec. 2008.

[52] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials II—Orthogonal matching pursuit versus basis pursuit," *Found. Comput. Math.*, vol. 8, pp. 737–763, Dec. 2008.

[53] D. L. Donoho, I. Drori, Y. Tsaig, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Dept. of Statist., Stanford Univ., Stanford, CA, Tech. Rep. 2006-02, Mar. 2006.

[54] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[55] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.

[56] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc.*, vol. 58, pp. 267–288, 1994.

[57] I. Loris, "On the performance of algorithms for the minimization of ℓ_1-penalized functionals," *Inverse Problems* vol. 25, no. 3(035008), 2009 [Online]. Available: http://arxiv.org/abs/0710.4082

[58] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[59] P. Schniter, L. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Proc. Worksh. Inf. Theory Appl. (ITA)*, La Jolla, CA, Jan. 2008, pp. 326–333.

[60] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, pp. 2313–2351, Dec. 2007.

[61] S. Barbarossa and A. Scaglione, "Time-varying fading channels," in *Signal Processing Advances in Wireless & Mobile Communications—Trends in Single- and Multi-User Systems*, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Upper Saddle River, NJ: Prentice-Hall, 2000, vol. 2, ch. 1, pp. 1–57.

[62] G. Matz and F. Hlawatsch, "Time-varying communication channels: Fundamentals, recent developments, and open problems," in *Proc. EU-SIPCO-06*, Florence, Italy, Sep. 2006.

[63] P. Schniter and A. M. Sayeed, "A sparseness-preserving virtual MIMO channel model," in *Proc. 38th Annu. Conf. Inf. Sci. Syst. (CISS'04)*, Princeton, NJ, Mar. 2004, pp. 36–41.

[64] Z. Wang, G. R. Arce, B. M. Sadler, J. L. Paredes, S. Hoyos, and Z. Yu, "Compressed UWB signal detection with narrowband interference mitigation," in *Proc. IEEE ICUWB-'08*, Hannover, Germany, Sep. 2008, pp. 157–160.

[65] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[66] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, Dec. 2004.

[67] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming (Web Page and Software), Stanford Univ., Stanford, CA [Online]. Available: http://stanford.edu/~boyd/cvx

[68] P. Schniter, "Low-complexity equalization of OFDM in doubly-selective channels," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1002–1011, Apr. 2004.

[69] L. Rugini, P. Banelli, and G. Leus, "Simple equalization of time-varying channels for OFDM," *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 619–621, Jul. 2005.

[70] G. Tauböck, M. Hampejs, G. Matz, F. Hlawatsch, and K. Gröchenig, "LSQR-based ICI equalization for multicarrier communications in strongly dispersive and highly mobile environments," in *Proc. IEEE SPAWC-2007*, Helsinki, Finland, Jun. 2007, pp. 1–5.

[71] M. Hampejs, P. Svac, G. Tauböck, K. Gröchenig, F. Hlawatsch, and G. Matz, "Sequential LSQR-based ICI equalization and decision-feedback ISI cancellation in pulse-shaped multicarrier systems," in *Proc. IEEE SPAWC-2009*, Perugia, Italy, June 2009, pp. 1–5.

[72] M. Friedlander and E. van den Berg, "Toolbox SPGL1," Univ. British Columbia. Vancouver, BC, Canada [Online]. Available: http://www.cs.ubc.ca/labs/scl/spgl1/

[73] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

Franz Hlawatsch (S'85–M'88–SM'00) received the Diplom-Ingenieur. Dr.techn., and Univ.-Dozent (habilitation) degrees in electrical engineering/signal processing from Vienna University of Technology, Vienna, Austria, in 1983, 1988, and 1996, respectively.

Since 1983, he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, where he is currently an Associate Professor. During 1991–1992, as a recipient of an Erwin Schrödinger Fellowship, he spent a sabbatical year with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI. In 1999, 2000, and 2001, he held one-month Visiting Professor positions with INP/ENSEEIHT/TeSA, Toulouse, France, and IRCCyN, Nantes, France. He (co)authored a book, a review paper that appeared in the *IEEE Signal Processing Magazine*, about 180 refereed scientific papers and book chapters, and three patents. He coedited two books. His research interests include signal processing for wireless communications, statistical signal processing, and compressive signal processing.

Prof. Hlawatsch was Technical Program Co-Chair of EUSIPCO 2004 and served on the technical committees of numerous IEEE conferences. From 2003 to 2007, he served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and since 2008, he has served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. From 2004 to 2009, he was a member of the IEEE SPCOM Technical Committee. He is coauthor of a paper that won an IEEE Signal Processing Society Young Author Best Paper Award.

Daniel Eiwen (S'10) received the diploma degree in mathematics from the University of Vienna, Vienna, Austria, in 2008. He is currently pursuing the Ph.D. degree at the Numerical Harmonic Analysis Group (NuHAG), Faculty of Mathematics, University of Vienna.

His research interests include compressed sensing, sparse approximation, and time–frequency analysis, as well as their application in signal processing.

Georg Tauböck (S'01–M'07) received the Dipl.-Ing. degree and the Dr.techn. degree (with highest honors) in electrical engineering and the Dipl.-Ing. degree in mathematics (with highest honors) from Vienna University of Technology, Vienna, Austria, in 1999, 2005, and 2008, respectively. He also received the diploma in violoncello from the Conservatory of Vienna, Vienna, Austria, in 2000.

From 1999 to 2005, he was with the FTW Telecommunications Research Center Vienna, Vienna, and since 2005, he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology. His research interests include wireline and wireless communications, compressed sensing, signal processing, and information theory.

Holger Rauhut received the diploma degree in mathematics from the Technical University of Munich, Munich, Germany, in 2001 and the Dr. rer. nat. degree in mathematics from the Applied Algorithmic Mathematics program, Technical University of Munich, in 2004.

From 2005 until 2008, he was with the Numerical Harmonic Analysis Group, Faculty of Mathematics, University of Vienna, Vienna, Austria, as a PostDoc. Since March 2008, he has been a Professor for mathematics (Bonn Junior Fellow) with the Hausdorff Center for Mathematics and the Institute for Numerical Simulation, University of Bonn, Bonn, Germany. His research interests include compressed sensing, sparse approximation, random matrices, and time–frequency and wavelet analysis.