

# Compressive Estimation of Doubly Selective Channels: Exploiting Channel Sparsity to Improve Spectral Efficiency in Multicarrier Transmissions\*

Georg Tauböck<sup>a</sup> (corresponding author), Franz Hlawatsch<sup>a</sup>, and Holger Rauhut<sup>b</sup>

<sup>a</sup>Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology  
Gusshausstrasse 25/389, 1040 Vienna, Austria

Phone: +43 1 58801 38962, Fax: +43 1 58801 38999, E-mail: gtauboec@nt.tuwien.ac.at

<sup>b</sup>Hausdorff Center for Mathematics and Institute for Numerical Simulation, University of Bonn  
Endenicher Allee 60, 53115 Bonn, Germany

Phone: +49 228 73 62245, Fax: +49 228 73 62251, E-mail: rauhut@hcm.uni-bonn.de

submitted to the IEEE Journal of Selected Topics in Signal Processing, March 16, 2009

**Abstract**—We consider the estimation of doubly selective wireless channels within pulse-shaping multicarrier systems (which include OFDM systems as a special case). A pilot-assisted channel estimation technique using the methodology of *compressed sensing* (CS) is proposed. By exploiting a channel’s delay-Doppler sparsity, CS-based channel estimation allows an increase in spectral efficiency through a reduction of the number of pilot symbols that have to be transmitted. We also present an extension of our basic channel estimator that employs a sparsity-improving basis expansion. We propose a framework for optimizing the basis and an iterative approximate basis optimization algorithm. Simulation results using three different CS recovery algorithms demonstrate significant performance gains (in terms of improved estimation accuracy or reduction of the number of pilots) relative to conventional least-squares estimation, as well as substantial advantages of using an optimized basis.

**Index terms**—OFDM, multicarrier modulation, channel estimation, doubly selective channel, compressed sensing, sparse reconstruction, basis pursuit, orthogonal matching pursuit, CoSaMP, dictionary learning.

---

\*This work was supported by WWTF projects MOHAWI (MA 44) and SPORTS (MA 07-004) and by FWF project “Statistical Inference” (S10603-N13) within the National Research Network SISE. HR acknowledges support by the Hausdorff Center for Mathematics, University of Bonn. Parts of this work have previously been published in [1, 2].

# 1 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) [3, 4]. CS has gained a fast-growing interest in applied mathematics and signal processing. Applications of CS have been proposed in areas as diverse as coding, information theory, high-dimensional geometry, statistical signal processing, machine learning, Bayesian inference, multi-band signal processing, imaging, analog-to-information conversion, biosensing, geophysical data analysis, radar, astronomy, metrology, and communications, see e.g. [5]. 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 [6, 7]. 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 mobile communication systems (UMTS LTE) [8–13].

Coherent detection in such systems requires accurate channel state information (CSI) at the receiver. Usually, CSI is obtained by embedding training data (pilot symbols) into the transmit signal and using a standard least-squares (LS) [14] or minimum mean-square error (MMSE) [15] estimator. More advanced channel estimators for MC transmissions include estimators employing one-dimensional (1-D), double 1-D, or two-dimensional (2-D) MMSE filtering algorithms [16–18]; 2-D irregular sampling techniques [19]; or basis expansion models [20–22].

Wireless multipath channels tend to be dominated by a relatively small number of clusters of significant paths, especially when transmitting over large bandwidths and large signaling durations [23]. Conventional methods for channel estimation do not take advantage of this *inherent sparsity* of the transmission channel. In [1, 2], we proposed CS-based techniques for the estimation of delay-Doppler sparse, doubly selective channels within pulse-shaping MC systems. We were able to demonstrate that CS provides a constructive 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.

The modeling and estimation of sparse channels have received some attention recently [1, 2, 24–32]. 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, more commonly studied under the rubric of CS these days. Both [24] and [27] considered single-carrier signaling and proposed variants of the matching pursuit algorithm [33] 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, 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 and sparsity in the delay-Doppler domain, somewhat similar to [1]. However, [29] studied also single-carrier signaling and applied a different recovery technique (the Dantzig selector [34]).

In this paper, building on our work in [1, 2], we propose and study CS-based estimation techniques for doubly selective channels that exploit a channel's delay-Doppler sparsity to reduce the number of pilot symbols and, hence, increase the spectral efficiency of MC transmissions. These techniques use three alternative nonlinear sparse recovery algorithms that have been extensively studied in the recent CS literature, namely,  $\ell_1$ -minimization (also known as *basis pursuit*) [35, 36], *orthogonal matching pursuit* (see [37] and references therein), and *CoSaMP* [38].

Besides the presentation of our basic compressive channel estimation scheme, this paper offers the following contributions.

- We propose a time-frequency subsampling that leads to reduced complexity and improved performance.
- We present generalized CS-based channel estimators using a *sparsity-improving basis expansion* of the channel's time-frequency coefficients. The improved sparsity is due to a reduction of leakage effects.
- We propose an iterative basis optimization procedure that aims to minimize sparsity, and we demonstrate significant performance gains obtained with the optimized basis. Our basis optimization is similar in spirit—but not algorithmically—to dictionary learning techniques recently proposed in [39–41].
- We demonstrate that CS-based channel estimation can exploit a channel's delay-Doppler sparsity for increasing spectral efficiency through a reduction of the number of pilot symbols.

This paper is organized as follows. In Section 2, we describe the MC system model, the general setup of pilot-assisted channel estimation, and the time-frequency subsampling. Section 3 provides a brief review of the basic concept of CS and of three recovery algorithms used later. In Section 4, we analyze the sparsity of the channel's delay-Doppler representation and present the basic CS-based channel estimation method. In Section 5, we develop the generalized CS-based channel estimation method using a

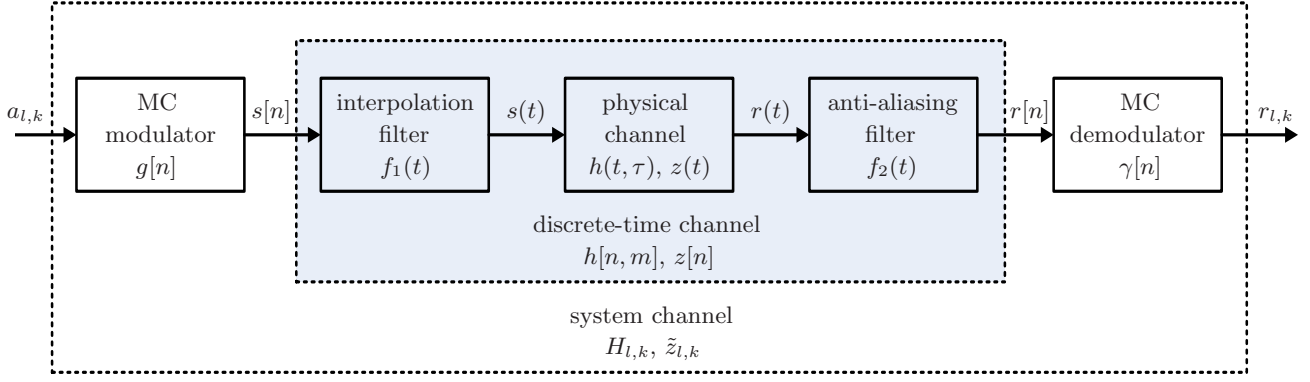


Figure 1: MC system and equivalent channels.

sparsity-improving basis expansion. A framework and an iterative algorithm for optimizing the sparsity-improving basis are proposed in Section 6. Finally, simulation results presented in Section 7 assess the performance gains achieved with CS-based channel estimation.

## 2 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 [6, 42]. This framework includes CP-OFDM as a special case. The complex baseband domain is considered throughout.

### 2.1 Modulator, Channel, Demodulator

The block diagram of an MC system is depicted in Fig. 1. The MC modulator generates the discrete-time transmit signal [6]

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

where  $K$  and  $L$  denote the numbers of subcarriers and transmitted MC symbols, respectively;  $a_{l,k} \in \mathcal{A}$  ( $l = 0, \dots, L-1$ ;  $k = 0, \dots, K-1$ ) denotes the complex data symbols, drawn from a finite symbol alphabet  $\mathcal{A}$ ; and

$$g_{l,k}[n] \triangleq g[n - lN] e^{j2\pi \frac{k}{K}(n - lN)}$$

is a time-frequency shifted version 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

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

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

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

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

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

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

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

Here,  $\gamma_{l,k}[n] \triangleq \gamma[n - lN] e^{j2\pi \frac{k}{K}(n - lN)}$  is a time-frequency shifted version of a receive pulse  $\gamma[n]$ . Finally, the demodulated symbols  $r_{l,k}$  are equalized (i.e., divided by estimates of the corresponding channel coefficients) and quantized according to the data symbol alphabet  $\mathcal{A}$ .

Combining (2)–(4), we obtain an equivalent discrete-time channel (see Fig. 1) that is described by the following equation relating the discrete-time signals  $s[n]$  and  $r[n]$ :

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

with the discrete-time time-varying impulse response

$$h[n, m] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t + nT_s, \tau) f_1(t - \tau + mT_s) f_2^*(t) dt d\tau \quad (7)$$

and the discrete-time noise  $z[n] = \int_{-\infty}^{\infty} z(t) f_2^*(t - nT_s) dt$ .

Despite the superior spectral concentration of smooth (nonrectangular) pulses  $g[n]$  and  $\gamma[n]$ , in current MC applications CP-OFDM is almost exclusively used [8–13]. 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  $n = 0, \dots, N-1$  and 0 otherwise, and a rectangular receive pulse  $\gamma[n]$  that is 1 for  $n = N-K, \dots, N-1$  and 0 otherwise ( $N-K \geq 0$  is the CP length).

## 2.2 System Channel

Next, we consider the equivalent *system channel* that subsumes the MC modulator, interpolation filter, physical channel, anti-aliasing filter, and MC demodulator (see Fig. 1). Combining (5), (6), and (1), we

obtain a relation of the form

$$r_{l,k} = \sum_{l'=0}^{L-1} \sum_{k'=0}^{K-1} H_{l,k;l',k'} a_{l',k'} + z_{l,k}, \quad l = 0, \dots, L-1, \quad k = 0, \dots, K-1, \quad (8)$$

with  $z_{l,k} = \langle z, \gamma_{l,k} \rangle = \sum_{n=-\infty}^{\infty} z[n] \gamma_{l,k}^*[n]$ . Assuming that the intersymbol and intercarrier interference characterized by  $H_{l,k;l',k'}$  with  $(l', k') \neq (l, k)$  is small (which is justified even in highly mobile environments if  $g[n]$  and  $\gamma[n]$  are properly designed [42]), we rewrite (8) as

$$r_{l,k} = H_{l,k} a_{l,k} + \tilde{z}_{l,k}, \quad l = 0, \dots, L-1, \quad k = 0, \dots, K-1, \quad (9)$$

where  $H_{l,k} \triangleq H_{l,k;l,k}$  and all intersymbol and intercarrier interference is absorbed in the noise/interference term  $\tilde{z}_{l,k}$ . The *system channel coefficients*  $H_{l,k}$  can be expressed in terms of  $g[n]$ ,  $h[n, m]$ , and  $\gamma[n]$  as [6]

$$H_{l,k} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h[n+lN, m] g[n-m] \gamma^*[n] e^{-j\frac{2\pi}{K}km}.$$

We will need an expression of the system channel coefficients  $H_{l,k}$  in a joint delay-Doppler-domain. Let us suppose that the receive pulse  $\gamma[n]$  is zero outside  $[0, L_\gamma]$ . To compute  $r_{l,k}$  in (5) for  $l = 0, \dots, L-1$ , we need to know  $r[n]$  for  $n = 0, \dots, N_r-1$ , where

$$N_r \triangleq (L-1)N + L_\gamma + 1.$$

In this interval, we can rewrite the discrete-time channel (6) as<sup>1</sup>

$$r[n] = \sum_{m=-\infty}^{\infty} \sum_{i=0}^{N_r-1} S_h[m, i] s[n-m] e^{j2\pi \frac{ni}{N_r}} + z[n], \quad (10)$$

with the *discrete-delay-Doppler spreading function* [45]

$$S_h[m, i] \triangleq \frac{1}{N_r} \sum_{n=0}^{N_r-1} h[n, m] e^{-j2\pi \frac{in}{N_r}}, \quad m, i \in \mathbb{Z}. \quad (11)$$

Note that  $S_h[m, i]$  represents the discrete-time channel (previously characterized by the discrete-time time-varying impulse response  $h[n, m]$ ) in terms of discrete delay (time shift)  $m$  and discrete Doppler frequency shift  $i$ . Combining (5), (10), and (1) and again including intersymbol and intercarrier interference in a noise/interference term  $\tilde{z}_{l,k}$ , we reobtain the system channel relation (9), however with the system channel coefficients  $H_{l,k}$  now expressed as

$$H_{l,k} = \sum_{m=-\infty}^{\infty} \sum_{i=0}^{N_r-1} F[m, i] e^{-j2\pi(\frac{km}{K} - \frac{Nli}{N_r})}, \quad l = 0, \dots, L-1, \quad k = 0, \dots, K-1, \quad (12)$$

---

<sup>1</sup>This expression is effectively equivalent to a basis expansion model using an exponential (Fourier) basis [43, 44].

where

$$F[m, i] \triangleq S_h[m, i] A_{\gamma, g}^* \left( m, \frac{i}{N_r} \right) \quad (13)$$

with the *cross-ambiguity function* [46] of the pulses  $\gamma[n], g[n]$

$$A_{\gamma, g}(m, \xi) \triangleq \sum_{n=-\infty}^{\infty} \gamma[n] g^*[n-m] e^{-j2\pi\xi n}, \quad m \in \mathbb{Z}, \xi \in \mathbb{R}.$$

Note that the function  $F[m, i]$  is a discrete-delay-Doppler representation of the system channel coefficients  $H_{l,k}$ ; it equals the discrete-delay-Doppler spreading function of the discrete-time channel,  $S_h[m, i]$ , up to a weighting by the pulse-dependent factor  $A_{\gamma, g}^* \left( m, \frac{i}{N_r} \right)$ .

For CS-based channel estimation, we need to reformulate the relation (12) as a 2-D discrete Fourier transform (DFT). Let us assume that  $h[n, m]$  is causal with maximum delay at most  $K-1$ , i.e.,  $h[n, m]$  is zero for  $m \notin \{0, \dots, K-1\}$ . Using the approximation  $N_r \approx LN$  (which, by the way, is exact for CP-OFDM), the system channel coefficients  $H_{l,k}$  can then be expressed as

$$H_{l,k} = \sum_{m=0}^{K-1} \sum_{i=0}^{L-1} \tilde{F}[m, i] e^{-j2\pi \left( \frac{km}{K} - \frac{li}{L} \right)}, \quad l = 0, \dots, L-1, \quad k = 0, \dots, K-1, \quad (14)$$

with the “pre-aliased” version of  $F[m, i]$

$$\tilde{F}[m, i] \triangleq \sum_{q=0}^{N-1} F[m, i + qL], \quad i \in \mathbb{Z}. \quad (15)$$

Since both  $S_h[m, i]$  and  $A_{\gamma, g}^* \left( m, \frac{i}{N_r} \right)$  are  $N_r$ -periodic with respect to  $i$ ,  $\tilde{F}[m, i]$  is  $L$ -periodic with respect to  $i$ . Assuming that  $L$  is even, we can hence rewrite (14) as

$$H_{l,k} = \sum_{m=0}^{K-1} \sum_{i=-L/2}^{L/2-1} \tilde{F}[m, i] e^{-j2\pi \left( \frac{km}{K} - \frac{li}{L} \right)}, \quad l = 0, \dots, L-1, \quad k = 0, \dots, K-1. \quad (16)$$

According to (16),  $H_{l,k}$  is the 2-D DFT of the discrete-delay-Doppler representation  $\tilde{F}[m, i]$ , with Fourier-dual variables  $l$  (symbol time)  $\leftrightarrow i$  (Doppler frequency shift) and  $k$  (subcarrier frequency)  $\leftrightarrow m$  (delay).

### 2.3 Time-Frequency Subsampling and Pilot-Based Channel Estimation

For practical wireless channels as well as transmit and receive pulses, the function  $F[m, i]$  in (13) is effectively supported in a subregion of the delay-Doppler plane. This allows us to perform a subsampling in the time-frequency domain. Thus, hereafter we assume that the support of  $F[m, i]$  (within the fundamental  $i$  period  $[-N_r/2, N_r/2 - 1]$ ) is contained in  $[0, D-1] \times [-I/2, I/2 - 1]$ , where  $D \leq K$  and  $I \leq L$ . For mathematical convenience,  $I$  is chosen even; furthermore,  $D$  and  $I$  are chosen such that

$\Delta K \triangleq K/D$  and  $\Delta L \triangleq L/I$  are integers. It follows that the support of  $\tilde{F}[m, i]$  (within the fundamental  $i$  period  $[-L/2, L/2-1]$ ) is contained in  $[0, D-1] \times [-I/2, I/2-1]$ . Because of (16), the channel coefficients  $H_{l,k}$  are then uniquely specified by their values on the *subsampled time-frequency grid*

$$\mathcal{G} \triangleq \{(l, k) = (l' \Delta L, k' \Delta K) : l' = 0, \dots, I-1, k' = 0, \dots, D-1\}.$$

Furthermore, (16) entails the relation

$$H_{l' \Delta L, k' \Delta K} = \sum_{m=0}^{D-1} \sum_{i=-I/2}^{I/2-1} \tilde{F}[m, i] e^{-j2\pi(\frac{k'm}{D} - \frac{l'i}{I})}, \quad l' = 0, \dots, I-1, k' = 0, \dots, D-1. \quad (17)$$

Note that we also allow the limiting case of no subsampling in either or both dimensions, that is,  $D = K$  (i.e.,  $\Delta K = 1$ ) and/or  $I = L$  (i.e.,  $\Delta L = 1$ ). However, subsampling is beneficial as a means to improve performance and reduce complexity.

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  $\mathcal{G}$ . The pilots  $p_{l,k}$  and their time-frequency positions are assumed to be known to the receiver. From (9),  $r_{l,k} = H_{l,k} p_{l,k} + \tilde{z}_{l,k}$  for  $(l, k) \in \mathcal{P}$ . 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

$$\hat{H}_{l,k} \triangleq \frac{r_{l,k}}{p_{l,k}} = H_{l,k} + \frac{\tilde{z}_{l,k}}{p_{l,k}}, \quad (l, k) \in \mathcal{P}. \quad (18)$$

Thus, the  $H_{l,k}$  for  $(l, k) \in \mathcal{P}$  are known up to additive noise terms  $\tilde{z}_{l,k}/p_{l,k}$ . Next, the receiver calculates channel estimates  $\hat{H}_{l,k}$  for all  $(l, k) \in \mathcal{G}$ , i.e., on the whole subsampled grid, from the  $\hat{H}_{l,k}$  for  $(l, k) \in \mathcal{P}$ . Some interpolation technique is usually employed for this task. Subsequently, the inverse of the DFT relation (17) is used to calculate an estimate of  $\tilde{F}[m, i]$ . Finally, estimates of all channel coefficients  $H_{l,k}$  are obtained from (16). In Section 4, we will replace the conventional interpolation-based approach by a CS-based method.

### 3 Review of Compressed Sensing

Before presenting the CS-based channel estimation technique, we need to review some CS basics. CS provides a theoretical framework and efficient algorithms for reconstructing sparse signals from a comparatively small number of measurements (samples) [3, 4]. More specifically, CS considers the *sparse reconstruction problem* of estimating an unknown 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”)

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



Here,  $\Phi \in \mathbb{C}^{Q \times M}$  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 *S-sparse*, i.e., at most  $S$  of its entries are nonzero. The positions (indices) of the nonzero entries of  $\mathbf{x}$  are unknown. Typically, the number of unknowns to be estimated is much larger than the number of known measurements, i.e.,  $M \gg Q$ . Thus,  $\Phi$  is a fat matrix.

We shall now review three CS recovery methods that will be used in Section 4.

### 3.1 Basis Pursuit

We first consider an estimator of  $\mathbf{x}$  that is given by the following version of *basis pursuit* (BP) [35, 36]:

$$\hat{\mathbf{x}} \triangleq \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\Phi \mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \quad (20)$$

for a given  $\epsilon > 0$  [35]. This is a convex program that can be solved efficiently by interior-point methods [47]. The classical BP is reobtained for  $\epsilon = 0$ , in which case the constraint becomes  $\Phi \mathbf{x} = \mathbf{y}$ . Note that the estimate  $\hat{\mathbf{x}}$  is not necessarily  $S$ -sparse with a given  $S$ ; however, in general it will be approximately  $S$ -sparse if  $\mathbf{x}$  is approximately  $S$ -sparse, as discussed below.

A performance analysis of BP can be based on the assumption that the measurement matrix  $\Phi$  obeys a “restricted isometry hypothesis” [35]. The *S-restricted isometry constant*  $\delta_S$  of  $\Phi$  is defined as the smallest positive number  $\delta$  such that

$$(1 - \delta) \|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2 \leq (1 + \delta) \|\mathbf{x}\|_2^2 \quad (21)$$

for all  $S$ -sparse vectors  $\mathbf{x} \in \mathbb{C}^M$ . Using the restricted isometry constant  $\delta_{2S}$ , the performance of BP (20) is characterized by the following theorem<sup>2</sup> [35, 48].

*For a given  $S$ , assume that the  $2S$ -restricted isometry constant of  $\Phi$  satisfies*

$$\delta_{2S} < \sqrt{2} - 1. \quad (22)$$

*Let  $\mathbf{x} \in \mathbb{C}^N$  (not necessarily sparse) and  $\mathbf{y} = \Phi \mathbf{x} + \mathbf{z}$  with  $\|\mathbf{z}\|_2 \equiv \|\mathbf{y} - \Phi \mathbf{x}\|_2 \leq \epsilon$ , and let  $\mathbf{x}_S \in \mathbb{C}^M$  contain the  $S$  components of  $\mathbf{x}$  with largest absolute values, the remaining  $M - S$  components being zero (thus,  $\mathbf{x}_S$  is the best  $S$ -sparse approximation to  $\mathbf{x}$ ). Then the reconstruction error of BP (20) is bounded as*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_1 \epsilon + C_2 \frac{\|\mathbf{x} - \mathbf{x}_S\|_1}{\sqrt{S}}, \quad (23)$$

*where the constants  $C_1$  and  $C_2$  depend only on  $\delta_{2S}$ .*

---

<sup>2</sup>In [35, 48, 49], this theorem is formulated for the real-valued case. However, it also holds for the complex case, with the same proof.

In particular, (23) shows that if  $\mathbf{x}$  is approximately  $S$ -sparse (i.e.,  $\mathbf{x} \approx \mathbf{x}_S$ ) and the norm of the noise vector  $\mathbf{z}$  is small, then the reconstruction error  $\|\hat{\mathbf{x}} - \mathbf{x}\|_2$  is small and thus also the estimate  $\hat{\mathbf{x}}$  is approximately  $S$ -sparse. In the noiseless case ( $\epsilon=0$ ), (23) shows that  $\hat{\mathbf{x}} = \mathbf{x}$  for any  $S$ -sparse vector  $\mathbf{x}$ .

Conditions under which the measurement matrix  $\Phi$  satisfies (22) are obviously of interest. It has been shown [3, 50, 51] that if  $\Phi \in \mathbb{C}^{Q \times M}$  is constructed by selecting uniformly at random  $Q$  rows<sup>3</sup> from a unitary  $M \times M$  matrix  $\mathbf{U}$  and normalizing the columns (so that they have unit  $\ell_2$  norms), a sufficient condition for  $\delta_S \leq \delta$  to be true with probability  $1-\eta$  is provided by the following lower bound on the number of observations:

$$Q \geq C_3 \delta^{-2} (\ln M)^4 \mu_{\mathbf{U}}^2 S \ln(1/\eta). \quad (24)$$

Here,  $\mu_{\mathbf{U}} \triangleq \sqrt{M} \max_{i,j} |U_{i,j}|$  (known as the *coherence* of  $\mathbf{U}$ ) and  $C_3$  is a constant.

### 3.2 Orthogonal Matching Pursuit

The second estimator of  $\mathbf{x}$  we are considering is *orthogonal matching pursuit* (OMP) [37]. This is an iterative algorithm that calculates sequences of partial estimates  $\hat{\mathbf{x}}_j$ , approximations  $\mathbf{y}_j$  to the observation vector  $\mathbf{y}$ , and residuals  $\mathbf{r}_j = \mathbf{y} - \mathbf{y}_j$  as summarized in the following.

**Initialization** ( $j=0$ ): Define the residual  $\mathbf{r}_0 = \mathbf{y}$  and the empty index set  $\mathcal{S}_0 = \emptyset$ .

**Steps at the  $j$ th iteration** ( $j=1, 2, \dots$ ):

1. Determine an index  $s_j$  that satisfies

$$|\langle \mathbf{r}_{j-1}, \boldsymbol{\phi}_{s_j} \rangle| = \max_{s \in \{1, \dots, M\} \setminus \mathcal{S}_{j-1}} |\langle \mathbf{r}_{j-1}, \boldsymbol{\phi}_s \rangle|,$$

where  $\boldsymbol{\phi}_s$  denotes the  $s$ th column of  $\Phi$ .

2. Augment the index set as  $\mathcal{S}_j = \mathcal{S}_{j-1} \cup \{s_j\}$ . (Note that  $|\mathcal{S}_j| = j$ .)
3. Calculate a new estimate  $\hat{\mathbf{x}}_j$  such that it has zero entries outside the index set  $\mathcal{S}_j$ , i.e.,  $\text{supp}\{\hat{\mathbf{x}}_j\} = \mathcal{S}_j$ , and the nonzero entries (combined into a  $j$ -dimensional vector denoted by  $\hat{\mathbf{x}}_{j|\mathcal{S}_j}$ ) are given by the solution of a least-squares problem:

$$\hat{\mathbf{x}}_{j|\mathcal{S}_j} = \arg \min_{\mathbf{x} \in \mathbb{C}^j} \|\mathbf{y} - \Phi_{\mathcal{S}_j} \mathbf{x}\|_2 = \Phi_{\mathcal{S}_j}^\dagger \mathbf{y}.$$

Here, the  $Q \times j$  matrix  $\Phi_{\mathcal{S}_j}$  comprises the columns of  $\Phi$  indexed by  $\mathcal{S}_j$  and  $\Phi_{\mathcal{S}_j}^\dagger$  is its Moore-Penrose pseudoinverse.

---

<sup>3</sup>That is, all possible subsets of  $Q$  rows are equally likely.

4. Calculate a new approximation to the observation vector  $\mathbf{y}$  and a new residual:

$$\mathbf{y}_j = \Phi \hat{\mathbf{x}}_j = \Phi_{\mathcal{S}_j} \hat{\mathbf{x}}_{j|\mathcal{S}_j}, \quad \mathbf{r}_j = \mathbf{y} - \mathbf{y}_j.$$

These steps are repeated until a stopping criterion is satisfied. This may be a predefined number of iterations (corresponding to a fixed known sparsity  $S$ ) or a required approximation quality (e.g., the norm of the residual  $\mathbf{r}_j$  is required to be below a given threshold). Upon termination (at the final iteration, say,  $j = J$ ), OMP outputs the  $J$ -sparse vector  $\hat{\mathbf{x}} = \hat{\mathbf{x}}_J$ .

Some theoretical results on the performance of OMP are available [52] but they are not as deep as for BP [53]. However, simulation results demonstrate a similar or even superior average performance; furthermore, OMP can be implemented more efficiently.

### 3.3 CoSaMP

The recently proposed *CoSaMP* algorithm [38] combines the near-optimal (uniform) recovery guarantees of BP with the advantages of greedy algorithms like OMP, i.e., speed and easy implementation. It is an iterative algorithm that refines the current estimate  $\hat{\mathbf{x}}_j$  in each iteration. (Note that *subspace pursuit* [54] is a very similar method.) The algorithm works as follows.

**Initialization** ( $j = 0$ ): Choose a sparsity level  $S$ . Let  $\hat{\mathbf{x}}_0$  be the  $M$ -dimensional zero vector, and set  $\mathbf{r}_0 = \mathbf{y}$ .

**Steps at the  $j$ th iteration** ( $j = 1, 2, \dots$ ):

1. Let  $\Omega_j$  be the index set corresponding to the  $2S$  largest of the correlations  $|\langle \mathbf{r}_{j-1}, \boldsymbol{\phi}_s \rangle|$ ,  $s = 1, \dots, M$ , where  $\boldsymbol{\phi}_s$  denotes the  $s$ th column of  $\Phi$ . Merge supports according to

$$\mathcal{T}_j = \Omega_j \cup \text{supp}\{\hat{\mathbf{x}}_{j-1}\},$$

where  $\text{supp}\{\hat{\mathbf{x}}_{j-1}\}$  is the set of the indices of all nonzero entries of  $\hat{\mathbf{x}}_{j-1}$ .

2. Calculate the vector  $\mathbf{b}_j$  that has zero entries outside the index set  $\mathcal{T}_j$ , i.e.,  $\text{supp}\{\mathbf{b}_j\} = \mathcal{T}_j$ , and whose nonzero entries (combined into a  $|\mathcal{T}_j|$ -dimensional vector denoted by  $\mathbf{b}_{j|\mathcal{T}_j}$ ) are given by the solution of a least-squares problem:

$$\mathbf{b}_{j|\mathcal{T}_j} = \arg \min_{\mathbf{b} \in \mathbb{C}^{|\mathcal{T}_j|}} \|\mathbf{y} - \Phi_{\mathcal{T}_j} \mathbf{b}\|_2 = \Phi_{\mathcal{T}_j}^\dagger \mathbf{y}.$$

Here, the  $Q \times |\mathcal{T}_j|$  matrix  $\Phi_{\mathcal{T}_j}$  comprises the columns of  $\Phi$  indexed by  $\mathcal{T}_j$ .

3. Calculate the new estimate  $\hat{\mathbf{x}}_j$  as the best  $S$ -sparse approximation to  $\mathbf{b}_j$ , i.e., the vector that agrees with  $\mathbf{b}_j$  in its  $S$  largest entries and is zero elsewhere.

4. Update the residual as  $\mathbf{r}_j = \mathbf{y} - \Phi \hat{\mathbf{x}}_j$ .

These steps are repeated until a stopping criterion is triggered, which may be a fixed number of iterations or a required approximation quality. CoSaMP then outputs the last iterate, i.e.,  $\hat{\mathbf{x}} = \hat{\mathbf{x}}_J$ , where  $J$  is the index of the last iteration. Based on the restricted isometry constant  $\delta_{4S}$  of  $\Phi$  (see (21)), the following result on the performance of CoSaMP was stated in [38, Theorem A].

*For a given  $S$ , assume that the  $4S$ -restricted isometry constant of  $\Phi$  satisfies  $\delta_{4S} \leq 0.1$ . Let  $\mathbf{x} \in \mathbb{C}^M$  (not necessarily sparse) and  $\mathbf{y} = \Phi \mathbf{x} + \mathbf{z}$ . For a given precision parameter  $\xi$ , CoSaMP produces an  $S$ -sparse approximation  $\hat{\mathbf{x}}$  that satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq C \max \left\{ \xi, \frac{\|\mathbf{x} - \mathbf{x}_{S/2}\|_1}{\sqrt{S}} + \|\mathbf{z}\|_2 \right\}, \quad (25)$$

*where  $\mathbf{x}_{S/2} \in \mathbb{C}^M$  contains the  $S/2$  components of  $\mathbf{x}$  with largest absolute values, the remaining  $M - S/2$  components being zero (thus,  $\mathbf{x}_{S/2}$  is the best  $S/2$ -sparse approximation to  $\mathbf{x}$ ). The computational complexity is  $\mathcal{O}(B \log(\|\mathbf{x}\|_2/\xi))$ , where  $B$  is a bound on the cost of a matrix-vector multiplication with  $\Phi$  or  $\Phi^*$ .*

The condition  $\delta_{4S} \leq 0.1$  is satisfied if the measurement matrix  $\Phi$  is constructed from a unitary matrix as it was described for BP in Section 3.1, provided that (24) is satisfied with an appropriately chosen  $\delta$ . The computational complexity of CoSaMP is slightly lower than that of OMP in practice.

### 3.4 Discussion

BP and OMP are probably the most popular recovery algorithms in the CS literature. Whereas for BP theoretical performance guarantees are available, OMP lacks similar results. On the other hand, OMP allows a faster implementation, and simulation results even demonstrate a better recovery performance. Low computational complexity is very important for our application, since the channel has to be estimated in real time during data transmission. The third recovery algorithm we are considering, CoSaMP, allows an even faster implementation than OMP. Using an efficient implementation of the pseudoinverse by means of the LSQR algorithm [55], we observed a running time that was only half or even less than half that of OMP, and a performance that was only slightly poorer than that of OMP. A major advantage of CoSaMP is the existence of performance bounds analogous to those for BP. Hence, CoSaMP offers arguably the best compromise between low complexity, good practical performance, and provable performance guarantees.

While there are thus good reasons for using BP, OMP, and CoSaMP for sparse channel estimation, many other algorithms exist, such as simple *thresholding* [56], the *stagewise OMP* [57], the *LARS method* [58, 59], etc. In [29, 30], application of the *Dantzig selector* (DS) [34] to sparse channel estimation was proposed. DS is especially interesting when the noise vector  $\mathbf{z}$  is modeled as random, because it satisfies optimal asymptotic performance bounds in that case. However, for the practically relevant case of finite (moderate)  $Q$  and  $M$ , the performance of DS is not necessarily superior. Furthermore, DS assumes knowledge of the noise statistics, and thus requires an additional estimation step. In our experiments, we did not observe any performance or complexity advantages of DS over BP.

## 4 Compressive Channel Estimation

In this section, we first analyze the sparsity of the channel's delay-Doppler representation for a simple time-varying multipath channel model. Motivated by this analysis, we then propose CS-based channel estimators. We also discuss implications regarding the number and time-frequency positions of the pilots.

### 4.1 Delay-Doppler Sparsity

We assume that the doubly selective wireless channel comprises  $P$  propagation paths corresponding to  $P$  specular (point) scatterers with fixed delays  $\tau_p$  and Doppler frequency shifts  $\nu_p$  for  $p = 1, \dots, P$ . This simple model is often a good approximation to real mobile radio channels [23, 60]. The channel impulse response thus has the form

$$h(t, \tau) = \sum_{p=1}^P \eta_p \delta(\tau - \tau_p) e^{j2\pi\nu_p t}, \quad (26)$$

where  $\eta_p$  characterizes the attenuation and initial phase of the  $p$ th propagation path and  $\delta(\cdot)$  denotes the Dirac delta. The discrete-time impulse response (7) then becomes

$$\begin{aligned} h[n, m] &= \sum_{p=1}^P \eta_p e^{j2\pi\nu_p n T_s} \int_{-\infty}^{\infty} e^{j2\pi\nu_p t} f_1(t - \tau_p + m T_s) f_2^*(t) dt \\ &= \sum_{p=1}^P \eta_p e^{j2\pi\nu_p n T_s} \phi^{(\nu_p)}\left(m - \frac{\tau_p}{T_s}\right), \end{aligned} \quad (27)$$

with

$$\phi^{(\nu)}(x) \triangleq \int_{-\infty}^{\infty} e^{j2\pi\nu t} f_1(t + T_s x) f_2^*(t) dt.$$

Furthermore, inserting (27) into (11), we obtain for the discrete-delay-Doppler spreading function

$$S_h[m, i] = \frac{1}{N_r} \sum_{p=1}^P \eta_p \phi^{(\nu_p)}\left(m - \frac{\tau_p}{T_s}\right) \sum_{n=0}^{N_r-1} e^{j2\pi(\nu_p T_s - \frac{i}{N_r})n}.$$

Applying the geometric sum formula, this becomes

$$S_h[m, i] = \sum_{p=1}^P \eta_p e^{j\pi(\nu_p T_s - \frac{i}{N_r})(N_r-1)} \Lambda^{(\nu_p)}\left(m - \frac{\tau_p}{T_s}, i - \nu_p T_s N_r\right), \quad (28)$$

with

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

where

$$\psi(y) \triangleq \frac{1}{N_r} e^{j\pi \frac{y}{N_r}(N_r-1)} \sum_{n=0}^{N_r-1} e^{-j2\pi \frac{y}{N_r} n} = \frac{\sin(\pi y)}{N_r \sin(\pi y/N_r)}. \quad (29)$$

The function  $\Lambda^{(\nu)}(x, y) = \phi^{(\nu)}(x) \psi(y)$  describes the *leakage effect* that is due to the finite transmit bandwidth ( $\approx 1/T_s$ ) and the finite blocklength ( $N_r \approx LN$ ).

We next investigate the sparsity of  $S_h[m, i]$ . In view of expression (28), this essentially amounts to studying the sparsity of  $\Lambda^{(\nu_p)}(m - \tau_p/T_s, i - \nu_p T_s N_r) = \phi^{(\nu_p)}(m - \tau_p/T_s) \psi(i - \nu_p T_s N_r)$ . To this end, we first consider the energy of those samples of  $\phi^{(\nu_p)}(m - \tau_p/T_s)$  whose distance from  $\tau_p/T_s$  is greater than  $\Delta m \in \{1, 2, \dots\}$ , i.e.,  $|m - \tau_p/T_s| > \Delta m$ . We assume that  $\phi^{(\nu)}(x)$  exhibits a polynomial decay,<sup>4</sup> i.e.,  $|\phi^{(\nu)}(x)| \leq C(1 + |x/x_0|)^{-s}$  with  $s \geq 1$ , for some positive constants  $C$  and  $x_0$ . We can then develop the following bound on the energy of all  $\phi^{(\nu_p)}(m - \tau_p/T_s)$  with  $|m - \tau_p/T_s| > \Delta m$ :

$$\begin{aligned} \sum_{|m - \tau_p/T_s| > \Delta m} \left| \phi^{(\nu_p)}\left(m - \frac{\tau_p}{T_s}\right) \right|^2 &\leq C^2 \sum_{|m - \tau_p/T_s| > \Delta m} \left(1 + \left|\frac{m - \tau_p/T_s}{x_0}\right|\right)^{-2s} \\ &\leq 2C^2 \sum_{m = \Delta m}^{\infty} \left(1 + \frac{m}{x_0}\right)^{-2s} \\ &\leq 2C^2 \int_{\Delta m-1}^{\infty} \left(1 + \frac{x}{x_0}\right)^{-2s} dx \\ &= \frac{2C^2 x_0}{2s-1} \left(1 + \frac{\Delta m - 1}{x_0}\right)^{-2s+1}. \end{aligned}$$

This shows that the energy of  $\phi^{(\nu_p)}(m - \tau_p/T_s)$  outside the interval  $[\tau_p/T_s - \Delta m, \tau_p/T_s + \Delta m]$  decays polynomially of order  $2s - 1$  with respect to  $\Delta m$ .

In a similar manner, we consider the energy of those samples of  $\psi(i - \nu_p T_s N_r)$  whose distance (up to the modulo- $N_r$  operation, see below) from  $\nu_p T_s N_r$  is greater than  $\Delta i \in \{2, \dots, N_r/2\}$  ( $N_r$  is assumed even). Let  $\mathcal{I}$  denote the set  $\{0, \dots, N_r-1\}$  with the exception of all  $i = i_{\mathbb{Z}} \bmod N_r$ , where  $i_{\mathbb{Z}}$  is any integer with  $|i_{\mathbb{Z}} - \nu_p T_s N_r| \leq \Delta i$ . From (29), we then obtain the bound

---

<sup>4</sup>This includes the following important special cases: (i) the ideal lowpass filter, i.e.,  $f_1(t) = f_2(t) = \sqrt{1/T_s} \text{sinc}(t/T_s)$  with  $\text{sinc}(x) \triangleq \frac{\sin(\pi x)}{\pi x}$ ; here  $s = 1$ ; and (ii) 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  $\rho$ , then, for  $\nu$  not too large,  $\phi^{(\nu)}(x) \approx \text{sinc}(x) \cos(\rho\pi x)/[1 - (2\rho x)^2]$  and  $s = 3$ .

$$\begin{aligned}
\sum_{i \in \mathcal{I}} |\psi(i - \nu_p T_s N_r)|^2 &\leq \frac{2}{N_r^2} \sum_{i=\Delta i}^{N_r/2} \frac{1}{\sin^2(\frac{\pi}{N_r} i)} \\
&\leq \frac{2}{N_r^2} \int_{\Delta i-1}^{N_r/2} \frac{dx}{\sin^2(\frac{\pi}{N_r} x)} \\
&= \frac{2}{N_r \pi} \cot\left(\frac{\pi}{N_r} (\Delta i - 1)\right) \\
&\leq \frac{1}{\pi(\Delta i - 1)},
\end{aligned}$$

where we have used  $\sin^2 x \leq 1$ , some monotonicity arguments, and the fact that  $\cot x \leq \frac{\pi}{2x}$  within  $[0, \frac{\pi}{2}]$ . The above bound shows that the energy of  $\psi(i - \nu_p T_s N_r)$  outside the interval  $[\lfloor \nu_p T_s N_r - \Delta i \rfloor, \lceil \nu_p T_s N_r + \Delta i \rceil]$  (modulo  $N_r$ ) decays linearly (polynomially of order 1) with respect to  $\Delta i$ .

From these decay results, it follows that  $\Lambda^{(\nu_p)}(m - \tau_p/T_s, i - \nu_p T_s N_r) = \phi^{(\nu_p)}(m - \tau_p/T_s) \psi(i - \nu_p T_s N_r)$  can be considered as an *approximately sparse* (or *compressible*, in CS terminology [35]) function. Thus, as an approximation, we can model  $\Lambda^{(\nu_p)}(m - \tau_p/T_s, i - \nu_p T_s N_r)$  as  $N_\Lambda$ -sparse, i.e., at most  $N_\Lambda$  values of  $\Lambda^{(\nu_p)}(m - \tau_p/T_s, i - \nu_p T_s N_r)$  are nonzero, with an appropriately chosen sparsity parameter  $N_\Lambda$ . It then follows from (28) that  $S_h[m, i]$  is  $PN_\Lambda$ -sparse, and the same is true for  $F[m, i]$  in (13) and for  $\tilde{F}[m, i]$  in (15) (in the respective fundamental  $i$  period). In the next subsection, we propose a CS-based channel estimator that exploits this sparsity. Unfortunately,  $N_\Lambda$  cannot be chosen extremely small because of the strong leakage due to the slowly (only linearly) decaying factor  $\psi(i - \nu_p T_s N_r)$ . In Section 5, this limitation will motivate the introduction of a sparsity-improving basis expansion.

We emphasize that the channel model (26) is only used for motivating our CS-based channel estimators via the sparsity arguments described above; it is not used in the formulation of the estimation algorithms. Thus, our estimators are not restricted to the channel model (26); all they require is the approximate sparsity of  $\tilde{F}[m, i]$ .

## 4.2 CS-based Channel Estimators

We now combine pilot-assisted channel estimation with CS-based signal reconstruction. Our starting-point is the 2-D DFT relation (17), repeated here for convenience:

$$H_{l' \Delta L, k' \Delta K} = \sum_{m=0}^{D-1} \sum_{i=-I/2}^{I/2-1} \tilde{F}[m, i] e^{-j2\pi(\frac{k'm}{D} - \frac{l'i}{T})}, \quad l' = 0, \dots, I-1, \quad k' = 0, \dots, D-1. \quad (30)$$

This can be written as the 2-D expansion

$$H_{l' \Delta L, k' \Delta K} = \sum_{m=0}^{D-1} \sum_{i=-I/2}^{I/2-1} \alpha_{m,i} u_{m,i}[l', k'], \quad (31)$$

with

$$\alpha_{m,i} \triangleq \sqrt{ID} \tilde{F}[m,i], \quad u_{m,i}[l',k'] \triangleq \frac{1}{\sqrt{ID}} e^{-j2\pi(\frac{k'm}{D} - \frac{l'i}{T})}. \quad (32)$$

The functions  $H_{l'\Delta L, k'\Delta K}$  and  $u_{m,i}[l',k']$  are both defined for  $l' = 0, \dots, I-1$  and  $k' = 0, \dots, D-1$ ; we may thus consider them as  $I \times D$  matrices. Let  $\mathbf{h} \triangleq \text{vec}\{H_{l'\Delta L, k'\Delta K}\}$  and  $\mathbf{u}_{m,i} \triangleq \text{vec}\{u_{m,i}[l',k']\}$  denote the vectors of length  $ID$  obtained by stacking all columns of these matrices (e.g.,  $\mathbf{h} = [h_1 \dots h_{ID}]^T$  with  $h_{k'I+l'+1} = H_{l'\Delta L, k'\Delta K}$ ). We can then rewrite (31) as

$$\mathbf{h} = \sum_{m=0}^{D-1} \sum_{i=-I/2}^{I/2-1} \alpha_{m,i} \mathbf{u}_{m,i} = \mathbf{U} \boldsymbol{\alpha}, \quad (33)$$

where  $\boldsymbol{\alpha} \triangleq \text{vec}\{\alpha_{m,i}\}$  and  $\mathbf{U}$  is the  $ID \times ID$  matrix whose  $((i + I/2)D + m + 1)$ 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 2.3, there are  $|\mathcal{P}|$  pilot symbols at time-frequency positions  $(l, k) \in \mathcal{P}$ . Thus,  $|\mathcal{P}|$  of the  $ID$  entries of the channel vector  $\mathbf{h}$  are given by the channel coefficients  $H_{l,k}$  at the pilot positions  $(l, k) \in \mathcal{P}$ . Let  $\mathbf{h}^{(p)}$  denote the corresponding length- $|\mathcal{P}|$  subvector of  $\mathbf{h}$ , and let  $\mathbf{U}^{(p)}$  denote the  $|\mathcal{P}| \times ID$  submatrix of  $\mathbf{U}$  constituted by the corresponding  $|\mathcal{P}|$  rows of  $\mathbf{U}$ . Reducing (33) to the pilot positions, we obtain

$$\mathbf{h}^{(p)} = \mathbf{U}^{(p)} \boldsymbol{\alpha}. \quad (34)$$

Equivalently,

$$\mathbf{h}^{(p)} = \boldsymbol{\Phi} \mathbf{x}, \quad (35)$$

with

$$\boldsymbol{\Phi} \triangleq \sqrt{\frac{ID}{P}} \mathbf{U}^{(p)} \quad \text{and} \quad \mathbf{x} \triangleq \sqrt{\frac{P}{ID}} \boldsymbol{\alpha}. \quad (36)$$

Note that  $\boldsymbol{\Phi}$  is normalized such that its columns have unit  $\ell_2$ -norm, and that the length- $ID$  vector  $\mathbf{x}$  is, up to a constant factor, the vector form of the function  $\tilde{F}[m,i]$  (on the fundamental  $i$  period).

Our task is to estimate  $\mathbf{x}$  based on relation (35). The vector  $\mathbf{h}^{(p)}$  is unknown, but estimates of its entries—the channel coefficient estimates  $\hat{H}_{l,k} = r_{l,k}/p_{l,k}$  at the pilot positions  $(l, k) \in \mathcal{P}$ , see (18)—are available. Therefore, let us replace  $\mathbf{h}^{(p)}$  by the corresponding vector of estimates  $\hat{H}_{l,k}|_{(l,k) \in \mathcal{P}}$ . For consistency with the notation used in Section 3, this latter vector will be denoted as  $\mathbf{y}$  (rather than  $\hat{\mathbf{h}}^{(p)}$ ). According to (18),

$$\mathbf{y} = \mathbf{h}^{(p)} + \mathbf{z}, \quad (37)$$

where  $\mathbf{z}$  is the vector of noise terms  $\tilde{z}_{l,k}/p_{l,k}|_{(l,k) \in \mathcal{P}}$ . Inserting (35) into (37), we finally obtain the “measurement equation”

$$\mathbf{y} = \boldsymbol{\Phi} \mathbf{x} + \mathbf{z}. \quad (38)$$



According to Section 4.1,  $\mathbf{x}$  is modeled as  $PN_\Lambda$ -sparse. Thus, (38) is seen to be a sparse reconstruction problem of the form (19), with dimensions  $M = \dim\{\mathbf{x}\} = ID$  and  $Q = \dim\{\mathbf{y}\} = |\mathcal{P}|$  and sparsity  $S = PN_\Lambda$ . We can hence use the CS recovery techniques reviewed in Section 3 to obtain an estimate of  $\mathbf{x}$  or, equivalently, of  $\boldsymbol{\alpha} = \sqrt{ID/P} \mathbf{x}$  or of  $\tilde{F}[m, i] = \alpha_{m,i}/\sqrt{ID}$ . From the estimate of  $\tilde{F}[m, i]$ , estimates of all channel coefficients  $H_{l,k}$  are finally obtained via (14).

### 4.3 Measurement Matrix and Pilots

Next, we discuss the construction of the measurement matrix  $\Phi$  and some implications regarding the number and positions of the pilots. According to (36),  $\Phi$  is constructed by selecting  $|\mathcal{P}|$  rows of the unitary  $ID \times ID$  matrix  $\mathbf{U}$  and normalizing the resulting columns. This agrees with the construction described in Sections 3.1 and 3.3 in the context of BP and CoSaMP, respectively. 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  $\{1, \dots, ID\}$  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  $\{1, \dots, ID\}$  of  $\mathbf{h}$  are selected uniformly at random.

For BP and CoSaMP, in order to achieve a good approximation quality in the sense of (23) and (25), the number of pilots should satisfy condition (24). In our case, this condition becomes

$$|\mathcal{P}| \geq C_3 \delta^{-2} (\ln(ID))^4 PN_\Lambda \ln(1/\eta), \quad (39)$$

with an appropriately chosen  $\delta$  (note that  $\mu_{\mathbf{U}} = 1$ ). This bound suggests that the required number of pilots scales only linearly with the number of channel paths (scatterers)  $P$  and the sparsity parameter  $N_\Lambda$ , and poly-logarithmically with the system design parameters  $I$  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. If the number of pilots is sufficiently large in the sense of (39), then with high probability, this random choice of the pilot positions will yield good estimation performance for arbitrary channels with at most  $P$  paths. This will be verified experimentally in Section 7.

## 5 Sparsity-Improving Basis Expansion

The compressive channel estimation scheme presented in the previous section was based on the 2-D DFT relation (30), which is an expansion of the subsampled channel coefficients  $H_{l'\Delta L, k'\Delta K}$  in terms of the 2-D DFT basis  $u_{m,i}[l', k'] = \frac{1}{\sqrt{ID}} e^{-j2\pi(\frac{k'm}{D} - \frac{l'i}{T})}$  (see (32)). The expansion coefficients,  $\alpha_{m,i} =$

$\sqrt{ID} \tilde{F}[m, i]$ , were shown in Section 4.1 to be approximately sparse, and this property was utilized by the compressive channel estimator. However, the sparsity is limited by the slowly (only linearly) decaying factor  $\psi(i - \nu_p T_s N_r)$  of the function  $\Lambda^{(\nu_p)}(m - \tau_p/T_s, i - \nu_p T_s N_r) = \phi^{(\nu_p)}(m - \tau_p/T_s) \psi(i - \nu_p T_s N_r)$  in (28).

## 5.1 2-D and 1-D Basis Expansions

In order to improve the sparsity of the expansion coefficients, we introduce a generalized 2-D expansion of the subsampled channel coefficients  $H_{l' \Delta L, k' \Delta K}$  into orthonormal basis functions  $v_{m,i}[l', k']$ :

$$H_{l' \Delta L, k' \Delta K} = \sum_{m=0}^{D-1} \sum_{i=-I/2}^{I/2-1} \beta_{m,i} v_{m,i}[l', k'], \quad l' = 0, \dots, I-1, \quad k' = 0, \dots, D-1. \quad (40)$$

Clearly, our previous 2-D DFT expansion (30), (31) is a special case of (40). The advantage of the generalized basis expansion (40) is the possibility of using a basis  $\{v_{m,i}[l', k']\}$  for which the coefficients  $\beta_{m,i}$  are sparser than our previous coefficients  $\alpha_{m,i} = \sqrt{ID} \tilde{F}[m, i]$ .

We will choose a 2-D basis that is adapted to the channel model (26) (but not to the specific channel parameters  $P$ ,  $\eta_p$ ,  $\tau_p$ , and  $\nu_p$  involved in (26)); this choice will be seen to reduce to the choice of a family of 1-D bases. Equation (26) suggests that the coefficients  $\beta_{m,i}$  should be sparse for the elementary single-scatterer channel

$$h^{(\tau_1, \nu_1)}(t, \tau) \triangleq \delta(\tau - \tau_1) e^{j2\pi \nu_1 t}, \quad (41)$$

for all possible choices of  $\tau_1 \in [0, \tau_{\max}]$  and  $\nu_1 \in [-\nu_{\max}, \nu_{\max}]$ . Specializing (28) to  $P = 1$  and  $\eta_1 = 1$ , and using (13) and (15), we obtain

$$\tilde{F}[m, i] = \sum_{q=0}^{N-1} e^{j\pi(\nu_1 T_s - \frac{i+qL}{N_r})(N_r-1)} \phi^{(\nu_1)}\left(m - \frac{\tau_1}{T_s}\right) \psi(i + qL - \nu_1 T_s N_r) A_{\gamma,g}^*\left(m, \frac{i+qL}{N_r}\right). \quad (42)$$

Inserting (42) into the 2-D DFT expansion (30) yields

$$H_{l' \Delta L, k' \Delta K} = \sum_{m=0}^{D-1} \phi^{(\nu_1)}\left(m - \frac{\tau_1}{T_s}\right) C^{(\nu_1)}[m, l'] e^{-j2\pi \frac{k'm}{D}}. \quad (43)$$

Here, we have set

$$C^{(\nu_1)}[m, l'] \triangleq \sum_{i=-I/2}^{I/2-1} \tilde{\alpha}_{m,i}^{(\nu_1)} \frac{1}{\sqrt{I}} e^{j2\pi \frac{l'i}{I}}, \quad (44)$$

where

$$\tilde{\alpha}_{m,i}^{(\nu_1)} \triangleq \sqrt{I} \sum_{q=0}^{N-1} \psi^{(\nu_1)}(i + qL) A_{\gamma,g}^*\left(m, \frac{i+qL}{N_r}\right), \quad (45)$$

with

$$\psi^{(\nu_1)}(i) \triangleq e^{j\pi(\nu_1 T_s - \frac{i}{N_r})(N_r-1)} \psi(i - \nu_1 T_s N_r). \quad (46)$$

According to (45) and (46), 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 (44) by a general 1-D basis expansion

$$C^{(\nu_1)}[m, l'] = \sum_{i=-I/2}^{I/2-1} \tilde{\beta}_{m,i}^{(\nu_1)} b_{m,i}[l'], \quad m = 0, \dots, D-1, \quad l' = 0, \dots, I-1, \quad (47)$$

with a family of bases  $\{b_{m,i}[l']\}_{i=-I/2, \dots, I/2-1}$ ,  $m = 0, \dots, D-1$  that are orthonormal (i.e.,  $\sum_{l'=0}^{I-1} b_{m,i_1}[l'] b_{m,i_2}^*[l'] = \delta[i_1 - i_2]$  for all  $m$ ) and do not depend on the value of  $\nu_1$  in  $C^{(\nu_1)}[m, l']$ . The idea is to choose the 1-D basis  $\{b_{m,i}[l']\}_{i=-I/2, \dots, I/2-1}$  such that the coefficient vector  $[\tilde{\beta}_{m,-I/2}^{(\nu_1)} \cdots \tilde{\beta}_{m,I/2-1}^{(\nu_1)}]^T$  is sparse for all  $m$  and all  $\nu_1 \in [-\nu_{\max}, \nu_{\max}]$ . Substituting (47) back into (43), we obtain

$$H_{l'\Delta L, k'\Delta K} = \sum_{m=0}^{D-1} \sum_{i=-I/2}^{I/2-1} \phi^{(\nu_1)}\left(m - \frac{\tau_1}{T_s}\right) \tilde{\beta}_{m,i}^{(\nu_1)} b_{m,i}[l'] e^{-j2\pi \frac{k'm}{D}}.$$

This can now be identified with the 2-D basis expansion (40), i.e.,  $H_{l'\Delta L, k'\Delta K} = \sum_{m=0}^{D-1} \sum_{i=-I/2}^{I/2-1} \beta_{m,i}^{(\tau_1, \nu_1)} v_{m,i}[l', k']$ , with the orthonormal 2-D basis

$$v_{m,i}[l', k'] \triangleq \frac{1}{\sqrt{D}} b_{m,i}[l'] e^{-j2\pi \frac{k'm}{D}} \quad (48)$$

and the 2-D coefficients

$$\beta_{m,i}^{(\tau_1, \nu_1)} \triangleq \sqrt{D} \phi^{(\nu_1)}\left(m - \frac{\tau_1}{T_s}\right) \tilde{\beta}_{m,i}^{(\nu_1)}.$$

The basis functions  $v_{m,i}[l', k']$  are seen to agree with our previous 2-D DFT basis functions  $u_{m,i}[l', k'] = \frac{1}{\sqrt{ID}} e^{-j2\pi(\frac{k'm}{D} - \frac{l'i}{I})}$  in (32) with respect to  $k'$ , but they are different with respect to  $l'$  because  $\frac{1}{\sqrt{I}} e^{j2\pi \frac{l'i}{I}}$  is replaced by  $b_{m,i}[l']$ . 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 (44) that were based on the 1-D DFT basis  $\left\{ \frac{1}{\sqrt{I}} e^{j2\pi \frac{l'i}{I}} \right\}$ .

The above considerations were based on the elementary single-scatterer channel (41), but they can be immediately extended to the multiple-scatterer case. When the channel comprises  $P$  scatterers as in (26), 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  $\{v_{m,i}[l', k']\}$  does not depend on the channel parameters  $P$ ,  $\eta_p$ ,  $\tau_p$ , and  $\nu_p$ . While it was motivated by (26), its formulation does not explicitly use that channel model. An optimal design of the 1-D basis  $\{b_{m,i}[l']\}_{i=-I/2, \dots, I/2-1}$  will be presented in Section 6.

We note that the use of the 2-D basis  $v_{m,i}[l', k']$  in (48) comes at the cost of an increased computational complexity compared with the 2-D DFT basis  $u_{m,i}[l', k']$ , for both applying the CS recovery algorithms and for computing the channel coefficients on the subsampled grid. This is because efficient FFT algorithms can only be applied with respect to  $k'$ , but not with respect to  $l'$ . However, if  $I$  is not too large (in our simulations  $I = 16$ , see Section 7.1), the additional complexity is small.

## 5.2 CS-based Channel Estimators Using the Basis Expansion

A CS-based channel estimation scheme that uses the generalized basis expansion (40) can now be developed as in Section 4.2. We can write (40) in matrix form as (cf. (33))

$$\mathbf{h} = \mathbf{V}\boldsymbol{\beta}, \quad (49)$$

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 4.2. Reducing (49) to the pilot positions yields (cf. (34), (35))

$$\mathbf{h}^{(\text{p})} = \mathbf{V}^{(\text{p})}\boldsymbol{\beta} = \boldsymbol{\Phi}\mathbf{x},$$

with  $\boldsymbol{\Phi} \triangleq \mathbf{V}^{(\text{p})}\mathbf{D}$  and  $\mathbf{x} \triangleq \mathbf{D}^{-1}\boldsymbol{\beta}$  (cf. (36)), where the diagonal matrix  $\mathbf{D}$  is chosen such that all columns of  $\boldsymbol{\Phi}$  have unit  $\ell_2$ -norm. Finally, we replace the unknown vector  $\mathbf{h}^{(\text{p})}$  by its pilot-based estimate, again denoted as  $\mathbf{y}$ . Using (18), we then obtain the “measurement equation” (cf. (38))

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{z}, \quad (50)$$

where  $\mathbf{z}$  is again the vector with entries  $\tilde{z}_{l,k}/p_{l,k}|_{(l,k) \in \mathcal{P}}$ . As in Section 4.2, our task is to recover the length- $ID$  vector  $\mathbf{x}$  from the known length- $|\mathcal{P}|$  vector  $\mathbf{y}$ , based on the measurement equation (50). From the resulting estimate of  $\mathbf{x}$ , estimates of all channel coefficients  $H_{l,k}$  are finally obtained as explained in Section 4.2. As discussed further above, we can expect  $\boldsymbol{\beta}$  to be (approximately) sparse provided the 1-D basis  $\{b_{m,i}[l']\}$  is chosen appropriately. Then  $\mathbf{x}$  is sparse as well, and our channel estimation problem is again recognized to be a sparse reconstruction problem of the form (19), with dimensions  $M = \dim\{\mathbf{x}\} = ID$  and  $Q = \dim\{\mathbf{y}\} = |\mathcal{P}|$ . We can thus use the CS recovery techniques summarized in Section 3 to obtain an estimate of  $\mathbf{x}$ .

As in Section 4.3, to be consistent with the CS framework of Sections 3.1 and 3.3, we select the pilot positions uniformly at random within the subsampled time-frequency grid  $\mathcal{G}$ . For BP and CoSaMP, to achieve a good approximation in the sense of (23) and (25), the number of pilots should satisfy condition (24), i.e.,

$$|\mathcal{P}| \geq C_3 \delta^{-2} (\ln(ID))^4 \mu_{\mathbf{V}}^2 S \ln(1/\eta),$$

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  $\{v_{m,i}[l',k']\}$ ; furthermore,  $\mu_{\mathbf{V}} \geq 1$  (for the DFT basis, we had  $\mu_{\mathbf{U}} = 1$ ). Thus, the potentially better sparsity  $S$  comes at the cost of an increased coherence. For improved performance, the gain due to the better sparsity should be larger than the loss due to the larger coherence.

## 6 Basis Optimization

We now discuss the optimal design of the 1-D basis functions  $\{b_{m,i}[l']\}_{i=-I/2,\dots,I/2-1}$  in the general 1-D expansion (47). We will propose a framework and, subsequently, an algorithm for basis optimization.

### 6.1 Basis Optimization Framework

Ideally, the  $m$  orthonormal 1-D bases  $\{b_{m,i}[l']\}_{i=-I/2,\dots,I/2-1}$ ,  $m = 0, \dots, D-1$  should be such that the coefficient vectors  $[\tilde{\beta}_{m,-I/2}^{(\nu)} \cdots \tilde{\beta}_{m,I/2-1}^{(\nu)}]^T$  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  $\nu_{\Delta}$ .

Regarding the choice of  $\nu_{\Delta}$ , it is interesting to note that for the “canonical spacing” given by  $\nu_{\Delta} = 1/(T_s N_r)$ , the coefficients  $\tilde{\alpha}_{m,i}^{(\nu_{\Delta}d)}$  in the 1-D DFT expansion (44) are 1-sparse with respect to  $i$ . Indeed, (46) here simplifies to

$$\psi^{(\nu_{\Delta}d)}(i) = e^{j\pi(d-i)\frac{N_r-1}{N_r}} \psi(i-d) = \delta_{N_r}[i-d],$$

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 (45) then reduces to

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

where  $\tilde{d}$  depends on  $d$  but not on  $i$ . Thus, for  $\nu_{\Delta} = 1/(T_s N_r)$ , the coefficients obtained using the 1-D DFT basis  $\{b_{m,i}[l'] = \frac{1}{\sqrt{I}} e^{j2\pi \frac{l'i}{I}}\}$  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.,  $\nu_{\Delta} = 1/(2T_s N_r)$ . In this case,  $\mathcal{D}$  includes also the Doppler frequencies located midway between any two adjacent canonical sampling points, for which the DFT basis results in maximum leakage (these frequencies are given by  $\nu_{\Delta}d$  for odd  $d$ ).

Because the basis  $\{b_{m,i}[l']\}$  is orthonormal, the expansion coefficients defined by (47) can be calculated as the inner products

$$\tilde{\beta}_{m,i}^{(\nu)} = \sum_{l'=0}^{I-1} C^{(\nu)}[m, l'] b_{m,i}^*[l'], \quad i = -I/2, \dots, I/2-1.$$

This can be rewritten as

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

with the length- $I$  vectors  $\tilde{\boldsymbol{\beta}}_m^{(\nu)} \triangleq [\tilde{\beta}_{m,-I/2}^{(\nu)} \cdots \tilde{\beta}_{m,I/2-1}^{(\nu)}]^T$  and  $\boldsymbol{\gamma}_m^{(\nu)} \triangleq [C^{(\nu)}[m, -I/2] \cdots C^{(\nu)}[m, I/2-1]]^T$  and the unitary  $I \times I$  matrix  $\mathbf{B}_m$  with elements  $(\mathbf{B}_m)_{i+1,l'+1} = b_{m,i-I/2}^*[l']$ . We can now state the basis optimization problem as follows.

*For given vectors  $\boldsymbol{\gamma}_m^{(\nu)}$ ,  $m = 0, \dots, D-1$ , with  $\boldsymbol{\gamma}_m^{(\nu)}$  defined as described above, find  $I \times I$  unitary matrices  $\mathbf{B}_m$  not dependent on  $\nu$  such that the vectors  $\tilde{\boldsymbol{\beta}}_m^{(\nu)} = \mathbf{B}_m \boldsymbol{\gamma}_m^{(\nu)}$  are maximally sparse for all  $\nu \in \mathcal{D}$ .*

As is usual in the CS framework, we measure the sparsity of  $\tilde{\boldsymbol{\beta}}_m^{(\nu)}$  by the  $\ell_1$ -norm or, more precisely, by the  $\ell_1$ -norm averaged over all  $\nu \in \mathcal{D}$ , i.e.,

$$\frac{1}{|\mathcal{D}|} \sum_{\nu \in \mathcal{D}} \|\tilde{\boldsymbol{\beta}}_m^{(\nu)}\|_1 = \frac{1}{|\mathcal{D}|} \sum_{\nu \in \mathcal{D}} \|\mathbf{B}_m \boldsymbol{\gamma}_m^{(\nu)}\|_1.$$

Thus, our basis optimization problem is formulated as the  $D$  constrained minimization problems<sup>5</sup>

$$\hat{\mathbf{B}}_m = \arg \min_{\mathbf{B}_m \in \mathcal{U}} \sum_{\nu \in \mathcal{D}} \|\mathbf{B}_m \boldsymbol{\gamma}_m^{(\nu)}\|_1, \quad m = 0, \dots, D-1, \quad (51)$$

where  $\mathcal{U}$  denotes the set of all unitary  $I \times I$  matrices. Note that the vectors  $\boldsymbol{\gamma}_m^{(\nu)}$  are known because they follow from the function  $C^{(\nu)}[m, l']$ , which is given by (see (44), (45), (46))

$$C^{(\nu)}[m, l'] = \sum_{i=-I/2}^{I/2-1} \sum_{q=0}^{N-1} \psi^{(\nu)}(i + qL) A_{\gamma,g}^* \left( m, \frac{i + qL}{N_r} \right) e^{j2\pi \frac{l' i}{T}}, \quad (52)$$

with  $\psi^{(\nu)}(i) = e^{j\pi(\nu T_s - \frac{i}{N_r})(N_r-1)} \psi(i - \nu T_s N_r)$ . It is seen that the optimal bases characterized by the matrices  $\hat{\mathbf{B}}_m$  depend on the pulses  $g[n]$ ,  $\gamma[n]$  and the maximum Doppler  $\nu_{\max}$  (via the definition of  $\mathcal{D}$ ). They do not depend on any other channel characteristics.

## 6.2 Basis Optimization Algorithm

Let us consider the minimization problem (51) for a fixed delay index  $m$ . Unfortunately, this problem is nonconvex (since  $\mathcal{U}$  is not a convex set), so we cannot use standard convex optimization techniques. We therefore propose an approximate iterative algorithm that relies on the following facts [61]. (i) Every unitary  $I \times I$  matrix  $\mathbf{B}$  can be represented in terms of a Hermitian  $I \times I$  matrix  $\mathbf{A}$  as

$$\mathbf{B} = e^{j\mathbf{A}}.$$

---

<sup>5</sup>We note that the optimization problem (51) is similar to *dictionary learning* problems that have recently been considered in CS theory [39–41]. In [41], conditions for the local identifiability of orthonormal bases by means of  $\ell_1$  minimization are derived.

(ii) The matrix exponential  $\mathbf{B} = e^{j\mathbf{A}}$  can be approximated by its first-order Taylor expansion, i.e.,

$$\mathbf{B} \approx \mathbf{I}_I + j\mathbf{A}, \quad (53)$$

where  $\mathbf{I}_I$  is the  $I \times I$  identity matrix. Even though  $\mathbf{I}_I + j\mathbf{A}$  is not a unitary matrix, the approximation (53) will be good if  $\mathbf{A}$  is small. 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 (53) 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  $r$ th 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 (51), i.e., by minimizing

$$\sum_{\nu \in \mathcal{D}} \|\mathbf{B}_m^{(r+1)} \boldsymbol{\gamma}_m^{(\nu)}\|_1 = \sum_{\nu \in \mathcal{D}} \|e^{j\mathbf{A}_m^{(r)}} \mathbf{B}_m^{(r)} \boldsymbol{\gamma}_m^{(\nu)}\|_1.$$

Since this problem is still nonconvex, we use the approximation (53), and thus the final minimization problem at the  $r$ th iteration is

$$\hat{\mathbf{A}}_m^{(r)} = \arg \min_{\mathbf{A} \in \mathcal{A}_r} \sum_{\nu \in \mathcal{D}} \|(\mathbf{I}_I + j\mathbf{A}) \mathbf{B}_m^{(r)} \boldsymbol{\gamma}_m^{(\nu)}\|_1. \quad (54)$$

Here,  $\mathcal{A}_r$  is the set of all Hermitian  $I \times I$  matrices  $\mathbf{A}$  that are small in the sense that  $\|\mathbf{A}\|_\infty \leq \lambda_r$ , where  $\|\mathbf{A}\|_\infty$  denotes the largest modulus of all elements of  $\mathbf{A}$  and  $\lambda_r$  is a positive constraint level (a small  $\lambda_r$  ensures a good approximation accuracy in (53) and also that the unitary matrix  $e^{j\hat{\mathbf{A}}_m^{(r)}}$  is close to  $\mathbf{I}_I$ ). The problem (54) is convex and thus can be solved by standard convex optimization techniques [47].

The next step at the  $r$ th 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

$$\sum_{\nu \in \mathcal{D}} \|e^{j\hat{\mathbf{A}}_m^{(r)}} \mathbf{B}_m^{(r)} \boldsymbol{\gamma}_m^{(\nu)}\|_1 < \sum_{\nu \in \mathcal{D}} \|\mathbf{B}_m^{(r)} \boldsymbol{\gamma}_m^{(\nu)}\|_1. \quad (55)$$

In the positive case, we actually perform the update of  $\mathbf{B}_m^{(r)}$  and we retain the constraint level  $\lambda_r$  for the next iteration:

$$\mathbf{B}_m^{(r+1)} = e^{j\hat{\mathbf{A}}_m^{(r)}} \mathbf{B}_m^{(r)}, \quad \lambda_{r+1} = \lambda_r.$$

Otherwise, we reject the update of  $\mathbf{B}_m^{(r)}$  and reduce the constraint level  $\lambda_r$ :

$$\mathbf{B}_m^{(r+1)} = \mathbf{B}_m^{(r)}, \quad \lambda_{r+1} = \frac{\lambda_r}{2}.$$

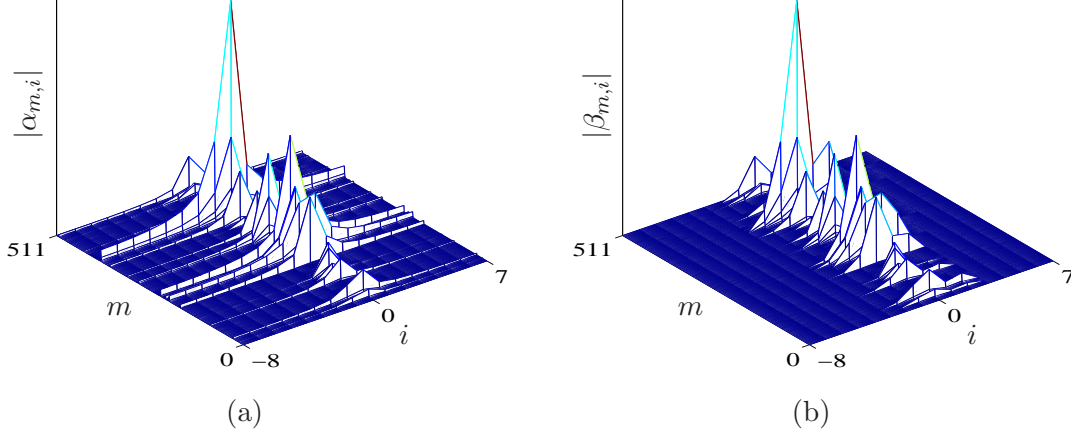


Figure 2: Sparsity improvement obtained with the proposed iterative basis optimization algorithm: (a) Modulus of the expansion coefficients for the DFT basis, (b) modulus of the expansion coefficients for the optimized basis.

This iteration process is terminated either if  $\lambda_r$  falls below a prescribed threshold or if the number of iterations exceeds a certain value. The iteration process is initialized by the  $I \times I$  DFT matrix  $\mathbf{F}_I$ , i.e.,  $\mathbf{B}_m^{(0)} = \mathbf{F}_I$ . As mentioned in Section 4.1, the DFT basis already yields a relatively sparse coefficient vector, and thus it provides a reasonable initial choice. We note that efficient algorithms for computing the matrix exponentials  $e^{j\hat{\mathbf{A}}_m^{(r)}}$  exist [61]. In any case, the bases  $\{b_{m,i}[l']\}$  (or, equivalently, the basis matrices  $\mathbf{B}_m$ ) have to be optimized only once before the actual channel estimation starts; the optimal bases are independent of the received signal and can thus be precomputed.

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  $m = 1, \dots, D - 1$ , so  $C^{(\nu)}[m, l'] = C^{(\nu)}[0, l']$  (see (52)) and thus  $\boldsymbol{\gamma}_m^{(\nu)} = \boldsymbol{\gamma}_0^{(\nu)}$ . Because  $\boldsymbol{\gamma}_m^{(\nu)}$  no longer depends on  $m$ , only one basis  $\mathbf{B}$  (instead of  $D$  different bases  $\mathbf{B}_m$ ,  $m = 0, \dots, D - 1$ ) has to be optimized.

In Fig. 2, we compare the expansion coefficients  $\alpha_{m,i}$  obtained with the DFT basis (see (31), (32)) and the expansion coefficients  $\beta_{m,i}$  obtained with the optimized basis (see (40), (48)) for one channel realization. The system parameters are as in Sections 7.1 and 7.2. The minimization (54) (not  $m$ -dependent, since we are simulating a CP-OFDM system) was carried out using the convex optimization package CVX [62]. It is seen that the basis optimization yields a significant improvement of sparsity.

## 7 Simulation Results

In this section, we assess the performance of our CS-based channel estimators and compare it with that of classical least-squares (LS) channel estimation. For CS-based channel estimation, we employ BP, OMP, and CoSaMP as recovery algorithms, and we use both the DFT basis and the optimized basis



obtained with the algorithm of Section 6.2.

## 7.1 Simulation Setup

In accordance with the DVB-T standard [10], we simulated a CP-OFDM system with  $K = 2048$  sub-carriers and CP length  $N - K = 512$ , whence  $N = 2560$ . The system employed 4-QAM symbols with Gray labeling, a rate-1/2 convolutional code (generator polynomials  $(13_8, 15_8)$ ), and  $32 \times 16$  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  $\rho = 1/4$ .

For BP, we used  $\epsilon = 0$  even though noise was present in our system, in order to avoid assuming knowledge or estimation of the noise characteristics and to obtain a smaller computational complexity. (Recent results [63] suggest that BP is robust to noise also for  $\epsilon = 0$ .) We used the MATLAB function `l1eq_pd()` from the toolbox  $\ell_1$ -MAGIC [64], which supports a large-scale modulus that can handle large systems efficiently (our complex-valued measurement matrices are typically of size  $2048 \times 8192$ ). This function requires that all vectors and matrices involved are real-valued. Therefore, for BP, we considered the real sparse reconstruction problem  $\mathbf{y}_r = \Phi_r \mathbf{x}_r + \mathbf{z}_r$ , with  $\mathbf{x}_r \triangleq [\Re\{\mathbf{x}^T\} \ \Im\{\mathbf{x}^T\}]^T$ ,  $\mathbf{y}_r \triangleq [\Re\{\mathbf{y}^T\} \ \Im\{\mathbf{y}^T\}]^T$ ,  $\mathbf{z}_r \triangleq [\Re\{\mathbf{z}^T\} \ \Im\{\mathbf{z}^T\}]^T$ , and  $\Phi_r \triangleq \begin{bmatrix} \Re\{\Phi\} & -\Im\{\Phi\} \\ \Im\{\Phi\} & \Re\{\Phi\} \end{bmatrix}$ . This is *not* fully equivalent to the original complex problem  $\mathbf{y} = \Phi \mathbf{x} + \mathbf{z}$ , since real and imaginary parts are treated separately. If  $\mathbf{x}$  is constrained to be (approximately)  $S$ -sparse, we now constrain  $\mathbf{x}_r$  to be (approximately)  $2S$ -sparse. Furthermore, if  $\Phi$  satisfies the restricted isometry hypothesis (21) with  $S$ -restricted isometry constant  $\delta_S$ , the same is true for  $\Phi_r$  (this follows from the special structure of  $\Phi_r$ —see, e.g., [65, Lemma 1]—and the fact that if a real-valued vector  $\mathbf{a}_r = [a_1 \cdots a_{2M}]^T$  is  $S$ -sparse, then so is the complex-valued vector  $\mathbf{a} \triangleq [a_1 \cdots a_M]^T + j[a_{M+1} \cdots a_{2M}]^T$ ).

CoSaMP requires a prior estimate of the sparsity of  $\mathbf{x}$ . In all simulations, we used the fixed sparsity estimate  $S = 262$ . This estimate was determined from the formula  $S = \lceil Q/(2 \log M) \rceil$  suggested in [38], in which we set  $Q \equiv |\mathcal{P}| = 2048$  regardless of the actual number of pilots used. (Note, however, that in most scenarios we actually used 2048 pilots for CS-based channel estimation.) The number of CoSaMP iterations was  $J = 15$ . For OMP, on the other hand, we always used  $J = S = 262$  iterations. Therefore, the coefficient vectors produced by OMP and CoSaMP were exactly  $S$ -sparse with  $S = 262$ .

For all channel estimators considered (CS-based and LS), we simulated and estimated the channel for blocks of  $L = 16$  transmitted OFDM symbols. All CS-based estimators employed a subsampling with  $\Delta K = 4$  and  $\Delta L = 1$ , whence  $D = 512$  and  $I = 16$ .

## 7.2 MSE and BER versus SNR

We first compare the normalized<sup>6</sup> mean square error (MSE) and bit error rate (BER) performance of CS-based channel estimation with that of LS channel estimation for different signal-to-noise ratios (SNRs). For LS channel estimation, we used two alternative rectangular pilot constellations: (i) a pilot on every fourth subcarrier for each OFDM symbol, corresponding to 8192 pilots or 25% of all transmit symbols, and (ii) a pilot on every fourth subcarrier for every second OFDM symbol, corresponding to 4096 pilots or 12.5% of all symbols. For all CS-based channel estimators, we placed uniformly at random 2048 pilots on the subsampled grid  $\mathcal{G}$  specified above (i.e.,  $\Delta K = 4$  and  $\Delta L = 1$ ). This corresponds to 6.25% of all symbols, which is one quarter of the number of pilots used for LS estimation with pilot constellation (i) and half that used with pilot constellation (ii). All CS-based methods used the same pilot constellation.

During blocks of  $L = 16$  transmitted OFDM symbols, we simulated a noisy doubly selective channel whose discrete-delay-Doppler spreading function  $S_h[m, i]$  was computed from (28). We assumed  $P = 20$  propagation paths with scatterer delay-Doppler positions  $(\tau_p/T_s, \nu_p T_s)$  chosen uniformly at random within  $[0, 511] \times [-0.03/K, 0.03/K]$  for each block of 16 OFDM symbols (hence, the maximum Doppler normalized by the subcarrier spacing was  $\pm 3\%$ ). The complex scatterer amplitudes  $\eta_p$  were randomly drawn from zero-mean, complex Gaussian distributions with three different variances (3 strong scatterers of equal mean power, 7 medium scatterers with 10 dB less mean power, and 10 weak scatterers with 20 dB less mean power).

Fig. 3 shows the normalized MSE of the channel estimates and the resulting BER of the overall receiver versus the channel SNR. It is seen that all CS estimators (with 6.25% pilots) significantly outperform the LS estimator with 12.5% pilots. The extremely poor performance of the LS estimator with 12.5% pilots is due to the fact that the Shannon sampling theorem is violated by the pilot grid. In contrast, the CS estimators are able to produce reliable channel estimates even far below the Shannon sampling rate. Compared with the LS estimator with 25% pilots, the CS estimators (with 6.25% pilots) using the DFT basis exhibit a performance degradation that is relatively small for low-to-medium SNR but larger for high SNR. The performance degradation at high SNR has two reasons: (i) the number of pilots used for the CS estimators is too small for the channel's sparsity, and (ii) the OMP-based and CoSaMP-based CS estimators produce  $S$ -sparse signals with  $S = 262$ , which is too small for the channel's sparsity.

It is furthermore seen that the CS estimators using the optimized basis clearly outperform those using the DFT basis. The BP-based and CoSaMP-based CS estimators using the optimized basis show

---

<sup>6</sup>The mean square error was normalized by the energy of all channel coefficients.

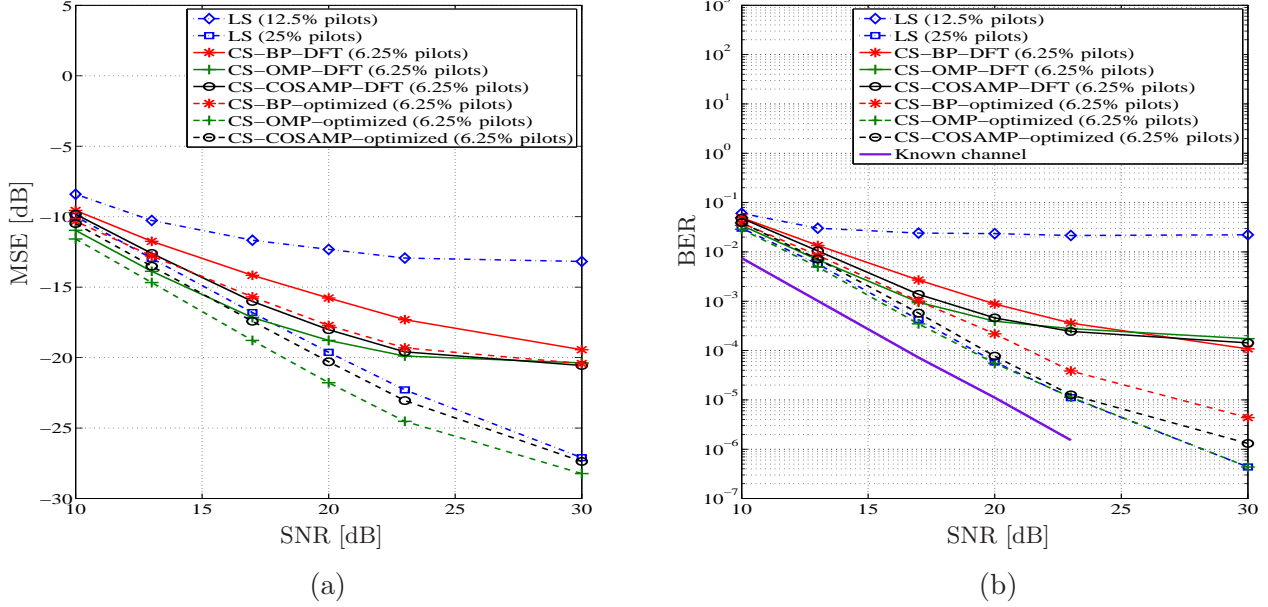


Figure 3: Performance of CS-based and LS channel estimators versus the SNR: (a) MSE, (b) BER.

only a small performance degradation with respect to the LS estimator with 25% pilots even for higher SNRs, while the OMP-based CS estimator is as good as or even better than the LS estimator with 25% pilots for all SNRs (note that the LS estimator uses four times as many pilots!). This performance gain is due to the better sparsity produced by the optimized basis, 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 ( $\mu_{\mathbf{U}} = 1$ ).

Comparing the different recovery algorithms, we see that OMP has the best performance, followed by CoSaMP and, in turn, BP (over a wide SNR range).

### 7.3 MSE and BER versus Number of Propagation Paths

Next, we compare the performance of the CS and LS estimators for different numbers  $P$  of propagation paths (scatterers) of the channel, which were chosen between 15 and 46. In each case, the scatterers' delay-Doppler positions  $(\tau_p/T_s, \nu_p T_s)$  were chosen uniformly at random within  $[0, 511] \times [-0.03/K, 0.03/K]$  for each block of 16 OFDM symbols (hence, the maximum normalized Doppler was again  $\pm 3\%$ ). The complex scatterer amplitudes  $\eta_p$  were randomly drawn from zero-mean, complex Gaussian distributions with three different variances—strong scatterers of equal power, medium scatterers with 10 dB less power, and weak scatterers with 20 dB less power—with (approximately) a ratio of 3:7:10 between their numbers. The SNR was fixed to 17 dB. As before, we placed uniformly at random 2048 pilots (6.25% of all symbols) on the subsampled grid  $\mathcal{G}$  specified above and used the resulting pilot constellation for all CS estimators. The pilot grids for the LS estimators were the same as in the previous

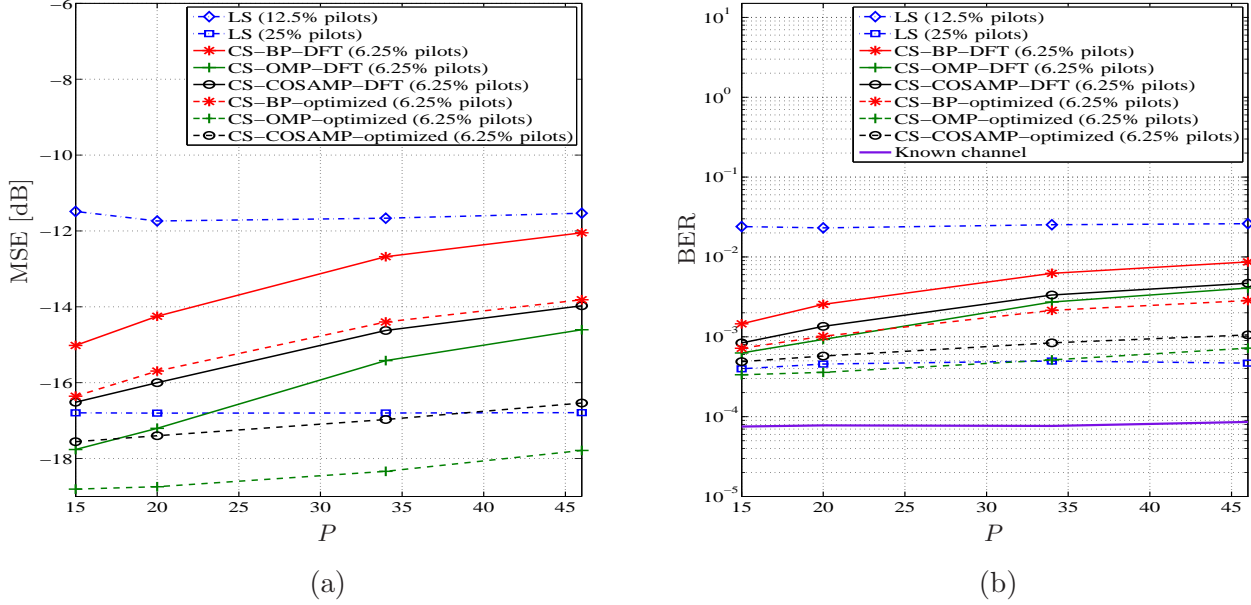


Figure 4: Performance of CS-based and LS channel estimation versus the number of propagation paths (scatterers): (a) MSE, (b) BER.

subsection (i.e., 25% and 12.5% of all symbols).

Fig. 4 depicts the normalized MSE and the BER versus  $P$ . All CS estimators perform better for a smaller  $P$ , due to the better sparsity. The performance of the LS estimators is effectively independent of  $P$ . Again, we observe significant gains resulting from the use of the optimized basis.

#### 7.4 MSE and BER versus Number of Pilots

An important design parameter influencing the performance of the CS estimators is the number of pilots,  $|\mathcal{P}|$ , since it equals the number of measurements available for sparse reconstruction. To study this issue, we placed uniformly at random  $|\mathcal{P}| \in \{512, \dots, 8192\}$  pilots on the subsampled grid  $\mathcal{G}$  (corresponding to 1.5625% ... 25% of all symbols) and used the resulting pilot constellations for all CS estimators. We also considered LS estimators with the number of pilots in the same range (but placed on rectangular subgrids of  $\mathcal{G}$ ). The channel was simulated as in Section 7.2. The SNR was fixed to 17 dB.

Fig. 5 depicts the normalized MSE and the BER of all estimators versus  $|\mathcal{P}|$ ; as a reference, the known-channel BER is also plotted as a horizontal line. It is seen that, as expected, the performance of all estimators improves with growing  $|\mathcal{P}|$ . We again observe large performance gains of the CS estimators over the LS estimators, and of the CS estimators using the optimized basis over those using the DFT basis. Note that the two best CS estimators almost achieve the known-channel BER.

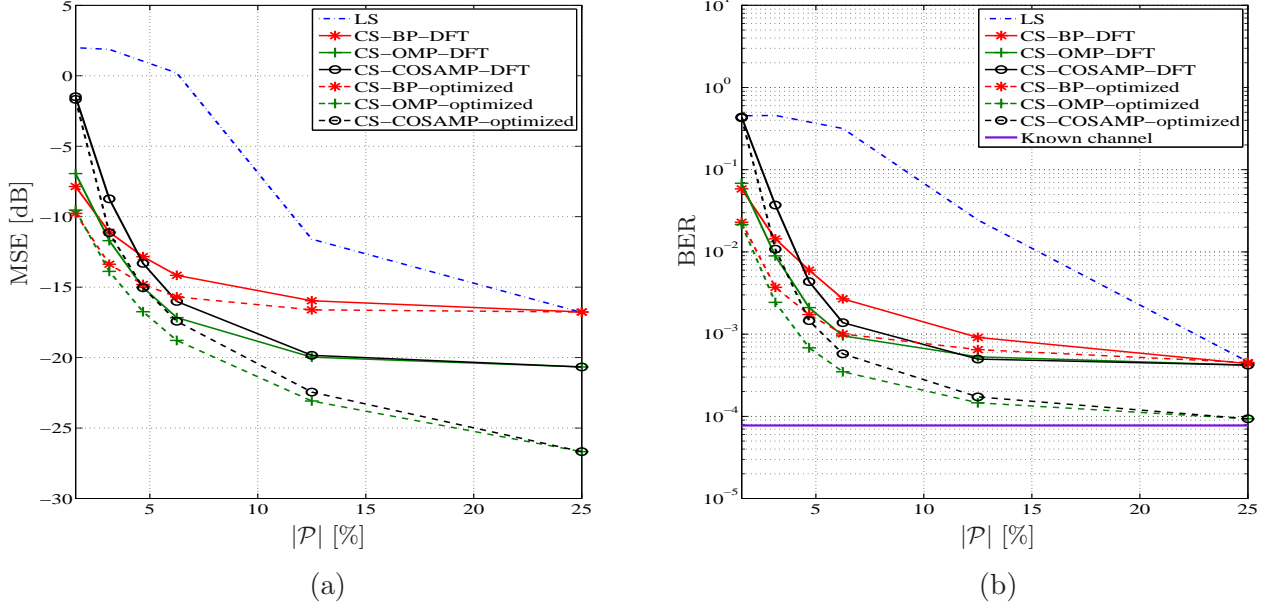


Figure 5: Performance of CS-based and LS channel estimators versus the number of pilots: (a) MSE, (b) BER.

## 7.5 MSE and BER versus Maximum Doppler Frequency

Finally, we compare the performance of the CS and LS estimators for different maximum Doppler frequencies  $\nu_{\max}$ . We simulated a channel consisting of  $P=20$  propagation paths with scatterer positions  $(\tau_p/T_s, \nu_p T_s)$  chosen uniformly at random within  $[0, 511] \times [-\nu_{\max} T_s, \nu_{\max} T_s]$  for each block of 16 OFDM symbols, where  $\nu_{\max} T_s K \in [0.0025, 0.03]$ . Hence, the maximum Doppler frequency  $\nu_{\max}$  normalized by the subcarrier spacing varied between 0.25% and 3%. As before, the complex scatterer amplitudes  $\eta_p$  were randomly drawn from zero-mean, complex Gaussian distributions with three different variances (3 strong, 7 medium, 10 weak). The SNR was 17 dB. All CS estimators used the same pilot constellation of 2048 pilots (6.25% of all symbols) placed uniformly at random on the subsampled grid  $\mathcal{G}$ . The pilot grids of the LS estimators were the same as in Subsection 7.2 (i.e., 25% and 12.5% of all symbols).

Fig. 6 depicts the normalized MSE and the BER versus the maximum normalized Doppler frequency  $\nu_{\max} T_s K$ . It can be seen that the LS estimator with 12.5% pilots that fails completely for large  $\nu_{\max}$  performs better for smaller  $\nu_{\max}$ . This is because for smaller maximum Doppler frequency, the Shannon sampling theorem is better satisfied by the pilot grid. A similar behavior can also be observed for the CS estimators. However, the reason now is that for smaller maximum Doppler frequency, the effective delay-Doppler support regions of the individual propagation paths (given by the supports of the leakage functions  $\Lambda^{(\nu_p)}(m - \frac{\tau_p}{T_s}, i - \nu_p T_s N_r)$ ) overlap more, which results in better sparsity. This effect is seen to be weaker when the optimized basis is used. Again, basis optimization results in improved overall

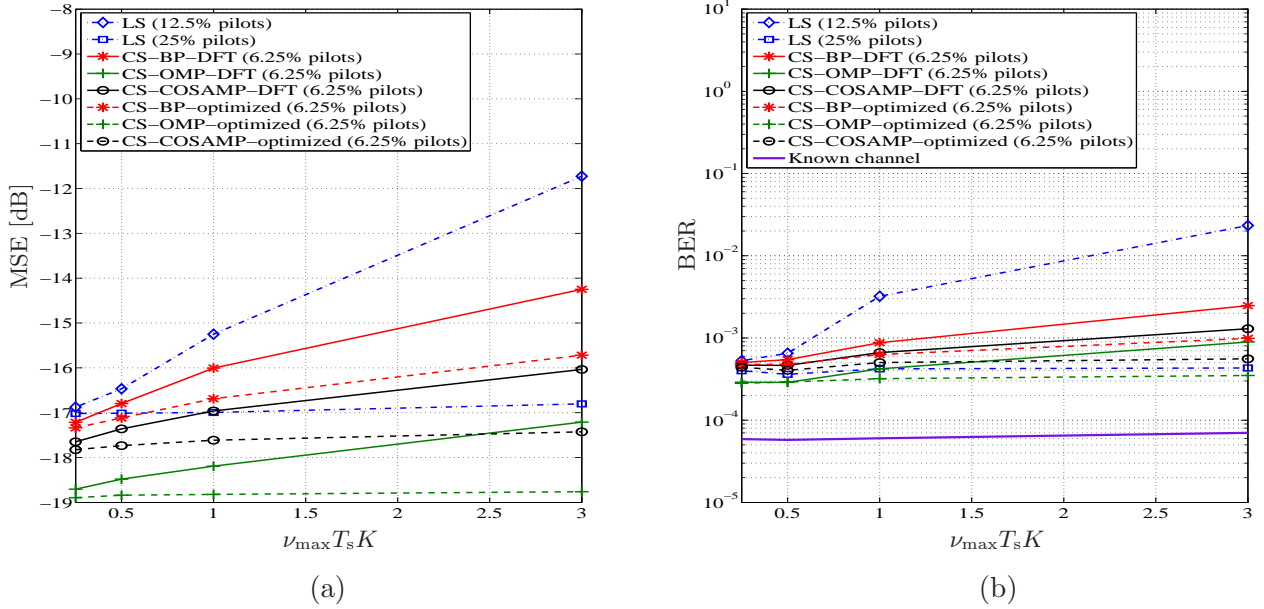


Figure 6: Performance of CS-based and LS channel estimators versus the maximum normalized Doppler frequency: (a) MSE, (b) BER.

performance. We note that the optimized basis depends on  $\nu_{\max}$ .

## 8 Conclusion

We proposed the application of compressed sensing (CS) methods to the estimation of doubly selective wireless 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 formulated this estimation task as a sparse reconstruction problem, for which several efficient CS recovery methods are available. The measurements used for sparse reconstruction are given by a small subset of the channel coefficients, which are calculated by means of pilot symbols. We demonstrated that, compared with classical least-squares estimation, our CS-based estimators allow an increase in spectral efficiency through a reduction of the number of pilot symbols that have to be transmitted. Alternatively, for a given number of pilot symbols, they yield an improved accuracy of channel estimation. Knowledge of channel or noise statistics is not required.

We also presented generalized CS-based channel estimators that use an explicit basis expansion of the channel's time-frequency coefficients. With a suitably designed basis, the sparsity is improved due to a reduction of leakage effects, which results in a further improvement of channel estimation performance.

We proposed a framework for optimizing the basis and an iterative approximate basis optimization algorithm. Simulation results demonstrated large performance gains obtained with the optimized basis. The additional computational complexity is moderate; in particular, the basis can be precomputed before the start of data transmission. We expect that our basis optimization algorithm may also be useful for dictionary learning in other applications besides sparse channel estimation.

Additional work is needed to better assess the potential of the proposed methods. Possible avenues for future research include the use of other sparse recovery methods (besides BP, OMP, and CoSaMP); a comparison with other channel estimators (besides least-squares methods); a performance evaluation using other methods of channel simulation and real wireless channels; the application of other dictionary learning methods for basis optimization, and a reduction of computational complexity.

## Acknowledgments

The authors would like to thank G. Matz and P. Fertl for helpful discussions.

## References

- [1] 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-2008*, (Las Vegas, NV), pp. 2885–2888, March/Apr. 2008.
- [2] G. Tauböck and F. Hlawatsch, “Compressed sensing based estimation of doubly selective channels using a sparsity-optimized basis expansion,” in *Proc. EUSIPCO 2008*, (Lausanne, Switzerland), Aug. 2008.
- [3] 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, pp. 489–509, Feb. 2006.
- [4] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, April 2006.
- [5] <http://www.dsp.ece.rice.edu/cs/>.
- [6] W. Kozek and A. F. Molisch, “Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels,” *IEEE J. Sel. Areas Comm.*, vol. 16, pp. 1579–1589, Oct. 1998.
- [7] J. A. C. Bingham, “Multicarrier modulation for data transmission: An idea whose time has come,” *IEEE Comm. Mag.*, vol. 28, pp. 5–14, May 1990.
- [8] IEEE P802 LAN/MAN Committee, “The working group for wireless local area networks (WLANs).” <http://grouper.ieee.org/groups/802/11/index.html>.
- [9] IEEE P802 LAN/MAN Committee, “The working group on broadband wireless access standards.” <http://grouper.ieee.org/groups/802/16/index.html>.
- [10] ETSI, “Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television.” EN 300 744, V1.4.1, 2001 (<http://www.etsi.org>).
- [11] ETSI, “Digital audio broadcasting (DAB) to mobile, portable and fixed receivers.” ETS 300 401, 1995. <http://www.etsi.org>.
- [12] ETSI, “Digital radio mondiale (DRM): System specification.” EN 201 980, V2.1.1, 2004 (<http://www.etsi.org>).



- [13] <http://www.3gpp.org/article/lte>.
- [14] E. G. Larsson, G. Liu, J. Li, and G. B. Giannakis, "Joint symbol timing and channel estimation for OFDM based WLANs," *IEEE Comm. Letters*, vol. 5, pp. 325–327, Aug. 2001.
- [15] Y. Li, L. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Comm.*, vol. 46, pp. 902–915, July 1998.
- [16] 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. Comm.*, vol. 46, pp. 931–939, July 1998.
- [17] P. Hoeher, S. Kaiser, and P. Robertson, "Pilot-symbol-aided channel estimation in time and frequency," in *Proc. IEEE GLOBECOM-97*, (Phoenix, AZ), pp. 90–96, Nov. 1997.
- [18] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEEVT*, vol. 49, pp. 1207–1215, July 2000.
- [19] P. Fertl and G. Matz, "Efficient OFDM channel estimation in mobile environments based on irregular sampling," in *Proc. Asilomar Conf. Signals, Systems, Computers*, (Pacific Grove, CA), Okt.-Nov. 2006.
- [20] G. Leus, "On the estimation of rapidly time-varying channels," in *Proc. EUSIPCO 2004*, (Vienna, Austria), pp. 2227–2230, Sept. 2004.
- [21] D. K. Borah and B. T. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Comm.*, vol. 47, pp. 862–873, June 1999.
- [22] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Processing*, vol. 53, pp. 3597–3607, Sept. 2005.
- [23] A. F. Molisch, ed., *Wideband Wireless Digital Communications*. Englewood Cliffs (NJ): Prentice Hall, 2001.
- [24] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Comm.*, vol. 50, pp. 374–377, March 2002.
- [25] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *IEEE J. Sel. Areas Comm.*, vol. 1, pp. 357–371, Oct. 2007.
- [26] O. Rabaste and T. Chonavel, "Estimation of multipath channels with long impulse response at low SNR via an MCMC method," *IEEE Trans. Signal Processing*, vol. 55, pp. 1312–1325, Apr. 2007.
- [27] W. Li and J. C. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Oceanic Eng.*, vol. 32, pp. 927–939, Oct. 2007.
- [28] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. 42nd Annu. Conf. Information Sciences and Systems (CISS'08)*, (Princeton, NJ), pp. 5–10, March 2008.
- [29] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proc. 46th Annu. Allerton Conf. Communication, Control, and Computing*, (Monticello, IL), Sept. 2008.
- [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. Signals, Systems, and Computers*, (Pacific Grove, CA), Oct. 2008.
- [31] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Sparse multipath channels: Modeling and estimation," in *Proc. 13th IEEE Digital Signal Processing Workshop*, (Marco Island, FL), Jan. 2009.
- [32] H. Rauhut, "Circulant and Toeplitz matrices in compressed sensing," in *Proc. SPARS'09*, (Saint-Malo, France), April 2009.
- [33] S. G. Mallat and Z. Zhang, "Matching pursuits and time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.
- [34] 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.



- [35] E. J. Candès, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comm. Pure Appl. Math.*, vol. 59, March 2006.
- [36] 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.
- [37] J. A. Tropp, “Greed is good: algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, Oct. 2004.
- [38] J. Tropp and D. Needell, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Comput. Harmon. Anal.*, to appear.
- [39] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Proc.*, vol. 11, no. 54, pp. 4311–4322, 2006.
- [40] K. Kreutz Delgado, J. Murray, and B. Rao, “Dictionary learning algorithms for sparse representation,” *Neural Computation*, vol. 15, pp. 349–396, 2003.
- [41] R. Gribonval and K. Schnass, “Dictionary identifiability from few training samples,” in *Proc. EUSIPCO 2008*, (Lausanne, Switzerland), Aug. 2008.
- [42] G. Matz, D. Schafhuber, K. Gröchenig, M. Hartmann, and F. Hlawatsch, “Analysis, optimization, and implementation of low-interference wireless multicarrier systems,” *IEEE Trans. Wireless Comm.*, vol. 6, pp. 1921–1931, May 2007.
- [43] M. K. Tsatsanis and G. B. Giannakis, “Modeling and equalization of rapidly fading channels,” *Int. J. Adaptive Control and Signal Processing*, vol. 10, pp. 159–176, March 1996.
- [44] G. B. Giannakis and C. Tepedelenlioglu, “Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels,” *Proc. IEEE*, vol. 86, pp. 1969–1986, Oct. 1998.
- [45] P. A. Bello, “Characterization of randomly time-variant linear channels,” *IEEE Trans. Comm. Syst.*, vol. 11, pp. 360–393, 1963.
- [46] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. San Diego (CA): Academic Press, 1999.
- [47] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge (UK): Cambridge Univ. Press, Dec. 2004.
- [48] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *C. R. Acad. Sci. Paris S’er. I Math.*, vol. 346, pp. 589–592, 2008.
- [49] S. Foucart and M. Lai, “Sparsest solutions of underdetermined linear systems via  $\ell_q$ -minimization for  $0 < q \leq 1$ ,” *Appl. Comput. Harmon. Anal.*, to appear.
- [50] M. Rudelson and R. Vershynin, “Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements,” in *Proc. 40th Annual Conf. Inform. Sci. Syst.*, (Princeton, NJ), pp. 207–212, March 2006.
- [51] H. Rauhut, “Stability results for random sampling of sparse trigonometric polynomials,” *IEEE Trans. Information Theory*, vol. 54, no. 12, pp. 5661–5670, 2008.
- [52] A. C. Gilbert and J. A. Tropp, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [53] H. Rauhut, “On the impossibility of uniform sparse reconstruction using greedy methods,” *Sampl. Theory Signal Image Process.*, vol. 7, pp. 197–215, May 2008.
- [54] W. Dai and O. Milenkovic, “Subspace Pursuit for Compressive Sensing Signal Reconstruction,” *preprint*, 2008.
- [55] 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, March 1982.

- [56] 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.
- [57] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, “Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit,” *Preprint*, March 2006.
- [58] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.
- [59] D. L. Donoho and Y. Tsaig, “Fast solution of  $\ell_1$ -norm minimization problems when the solution may be sparse,” Tech. Rep. 2006-18, Stanford University Department of Statistics, 2006.
- [60] G. Matz and F. Hlawatsch, “Time-varying communication channels: Fundamentals, recent developments, and open problems,” in *Proc. EUSIPCO-06*, (Florence, Italy), Sept. 2006.
- [61] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore: Johns Hopkins University Press, 2nd ed., 1989.
- [62] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming (web page and software), Stanford University, CA (<http://stanford.edu/~boyd/cvx>).
- [63] P. Wojtaszczyk, “Stability and instance optimality for Gaussian measurements in compressed sensing,” *preprint*, 2008.
- [64] E. J. Candès and J. Romberg. Toolbox  $\ell_1$ -MAGIC, California Inst. of Technol., Pasadena, CA (<http://www.acm.caltech.edu/l1magic/>).
- [65] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Trans. Telecomm.*, vol. 10, pp. 585–596, Nov. 1999.