INVITED PAPER

Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels

High-rate wireless data communication can usually be achieved by collecting a relatively small sample of the available information about the communications channel.

By Waheed U. Bajwa, *Member IEEE*, Jarvis Haupt, Akbar M. Sayeed, *Senior Member IEEE*, and Robert Nowak, *Fellow IEEE*

ABSTRACT | High-rate data communication over a multipath wireless channel often requires that the channel response be known at the receiver. Training-based methods, which probe the channel in time, frequency, and space with known signals and reconstruct the channel response from the output signals, are most commonly used to accomplish this task. Traditional training-based channel estimation methods, typically comprising linear reconstruction techniques, are known to be optimal for rich multipath channels. However, physical arguments and growing experimental evidence suggest that many wireless channels encountered in practice tend to exhibit a sparse multipath structure that gets pronounced as the signal space dimension gets large (e.g., due to large bandwidth or large number of antennas). In this paper, we formalize the notion of multipath sparsity and present a new approach to estimating sparse (or effectively sparse) multipath channels that is based on some of the recent advances in the theory of compressed sensing. In particular, it is shown in the paper that the proposed approach, which is termed as compressed channel sensing (CCS), can potentially achieve a target reconstruction error using far less energy and, in many instances, latency and bandwidth than that dictated by the traditional least-squares-based training methods.

KEYWORDS | Channel estimation; compressed sensing; Dantzig selector; least-squares estimation; multiple-antenna channels; orthogonal frequency division multiplexing; sparse channel modeling; spread spectrum; training-based estimation

I. INTRODUCTION

A. Motivation and Background

In a typical scattering environment, a radio signal emitted from a transmitter is reflected, diffracted, and scattered from the surrounding objects, and arrives at the receiver as a superposition of multiple attenuated, delayed, and phase- and/or frequency-shifted copies of the original signal. This superposition of multiple copies of the transmitted signal, called multipath signal components, is the defining characteristic of many wireless systems, and is both a curse and a blessing from a communications viewpoint. On the one hand, this *multipath signal propagation* leads to fading—fluctuations in the received signal strength—that severely affects the rate and reliability of communication [1]. On the other hand, research in the last decade has shown that multipath propagation also results in an increase in the number of degrees of freedom (DoF) available for communication, which—if utilized effectively—can lead to significant gains in the rate (multiplexing gain) and/or reliability (diversity gain) of communication [2]. The impact of fading versus diversity/multiplexing gain on performance critically depends on the amount of channel state information (CSI) available to the system. For example, knowledge of instantaneous CSI at the receiver (coherent reception) enables exploitation of delay, Doppler, and/or spatial diversity to combat fading,

Manuscript received April 21, 2009; revised November 19, 2009; accepted January 17, 2010. Date of publication April 26, 2010; date of current version May 19, 2010.
W. U. Bajwa is with the Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544 USA (e-mail: wbajwa@math.princeton.edu).
J. Haupt is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: jdhaupt@rice.edu).
A. M. Sayeed and **R. Nowak** are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: akbar@engr.wisc.edu; nowak@engr.wisc.edu).

Digital Object Identifier: 10.1109/JPROC.2010.2042415

while further gains in rate and reliability are possible if (even partial) CSI is available at the transmitter as well [1].

In practice, CSI is seldom—if ever—available to communication systems *a priori* and the channel needs to be (periodically) estimated at the receiver in order to reap the benefits of additional DoF afforded by multipath propagation. As such, two classes of methods are commonly employed to estimate multipath channels at the receiver. In *training-based channel estimation* methods, the transmitter multiplexes signals that are known to the receiver (henceforth referred to as training signals) with data-carrying signals in time, frequency, and/or code domain, and CSI is obtained at the receiver from knowledge of the training and received signals. In *blind channel estimation* methods, CSI is acquired at the receiver by making use of the statistics of data-carrying signals only. Although theoretically feasible, blind estimation methods typically require complex signal processing at the receiver and often entail inversion of large data-dependent matrices, which also makes them highly prone to error propagation in rapidly varying channels. Training-based methods, on the other hand, require relatively simple receiver processing and often lead to decoupling of the data-detection module from the channel-estimation module at the receiver, which reduces receiver complexity even further. As such, training-based methods are widely prevalent in modern wireless systems [3] and we therefore focus exclusively on them throughout the rest of the paper; see [4] for an overview of blind approaches to channel estimation.

One of the first analytical studies of training-based estimation methods for multipath channels was authored by Cavers in 1991 [5]. Since then, there has been a growing body of literature devoted to the design and analysis of training-based methods for various classes of channels. These works often highlight two salient aspects of training-based methods, namely, *sensing* and *reconstruction*. Sensing corresponds to the design of training signals used by the transmitter to probe the channel, while reconstruction is the problem of processing the corresponding channel output at the receiver to recover the CSI. The ability of a training-based method to accurately estimate the channel depends critically on both the design of training signals and the application of effective reconstruction strategies. Much of the work in the channel estimation literature is based on the implicit assumption of a *rich* underlying multipath environment in the sense that the number of DoF in the channel is expected to scale linearly with the signal space dimension (product of signaling bandwidth, symbol duration, and minimum of the number of transmit and receive antennas). As a result, training-based methods proposed in such works are mainly composed of linear reconstruction techniques, which are known to be optimal for rich multipath channels, thereby more or less reducing the problem of channel estimation to that of designing optimal training signals for various channel classes [5]–[12].

Numerous experimental studies undertaken by various researchers in the recent past have shown though that wireless channels associated with a number of scattering environments tend to exhibit *sparse* structures at high signal space dimension in the sense that majority of the channel DoF end up being either zero or below the noise floor when operating at large bandwidths and symbol durations and/or with large plurality of antennas [13]–[18]. However, traditional training-based methods that rely on linear reconstruction schemes at the receiver seem incapable of exploiting the inherent low dimensionality of such sparse channels, thereby leading to overutilization of the key communication resources of energy, latency, and bandwidth. A number of researchers have tried to address this problem in the recent past and proposed training signals and reconstruction strategies that are tailored to the anticipated characteristics of sparse multipath channels [19]–[27]. But much of the emphasis in these studies has been directed towards establishing the feasibility of the proposed sparse-channel estimation methods numerically rather than analytically. A major drawback of this approach is that the methods detailed in the previous investigations lack a quantitative theoretical analysis of their performance in terms of the reconstruction error.

B. Historical Developments

As is the case with so many other research problems in wireless communications, the area of sparse-channel estimation has a history that dates back to the early 1990s. Historically, the problem of sparse-channel estimation using training-based methods was first explored in the literature in the context of underwater acoustic communications. Specifically, prompted by the fact that typical underwater acoustic channels have impulse responses with large delay and Doppler spreads but only a few dominant echoes, an adjustable complexity, recursive least-squares estimation algorithm that ignores the weakest dimensions ("taps") of the channel was proposed in [19] for doubly-selective single-antenna channels using single-carrier waveforms. Afterwards, inspired by the fact that digital television channels and broadband channels in hilly terrains also exhibit sparse structures, Cotter and Rao proposed a sparse-channel estimation method based on the matching pursuit (MP) algorithm [28] for frequency-selective single-antenna channels using single-carrier waveforms [20], [22]. Later, the MP-based sparse-channel estimation method of [20] was extended to frequency-selective single- and multiple-antenna channels using multicarrier waveforms in [25] and [23], respectively, and to doubly-selective single-antenna channels using single-carrier waveforms in [21] and [27]. Here, one of the main differences between the approaches of [21] and [27] is that [21] attempts to exploit the sparsity of the channel in the delay domain only.

In contrast to the MP-based approach of [20]–[23], [25], and [27], Raghavendra and Giridhar proposed a modified least-squares (LS) estimator in [24] for sparse frequency-selective single-antenna channels using multicarrier waveforms. The idea behind the approach in [24] was to reduce the signal space of the LS estimator by using a generalized Akaike information criterion to estimate the locations of nonzero channel taps. Finally, a somewhat similar idea was employed most recently in [26] for sparse frequency-selective single-antenna channels using single-carrier waveforms. In particular, one of the key differences between [24] and [26] is that [26] attempts to estimate the locations of nonzero channel taps by transforming the tap detection problem into an equivalent ON–OFF keying detection problem.

C. Scope of This Paper

By leveraging key ideas from the theory of compressed sensing [29], various researchers have recently proposed new training-based estimation methods for different classes of sparse single- and multiple-antenna channels that are *provably* more effective than their LS-based counterparts in the limit of large signal space dimension [30]–[38]. In particular, the training-based methods detailed in [30], [34], [35], and [38] have been analytically shown to achieve a target reconstruction-error scaling using far less energy and, in many instances, latency and bandwidth than that dictated by the LS-based methods. As in the case of previous research, the exact nature of training signals employed by the proposed methods in [38] varies with the type of signaling waveforms used for sensing (e.g., single-carrier or multicarrier signaling waveforms) and the class to which the underlying multipath channel belongs (e.g., frequency- or doubly-selective channel). However, a common theme underlying all these training-based methods is the use of sparsity-inducing mixed-norm optimization criteria, such as the basis pursuit [39], Dantzig selector [40], and lasso [41], for reconstruction at the receiver. These criteria have arisen out of recent advances in the theory of sparse signal recovery, which is more commonly studied under the rubric of compressed sensing these days. In the spirit of compressed sensing, we term this particular approach to estimating sparse multipath channels as compressed channel sensing (CCS); the analogy here being that CCS requires far fewer communication resources to estimate sparse channels than do the traditional LS-based training methods.

The goal of this paper is to complement this existing work on sparse-channel estimation by providing a unified summary of the key ideas underlying the theory of CCS. In order to accomplish this goal, we focus on four specific classes of multipath channels within the paper, namely, frequency- and doubly-selective single-antenna channels, and nonselective and frequency-selective multiple-antenna channels. For each of these four channel classes, the discussion in the paper focuses on the nature of the training signals used for probing a sparse channel, the reconstruction method used at the receiver for recovering the CSI, and quantification of the reconstruction error in the resulting estimate. In terms of modeling of the sparse channels within each channel class, we use a virtual representation of physical multipath channels that represents the expansion of the time-frequency response of a channel in terms of multidimensional Fourier basis functions. It is worth noting though that the main ideas presented in the paper can be generalized to channel models that make use of a basis other than the Fourier one, provided the expansion basis effectively exposes the sparse nature of the underlying multipath environment and can be made available to the receiver *a priori*. Finally, most of the mathematical claims in the paper are stated without accompanying proofs in order to keep the exposition short and accessible to general audience. Extensive references are made to the original papers in which the claims first appeared for those interested in further details.

D. Organization

The rest of this paper is organized as follows. In Section II, we review a widely used modeling framework in the communications literature that provides a discretized approximation of the time-frequency response of a physical channel. This framework plays a key role in subsequent developments in the paper since it not only exposes the relationship between the distribution of physical paths within the angle-delay-Doppler space and the sparsity of channel DoF, but also sets the stage for the application of compressed sensing theory to sparse-channel estimation. In Section III, we first formalize the notion of sparse multipath channels by making use of the modeling framework of Section II and then characterize the performance of LS-based training methods for various classes of sparse channels. In Section IV, we succinctly summarize the performance advantages of CCS over traditional LS-based methods and provide a brief review of the theory of compressed sensing. We devote the discussion in Sections V and VI to exploring the specifics of CCS for sparse frequency-/doubly-selective, single-antenna channels and sparse nonselective/frequency-selective, multiple-antenna channels, respectively. Finally, we conclude in Section VII by discussing some of the finer technical details pertaining to the results presented in the paper.

II. MULTIPATH WIRELESS CHANNEL MODELING

A. Physical Model

Consider, without loss of generality, a multiple-antenna channel with half-wavelength spaced linear arrays at the transmitter and the receiver. Let N_T and N_R denote the number of transmit and receive antennas, respectively. It is customary to model a multipath wireless channel $\boldsymbol{\mathcal{H}}$ as a linear, time-varying system [1], [42]. The corresponding (complex) baseband transmitted signal and channel output are related as

$$\boldsymbol{\mathcal{H}}(\mathbf{x}(t)) = \int_{\mathbb{R}} \mathbf{H}(t,f)\mathbf{X}(f)e^{j2\pi ft}\,df \qquad (1)$$

where $\mathcal{H}(\mathbf{x}(t))$ is the N_R-dimensional channel output, $\mathbf{X}(f)$ is the (elementwise) Fourier transform of the N_T-dimensional transmitted signal $\mathbf{x}(t)$, and $\mathbf{H}(t,f)$ is the $N_R \times N_T$ time-varying frequency response matrix of the channel. The matrix $\mathbf{H}(t,f)$ can be further expressed in terms of the underlying physical paths as

$$\mathbf{H}(t,f) = \sum_{n=1}^{N_p} \beta_n \mathbf{a}_R(\theta_{R,n}) \mathbf{a}_T^{\mathrm{H}}(\theta_{T,n}) e^{-j2\pi\tau_n f} e^{j2\pi\nu_n t} \tag{2}$$

which represents signal propagation over N_p paths[1]; here, $(\cdot)^{\mathrm{H}}$ denotes the Hermitian operation and β_n is the complex path gain, $\theta_{R,n}$ is the angle of arrival (AoA) at the receiver, $\theta_{T,n}$ is the angle of departure (AoD) at the transmitter, τ_n is the (relative) delay, and ν_n is the Doppler shift associated with the nth path. The $N_T \times 1$ vector $\mathbf{a}_T(\theta_T)$ and the $N_R \times 1$ vector $\mathbf{a}_R(\theta_R)$ denote the array steering and response vectors, respectively, for transmitting/receiving a signal in the direction θ_T/θ_R and are periodic in θ with unit period [45].[2] We assume that the channel is maximally spread in the angle space, $(\theta_{R,n}, \theta_{T,n}) \in [-1/2, 1/2] \times [-1/2, 1/2]$, while $\tau_n \in [0, \tau_{\max}]$ and $\nu_n \in [-(\nu_{\max}/2), \nu_{\max}/2]$ in the delay and Doppler space, respectively. Here, $\tau_{\max}$ and $\nu_{\max}$ are termed as the delay spread and (two-sided) Doppler spread of the channel, respectively. Estimating a channel having $\tau_{\max}\nu_{\max} > 1$ can often be an ill-posed problem even in the absence of noise [46]. Instead, we limit the discussion in this paper to underspread channels, characterized by $\tau_{\max}\nu_{\max} \ll 1$, which is fortunately true of many wireless channels [47, Sec. 14.2].[3]

Finally, throughout the paper, we implicitly consider signaling over wireless channels using symbols of duration T and (two-sided) bandwidth W, $\mathbf{x}(t) = \mathbf{0}_{N_T} \forall t \notin [0, T]$ and $\mathbf{X}(f) = \mathbf{0}_{N_T} \forall f \notin [-W/2, W/2]$, thereby giving rise to a *temporal signal space* of dimension $N_o = TW$ [49]. Note that these signaling parameters, together with the delay and Doppler spreads of a channel, can be used to broadly classify wireless channels as nonselective, frequency selective, time selective, or doubly selective; see Table 1 for a definition of each of these classes. As noted earlier, we limit ourselves in this paper to primarily discussing frequency- and doubly-selective channels in the single-antenna setting ($N_T = N_R = 1$) and to nonselective and frequency-selective channels in the multiple-antenna setting.

[1]Time-varying frequency responses of multipath channels often correspond to superposition of a small number of strong paths (specular scattering) and a huge number of weak paths (diffuse scattering) [43]. From an analytical viewpoint, as opposed to the channel measurement viewpoint, (2) accurately captures the effects of both specular and diffuse scattering in the limit of large N_p (also, see [44]).

[2]The normalized angle θ is related to the physical angle ϕ (measured with respect to array broadside) as $\theta = d\sin(\phi)/\lambda$, where d is the antenna spacing and λ is the wavelength of propagation; see [45] for further details.

[3]It is worth mentioning here though that part of the discussion in this paper is also applicable to underwater acoustic communication channels, even though they may not always be underspread [48].

Table 1 Classification of Wireless Channels on the Basis of Channel and Signaling Parameters

Channel Classification	$W\tau_{\max}$	$T\nu_{\max}$
Nonselective Channels	$\ll 1$	$\ll 1$
Frequency-Selective Channels	≥ 1	$\ll 1$
Time-Selective Channels	$\ll 1$	≥ 1
Doubly-Selective Channels	≥ 1	≥ 1

B. Virtual Representation

While the physical model (2) is highly accurate, it is difficult to analyze and estimate owing to its *nonlinear* dependence on a potentially large number of parameters $\{(\beta_n, \theta_{R,n}, \theta_{T,n}, \tau_n, \nu_n)\}$. However, because of the finite (transmit and receive) array apertures, signaling bandwidth, and symbol duration, it can be well approximated by a linear (in parameters) counterpart, known as a *virtual* or *canonical channel model*, with the aid of a 4-D Fourier series expansion [42], [45], [50]–[52].

On an abstract level, virtual representation of a multipath channel $\mathcal{H}$ provides a discretized approximation of its time-varying frequency response $\mathbf{H}(t,f)$ by uniformly sampling the angle-delay-Doppler space at the Nyquist rate: $(\Delta\theta_R, \Delta\theta_T, \Delta\tau, \Delta\nu) = (1/N_R, 1/N_T, 1/W, 1/T)$. Specifically, partition the N_p paths into the following subsets:

$$S_{R,i} = \{n : \theta_{R,n} \in (i/N_R - \Delta\theta_R/2, i/N_R + \Delta\theta_R/2]\}$$
$$S_{T,k} = \{n : \theta_{T,n} \in (k/N_T - \Delta\theta_T/2, k/N_T + \Delta\theta_T/2]\}$$
$$S_{\tau,\ell} = \{n : \tau_n \in (\ell/W - \Delta\tau/2, \ell/W + \Delta\tau/2]\}$$

and

$$S_{\nu,m} = \{n : \nu_n \in (m/T - \Delta\nu/2, m/T + \Delta\nu/2]\}.$$

Then, the virtual representation of $\mathcal{H}$ is given by

$$\tilde{\mathbf{H}}(t,f) = \sum_{i=1}^{N_R}\sum_{k=1}^{N_T}\sum_{\ell=0}^{L-1}\sum_{m=-M}^{M} H_v(i,k,\ell,m)\mathbf{a}_R\left(\frac{i}{N_R}\right) \times \mathbf{a}_T^{\mathrm{H}}\left(\frac{k}{N_T}\right) e^{-j2\pi\frac{\ell}{W}f} e^{j2\pi\frac{m}{T}t} \tag{3}$$

$$H_v(i,k,\ell,m) \approx \sum_{n\in S_{R,i}\cap S_{T,k}\cap S_{\tau,\ell}\cap S_{\nu,m}} \beta_n f_{N_R}(i/N_R - \theta_{R,n}) \times f_{N_T}^{*}(k/N_T - \theta_{T,n}) \times \mathrm{sinc}(m - T\nu_n, \ell - W\tau_n) \tag{4}$$

and it approximates $\mathbf{H}(t,f)$ in the sense [42], [45], [50]–[52]

$$\int_{\mathbb{R}} \mathbf{H}(t,f)\mathbf{X}(f)e^{j2\pi ft}df \approx \int_{\mathbb{R}} \tilde{\mathbf{H}}(t,f)\mathbf{X}(f)e^{j2\pi ft}\,df.$$

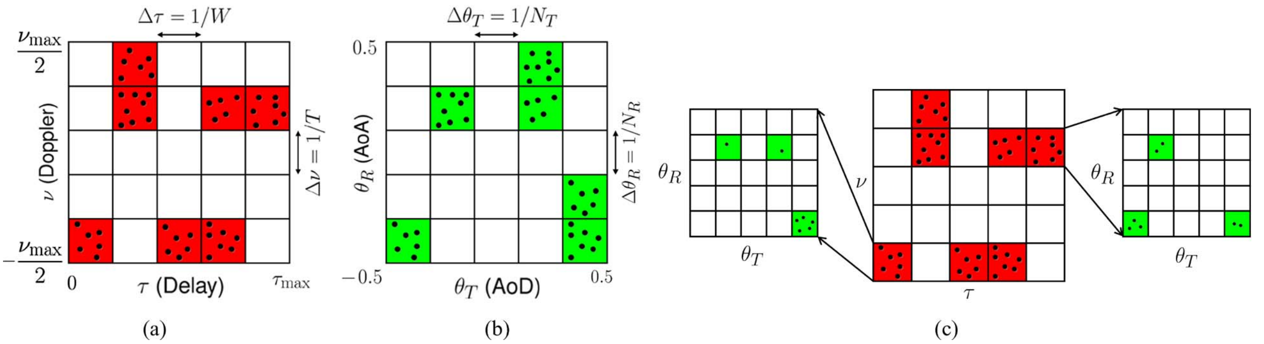

Fig. 1. ***An idealized illustration of the virtual channel representation (VCR) and the channel sparsity pattern (SP). Each square represents a resolution bin associated with a distinct virtual channel coefficient. The total number of these squares equals D. The shaded squares represent the SP, $\mathcal{S}_d$, corresponding to the $d \ll D$ dominant channel coefficients, and the dots represent the paths contributing to each dominant coefficient. (a) VCR and SP in delay-Doppler: $\{H_v(\ell, m)\}_{\mathcal{S}_d}$. (b) VCR and SP in angle: $\{H_v(i, k)\}_{\mathcal{S}_d}$. (c) VCR and SP in angle-delay-Doppler: $\{H_v(i, k, \ell, m)\}_{\mathcal{S}_d}$. The paths contributing to a fixed dominant delay-Doppler coefficient, $H_v(\ell_o, m_o)$, are further resolved in angle to yield the conditional SP in angle: $\{H_v(i, k, \ell_o, m_o)\}_{\mathcal{S}_d(\ell_o, m_o)}$.***

Here, the smoothing kernels $f_{N_R}(\theta_R)$ and $f_{N_T}(\theta_T)$ in (4) are the Dirichlet kernels, $f_N(\theta) = (1/N)\sum_{i=0}^{N-1} e^{-j2\pi i\theta}$, while the 2-D sinc kernel is defined as $\mathrm{sinc}(x, y) = e^{-j\pi x}\sin(\pi x)\sin(\pi y)/(\pi^2 xy)$. The approximation in (4) is due to the sidelobes of the Dirichlet and sinc kernels induced by the finite signaling parameters, and the approximation gets more accurate with increasing T, W, N_R, and N_T.[4]

Note that due to the fixed angle-delay-Doppler sampling of (2), which defines the spatio–temporal Fourier basis functions in (3), $\tilde{\mathbf{H}}(t, f)$ is a *linear* channel representation that is completely characterized by the *virtual channel coefficients* $\{H_v(i, k, \ell, m)\}$. From (3), the total number of these coefficients is given by $D = N_R N_T L(2M + 1)$, where $N_R, N_T, L = \lceil W\tau_{\max}\rceil + 1$, and $M = \lceil T\nu_{\max}/2\rceil$ represent the maximum number of *resolvable* AoAs, AoDs, delays, and (one-sided) Doppler shifts within the angle-delay-Doppler spread of the channel, respectively.[5] Further, the notation in (4) signifies that each coefficient $H_v(i, k, \ell, m)$ is approximately equal to the sum of the complex gains of all physical paths whose angles, delays, and Doppler shifts lie within an *angle-delay-Doppler resolution bin* of size $\Delta\theta_R \times \Delta\theta_T \times \Delta\tau \times \Delta\nu$ centered around the sampling point $(\hat{\theta}_{R,i}, \hat{\theta}_{T,k}, \hat{\tau}_\ell, \hat{\nu}_m) = (i/N_R, k/N_T, \ell/W, m/T)$ in the angle-delay-Doppler space; we refer the reader to [51] for further details (also, see Fig. 1). In other words, the virtual representation $\tilde{\mathbf{H}}(t, f)$ effectively captures the underlying multipath environment comprising N_p physical paths through D resolvable paths, thereby reducing the task of estimating $\boldsymbol{\mathcal{H}}$ to that of reconstructing the virtual channel coefficients $\{H_v(i, k, \ell, m)\}$.

[4]Note that in the case of a doubly-selective single-antenna channel, the virtual representation (3) is similar to the well-known exponential basis expansion model [53, Sec. III-A].

[5]With a slight abuse of notation, we define $\lceil W\tau_{\max}\rceil = 0$ and $\lceil T\nu_{\max}/2\rceil = 0$ for $W\tau_{\max} \ll 1$ and $T\nu_{\max} \ll 1$, respectively.

III. SPARSE MULTIPATH WIRELESS CHANNELS

A. Modeling

The virtual representation of a multipath wireless channel signifies that the maximum number of DoF in the channel is

$$D = N_R N_T L(2M + 1) \approx \tau_{\max}\nu_{\max} N_R N_T T W \qquad (5)$$

which corresponds to the maximum number of angle-delay-Doppler resolution bins in the virtual representation, and reflects the maximum number of resolvable paths within the 4-D channel spread. However, the actual or *effective* number of DoF d in the channel that govern its capacity and diversity corresponds to the number of *dominant* virtual channel coefficients: $d = |\{(i, k, \ell, m) : \ |H_v(i, k, \ell, m)| > \epsilon\}|$. Here, ϵ is an appropriately chosen parameter whose value depends upon the operating received signal-to-noise ratio (SNR). An intuitive choice for ϵ is the standard deviation of the receiver noise, meaning only channel coefficients with power above the noise floor contribute to the DoF. Trivially, we have $d \le D$ and, by virtue of (4), $d = D$ if there are at least $N_p \ge D$ physical paths distributed in a way within the channel spread such that each angle-delay-Doppler resolution bin is either populated by 1) at least one strong (specular) path, and/or 2) numerous weak (diffuse) paths whose aggregate energy is above the noise floor (see Fig. 1), or it is in the (close) proximity of another such resolution bin.

Much of the work in the existing channel estimation literature is based on the implicit assumption of a rich scattering environment in which there are sufficiently many specular and diffuse paths uniformly distributed within the angle-delay-Doppler spread of the channel so that $d \approx D$ for any choice of the received SNR and the signaling parameters. Numerous past and recent channel measurement campaigns have shown, however, that propagation paths in many physical channels tend to be distributed as clusters within their respective channel spreads [13]–[18]. Consequently, as we vary the spatio–temporal signaling parameters in such channels by increasing the number of antennas, signaling bandwidth, and/or symbol duration, a point comes where $\Delta\theta_R$, $\Delta\theta_T$, $\Delta\tau$, and/or $\Delta\nu$ become smaller than the interspacings between the multipath clusters, thereby leading to the situation depicted in Fig. 1 where not every resolution bin of size $\Delta\theta_R \times \Delta\theta_T \times \Delta\tau \times \Delta\nu$ contains significant contributions from physical paths.[6] This implies that wireless channels with clustered multipath components tend to have far fewer than D dominant virtual channel coefficients when operated at large bandwidths and symbol durations and/or with large plurality of antennas. We refer to such channels as *sparse multipath channels* and formalize this notion of multipath sparsity in the following definition.

Definition 1 (d-Sparse Multipath Channels): Suppose that $\mathcal{S}_d = \{(i,k,\ell,m) : |H_v(i,k,\ell,m)| > \epsilon\}$ denotes the set of indices of dominant virtual channel coefficients of a multipath wireless channel for some appropriately chosen ϵ. We say that the channel is *effectively* d-sparse if the number of its effective DoF satisfies $d = |\mathcal{S}_d| \ll D$. Similarly, we say that the channel is *exactly* d-sparse if the same holds for $\epsilon = 0$. In either case, the corresponding set of indices $\mathcal{S}_d$ is termed as the *channel sparsity pattern*.

It is worth mentioning here that, even in the best of scenarios, real-world multipath wireless channels can never be exactly d-sparse due to a multitude of reasons. Nevertheless, analyzing training-based methods for exactly sparse channels enables us to develop insight into the estimation of effectively sparse channels. As an illustration, rearrange the D virtual channel coefficients of $\boldsymbol{\mathcal{H}}$ by decreasing order of magnitude: $|H_v(\pi(1))| \geq |H_v(\pi(2))| \geq \cdots \geq |H_v(\pi(D))|$. Now suppose that the jth largest rearranged coefficient obeys

$$|H_v(\pi(j))| \leq R j^{-1/s} \tag{6}$$

for some $R > 0$ and $s \leq 1$. In the literature, objects that satisfy (6) are termed as *s-compressible* and it is an easy exercise to show that s-compressible objects are effectively $(R/\epsilon)^{1/s}$-sparse for any $\epsilon > 0$ [40]. In other words, results obtained for exactly d-sparse objects can always be extended to s-compressible objects by taking $d = (R/\epsilon)^{1/s}$. As such, we limit ourselves in the rest of this paper to discussing exactly d-sparse channels; the understanding here being that the ensuing analysis can be generalized to effectively d-sparse channels in general and s-compressible channels in particular.

Finally, while statistical characterization of a sparse channel $\boldsymbol{\mathcal{H}}$ is critical from a communication-theoretic viewpoint, either Bayesian (random) or non-Bayesian formulation of $\boldsymbol{\mathcal{H}}$ suffices from the channel estimation perspective. In this paper, we stick to the non-Bayesian paradigm and assume that both the channel sparsity pattern $\mathcal{S}_d$ and the corresponding coefficients $\{H_v(i,k,\ell,m)\}_{\mathcal{S}_d}$ are deterministic but unknown.[7]

B. Sensing and Reconstruction

In wireless systems that rely on training-based methods for channel estimation, the transmitted symbol takes the form

$$\mathbf{x}(t) = \mathbf{x}_{\rm tr}(t) + \mathbf{x}_{\rm data}(t), \qquad 0 \leq t \leq T \tag{7}$$

where $\mathbf{x}_{\rm tr}(t)$ and $\mathbf{x}_{\rm data}(t)$ represent the training signal and data-carrying signal, respectively. Because of the linearity of $\boldsymbol{\mathcal{H}}$, and under the assumption of $\mathbf{x}_{\rm tr}(t)$ being orthogonally multiplexed with $\mathbf{x}_{\rm data}(t)$ in time, frequency, and/or code domain, the resulting signal at the receiver can often be partitioned into two noninterfering components: one corresponding to $\mathbf{x}_{\rm tr}(t)$ and the other corresponding to $\mathbf{x}_{\rm data}(t)$. In order to estimate $\boldsymbol{\mathcal{H}}$, training-based methods ignore the received data and focus only on the training component of the received signal, given by

$$\mathbf{y}_{\rm tr}(t) = \boldsymbol{\mathcal{H}}(\mathbf{x}_{\rm tr}(t)) + \mathbf{z}_{\rm tr}(t), \qquad 0 \leq t \leq T + \tau_{\max} \tag{8}$$

where $\mathbf{z}_{\rm tr}(t)$ is an N_R-dimensional complex additive white Gaussian noise (AWGN) signal that is introduced by the receiver circuitry.

As a first step towards estimating $\boldsymbol{\mathcal{H}}$, the (noisy) received training signal $\mathbf{y}_{\rm tr}(t)$ is matched filtered with the transmitted waveforms at the receiver to obtain an equivalent discrete-time representation of (8). The exact form of this representation depends on a multitude of factors such as selectivity of the channel (nonselective, frequency selective, etc.), type of the signaling waveform used for sensing (single carrier or multicarrier), and number of transmit antennas. While this gives rise to a large number of possible scenarios to be examined, each one corresponding to a different combination of these factors, it turns out that in each case elementary algebraic manipulations of the matched-filtered output result in the

[6]Note that Fig. 1 is an idealized representation in which the *leakage effects* due to the Dirichlet and sinc kernels in (4) have been ignored.

[7]We refer the reader to [54] for a Bayesian formulation of sparse channels.

following general linear form at the receiver [30], [32], [34], [35]:

$$\underbrace{\begin{bmatrix}\mathbf{y}_1 & \dots & \mathbf{y}_{N_R}\end{bmatrix}}_{\mathbf{Y}} = \sqrt{\frac{\mathcal{E}}{N_T}}\mathbf{X}\underbrace{\begin{bmatrix}\mathbf{h}_{v,1} & \dots & \mathbf{h}_{v,N_R}\end{bmatrix}}_{\mathbf{H}_v} + \underbrace{\begin{bmatrix}\mathbf{z}_1 & \dots & \mathbf{z}_{N_R}\end{bmatrix}}_{\mathbf{Z}}. \quad (9)$$

Here, $\mathcal{E}/N_T$ is the average training energy budget per transmit antenna ($\mathcal{E}$ being defined as: $\mathcal{E} = \int_0^T \|\mathbf{x}_{\text{tr}}(t)\|_2^2 dt$), the vectors $\mathbf{h}_{v,i}$, $i = 1, \dots, N_R$, are $N_T L(2M+1)$-dimensional complex vectors comprising of the virtual channel coefficients $\{H_v(i,k,\ell,m)\}$, and we let the AWGN matrix $\mathbf{Z}$ have zero-mean, unit-variance, independent complex-Gaussian entries. Thus, $\mathcal{E}$ is a measure of the training SNR at each receive antenna. Finally, the *sensing matrix* $\mathbf{X}$ is a complex-valued matrix having $D/N_R = N_T L(2M+1)$ columns that are normalized in a way such that $\|\mathbf{X}\|_F^2 = D/N_R$, where $\|\cdot\|_F$ denotes the Frobenius norm. The exact form and dimensions of $\mathbf{X}$ (and hence the dimensions of $\mathbf{Y}$ and $\mathbf{Z}$) in (9) are completely determined by $\mathbf{x}_{\text{tr}}(t)$ and the class to which $\mathcal{H}$ belongs; concrete representations of $\mathbf{X}$ corresponding to the various training signals and channel configurations studied in the paper can be found in Sections V and VI.

As noted in Section I-A, training-based channel estimation methods are characterized by the two distinct—but highly intertwined—operations of sensing and reconstruction. The reconstruction aspect of a training-based method involves designing either a linear or a nonlinear procedure that produces an estimate of $\mathbf{H}_v$ at the receiver from the knowledge of $\mathcal{E}$, $\mathbf{X}$, and $\mathbf{Y}$: $\mathbf{H}_v^{\text{est}} = \mathbf{H}_v^{\text{est}}(\mathcal{E}, \mathbf{X}, \mathbf{Y})$, where the notation is meant tosignify the dependence of $\mathbf{H}_v^{\text{est}}$ on $\mathcal{E}$, $\mathbf{X}$, and $\mathbf{Y}$. The resulting estimate also has associated with it a reconstruction error given by $\mathbb{E}[\|\mathbf{H}_v - \mathbf{H}_v^{\text{est}}\|_F^2]$, where $\mathbb{E}$ denotes the expectation with respect to the distribution of $\mathbf{Z}$. The corresponding sensing component at the transmitter involves probing the channel with a training signal of fixed energy and temporal dimensions that reduces this reconstruction error the most. Specifically, a training signal $\mathbf{x}_{\text{tr}}(t)$ has associated with it the concepts of *temporal training dimensions*, defined as $M_{\text{tr}} = \#\{\text{temporal signal space dimensions occupied by } \mathbf{x}_{\text{tr}}(t)\}$, and *receive training dimensions*, defined as $N_{\text{tr}} = M_{\text{tr}} \times N_R$, as a measure of its spectral efficiency. Therefore, given fixed training energy budget $\mathcal{E}$ and receive training dimensions N_{tr} dedicated to $\mathbf{x}_{\text{tr}}(t)$, the effectiveness of any particular training-based method is measured in terms of the minimum reconstruction error, $\min_{\mathbf{x}_{\text{tr}}(t)} \mathbb{E}[\|\mathbf{H}_v - \mathbf{H}_v^{\text{est}}\|_F^2]$, achieved by that method.

Traditional training-based methods such as those in [5]–[12] have been developed under the implicit assumption that the number of DoF d in $\mathcal{H}$ is roughly the same as the maximum *possible* number of its DoF: $d \approx D$. One direct consequence of this assumption has been that linear procedures have become the *de facto* standard for reconstruction in much of the existing channel estimation literature. In particular, with some notable exceptions such as [19]–[27], many training-based methods proposed in the past make use of the minimum LS error criterion—or its Bayesian counterpart, the minimum mean squared error criterion, for a Bayesian formulation of $\mathcal{H}$—to obtain an estimate of $\mathbf{H}_v$ from $\mathbf{Y}$

$$\mathbf{H}_v^{\text{LS}} = \arg\min_{\mathbf{H}} \left\| \mathbf{Y} - \sqrt{\frac{\mathcal{E}}{N_T}}\mathbf{X}\mathbf{H} \right\|_F^2. \quad (10)$$

This is a well-known problem in the statistical estimation literature [55] and its closed-form solution is given by $\mathbf{H}_v^{\text{LS}} = \sqrt{N_T/\mathcal{E}}\mathbf{X}^\dagger\mathbf{Y}$, where $\mathbf{X}^\dagger$ is the Moore–Penrose pseudoinverse of $\mathbf{X}$. In order to ensure that (10) returns a physically meaningful estimate—in the sense that $\mathbf{H}_v^{\text{LS}}$ equals $\mathbf{H}_v$ in the noiseless setting—reconstruction based on the LS error criterion further requires that the sensing matrix $\mathbf{X}$ has at least as many rows as D/N_R, resulting in the following form for $\mathbf{H}_v^{\text{LS}}$:

$$\mathbf{H}_v^{\text{LS}} = \sqrt{\frac{N_T}{\mathcal{E}}}(\mathbf{X}^{\text{H}}\mathbf{X})^{-1}\mathbf{X}^{\text{H}}\mathbf{Y} \quad (11)$$

where it is assumed that the number of receive training dimensions N_{tr} is such that $\mathbf{X}$ has full column rank. It can be shown in this case that the accompanying reconstruction error of an LS-based channel estimation method is

$$\mathbb{E}\left[\Delta\left(\mathbf{H}_v^{\text{LS}}\right)\right] = \frac{\text{trace}\left(\left(\mathbf{X}^{\text{H}}\mathbf{X}\right)^{-1}\right) \cdot N_R N_T}{\mathcal{E}} \quad (12)$$

where we have used the notation $\Delta(\mathbf{H}) = \|\mathbf{H}_v - \mathbf{H}\|_F^2$ in the above equation. This expression can be simplified further through the use of the arithmetic-harmonic means inequality, resulting in the following lower bound for the reconstruction error (see, e.g., [56, Th. 4.7])

$$\mathbb{E}\left[\Delta\left(\mathbf{H}_v^{\text{LS}}\right)\right] \overset{(a)}{\geq} \frac{(D/N_R)^2 \cdot N_R N_T}{\text{trace}(\mathbf{X}^{\text{H}}\mathbf{X}) \cdot \mathcal{E}} \overset{(b)}{=} \frac{D \cdot N_T}{\mathcal{E}} \quad (13)$$

where the equality in (a) holds if and only if $\mathbf{X}$ has orthonormal columns, while (b) follows from the fact that $\text{trace}(\mathbf{X}^{\text{H}}\mathbf{X}) = \|\mathbf{X}\|_F^2 = D/N_R$. Consequently, an optimal training signal for LS-based estimation methods is the one

Table 2 Summary and Comparison of CCS Results for the Signaling and Channel Configurations Studied in the Paper[a]

Channel Classification	Signaling Waveform	Traditional LS-Based Methods		Compressed Channel Sensing[b, c]	
		Recon. Error	Condition	Recon. Error	Condition
Frequency-Selective Single-Antenna ($D = L$)	Single-Carrier [30]	$\succeq \frac{D}{\mathcal{E}}$	—	$\preceq \frac{d}{\mathcal{E}} \cdot \log D$	$N_o \succeq d^2 \cdot \log D$
	Multi-Carrier [30]–[32]	$\succeq \frac{D}{\mathcal{E}}$	$N_{tr} \succeq D$	$\preceq \frac{d}{\mathcal{E}} \cdot \log D$	$N_{tr} \succeq d \cdot \log^5 N_o$
Doubly-Selective Single-Antenna ($D = L(2M+1)$)	Single-Carrier [34]	$\succeq \frac{D}{\mathcal{E}}$	—	$\preceq \frac{d}{\mathcal{E}} \cdot \log D$	$N_o \succeq d^2 \cdot \log D$
	Multi-Carrier [32]–[34]	$\succeq \frac{D}{\mathcal{E}}$	$N_{tr} \succeq D$	$\preceq \frac{d}{\mathcal{E}} \cdot \log D$	$N_{tr} \succeq d \cdot \log^5 N_o$
Nonselective Multiple-Antenna ($D = N_R N_T$)	[35]	$\succeq \frac{D \cdot N_T}{\mathcal{E}}$	$N_{tr} \succeq D$	$\preceq \frac{d \cdot N_T}{\mathcal{E}} \cdot \log D$	$N_{tr} \succeq d \cdot \log \frac{D}{d}$
Frequency-Selective Multiple-Antenna ($D = N_R N_T L$)	Multi-Carrier [35], [38]	$\succeq \frac{D \cdot N_T}{\mathcal{E}}$	$N_{tr} \succeq D$	$\preceq \frac{d \cdot N_T}{\mathcal{E}} \cdot \log D$	$N_{tr} \succeq d \cdot \log^5 N_o$

[a] Displayed using Hardy's notation for compactness: $f_n \succeq g_n$ and $f_n \preceq g_n$ for $f_n = \Omega(g_n)$ and $f_n = O(g_n)$, respectively.
[b] The results in the first column hold with probability that approaches one with increasing channel and signal space dimensions.
[c] The last two conditions in the second column are for the case when the conditional sparsity of each AoA equals the average AoA sparsity.

that leads to $\mathbf{X}^{\mathrm{H}}\mathbf{X} = \mathbf{I}_{N_T L(2M+1)}$, and much of the emphasis in the previously proposed LS-based training methods has been on designing training signals that are not only optimal in the reconstruction error sense, but are also spectrally efficient in the receive training dimensions sense [5]–[12].

IV. COMPRESSED CHANNEL SENSING: MAIN RESULTS

The preceding discussion brings forth several salient characteristics of traditional training-based methods such as those in [5]–[12]. First, these methods more or less rely on LS-based linear reconstruction strategies, such as the one in (11), at the receiver to obtain an estimate of $\mathbf{H}_v$. Second, because of their reliance on linear reconstruction procedures, the training signals used in these methods must be such that the resulting sensing matrix $\mathbf{X}$ has at least D/N_R rows. As noted in Table 2, depending upon the type of signaling waveforms used for training and the channel class to which $\mathcal{H}$ belongs, this requirement often translates into the condition that the number of receive training dimensions dedicated to $\mathbf{x}_{\text{tr}}(t)$ must be at least as large as the maximum number of DoF in $\mathcal{H}$: $N_{\text{tr}} = \Omega(D)$[8]; see Sections V and VI for further details on this condition. Third, regardless of the eventual choice of training signals, the reconstruction error in these methods is given by $\mathbb{E}[\Delta(\mathbf{H}_v^{\text{LS}})] = \Omega(D \cdot (N_T/\mathcal{E}))$.

In the light of the above observations, a natural question to ask here is: How good is the performance of traditional LS-based training methods? In fact, if one assumes that $\mathcal{H}$ is not sparse (in other words, $d = D$) then it is easy to argue the optimality of these methods [55]: 1) $\mathbf{H}_v^{\text{LS}}$ in this case is also the maximum-likelihood estimate of $\mathbf{H}_v$, and 2) the reconstruction error lower bound (13) is also the Cramer–Rao lower bound, which, as noted earlier, can be achieved through an appropriate choice of the training signal. However, it is arguable whether LS-based channel estimation methods are also optimal for the case when $\mathcal{H}$ is either exactly or effectively d-sparse. In particular, note that exactly d-sparse channels are completely characterized by $2d$ parameters, which correspond to the locations and values of nonzero virtual channel coefficients. Our estimation theory intuition therefore suggests that perhaps $\mathbb{E}[\Delta(\mathbf{H}_v^{\text{est}})] = \Omega(d \cdot (N_T/\mathcal{E}))$ and, for signaling and channel configurations that require $N_{\text{tr}} = \Omega(D)$ in the case of LS-based estimation methods, $N_{\text{tr}} = \Omega(d)$ are the actual fundamental scaling limits in sparse-channel estimation.

In Sections V and VI, we present new training-based estimation methods for six particular signaling and channel configurations (see Table 2) and show that our intuition is indeed correct (modulo polylogarithmic factors). In particular, a key feature of the proposed approach to estimating sparse multipath channels—first presented in [30] and [32] for frequency- and doubly-selective single-antenna channels, respectively, and later generalized in [33]–[35], [37], and [38] to other channel classes—is the use of a sparsity-inducing mixed-norm optimization criterion for reconstruction at the receiver that is based on recent advances in the theory of compressed sensing [29]. This makes the proposed approach—termed as CCS—fundamentally different from the traditional LS-based training methods: the former relies on a nonlinear reconstruction procedure while the latter utilize linear reconstruction techniques. Note that a number of researchers in the recent past have also proposed various training-based methods for sparse multipath channels that are based on nonlinear reconstruction techniques [19]–[27]. The thing that distinguishes CCS from the prior work is that the CCS framework is highly amenable to (scaling) analysis. Specifically, in order to give a summary of the results to come, define the *conditional* sparsity pattern associated with the ith resolvable AoA to be $\mathcal{S}_d(i) = \{(i,k,\ell,m) : (i,k,\ell,m) \in \mathcal{S}_d\}$. Then, it is shown in

[8] Recall Landau's notation: $f_n = \Omega(g_n)$ if $\exists c_o > 0,\ n_o$: $\forall n \geq n_o$, $f_n \geq c_o g_n$; alternatively, we can also write $g_n = O(f_n)$.

Sections V and VI that in the limit of large signal space dimension, we have the following.

1) The performance of CCS in terms of the reconstruction error is provably better than the LS-based training methods. The training signals and reconstruction procedures specified by CCS for the signaling and channel configurations studied in the paper ensure that $\Delta(\mathbf{H}_v^{\text{CCS}}) = O(d \cdot (N_T/\mathcal{E}) \cdot \log D)$ with high probability.
2) CCS is often more spectrally efficient than the LS-based methods. Assume that the conditional sparsity of each AoA is equal to the average AoA sparsity: $|\mathcal{S}_d(i)| = d/N_R, i = 1, \ldots, N_R$. Then, while LS-based methods require that $N_{\text{tr}} = \Omega(D)$ for certain signaling and channel configurations, CCS only requires that $N_{\text{tr}} = \Omega(d \times \text{polylog factor})$ for the same configurations.

Conversely, 1) and 2) together imply that CCS achieves a target reconstruction error scaling using far less energy and, in many instances, latency and bandwidth than that dictated by the traditional LS-based training methods.

Table 2 provides a compact summary of the CCS scaling results as they pertain to the six signaling and channel configurations studied in the paper and compares them to the corresponding results for traditional LS-based training methods. One thing to point out in this table is the CCS condition $N_o = \Omega(d^2 \cdot \log D)$ when using single-carrier signaling waveforms for estimating single-antenna channels. This condition *seems* to be nonexistent for LS-based methods. Note, however, that in order to make the columns of $\mathbf{X}$ as close to orthonormal as possible—a necessary condition for the LS-based reconstruction to achieve the lower bound of (13)—traditional LS-based training methods implicitly require that the temporal signal space dimensions be as large as possible: $N_o \nearrow \infty$. As such, the CCS condition is in fact a relaxation of this implicit requirement for LS-based methods.

As is evident from the preceding discussion and analysis, the scaling performance of CCS is a significant improvement over that of traditional LS-based training methods when it comes to sparse-channel estimation. And while we have purposely avoided providing concrete details of the CCS framework up to this point so as not to clutter the presentation, the rest of the paper is primarily devoted to discussing the exact form of training signals and reconstruction procedures used by CCS for the configurations listed in Table 2. However, since CCS builds on top of the theoretical framework provided by compressed sensing, it is advantageous to briefly review some facts about compressed sensing before proceeding further.

A. Review of Compressed Sensing

Compressed sensing (CS) is a relatively new area of theoretical research that lies at the intersection of a number of other research areas such as signal processing, statistics, and computational harmonic analysis; see [57]–[59] for a tutorial overview of some of the foundational developments in CS. In order to review the theoretical underpinnings of CS, consider the following classical linear measurement model:

$$r_i = \psi_i^{\mathrm{T}} \boldsymbol{\theta} + \eta_i, \qquad i = 1, \ldots, n \tag{14}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose operation, $\psi_i \in \mathbb{C}^p$ is a known *measurement vector*, $\boldsymbol{\theta} \in \mathbb{C}^p$ is an unknown vector, and $\eta_i \in \mathbb{C}$ is either stochastic noise or deterministic perturbation. This measurement model can also be written compactly using the matrix–vector representation: $\mathbf{r} = \boldsymbol{\Psi}\boldsymbol{\theta} + \boldsymbol{\eta}$. Here, the measurement matrix $\boldsymbol{\Psi}$ is composed of the measurement vectors as its rows and the goal is to reliably reconstruct $\boldsymbol{\theta}$ from the knowledge of $\mathbf{r}$ and $\boldsymbol{\Psi}$.

One of the central tenets of CS theory is that if $\boldsymbol{\theta}$ is sparse (has only a few nonzero entries) or approximately sparse (when reordered by magnitude, its entries decay rapidly), then a relatively small number—typically much smaller than p—of appropriately designed measurement vectors can capture most of its salient information. In addition, recent theoretical results have established that $\boldsymbol{\theta}$ in this case can be reliably reconstructed from $\mathbf{r}$ by making use of either tractable mixed-norm optimization programs [39]–[41], efficient greedy algorithms [28], [60], or fast iterative thresholding methods [61], [62]; see [63] for the references of other relevant CS reconstruction procedures. As one would expect, proofs which establish that certain reconstruction procedures reliably reconstruct $\boldsymbol{\theta}$ in the end depend only upon some property of the measurement matrix $\boldsymbol{\Psi}$ and the level of sparsity (or approximate sparsity) of $\boldsymbol{\theta}$. In particular, one key property of $\boldsymbol{\Psi}$ that has been very useful in proving the optimality of a number of CS reconstruction procedures is the so-called restricted isometry property (RIP) [64].

Definition 2 (RIP): Let $\boldsymbol{\Psi}$ be an $n \times p$ (real- or complex-valued) matrix having unit ℓ_2-norm columns. For each integer $S \in \mathbb{N}$, we say that $\boldsymbol{\Psi}$ satisfies the RIP of order S with parameter $\delta_S \in (0, 1)$—and write $\boldsymbol{\Psi} \in \text{RIP}(S, \delta_S)$—if for all $\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_0 \le S$

$$(1 - \delta_S)\|\boldsymbol{\theta}\|_2^2 \le \|\boldsymbol{\Psi}\boldsymbol{\theta}\|_2^2 \le (1 + \delta_S)\|\boldsymbol{\theta}\|_2^2 \tag{15}$$

where $\|\cdot\|_2$ denotes the ℓ_2-norm of a vector and $\|\cdot\|_0$ counts the number of nonzero entries of its argument.

Note that the RIP of order S is essentially a statement about the singular values of all $n \times S$ submatrices of $\boldsymbol{\Psi}$. And while no algorithms are known to date that can explicitly check the RIP for a given matrix in polynomial time, one of the reasons that has led to the widespread applicability of CS theory in various application areas is the revelation that certain probabilistic constructions of matrices satisfy the RIP with high probability. For example, let the $n \times p$ matrix $\boldsymbol{\Psi}$ be such that either 1) its entries are drawn independently from a $\mathcal{N}(0, 1/n)$ distribution, or 2) its rows are first sampled uniformly at random (without replacement) from the set of rows of a $p \times p$ unitary matrix with entries of magnitude $O(1/\sqrt{p})$ and then scaled by a factor of $\sqrt{p/n}$. Then, for every $\delta_S \in (0,1)$, it has been established that $\boldsymbol{\Psi} \in \text{RIP}(S, \delta_S)$ with probability exceeding $1 - e^{-O(n)}$ in the former case if $n = \Omega(S \cdot \log(p/S))$ [65], while $\boldsymbol{\Psi} \in \text{RIP}(S, \delta_S)$ with probability exceeding $1 - p^{-O(\delta_S^2)}$ in the latter case as long as $n = \Omega(S \cdot \log^5 p)$ [66].[9]

As noted earlier, there exist a number of CS reconstruction procedures in the literature that are based on the RIP characterization of measurement matrices. The one among them that is the most relevant to our formulation of the sparse-channel estimation problem—and one that will be frequently referred to in the rest of this paper—goes by the name of *Dantzig selector* (DS) [40]. In particular, there are three main reasons that we have chosen to make the DS an integral part of our discussion on the CCS framework. First, it is one of the few reconstruction methods in the CS literature that are guaranteed to perform near-optimally vis-à-vis stochastic noise—the others being the risk minimization method of Haupt and Nowak [67] and the lasso [41], which also goes by the name of basis pursuit denoising [39]. Second, unlike the method of [67], it is highly computationally tractable since it can be recast as a linear program. Third, it comes with the cleanest and most interpretable reconstruction error bounds that we know for both sparse and approximately sparse signals. It is worth mentioning here though that some of the recent results in the literature seem to suggest that the lasso also enjoys many of the useful properties of the DS, including the reconstruction error bounds that appear very similar to those of the DS [68], [69]. As such, making use of the lasso in practical settings can sometimes be more computationally attractive because of the availability of a wide range of efficient software packages, such as GPSR [70] and SpaRSA [71], for solving it. However, since a RIP-based characterization of the lasso that parallels that of the DS does not exist to date, we limit ourselves in this paper to discussing the DS only. The following theorem, which is a slight modification of the results of [40], states the reconstruction error performance of the DS for *any* $\boldsymbol{\theta} \in \mathbb{C}^p$.[10]

Theorem 1 (The Dantzig Selector [40]): Let $\boldsymbol{\theta} \in \mathbb{C}^p$ be a deterministic but unknown signal and let $\boldsymbol{\Psi\theta} + \boldsymbol{\eta} = \mathbf{r} \in \mathbb{C}^n$ be a vector of noisy measurements, where the $n \times p$ matrix $\boldsymbol{\Psi}$ has unit ℓ_2-norm columns and the complex AWGN vector $\boldsymbol{\eta}$ is distributed as $\mathcal{CN}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$. Further, let $\boldsymbol{\Psi} \in \text{RIP}(2S, \delta_{2S})$ for some $\delta_{2S} < 1/3$ and choose $\lambda = \sqrt{2\sigma^2(1+a)\log p}$ for any $a \geq 0$. Then, the estimate $\boldsymbol{\theta}^{\mathsf{DS}}$ obtained as a solution to the optimization program

$$\boldsymbol{\theta}^{\mathsf{DS}} = \arg\min_{\tilde{\boldsymbol{\theta}} \in \mathbb{C}^p} \|\tilde{\boldsymbol{\theta}}\|_1 \quad \text{s.t.} \quad \left\|\boldsymbol{\Psi}^{\mathrm{H}}(\mathbf{r} - \boldsymbol{\Psi}\tilde{\boldsymbol{\theta}})\right\|_\infty \leq \lambda \qquad \text{(DS)}$$

satisfies

$$\|\boldsymbol{\theta}^{\mathsf{DS}} - \boldsymbol{\theta}\|_2^2 \leq c_1^2 \min_{1 \leq m \leq S} \left(\lambda\sqrt{m} + \frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_m\|_1}{\sqrt{m}}\right)^2 \qquad (16)$$

with probability at least $1 - 2\left(\sqrt{\pi(1+a)\log p} \cdot p^a\right)^{-1}$. Here, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the ℓ_1- and ℓ_∞-norm of a vector, respectively, $\boldsymbol{\theta}_m$ is the vector formed by setting all but the m largest (in magnitude) entries of the true signal $\boldsymbol{\theta}$ to zero, and the constant $c_1 = 16/(1 - 3\delta_{2S})^2$.

In the following, we will often make use of the shorthand notation $\boldsymbol{\theta}^{\mathsf{DS}} = DS(\boldsymbol{\Psi}, \mathbf{r}, \lambda)$ to denote a solution of the optimization program (DS) that takes as input $\boldsymbol{\Psi}$, $\mathbf{r}$, and λ. A few remarks are in order now concerning the performance of (DS). First, note that (16) is akin to saying that if $\boldsymbol{\Psi} \in \text{RIP}(2S, \delta_{2S})$ then $\boldsymbol{\theta}^{\mathsf{DS}}$ more or less recovers S largest (in magnitude) entries of $\boldsymbol{\theta}$ that are above the noise floor σ^2. In particular, (16) implies that: 1) if $\boldsymbol{\theta}$ is S-sparse, then $\|\boldsymbol{\theta}^{\mathsf{DS}} - \boldsymbol{\theta}\|_2^2 = O(S\sigma^2 \log p)$, which has roughly the same scaling behavior as the usual parametric error of $S\sigma^2$ [55]; and 2) if $\boldsymbol{\theta}$ is s-compressible and $S \geq (R/\sigma)^{1/s}$, then $\|\boldsymbol{\theta}^{\mathsf{DS}} - \boldsymbol{\theta}\|_2^2 = O(R^s(\sigma^2)^{1-s/2} \log p)$, which has the same scaling behavior as the minimax error for s-compressible objects [40]. Second, the probability deviation bound (16) can be converted into a similar-looking bound on the expected value of $\|\boldsymbol{\theta}^{\mathsf{DS}} - \boldsymbol{\theta}\|_2^2$ at the expense of some extra work. This is due to the well-known fact that, for positive random variables X, we have $\mathbb{E}[X] = \int_0^t \Pr\{X > t\} dt$. Third, although the parameter a in Theorem 1 affects the upper bound (16) and the accompanying probability of failure differently, it does not affect the scaling behavior of the reconstruction error. This is because of the fact that the upper bound (16) increases only linearly with increasing a, whereas the probability of failure decreases exponentially with increasing a. An obvious choice for a in this regard is $a = 1$, which results in the probability of failure $2\left(\sqrt{2\pi\log p} \cdot p\right)^{-1}$. Finally, note that the statement of Theorem 1 assumes that $\boldsymbol{\Psi} \in \text{RIP}(2S, \delta_{2S})$ almost surely. However, if this is not true then in this case Theorem 1 simply implies that (16) is satisfied with probability at least $1 - 2\max\{2\left(\sqrt{\pi(1+a)\log p} \cdot p^a\right)^{-1}, \Pr\{\boldsymbol{\Psi} \notin \text{RIP}(2S, \delta_{2S})\}\}$. We are now ready to discuss the specifics of CCS for sparse multipath channels.

[9]In fact, the actual condition in the latter case only requires that $n = \Omega(S \log^2(p) \log(S \log p) \log^2(S))$ [66]; for the sake of compactness, however, we use the lax requirement $n = \Omega(S \cdot \log^5 p)$ in the paper.

[10]We refer the reader to the discussion following [56, Th. 2.13] for an outline of the differences between Theorem 1 and the results stated in [40].

V. COMPRESSED CHANNEL SENSING: SINGLE-ANTENNA CHANNELS

A. Estimating Sparse Frequency-Selective Channels

For a single-antenna channel that is frequency selective, the virtual representation (3) of the channel reduces to

$$\tilde{H}(f) = \sum_{\ell=0}^{L-1} H_v(\ell) e^{-j2\pi\frac{\ell}{W}f} \tag{17}$$

and the corresponding received training signal is [cf., (8)]

$$y_{\mathrm{tr}}(t) \approx \sum_{\ell=0}^{L-1} H_v(\ell) x_{\mathrm{tr}}(t-\ell/W) + z_{\mathrm{tr}}(t), \quad 0 \le t \le T + \tau_{\max}. \tag{18}$$

In general, two types of signaling waveforms are commonly employed to communicate over a frequency-selective channel, namely, (single-carrier) spread spectrum (SS) waveforms and (multicarrier) orthogonal frequency division multiplexing (OFDM) waveforms. We begin our discussion of the CCS framework for sparse frequency-selective channels by focusing first on SS signaling and then on OFDM signaling.

1) Spread Spectrum Signaling: In the case of SS signaling, the training signal $x_{\mathrm{tr}}(t)$ can be represented as

$$x_{\mathrm{tr}}(t) = \sqrt{\mathcal{E}} \sum_{n=0}^{N_o-1} x_n g(t - nT_c), \qquad 0 \le t \le T \tag{19}$$

where $g(t)$ is a unit-energy *chip waveform* ($\int |g(t)|^2 dt = 1$), $T_c \approx 1/W$ is the chip duration, and $\{x_n\}$ represents the N_o-dimensional spreading code associated with the training signal that also has unit energy ($\sum_n |x_n|^2 = 1$). In this case, chip-matched filtering of the received training signal (18) yields the discrete-time representation [30]

$$\mathbf{y} = \sqrt{\mathcal{E}}(\mathbf{x} * \mathbf{h}_v) + \mathbf{z} \quad \Longrightarrow \quad \mathbf{y} = \sqrt{\mathcal{E}}\mathbf{X}\mathbf{h}_v + \mathbf{z} \tag{20}$$

where $*$ denotes discrete-time convolution, $\mathbf{h}_v \in \mathbb{C}^L$ is the vector of virtual channel coefficients $\{H_v(\ell)\}$, and $\mathbf{x} \in \mathbb{C}^{N_o}$ is composed of the spreading code $\{x_n\}$. Further, define $\tilde{N}_o = N_o + L - 1$. Then, $\mathbf{z}$ is an AWGN vector distributed as $\mathcal{CN}(\mathbf{0}_{\tilde{N}_o}, \mathbf{I}_{\tilde{N}_o})$, while $\mathbf{X}$ is an $\tilde{N}_o \times L$ Toeplitz (convolutional) matrix whose first row and first column can be explicitly written as $[x_0 \ \mathbf{0}_{L-1}^{\mathrm{T}}]$ and $[\mathbf{x}^{\mathrm{T}} \ \mathbf{0}_{L-1}^{\mathrm{T}}]^{\mathrm{T}}$, respectively.

Note that (20) is the single-antenna version of the standard form (9). Therefore, from (13), the reconstruction error scaling of LS-based training methods in this case is given by $\mathbb{E}[\Delta(\mathbf{h}_v^{\mathsf{LS}})] = \Omega(L/\mathcal{E})$. We now describe the CCS approach to estimating frequency-selective channels using SS signaling, which was first described in [30]. In particular, we show that for d-sparse channels it leads to an improvement of a factor of about L/d (modulo a log factor).

CCS-1: SS Training and Reconstruction

Training: Pick the spreading code $\{x_n\}_{n=0}^{N_o-1}$ associated with $x_{\mathrm{tr}}(t)$ to be a sequence of independent and identically distributed (i.i.d.) Rademacher variables taking values $+1/\sqrt{N_o}$ or $-1/\sqrt{N_o}$ with probability 1/2 each.

Reconstruction: Fix any $a \ge 0$ and choose the parameter $\lambda = \sqrt{2\mathcal{E}(1+a)\log L}$. The CCS estimate of $\mathbf{h}_v$ is then given as follows: $\mathbf{h}_v^{\mathsf{CCS}} = DS(\sqrt{\mathcal{E}}\mathbf{X}, \mathbf{y}, \lambda)$.

The following theorem summarizes the performance of CCS-1 in terms of the reconstruction error scaling.

Theorem 2: Pick $\delta_{2d} \in (0, 1/3)$, $c_2 \in (0, \delta_{2d}^2/4)$, and define $c_3 = 48/(\delta_{2d}^2 - 4c_2)$. Next, suppose that the number of temporal signal space dimensions $N_o (= TW) \ge c_3 d^2 \log L$. Then, under the assumption that $\|\mathbf{h}_v\|_0 \le d$, the CCS estimate of $\mathbf{h}_v$ satisfies

$$\Delta\left(\mathbf{h}_v^{\mathsf{CCS}}\right) \le c_0^2 \cdot \frac{d}{\mathcal{E}} \cdot \log L \tag{21}$$

with probability $\ge 1 - 2\max\{2(\pi(1+a)\log L \cdot L^{2a})^{-1/2}, \exp(-(c_2 N_o/4d^2))\}$. Here, the absolute constant $c_0 > 0$ is defined as $c_0 = 4\sqrt{2(1+a)}/(1 - 3\delta_{2d})$.

Note that the frequency-selective channel being d-sparse simply means that $\|\mathbf{h}_v\|_0 \le d \ll L$. Therefore, the proof of this theorem essentially follows from the statement of Theorem 1 and [30, Th. 2] (also, see [56, Th. 3.5]), where it was shown that $\Pr\{\mathbf{X} \notin \mathrm{RIP}(2d, \delta_{2d})\} \le \exp(-(c_2 N_o/4d^2))$ for any $\delta_{2d} \in (0, 1/3)$, provided $N_o \ge c_3 d^2 \log L$.

Remark 1: Theorem 2 is stated above in its entire generality and, in its current form, depends on three key parameters: δ_{2d}, c_2, and a. Nevertheless, it is instructive to specify one possible choice of these parameters so as to be a little more concrete regarding the scaling performance of CCS-1. In this regard, note that the first term in the max expression in Theorem 2 is upper bounded by L^{-a} for any $a \ge 1$. It therefore makes intuitive sense to choose $a = 1$ and pick c_2 such that we have $\exp(-(c_2 N_o/4d^2)) \le L^{-1}$. This in turn requires that $c_2 = \delta_{2d}^2/8$ and therefore if one picks $\delta_{2d} = 1/4$ then Theorem 2 implies that $\Delta(\mathbf{h}_v^{\mathsf{CCS}}) \le 1024 \cdot (d/\mathcal{E}) \cdot \log L$ with probability exceeding $1 - 2L^{-1}$ as long as $N_o \ge 1536 \cdot d^2 \log L$. Note that explicit values of the numerical constants can also be obtained in a similar manner for subsequent theorems in the paper. It is worth pointing out here though that extensive simulations carried out in [30]–[34], [36], and [37] suggest that actual values of the CCS numerical constants are in fact much smaller than the ones predicted by the CCS theory.

2) OFDM Signaling: If OFDM signaling is used for communication, then the training signal takes the form

$$x_{\rm tr}(t) = \sqrt{\frac{\mathcal{E}}{N_{\rm tr}}} \sum_{n \in \mathcal{S}_{\rm tr}} g(t) e^{j2\pi \frac{n}{T} t}, \qquad 0 \le t \le T \tag{22}$$

where $g(t)$ is simply a prototype pulse having unit energy, $\mathcal{S}_{\rm tr} \subset \mathcal{S} = \{0, 1, \ldots, N_o - 1\}$ is the set of indices of *pilot tones* used for training, and $N_{\rm tr}$—the number of receive training dimensions—denotes the total number of pilot tones in this case, $N_{\rm tr} = |\mathcal{S}_{\rm tr}|$, and is a measure of the spectral efficiency of $x_{\rm tr}(t)$. Finally, matched filtering the received training signal (18) with the OFDM basis waveforms $\{g(t)e^{j2\pi(n/T)t}\}_{\mathcal{S}_{\rm tr}}$ and collecting the output into a vector again yields the standard form [1]: $\mathbf{y} = \sqrt{\mathcal{E}}\mathbf{X}\mathbf{h}_v + \mathbf{z}$. The difference here is that $\mathbf{X}$ is now an $N_{\rm tr} \times L$ sensing matrix that comprises $\{(1/\sqrt{N_{\rm tr}})[1 \ \omega_{N_o}^{n\cdot 1} \ \ldots \ \omega_{N_o}^{n\cdot(L-1)}] : n \in \mathcal{S}_{\rm tr}\}$ as its rows, where $\omega_{N_o} = e^{-j(2\pi/N_o)}$, and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_{N_{\rm tr}}, \mathbf{I}_{N_{\rm tr}})$.[11]

Note that the form of $\mathbf{X}$ in the case of OFDM signaling imposes the condition that $N_{\rm tr} \ge L$ for $\mathbf{X}$ to have full column rank. In order to estimate a frequency-selective channel using OFDM signaling, LS-based methods, such as [6], therefore require that $N_{\rm tr} = \Omega(L)$ and, from (13), at best yield $\mathbb{E}[\Delta(\mathbf{h}_v^{\sf LS})] = \Omega(L/\mathcal{E})$. In contrast, we now outline the CCS approach to this problem, described initially in [30]–[32], and quantify its advantage over traditional LS-based training methods for sparse channels.

CCS-2: OFDM Training and Reconstruction

Training: Pick $\mathcal{S}_{\rm tr}$—the set of indices of pilot tones—to be a set of $N_{\rm tr}$ indices sampled uniformly at random (without replacement) from $\mathcal{S} = \{0, 1, \ldots, N_o - 1\}$.

Reconstruction: Same as in CCS-1 (but with the sensing matrix $\mathbf{X}$ specified as above).

Below, we summarize the performance of CCS-2 in terms of the reconstruction error scaling.

Theorem 3: Suppose N_o, $d > 2$ and pick $\delta_{2d} \in (0, 1/3)$. Next, let the number of pilot tones $N_{\rm tr} \ge 2c_4 d \log^5 N_o$. Then, the reconstruction error of $\mathbf{h}_v^{\sf CCS}$ satisfies (21) with probability at least $1 - 2\max\{2(\pi(1+a)\log L \cdot L^{2a})^{-1/2}, 10N_o^{-c_5\delta_{2d}^2}\}$. Here, $c_4, c_5 > 0$ are absolute numerical constants that do not depend on $N_{\rm tr}$, N_o, or d.

The proof of this theorem follows trivially from Theorem 1 and the fact that $\mathbf{X}$ in this case corresponds to a *column submatrix* of a matrix whose (appropriately normalized) rows are randomly sampled from an $N_o \times N_o$ unitary DFT matrix. Therefore, from the definition of RIP and [66, Th. 3.3], $\Pr\{\mathbf{X} \notin \text{RIP}(2d, \delta_{2d})\} \le 10N_o^{-c_5\delta_{2d}^2}$ for any $\delta_{2d} \in (0, 1/3)$, provided $N_{\rm tr} \ge 2c_4 d \log^5 N_o$.

[11]Note that $\mathbf{X}$ has this particular form as long as $T \gg \tau_{\max}$ [1], which also implies $N_o \gg L$.

B. Estimating Sparse Doubly-Selective Channels

In the case of a single-antenna channel that is doubly-selective, the virtual representation (3) reduces to

$$\tilde{H}(t,f) = \sum_{\ell=0}^{L-1} \sum_{m=-M}^{M} H_v(\ell, m) e^{-j2\pi \frac{\ell}{W} f} e^{j2\pi \frac{m}{T} t} \tag{23}$$

and the received training signal can be written as

$$y_{\rm tr}(t) \approx \sum_{\ell=0}^{L-1} \sum_{m=-M}^{M} H_v(\ell, m) e^{j2\pi \frac{m}{T} t} x_{\rm tr}(t - \ell/W) + z_{\rm tr}(t), \qquad 0 \le t \le T + \tau_{\max}. \tag{24}$$

Signaling waveforms that are often used to communicate over a doubly-selective channel can be broadly categorized as (single-carrier) SS waveforms and (multicarrier) short-time Fourier (STF) waveforms, which are a generalization of OFDM waveforms for doubly-selective channels [72], [73]. Below, we discuss the specifics of the CCS framework for sparse doubly-selective channels as it pertains to both SS and STF signaling waveforms.

1) Spread Spectrum Signaling: The SS training signal $x_{\rm tr}(t)$ in the case of a doubly-selective channel has the same form as in (19). The difference here is that the chip-matched-filtered output in this case looks different from the one in (20). Specifically, define again $\tilde{N}_o = N_o + L - 1$. Then, chip-matched filtering the received training signal (24) yields [34]

$$y_n = \sqrt{\mathcal{E}} \sum_{\ell=0}^{L-1} \sum_{m=-M}^{M} H_v(\ell, m) e^{j2\pi \frac{m}{N_o} n} x_{n-\ell} + z_n, \qquad n = 0, 1, \ldots, \tilde{N}_o - 1. \tag{25}$$

Nevertheless, it is established in [34, Sec. III-A] that this received training data can be represented into the standard form (9) by collecting it into a vector $\mathbf{y} \in \mathbb{C}^{\tilde{N}_o}$ and algebraically manipulating the right-hand side of (25). That is, $\mathbf{y} = \sqrt{\mathcal{E}}\mathbf{X}\mathbf{h}_v + \mathbf{z}$, where $\mathbf{h}_v \in \mathbb{C}^{L(2M+1)}$ is the vector of channel coefficients $\{H_v(\ell, m)\}$, $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_{\tilde{N}_o}, \mathbf{I}_{\tilde{N}_o})$, and the sensing matrix $\mathbf{X}$ is an $\tilde{N}_o \times L(2M+1)$ block matrix of the form

$$\mathbf{X} = [\mathbf{X}_{-M} \quad \ldots \quad \mathbf{X}_0 \quad \ldots \quad \mathbf{X}_M]. \tag{26}$$

Here, each block $\mathbf{X}_m$ has dimensions $\tilde{N}_o \times L$ and is of the form $\mathbf{X}_m = \mathbf{W}_m \mathbf{T}$, where $\mathbf{W}_m$ is an $\tilde{N}_o \times \tilde{N}_o$ diagonal matrix given by $\mathbf{W}_m = \mathrm{diag}(\omega_{N_o}^{-m\cdot 0}, \omega_{N_o}^{-m\cdot 1}, \ldots, \omega_{N_o}^{-m\cdot(\tilde{N}_o-1)})$ and $\mathbf{T}$ is an $\tilde{N}_o \times L$ Toeplitz matrix whose first row and first column are given by $[x_0 \ \mathbf{0}_{L-1}^{\mathrm{T}}]$ and $[\mathbf{x}^{\mathrm{T}} \ \mathbf{0}_{L-1}^{\mathrm{T}}]^{\mathrm{T}}$, respectively.

Note that under the assumption that the doubly-selective channel is underspread ($\tau_{\max}\nu_{\max} \ll 1$), we have the condition $TW \gg \tau_{\max}\nu_{\max}TW \Rightarrow \tilde{N}_o > L(2M+1)$. This, combined with the form of $\mathbf{X}$, ensures that the sensing matrix in this case has full column rank and training-based methods can use the LS criterion (10) without further conditions, resulting in $\mathbb{E}[\Delta(\mathbf{h}_v^{\mathsf{LS}})] = \Omega(L(2M+1)/\mathcal{E})$. Below, we describe the CCS approach to estimating doubly-selective channels using SS signaling, first presented in [34], and provide an upper bound on the corresponding reconstruction error for d-sparse channels that is significantly better than $\Omega(L(2M+1)/\mathcal{E})$.

CCS-3: SS Training and Reconstruction

Training: Same as specified in the case of CCS-1.

Reconstruction: Fix any $a \geq 0$ and choose the parameter $\lambda = \sqrt{2\mathcal{E}(1+a)\log L(2M+1)}$. The CCS estimate of $\mathbf{h}_v$ is then given as follows: $\mathbf{h}_v^{\mathsf{CCS}} = DS(\sqrt{\mathcal{E}}\mathbf{X}, \mathbf{y}, \lambda)$.

Theorem 4: Pick $\delta_{2d} \in (0, 1/3)$, $c_6 \in (0, \delta_{2d}^2/8)$, and define $c_7 = 128/(\delta_{2d}^2 - 8c_6)$. Next, suppose that the number of temporal signal space dimensions $N_o \geq c_7 d^2 \log L(2M+1)$. Then, under the assumption that $\|\mathbf{h}_v\|_0 \leq d$, the CCS estimate of $\mathbf{h}_v$ satisfies

$$\Delta\left(\mathbf{h}_v^{\mathsf{CCS}}\right) \leq c_0^2 \cdot \frac{d}{\mathcal{E}} \cdot \log L(2M+1) \qquad (27)$$

with probability $\geq 1 - 2\max\{2(\pi(1+a)\log L(2M+1)\cdot (L(2M+1))^{2a})^{-1/2}, \exp(-(c_6 N_o/4d^2))\}$. Here, the numerical constant $c_0 > 0$ is the same as defined in Theorem 2.

Note that the key ingredient in the proof of this theorem is characterizing the RIP of the sensing matrix given in (26). Therefore, this theorem in essence is a direct consequence of [34, Th. 2] (also, see [56, Th. 3.9]), where it was established that $\Pr\{\mathbf{X} \notin \mathrm{RIP}(2d, \delta_{2d})\} \leq \exp(-(c_6 N_o/4d^2))$ for any value of the parameter $\delta_{2d} \in (0, 1/3)$, provided $N_o \geq c_7 d^2 \log L$.

2) STF Signaling: In the case of STF signaling, which is a generalization of OFDM signaling to counteract the time selectivity of doubly-selective channels [72], [73], the training signal $x_{\mathrm{tr}}(t)$ is of the form

$$x_{\mathrm{tr}}(t) = \sqrt{\frac{\mathcal{E}}{N_{\mathrm{tr}}}} \sum_{(n,m)\in\mathcal{S}_{\mathrm{tr}}} g(t - nT_o)e^{j2\pi m W_o t}, \qquad t \in [0, T] \qquad (28)$$

where $g(t)$ is again a prototype pulse having unit energy, $\mathcal{S}_{\mathrm{tr}} \subset \mathcal{S} = \{0, 1, \ldots, N_t - 1\} \times \{0, 1, \ldots, N_f - 1\}$ is the set of indices of STF pilot tones used for training, and N_{tr}, a measure of the spectral efficiency of $x_{\mathrm{tr}}(t)$, denotes the total number of pilot tones: $N_{\mathrm{tr}} = |\mathcal{S}_{\mathrm{tr}}|$. Here, the parameters $T_o \in [\tau_{\max}, 1/\nu_{\max}]$ and $W_o \in [\nu_{\max}, 1/\tau_{\max}]$ correspond to the time and frequency separation of the STF basis waveforms $\{g(t - nT_o)e^{j2\pi m W_o t}\}$ in the time-frequency plane, respectively, and are chosen so that $T_o W_o = 1$ [73]. Finally, the total number of STF basis waveforms available for communication/training are $N_t N_f = N_o$, where $N_t = T/T_o$ and $N_f = W/W_o$.

For sufficiently underspread channels, corresponding to $\tau_{\max}\nu_{\max} < 0.01$, it has been shown in [73] that matched filtering the received training signal (24) with the STF basis waveforms $\{g(t - nT_o)e^{j2\pi m W_o t}\}_{\mathcal{S}_{\mathrm{tr}}}$ yields

$$y_{n,m} \approx \sqrt{\frac{\mathcal{E}}{N_{\mathrm{tr}}}} H_{n,m} + z_{n,m}, \qquad (n,m) \in \mathcal{S}_{\mathrm{tr}} \qquad (29)$$

where the STF channel coefficients are related to $\tilde{H}(t,f)$ as $H_{n,m} \approx \tilde{H}(t,f)|_{(t,f)=(nT_o, mW_o)}$. As shown in [32] and [34, Sec. IV-A], collection of this matched-filtered output into a vector $\mathbf{y} \in \mathbb{C}^{N_{\mathrm{tr}}}$ followed by simple manipulations yields the standard form $\mathbf{y} = \sqrt{\mathcal{E}}\mathbf{X}\mathbf{h}_v + \mathbf{z}$, where $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_{N_{\mathrm{tr}}}, \mathbf{I}_{N_{\mathrm{tr}}})$ and the $N_{\mathrm{tr}} \times L(2M+1)$ matrix $\mathbf{X}$ comprises

$$\left\{ \frac{1}{\sqrt{N_{\mathrm{tr}}}} \left[\omega_{N_t}^{n\cdot M} \quad \omega_{N_t}^{n\cdot(M-1)} \quad \ldots \quad \omega_{N_t}^{-n\cdot M}\right] \otimes \left[1 \quad \omega_{N_f}^{m\cdot 1} \quad \ldots \quad \omega_{N_f}^{m\cdot(L-1)}\right] : (n,m) \in \mathcal{S}_{\mathrm{tr}} \right\}$$

as its rows.[12] Consequently, traditional LS-based training methods impose the condition $N_{\mathrm{tr}} = \Omega(L(2M+1))$ in order to satisfy the requirement that $\mathbf{X}$ has full column rank in this setting and yield, at best, $\mathbb{E}[\Delta(\mathbf{h}_v^{\mathsf{LS}})] = \Omega(L(2M+1)/\mathcal{E})$. We now describe the CCS approach to estimating d-sparse doubly-selective channels using STF signaling, which not only has a lower reconstruction error scaling than the LS-based approach but is also spectrally more efficient in the limit of large signal space dimension.

Remark 2: Note that the main difference between [32] and [34, Sec. IV-A] is that [32] chooses T_o and W_o such that $T_o W_o > 1$ and uses two sets of biorthogonal waveforms at the transmitter and the receiver, the overall effect of this being that it makes the approximation in

[12] Here, $\otimes$ is used to denote the Kronecker product; also, since we have that $T_o \in [\tau_{\max}, 1/\nu_{\max}]$ and $W_o \in [\nu_{\max}, 1/\tau_{\max}]$, this implies that $N_t \geq 2M+1$ and $N_f \geq L$.

(29) more accurate at the expense of some loss in spectral efficiency [72].

CCS-4: STF Training and Reconstruction

Training: Pick $\mathcal{S}_{\rm tr}$, the set of indices of pilot tones, to be a set of $N_{\rm tr}$ indices sampled uniformly at random (without replacement) from $\mathcal{S} = \{0, 1, \ldots, N_t - 1\} \times \{0, 1, \ldots, N_f - 1\}$.

Reconstruction: Same as in CCS-3 (with the sensing matrix $\mathbf{X}$ specified as above).

Theorem 5: Suppose N_o, $d > 2$ and pick $\delta_{2d} \in (0, 1/3)$. Next, let the number of pilot tones $N_{\rm tr} \geq 2c_4 d \log^5 N_o$. Then, the reconstruction error of $\mathbf{h}_v^{\sf CCS}$ satisfies (27) with probability exceeding $1 - 2\max\{2(\pi(1+a)\log L(2M+1)\cdot (L(2M+1))^{2a})^{-1/2}, 10N_o^{-c_5\delta_{2d}^2}\}$. Here, the numerical constants $c_4, c_5 > 0$ are the same as described in Theorem 3.

Note that [32] and [34, Th. 3] specify the conditions under which the sensing matrix $\mathbf{X}$ arising in this setting satisfies RIP, and Theorem 5 follows immediately from that characterization. This concludes our discussion of the CCS framework for single-antenna channels; see Table 2 for a summary of the results presented in this section.

VI. COMPRESSED CHANNEL SENSING: MULTIPLE-ANTENNA CHANNELS

A. Estimating Sparse Nonselective Channels

The virtual representation of a nonselective multiple-input–multiple-output (MIMO) channel is of the form [cf., (3)]

$$\tilde{\mathbf{H}} = \underbrace{\sum_{i=1}^{N_R} \sum_{k=1}^{N_T} H_v(i,k) \mathbf{a}_R\left(\frac{i}{N_R}\right) \mathbf{a}_T^{\rm H}\left(\frac{k}{N_T}\right)}_{\mathbf{A}_R \mathbf{H}_v^{\rm T} \mathbf{A}_T^{\rm H}}. \tag{30}$$

Here, $\mathbf{A}_R$ and $\mathbf{A}_T$ are $N_R \times N_R$ and $N_T \times N_T$ unitary matrices (comprising $\{\mathbf{a}_R(i/N_R)\}$ and $\{\mathbf{a}_T(k/N_T)\}$ as their columns), respectively, and $\mathbf{H}_v = [\mathbf{h}_{v,1} \ldots \mathbf{h}_{v,N_R}]$ is an $N_T \times N_R$ matrix of virtual channel coefficients in which the ith column $\mathbf{h}_{v,i} \in \mathbb{C}^{N_T}$ consists of the coefficients $\{H_v(i,k)\}$ associated with the ith resolvable AoA.

Generally, the training signal used to probe a nonselective MIMO channel can be written as

$$\mathbf{x}_{\rm tr}(t) = \sqrt{\frac{\mathcal{E}}{N_T}} \sum_{n=0}^{M_{\rm tr}-1} \mathbf{x}_n g\left(t - \frac{n}{W}\right), \qquad 0 \leq t \leq \frac{M_{\rm tr}}{W} \tag{31}$$

where $g(t)$ is a unit-energy prototype pulse, the (vector-valued) training sequence is denoted by $\{\mathbf{x}_n \in \mathbb{C}^{N_T}\}$ and has energy $\sum_n \|\mathbf{x}_n\|_2^2 = N_T$, and $M_{\rm tr}$, the number of temporal training dimensions, denotes the total number of time slots dedicated to training in this setting. Trivially, matched filtering the received training signal $\mathbf{y}_{\rm tr}(t) = \tilde{\mathbf{H}}\mathbf{x}_{\rm tr}(t) + \mathbf{z}_{\rm tr}(t)$ in this case with time-shifted versions of the prototype pulse yields

$$\tilde{\mathbf{y}}_n = \sqrt{\frac{\mathcal{E}}{N_T}} \tilde{\mathbf{H}} \mathbf{x}_n + \tilde{\mathbf{z}}_n, \qquad n = 0, \ldots, M_{\rm tr} - 1 \tag{32}$$

where $\{\tilde{\mathbf{y}}_n \in \mathbb{C}^{N_R}\}$ is the (vector-valued) received training sequence and the AWGN vectors $\{\tilde{\mathbf{z}}_n\}$ are independently distributed as $\mathcal{CN}(\mathbf{0}_{N_R}, \mathbf{I}_{N_R})$. As shown in [35, Sec. III], premultiplying the $\tilde{\mathbf{y}}_n$'s in this case with $\mathbf{A}_R^{\rm H}$ and row-wise stacking the resulting vectors into an $M_{\rm tr} \times N_R$ matrix $\mathbf{Y}$ yields the standard linear form (9): $\mathbf{Y} = \sqrt{\mathcal{E}/N_T}\mathbf{X}\mathbf{H}_v + \mathbf{Z}$, where the entries of $\mathbf{Z}$ are independently distributed as $\mathcal{CN}(0, 1)$. Here, $\mathbf{X}$ is an $M_{\rm tr} \times N_T$ matrix of the form

$$\mathbf{X} = [\mathbf{x}_0 \quad \mathbf{x}_1 \quad \ldots \quad \mathbf{x}_{M_{\rm tr}-1}]^{\rm T} \mathbf{A}_T^* \tag{33}$$

where $(\cdot)^*$ denotes the conjugation operation. In order to estimate nonselective MIMO channels, traditional LS-based methods such as those in [9] and [10] therefore require that $M_{\rm tr} = \Omega(N_T)$ so as to ensure that $\mathbf{X}$ has full column rank and produce an estimate that satisfies $\mathbb{E}[\Delta(\mathbf{H}_v^{\sf LS})] = \Omega(N_R N_T^2/\mathcal{E})$. In particular, note that the condition $M_{\rm tr} = \Omega(N_T)$ means that LS-based methods in this case require the number of receive training dimensions to satisfy $N_{\rm tr} = M_{\rm tr} N_R = \Omega(N_R N_T)$. In contrast, we now describe the CCS approach to this problem for d-sparse channels and quantify its performance in terms of the reconstruction error scaling and receive training dimensions. Before proceeding further, however, recall that the conditional sparsity pattern associated with the ith resolvable AoA is $\mathcal{S}_d(i) = \{(i,k) : (i,k) \in \mathcal{S}_d\}$, and define the *maximum* conditional AoA sparsity as $\bar{d} = \max_i |\mathcal{S}_d(i)|$.

CCS-5: Training and Reconstruction

Training: Pick $\{\mathbf{x}_n, n = 0, \ldots, M_{\rm tr} - 1\}$ to be a training sequence of i.i.d. Rademacher vectors in which each entry independently takes the value $+1/\sqrt{M_{\rm tr}}$ or $-1/\sqrt{M_{\rm tr}}$ with probability 1/2 each.

Reconstruction: Fix any $a \geq 0$ and choose the parameter $\lambda = \sqrt{2\mathcal{E}(1+a)(\log N_R N_T)/N_T}$. The CCS estimate of the $N_T \times N_R$ matrix $\mathbf{H}_v$ is then given as follows: $\mathbf{H}_v^{\sf CCS} = [DS(\sqrt{\mathcal{E}/N_T}\mathbf{X}, \mathbf{y}_1, \lambda) \ldots DS(\sqrt{\mathcal{E}/N_T}\mathbf{X}, \mathbf{y}_{N_R}, \lambda)]$.

Theorem 6: Pick $\delta_{2\bar{d}} \in (0, 1/3)$, $c_8 \in (0, \delta_{2\bar{d}}^2(3 - \delta_{2\bar{d}})/48)$, and define $c_9 = 384 \log(12/\delta_{2\bar{d}})/(3\delta_{2\bar{d}}^2 - \delta_{2\bar{d}}^3 - 48c_8)$. Next, let the number of training time slots $M_{\text{tr}} \geq c_9 \bar{d} \log N_T$. Then, under the assumption that $\sum_{i=1}^{N_R} \|\mathbf{h}_{v,i}\|_0 \leq d$, the CCS estimate of $\mathbf{H}_v$ satisfies

$$\Delta\left(\mathbf{H}_v^{\text{CCS}}\right) \leq c_0^2 \cdot \frac{d \cdot N_T}{\mathcal{E}} \cdot \log N_R N_T \tag{34}$$

with probability exceeding $1 - 2\max\{2(\pi(1+a)\log N_R N_T \cdot (N_R N_T)^{2a})^{-1/2}, \exp(-c_8 M_{\text{tr}})\}$. Here, the constant $c_0 > 0$ is the same as in Theorem 2 (with $\delta_{2\bar{d}}$ in place of δ_{2d}).

The proof of this theorem is provided in [56, Th. 4.12], and is based on [65, Th. 5.2] and a slight modification of the proof of Theorem 1 in [40]. Before concluding this discussion, it is worth evaluating the minimum number of receive training dimensions required for the CCS approach to succeed in the case of sparse nonselective MIMO channels. From the structure of the training signal in CCS-5, we have that $N_{\text{tr}} = M_{\text{tr}} N_R = \Omega(\bar{d} N_R \cdot \log N_T/\bar{d})$ for CCS, which—modulo the logarithmic factor—always scales better than $N_{\text{tr}} = \Omega(N_R N_T)$ for traditional LS-based methods. In particular, for the case when the scattering geometry is such that the conditional AoA sparsity is equal to the average AoA sparsity ($\bar{d} = d/N_R$), we have from the previous arguments that CCS requires $N_{\text{tr}} = \Omega(d \cdot \log N_R N_T/d)$.

B. Estimating Sparse Frequency-Selective Channels

From (3), the virtual representation of a frequency-selective MIMO channel can be written as

$$\tilde{\mathbf{H}}(f) = \sum_{\ell=0}^{L-1} \mathbf{A}_R \mathbf{H}_v^{\mathrm{T}}(\ell) \mathbf{A}_T^{\mathrm{H}} e^{-j2\pi\frac{\ell}{W}f} \tag{35}$$

where the unitary matrices $\mathbf{A}_R$ and $\mathbf{A}_T$ are as given in (30), and $\mathbf{H}_v(\ell) = [\mathbf{h}_{v,1}(\ell)\ \ldots\ \mathbf{h}_{v,N_R}(\ell)]$ is an $N_T \times N_R$ matrix in which the ith column $\mathbf{h}_{v,i}(\ell) \in \mathbb{C}^{N_T}$ consists of the coefficients $\{H_v(i,k,\ell)\}$. As in the case of single-antenna channels, both SS and OFDM waveforms can be used to communicate over a frequency-selective MIMO channel. For the sake of this exposition, however, we limit ourselves to a block OFDM signaling structure similar to the one studied in [11, Sec. IV-B] and [12, Sec. IV].

Specifically, we assume that the N_o-dimensional symbol consists of $N_t \geq N_T$ (vector-valued) OFDM symbols. Since signaling using a block of N_t OFDM symbols is essentially STF signaling with parameters $T_o = T/N_t$ and $W_o = N_t/T$, we make use of the STF formulation developed in Section V to carry out the analysis in this section. In particular, the training signal in this case can be written using the notation in (28) as

$$\mathbf{x}_{\text{tr}}(t) = \sqrt{\frac{\mathcal{E}}{N_T}} \sum_{(n,m)\in\mathcal{S}_{\text{tr}}} \mathbf{x}_{n,m} g(t - nT_o) e^{j2\pi m W_o t} \tag{36}$$

where $\mathcal{S}_{\text{tr}} \subset \mathcal{S} = \{0, 1, \ldots, N_t - 1\} \times \{0, 1, \ldots, N_f - 1\}$ here is again the set of indices of pilot tones used for training, while $\{\mathbf{x}_{n,m} \in \mathbb{C}^{N_T}\}$ is the (vector-valued) training sequence having energy $\sum_{\mathcal{S}_{\text{tr}}} \|\mathbf{x}_{n,m}\|_2^2 = N_T$. The main difference here from the single-antenna formulation is that we use M_{tr}, instead of N_{tr}, to denote the total number of pilot tones (equivalently, the number of temporal training dimensions): $M_{\text{tr}} = |\mathcal{S}_{\text{tr}}|$.[13]

From [73], matched filtering $\mathbf{y}_{\text{tr}}(t) = \boldsymbol{\mathcal{H}}(\mathbf{x}_{\text{tr}}(t)) + \mathbf{z}_{\text{tr}}(t)$ in this case with $\{g(t - nT_o)e^{j2\pi m W_o t}\}_{\mathcal{S}_{\text{tr}}}$ yields

$$\mathbf{y}_{n,m} = \sqrt{\frac{\mathcal{E}}{N_T}} \mathbf{H}_m \mathbf{x}_{n,m} + \mathbf{z}_{n,m}, \qquad (n,m) \in \mathcal{S}_{\text{tr}} \tag{37}$$

where the AWGN vectors $\{\mathbf{z}_{n,m}\}$ are independently distributed as $\mathcal{CN}(\mathbf{0}_{N_R}, \mathbf{I}_{N_R})$, while the (matrix-valued) channel coefficients are related to $\tilde{\mathbf{H}}(f)$ as $\mathbf{H}_m \approx \tilde{\mathbf{H}}(f)|_{f=mW_o}$. As in the case of nonselective MIMO channels, we can premultiply the received training vectors $\mathbf{y}_{n,m}$'s with $\mathbf{A}_R^{\mathrm{H}}$ and row-wise stack the resulting vectors $\mathbf{A}_R^{\mathrm{H}} \mathbf{y}_{n,m}$ to yield an $M_{\text{tr}} \times N_R$ matrix $\mathbf{Y}$. Further, as in [35, Sec. IV], the right-hand side of (37) can be manipulated to express this matrix into the standard form (9): $\mathbf{Y} = \sqrt{\mathcal{E}/N_T}\mathbf{X}\mathbf{H}_v + \mathbf{Z}$. Here, $\mathbf{H}_v = [\mathbf{h}_{v,1}\ \ldots\ \mathbf{h}_{v,N_R}]$ is the $N_T L \times N_R$ channel matrix in which the ith column consists of the coefficients $\{H_v(i,k,\ell)\}$ associated with the ith resolvable AoA, while $\mathbf{X}$ is an $M_{\text{tr}} \times N_T L$ matrix comprising $\{[1\ \omega_{N_f}^{m \cdot 1}\ \ldots\ \omega_{N_f}^{m\cdot(L-1)}] \otimes \mathbf{x}_{n,m}^{\mathrm{T}} \mathbf{A}_T^* : (n,m) \in \mathcal{S}_{\text{tr}}\}$ as its rows.

Once again, the form of the sensing matrix $\mathbf{X}$ here dictates that $M_{\text{tr}} = \Omega(N_T L)$ for the traditional LS-based methods such as those in [11] and [12] to obtain a meaningful estimate of $\mathbf{H}_v$, and we have from (13) that $\mathbb{E}[\Delta(\mathbf{H}_v^{\text{LS}})] = \Omega(N_R N_T^2 L/\mathcal{E})$ in that case. Note that in terms of the receive training dimensions, this implies that the LS-based methods require $N_{\text{tr}} = \Omega(N_R N_T L)$ for frequency-selective MIMO channels. In contrast, we now provide the CCS approach to estimating d-sparse channels using block OFDM signaling and quantify its performance advantage over traditional methods. The following discussion once again makes use of the definition of maximum

[13]Note that the number of temporal and receive training dimensions is the same in the case of single-antenna channels.

conditional sparsity within the AoA spread of the channel: $\bar{d} = \max_i |\{(i,k,\ell) : (i,k,\ell) \in \mathcal{S}_d\}|$.

CCS-6: OFDM Training and Reconstruction

Training: Pick $\mathcal{S}_{\rm tr}$ to be a set of $M_{\rm tr}$ ordered pairs sampled uniformly at random (without replacement) from the set $\mathcal{S} = \{0,1,\ldots,N_T-1\} \times \{0,1,\ldots,N_f-1\}$ and define the corresponding sequence of $M_{\rm tr}$ training vectors as $\{\mathbf{x}_{n,m} = \sqrt{N_T/M_{\rm tr}}\mathbf{e}_{n+1} : (n,m) \in \mathcal{S}_{\rm tr}\}$, where $\mathbf{e}_i$ denotes the ith standard basis element of $\mathbb{C}^{N_T}$.

Reconstruction: Fix any $a \geq 0$ and choose the parameter $\lambda = \sqrt{2\mathcal{E}(1+a)(\log N_R N_T L)/N_T}$. The CCS estimate of the $N_T L \times N_R$ matrix $\mathbf{H}_v$ is then given as follows: $\mathbf{H}_v^{\sf CCS} = [DS(\sqrt{\mathcal{E}/N_T}\mathbf{X}, \mathbf{y}_1, \lambda) \ \ldots \ DS(\sqrt{\mathcal{E}/N_T}\mathbf{X}, \mathbf{y}_{N_R}, \lambda)]$.

Theorem 7: Suppose N_o, $\bar{d} > 2$ and pick $\delta_{2\bar{d}} \in (0, 1/3)$. Next, choose the number of pilot tones $M_{\rm tr} \geq 2c_4\bar{d}\log^5 N_o$. Then, under the assumption that $\sum_{i=1}^{N_R} \|\mathbf{h}_{v,i}\|_0 \leq d$, the CCS estimate of $\mathbf{H}_v$ satisfies

$$\Delta\left(\mathbf{H}_v^{\sf CCS}\right) \leq c_0^2 \cdot \frac{d \cdot N_T}{\mathcal{E}} \cdot \log N_R N_T L. \tag{38}$$

with probability at least $1 - 2\max\{2(\pi(1+a)\log N_R N_T L \cdot (N_R N_T L)^{2a})^{-1/2}, 10N_o^{-c_5\delta_{2\bar{d}}^2}\}$. Here, the absolute numerical constants $c_4, c_5 > 0$ are the same as described in Theorem 3, while $c_0 = 4\sqrt{2(1+a)}/(1-3\delta_{2\bar{d}})$.

The proof of this theorem is omitted here for brevity, but depends to a large extent on first characterizing the RIP of the sensing matrix $\mathbf{X}$ arising in this setting using the proof technique of [34, Th. 3] and then essentially follows along the lines of the proof of Theorem 6 in [56]. One key observation from the description of the training signal above is that $N_{\rm tr} = \Omega(\bar{d}N_R \cdot \log^5 N_T N_f)$ for CCS. In particular, for the case of conditional AoA sparsity being equal to the average AoA sparsity (and since $N_t \geq N_T$), this implies that CCS requires $N_{\rm tr} = \Omega(d \cdot \log^5 N_o)$ in this setting as opposed to $N_{\rm tr} = \Omega(N_R N_T L)$ for traditional LS-based training methods—a significant improvement in terms of the training spectral efficiency when operating at large bandwidths and with large plurality of antennas.

VII. DISCUSSION

There is a large body of physical evidence that suggests that multipath signal components in many wireless channels tend to be distributed as clusters within their respective channel spreads. Consequently, as the world transitions from single-antenna communication systems operating at small bandwidths (typically in the megahertz range) to multiple-antenna ones operating at large bandwidths (possibly in the gigahertz range), the representation of such channels in appropriate bases starts to look sparse. This has obvious implications for the design and implementation of training-based channel estimation methods. Since, by definition, the *effective* intrinsic dimension d of sparse multipath channels tends to be much smaller than their extrinsic dimension D, one expects to estimate them using far fewer communication resources than that dictated by traditional methods based on the LS criterion. Equally importantly, however, sparsity of multipath channels also has implications for the design and implementation of the communication aspects of a wireless system that is equipped with a limited-rate feedback channel. First, if the channel-estimation module at the receiver yields a sparse estimate of the channel (something which LS-based reconstruction fails to accomplish) then, even at a low rate, that estimate can also be reliably fed back to the transmitter. Second, this reliable knowledge of the channel sparsity structure at both the transmitter and the receiver can be exploited by agile transceivers, such as the ones in [74], for improved communication performance.

In this paper, we have described a new approach to estimating multipath channels that have a sparse representation in the Fourier basis. The proposed approach is based on some of the recent advances in the theory of compressed sensing and is accordingly termed as CCS. Ignoring polylogarithmic factors, two distinct features of CCS are as follows: 1) it has a reconstruction error that scales like $O(d)$ as opposed to $\Omega(D)$ for traditional LS-based methods, and 2) it requires the number of receive training dimensions $N_{\rm tr}$ to scale like $N_{\rm tr} = \Omega(d)$ for certain signaling and channel configurations as opposed to $N_{\rm tr} = \Omega(D)$ for LS-based methods. It is also worth pointing out here that the CCS results presented in Section V-B using STF signaling waveforms have been generalized in [33] to channel representations that make use of a basis other than the Fourier one. Similarly, the CCS results presented in Section VI-B have been generalized in [38] to frequency-selective MIMO channels using $N_t = 1$ OFDM symbols and in [56, Sec. 4.5.2] to doubly-selective MIMO channels using STF signaling waveforms.

Admittedly, there are several other theoretical and practical aspects of CCS that need discussing but space limitations forbid us from exploring them in detail in this paper. Below, however, we briefly comment on some of these aspects. First, while there is no discussion of the scaling optimality of CCS in this paper, it has been shown in [30] and [34] that its theoretical performance for single-antenna sparse channels comes within a (poly)-logarithmic factor of an (unrealizable) training-based method that clairvoyantly knows the channel sparsity pattern. Somewhat similar theoretical arguments can be made to argue the near-optimal scaling nature of CCS for multiple-antenna sparse channels also. Second, extensive numerical simulations carried out in [30]–[34], [36], and [37] for a number of practically relevant scenarios have established that the performance of CCS is markedly superior to that of traditional methods based on the LS criterion and of nontraditional methods based on MUSIC and ESPRIT algorithms [75]. Note that the fact that parametric methods such as MUSIC and ESPRIT are not optimal for estimating sparse channels is hardly surprising. This is because it is possible for a channel to have a small number of *resolvable* paths but still have a very large number of underlying *physical* paths, especially in the case of diffuse scattering. Third, as noted in Section III-A, one expects the representation of real-world multipath channels in certain bases to be only effectively sparse because of practical constraints such as the leakage effect, diffuse scattering, and nonideal filters at the transmitter and the receiver. While our primary focus in this paper has been on characterizing the scaling performance of CCS for exactly sparse channels, it works equally well for effectively sparse channels thanks to the near-optimal nature of the Dantzig selector. Finally, and perhaps most importantly for the success of the envisioned wireless systems, CCS can be leveraged to design efficient training-based methods for estimating sparse *network* channels—a critical component of the emerging area of cognitive radio in which wireless transceivers sense and adapt to the wireless environment for enhanced spectral efficiency and interference management. ■

Acknowledgment

The authors gratefully acknowledge the many helpful suggestions of the two anonymous reviewers and Guest Associate Editor Prof. R. Baraniuk.

REFERENCES

[1] A. Goldsmith, *Wireless Communications*. New York: Cambridge Univ. Press, 2005.

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

[3] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[4] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[5] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[6] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.

[7] M.-A. R. Baissas and A. M. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 2037–2049, Aug. 2002.

[8] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.

[9] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Sep. 1999, pp. 958–966.

[10] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" IEEE Trans. Inf. Theory, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[11] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.

[12] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[13] R. J.-M. Cramer, R. A. Scholtz, and M. Z. Win, "Evaluation of an ultra-wide-band propagation channel," *IEEE Trans. Antennas Propag.*, vol. 50, no. 5, pp. 561–570, May 2002.

[14] J. C. Preisig and G. B. Deane, "Surface wave focusing and acoustic communications in the surf zone," *J. Acoust. Soc. Amer.*, vol. 116, no. 4, pp. 2067–2080, Oct. 2004.

[15] A. F. Molisch, "Ultrawideband propagation channels—Theory, measurement, and modeling," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1528–1545, Sep. 2005.

[16] Z. Yan, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Univ. Wisconsin-Madison, Madison, Tech. Rep., Feb. 2007.

[17] N. Czink, X. Yin, H. Ozcelik, M. Herdin, E. Bonek, and B. H. Fleury, "Cluster characteristics in a MIMO indoor propagation environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1465–1475, Apr. 2007.

[18] L. Vuokko, V.-M. Kolmonen, J. Salo, and P. Vainikainen, "Measurement of large-scale cluster power characteristics for geometric channel models," *IEEE Trans. Antennas Propag.*, vol. 55, no. 11, pp. 3361–3365, Nov. 2007.

[19] M. Kocic, D. Brady, and S. Merriam, "Reduced-complexity RLS estimation for shallow-water channels," in *Proc. Symp. Autonom. Underwater Veh. Technol.*, Cambridge, MA, Jul. 1994, pp. 165–170.

[20] S. F. Cotter and B. D. Rao, "Matching pursuit based decision-feedback equalizers," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Istanbul, Turkey, Jun. 2000, pp. 2713–2716.

[21] S. F. Cotter and B. D. Rao, "The adaptive matching pursuit algorithm for estimation and equalization of sparse time-varying channels," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Oct./Nov. 2000, pp. 1772–1776.

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

[23] W. Dongming, H. Bing, Z. Junhui, G. Xiqi, and Y. Xiaohu, "Channel estimation algorithms for broadband MIMO-OFDM sparse channel," in *Proc. IEEE Int. Symp. Personal Indoor Mobile Radio Commun.*, Beijing, China, Sep. 2003, pp. 1929–1933.

[24] M. R. Raghavendra and K. Giridhar, "Improving channel estimation in OFDM systems for sparse multipath channels," *IEEE Signal Process. Lett.*, vol. 12, no. 1, pp. 52–55, Jan. 2005.

[25] C.-J. Wu and D. W. Lin, "A group matching pursuit algorithm for sparse channel estimation for OFDM transmission," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Toulouse, France, May 2006, pp. 429–432.

[26] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1743–1753, May 2007.

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

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

[29] *IEEE Signal Processing Mag.*, vol. 25, *Special Issue on Compressive Sampling*, no. 2, Mar. 2008.

[30] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. Annu. Conf. Inf. Sci. Syst.*, Princeton, NJ, Mar. 2008, pp. 5–10.

[31] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Las Vegas, NV, Apr. 2008, pp. 3469–3472.

[32] 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 Int. Conf. Acoust. Speech Signal Process.*, Las Vegas, NV, Apr. 2008, pp. 2885–2888.

[33] G. Tauböck and F. Hlawatsch, "Compressed sensing based estimation of doubly selective channels using a sparsity-optimized basis expansion," in *Proc. Eur. Signal Process. Conf.*, Lausanne, Switzerland, Aug. 2008.

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

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

[36] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," in *Proc. IEEE OCEANS Conf.*, Bremen, Germany, May 2009, DOI: 10.1109/OCEANSE.2009.5278228.

[37] D. Angelosante, E. Grossi, G. Giannakis, and M. Lops, "Sparsity-aware estimation of CDMA system parameters," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Perugia, Italy, Jun. 2009, pp. 697–701.

[38] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "A restricted isometry property for structurally-subsampled unitary matrices," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Sep./Oct. 2009, pp. 1005–1012.

[39] 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, Jan. 1998.

[40] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when*p* is much larger than*n*," *Ann. Stat.*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.

[41] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

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

[43] A. Richter, C. Schneider, M. Landmann, and R. Thomä, "Parameter estimation results of specular and dense multipath components in micro-cell scenarios," in *Proc. Int. Symp. Wireless Personal Multimedia Commun.*, Abano Terme, Italy, pp. V2-90–V2-94, Sep. 2004.

[44] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, Kyösti, D. Laurenson, G. Matz, A. F. Molisch, C. Oestges, and H. Özcelik, "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2007, Article ID 19070, 19 pages, 2007. DOI: 10.1155/2007/19070.

[45] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[46] P. Bello, "Measurement of random time-variant linear channels," *IEEE Trans. Inf. Theory*, vol. 15, no. 4, pp. 469–475, Jul. 1969.

[47] J. G. Proakis, *Digital Communications,* 4th ed. New York: McGraw-Hill, 2001.

[48] D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng., vol. 25, no. 1, pp. 4–27, Jan. 2000.

[49] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, no. 3, pp. 292–300, Mar. 1976.

[50] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York: Wiley-Interscience, 1969.

[51] A. M. Sayeed, "A virtual representation for time- and frequency-selective correlated MIMO channels," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process*, Hong Kong, Apr. 2003, pp. 648–651.

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

[53] 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, no. 10, pp. 1969–1986, Oct. 1998.

[54] Y. Selen and E. G. Larsson, "RAKE receiver for channels with a sparse impulse response," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3175–3180, Sep. 2007.

[55] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[56] W. U. Bajwa, "New information processing theory and methods for exploiting sparsity in wireless systems," Ph.D. dissertation, Univ. Wisconsin-Madison, Madison, 2009.

[57] J. A. Tropp, "Topics in sparse approximation," Ph.D. dissertation, Univ. Texas at Austin, Austin, 2004.

[58] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Mathematicians*, Madrid, Spain, Aug. 2006, vol. III, pp. 1433–1452.

[59] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[60] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[61] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.

[62] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[63] Rice Univ., *The compressive sensing resources webpage*, Houston, TX, 2009. [Online]. Available: http://www.dsp.ece.rice.edu/cs

[64] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci., Ser. I*, vol. 346, pp. 589–592, 2008.

[65] R. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," in *Constructive Approximation*. New York: Springer-Verlag, 2008.

[66] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, no. 8, pp. 1025–1045, Aug. 2008.

[67] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[68] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis of lasso and Dantzig selector," *Ann. Stat.*, vol. 37, no. 4, pp. 1705–1732, Aug. 2009.

[69] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Near-oracle performance of basis pursuit under random Noise. [Online]. Available: arXiv:0903.4579v1

[70] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[71] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.

[72] W. Kozek and A. F. Molisch, "Nonorthogonal pulse shapes for multicarrier communications in doubly dispersive channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1579–1589, Oct. 1998.

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

[74] A. M. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, Jun. 2007.

[75] H. L. Van Trees, *Optimum Array Processing: Detection, Estimation, and Modulation Theory (Part IV)*. New York: Wiley, 2002.

ABOUT THE AUTHORS

Waheed U. Bajwa (Member, IEEE) received the B.E. degree (with honors) in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2001 and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, Madison, in 2005 and 2009, respectively.

He was an Intern at Communications Enabling Technologies, Islamabad, Pakistan—the research arm of Avaz Networks Inc., Irvine, CA (now Quartics LLC)—during 2000–2001, a Design Engineer at Communications Enabling Technologies from 2001 to 2003, and an R&D Intern in the RF and Photonics Lab of GE Global Research, Niskayuna, NY, during summer 2006. He was also briefly affiliated with the Center for Advanced Research in Engineering, Islamabad, Pakistan, during 2003. He is currently a Postdoctoral Research Associate in the Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ. His research interests include high-dimensional inference, statistical signal processing, wireless communications, learning theory, and applications in biological sciences, radar and image processing, and cognitive radio/ad-hoc networks.

Dr. Bajwa received the Best in Academics Gold Medal and President's Gold Medal in Electrical Engineering from the National University of Sciences and Technology (NUST) in 2001, and the Morgridge Distinguished Graduate Fellowship from the University of Wisconsin-Madison in 2003. He was Junior NUST Student of the Year (2000), Wisconsin Union Poker Series Champion (Spring 2008), and President of the University of Wisconsin-Madison chapter of Golden Key International Honor Society (2009). He is currently a member of the Pakistan Engineering Council and Golden Key International Honor Society.

Jarvis Haupt received the B.S. (with highest distinction), M.S., and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, Madison, in 2002, 2003, and 2009, respectively.

He is currently a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering, Rice University, Houston, TX. His research interests include high-dimensional statistical inference, sparse recovery, adaptive sampling techniques, statistical signal processing and learning theory, and applications in communications, network science, remote sensing, and imaging.

Dr. Haupt has completed internships at Georgia Pacific, Domtar Industries, Cray, and L-3 Communications/Integrated Systems. He was the recipient of several academic awards, including the Wisconsin Academic Excellence Scholarship, the Ford Motor Company Scholarship, the Consolidated Papers Tuition Scholarship, the Frank D. Cady Mathematics Scholarship, and the Claude and Dora Richardson Distinguished Fellowship. He served as Co-Chair of the Teaching Improvement Program at the University of Wisconsin-Madison for two semesters, and was awarded Honorable Mention for the Gerald Holdridge Teaching Award for his work as a teaching assistant. Dr. Haupt is also a Certified Professional Locksmith.

Akbar M. Sayeed (Senior Member, IEEE) received the B.S. degree from the University of Wisconsin-Madison, Madison, in 1991 and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, Urbana, in 1993 and 1996, respectively, all in electrical engineering.

He is currently Professor of Electrical and Computer Engineering at the University of Wisconsin-Madison, Madison. He was a Postdoctoral Fellow at Rice University, Houston, TX, from 1996 to 1997. His current research interests include wireless communications, statistical signal processing, multidimensional communication theory, information theory, learning theory, time-frequency analysis, and applications in wireless communication networks and sensor networks.

Dr. Sayeed is a recipient of the Robert T. Chien Memorial Award (1996) for his doctoral work at Illinois, the National Science Foundation CAREER Award (1999), the Office of Naval Research Young Investigator Award (2001), and the UW Grainger Junior Faculty Fellowship (2003). He is currently serving on the signal processing for communications technical committee of the IEEE Signal Processing Society. He also served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002, and as the Technical Program Co-Chair for the 2007 IEEE Statistical Signal Processing Workshop and the 2008 IEEE Communication Theory Workshop.

Robert Nowak (Fellow, IEEE) received the B.S. (with highest distinction), M.S., and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, Madison, in 1990, 1992, and 1995, respectively.

He was a Postdoctoral Fellow at Rice University, Houston, TX, in 1995–1996, an Assistant Professor at Michigan State University, East Lansing, from 1996 to 1999, held Assistant and Associate Professor positions at Rice University from 1999 to 2003, and was a Visiting Professor at INRIA, France, in 2001. He is now the McFarland-Bascom Professor of Engineering at the University of Wisconsin-Madison. His research interests include signal processing, machine learning, imaging and network science, and applications in communications, bioimaging, and systems biology.

Dr. Nowak has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the Secretary of the SIAM Activity Group on Imaging Science, and is currently an Associate Editor for the *ACM Transactions on Sensor Networks*. He was General Chair for the 2007 IEEE Statistical Signal Processing Workshop and Technical Program Chair for the 2003 IEEE Statistical Signal Processing Workshop and the 2004 IEEE/ACM International Symposium on Information Processing in Sensor Networks. He received the General Electric Genius of Invention Award in 1993, the National Science Foundation CAREER Award in 1997, the Army Research Office Young Investigator Program Award in 1999, the Office of Naval Research Young Investigator Program Award in 2000, and the IEEE Signal Processing Society Young Author Best Paper Award in 2000.