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

Waheed U. Bajwa, Jarvis Haupt, Akbar Sayeed, and Robert Nowak

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 of linear reconstruction techniques (such as the maximum likelihood or the minimum mean squared error estimators), are known to be optimal for rich multipath channels. However, physical arguments and growing experimental evidence suggest that wireless channels encountered in practice 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 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, achieves a target reconstruction error using far less energy and, in many instances, latency and bandwidth than that dictated by the traditional training-based methods.

Index Terms

Channel estimation, compressed sensing, Dantzig selector, least-squares estimation, orthogonal frequency division multiplexing, restricted isometry property, sparse channel modeling, spread spectrum, training-based estimation.

I. Introduction

Wireless technology has had and continues to have a profound impact on our society. It is arguably one of the leading drivers of the Information Revolution in the 21st century and is expected to play an increasingly important role in the global economic growth as the world transitions from a manufacturing-based economy to an information-based economy. Despite having a history of more than a century of rapid technological advancements, the field of wireless communications remains far from mature. A number of key technical challenges still need to be overcome in order to realize our vision of a future with ubiquitous wireless connectivity. Foremost among these challenges

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mails: bajwa@cae.wisc.edu, jdhaupt@wisc.edu, akbar@engr.wisc.edu, nowak@engr.wisc.edu).

is designing wireless systems that not only support data rates comparable to that of wired systems, but also enable increased mobility while maintaining constant, reliable connectivity under resource constraints—energy, latency, and bandwidth constraints being the strictest among them. Successfully addressing this and similar challenges requires significant technical advances on multiple fronts. One such front is the development of signal processing techniques for estimating multipath wireless channels using minimal resources, and this paper summarizes the findings of some of our recent efforts in this direction.

A. Background

In a typical terrestrial 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 terrestrial 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 impacts 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, 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 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 in the sequel; 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 are 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 comprised 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 physical wireless channels encountered in practice 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 nearly zero when operating at large bandwidths and symbol durations and/or with large plurality of antennas [13]–[16]. However, traditional training-based methods—relying 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. Recently, a number of researchers have tried to address this problem and proposed training signals and reconstruction strategies that are tailored to the anticipated characteristics of sparse multipath channels [17]–[22]. 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.

Example 1 – Delay Sparsity

To motivate the idea of multipath sparsity, take the simple example of a single-antenna transmitter–receiver pair communicating at large bandwidth in a static environment. The underlying multipath channel in this setting is best described as a linear, time-invariant (LTI) system whose impulse response consists of only a few dominant echoes (due to the reflections of the transmitted signal from the surrounding objects such as buildings and hills). In this case, discretized approximation of the channel impulse response (CIR)—essentially obtained by sampling the CIR at the *Nyquist rate* (twice the one-sided bandwidth)—can be accurately described through a sparse set of coefficients since most of the samples of the continuous-time CIR would be zero because of the high sampling rate.

B. Scope of this Paper

By leveraging key ideas from the theory of compressed sensing [23], we have recently proposed new training-based estimation methods for various classes of sparse single- and multiple-antenna channels that are provably more effective than their traditional counterparts [24]–[26]. In particular, the training-based methods detailed in [24]–[26] have been analytically shown to achieve a target reconstruction error using far less energy and, in many instances, latency and bandwidth than that dictated by the traditional methods. As in the case of previous research, the exact nature of training signals employed by our proposed methods varies with the type of signaling waveforms used for sensing (e.g., single- or multi-carrier 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 our training-based methods is the use of sparsity-inducing mixed-norm optimization criteria such as the Dantzig selector [27] and the lasso [28] 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 training-based methods.

The goal of this paper is to complement our 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 focusses 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 multi-dimensional 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.

II. Multipath Wireless Channel Modeling

Signal propagation in a wireless channel over multiple paths gives rise to a large number of propagation parameters. However, exact knowledge of these parameters is not critical for reliable communication of data over the channel. Rather, from a communication-theoretic perspective, we are only interested in characterizing the *interaction* between the physical propagation environment and the transmitter/receiver signal space. In this section, we review a virtual modeling framework for multipath wireless channels that captures this interaction through Nyquist sampling of the angle-delay-Doppler space. As we will later see, 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

TABLE I

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

sensing theory to sparse-channel estimation.

A. Multipath Wireless Channels: Physical Model

Consider, without loss of generality, a multiple-antenna channel with half-wavelength spaced linear arrays at the transmitter and 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], [29]. The corresponding (complex) baseband transmitted signal and channel output are related as

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

where $\boldsymbol{\mathcal{H}}(\mathbf{x}(t))$ is the N_R-dimensional channel output, $\mathbf{X}(f)$ is the (element-wise) 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

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

which represents signal propagation over N_p paths; here, $(\cdot)^{\mathsf{H}}$ denotes the Hermitian operation and β_n is the complex path gain, $\theta_{R,n}$ the angle of arrival (AoA) at the receiver, $\theta_{T,n}$ the angle of departure (AoD) at the transmitter, τ_n the (relative) delay, and ν_n the Doppler shift associated with the n-th 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 [30].[1] 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 $\nu_n \in [-\frac{\nu_{max}}{2}, \frac{\nu_{max}}{2}]$ in the delay and Doppler space, respectively. Here, τ_{max} and ν_{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 [31]. Instead, we limit the discussion in this paper to underspread channels, characterized by $\tau_{max}\nu_{max} \ll 1$, which is fortunately true of most wireless channels [32].

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

[1]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 [30] for further details.

to a *temporal signal space* of dimension $N_o = TW$ [33]. 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 I for a definition of each of these classes. As noted earlier, we limit ourselves in the sequel 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.

B. Multipath Wireless Channels: 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 channel model*, with the aid of a four-dimensional Fourier series expansion [30], [34].

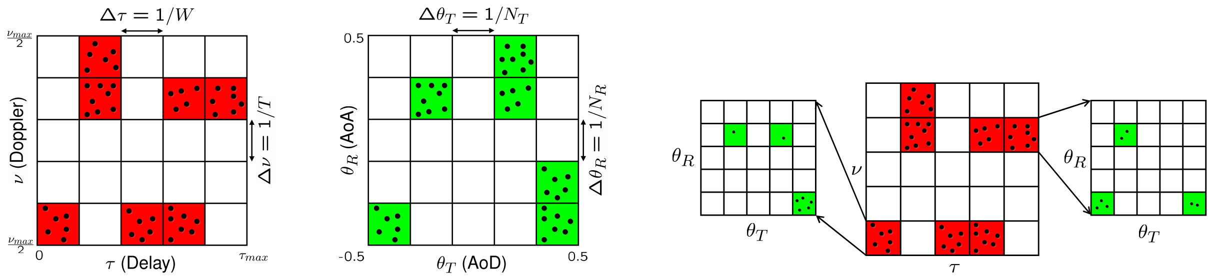
On an abstract level, virtual representation of a multipath channel $\boldsymbol{\mathcal{H}}$ provides a discretized approximation of its time-varying frequency response by uniformly sampling the angle-delay-Doppler space at the Nyquist rate: $(\Delta\theta_R, \Delta\theta_T, \Delta\tau, \Delta\nu)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

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

$$H_v(i,k,\ell,m) \approx \sum_{n\in S_{R,i}\cap S_{T,k}\cap S_{\tau,\ell}\cap S_{\nu,m}} \beta_n \tag{4}$$

approximates the actual time-varying frequency response matrix $\mathbf{H}(t,f)$ in the sense that $\int_{\mathbb{R}} \mathbf{H}(t,f)\mathbf{X}(f)e^{j2\pi ft}df \approx \int_{\mathbb{R}} \tilde{\mathbf{H}}(t,f)\mathbf{X}(f)e^{j2\pi ft}df$ [30], [34]. 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.[2] 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 [34] 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 of 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 $\{H_v(i,k,\ell,m)\}$.

[2] 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.



(a) (b) (c)

Fig. 1. 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 $d \ll D$ nonzero channel coefficients, and the dots represent the paths contributing to each nonzero 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 nonzero delay-Doppler coefficient, $H_v(\ell_o, m_o)$, are further resolved in angle to yield the conditional SP in angle: $\{H_v(i, k, \ell_o, m_o)\}_{\mathcal{S}_d(\ell_o, m_o)}$.

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

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

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 four-dimensional channel spread. However, the actual or *effective* number of DoF, d, in the channel corresponds to the number of nonvanishing virtual channel coefficients: $d = |\{(i, k, \ell, m) : |H_v(i, k, \ell, m)| > 0\}|$. 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 populated by at least one path (see Fig. 1).

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 paths uniformly distributed within the angle-delay-Doppler spread of the channel so that $d \approx D$ for any choice of 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]–[16], [35]. 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 a physical path. This implies that wireless channels with clustered multipath components tend to have far fewer than D nonzero 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 Wireless Channels): Let $\mathcal{S}_d = \{(i,k,\ell,m) : |H_v(i,k,\ell,m)| > 0\}$ denote the set of indices of nonzero virtual channel coefficients of a multipath wireless channel. We say that the channel is d-sparse if the number of its effective DoF satisfies $d = |\mathcal{S}_d| \ll D$. The corresponding set of indices $\mathcal{S}_d$ is termed as the *channel sparsity pattern*.

It is worth noting that sparsity in multipath wireless channels is inherently tied to the choice of signaling parameters: channels with small-enough values of N_R, N_T, T, and W are bound to have $d \approx D$. Nevertheless, the trend in modern wireless systems is to operate at high spatio-temporal signal space dimension (defined as: $N_s = \min\{N_T, N_R\} TW$). As such, sparse channels are becoming more and more ubiquitous in today's communications landscape. 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 $\boldsymbol{\mathcal{H}}$ suffice 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.[3]

Example 1 *(Continued)* – Delay Sparsity

In order to further illustrate the idea of multipath sparsity, consider again the single-antenna setup described in Example 1. In this case, the CIR is given by $h(\tau) = \sum_{n=1}^{N_p} \beta_n \delta(\tau - \tau_n)$ and its virtual representation can be expressed as $\tilde{h}(\tau) = \sum_{\ell=0}^{L-1} H_v(\ell)\delta(\tau - \frac{\ell}{W})$. Here, $H_v(\ell)$ is approximately equal to the sum of gains of all echoes (paths) whose delays lie within the interval $\mathcal{I}_\ell = (\frac{\ell}{W} - \frac{1}{2W}, \frac{\ell}{W} + \frac{1}{2W}]$ and there are a total of $D\ (= L) = \lceil W\tau_{max} \rceil + 1$ of these virtual channel coefficients. However, since the CIR consists of only a few dominant echoes (or clusters of echoes), a large number of the intervals $\{\mathcal{I}_\ell\}$ would contain no echoes under the assumption of large-enough signaling bandwidth. Therefore, majority of the channel coefficients in this case would be zero, $d = |\{\ell : |H_v(\ell)| > 0\}| \ll L$, and we accordingly term the underlying multipath channel as d-sparse.

B. Sensing and Reconstruction

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

$$\mathbf{x}(t) = \mathbf{x}_{tr}(t) + \mathbf{x}_{data}(t)\,,\ \ 0 \le t \le T \tag{6}$$

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

[3]We refer the reader to [36] for a Bayesian formulation of sparse channels.

ignore the received data and focus only on the training component of the received signal, given by

$$\mathbf{y}_{tr}(t) = \boldsymbol{\mathcal{H}}(\mathbf{x}_{tr}(t)) + \mathbf{z}_{tr}(t)\,, \ 0 \leq t \leq T + \tau_{max} \tag{7}$$

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

As a first step towards estimating $\boldsymbol{\mathcal{H}}$, the (noisy) received training signal $\mathbf{y}_{tr}(t)$ is matched filtered with the transmitted waveforms at the receiver to obtain an equivalent discrete-time representation of (7). 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- or multi-carrier), 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 following general linear form at the receiver [24]–[26]

$$\underbrace{\begin{bmatrix}\mathbf{y}_1 & \ldots & \mathbf{y}_{N_R}\end{bmatrix}}_{\mathbf{Y}} = \sqrt{\frac{\mathcal{E}}{N_T}}\,\mathbf{X}\underbrace{\begin{bmatrix}\mathbf{h}_{v,1} & \ldots & \mathbf{h}_{v,N_R}\end{bmatrix}}_{\mathbf{H}_v} + \underbrace{\begin{bmatrix}\mathbf{z}_1 & \ldots & \mathbf{z}_{N_R}\end{bmatrix}}_{\mathbf{Z}}\,. \tag{8}$$

Here, $\mathcal{E}/N_T$ is the average training energy budget per transmit antenna ($\mathcal{E}$ being defined as: $\mathcal{E} = \int_0^T \|\mathbf{x}_{tr}(t)\|_2^2 dt$), the vectors $\mathbf{h}_{v,i}, i = 1, \ldots, N_R$, are $N_T L(2M+1)$-dimensional complex vectors comprising of the 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 signal-to-noise ratio (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 (8) are completely determined by $\mathbf{x}_{tr}(t)$ and the class to which $\boldsymbol{\mathcal{H}}$ 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.

Example 1 *(Continued)* – Delay Sparsity

Continuing with our discussion of the single-antenna setup described in Example 1, the general linear form (8) in this case can be expressed as $\mathbf{y} = \sqrt{\mathcal{E}}\,\mathbf{X}\mathbf{h}_v + \mathbf{z}$. Here, $\mathbf{h}_v$ is an L-dimensional complex vector consisting of the channel coefficients $\{H_v(\ell)\}$. In addition, due to the LTI nature of the channel, it is easy to see that if single-carrier waveforms are used for signaling in this setting then $\mathbf{X}$ has a Toeplitz form. On the other hand, if multi-carrier waveforms are used for signaling then $\mathbf{X}$ corresponds to a submatrix of an $N_o \times N_o$ unitary discrete Fourier transform (DFT) matrix. The two signaling scenarios are discussed further in extensive details in Section V-A.

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 to signify 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 $\mathbb{E}\left[\|\mathbf{H}_v - \mathbf{H}_v^{\mathsf{est}}\|_F^2\right]$, 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 that minimizes this reconstruction error[4]

$$\mathbf{x}_{tr}^{\mathsf{opt}}(t) = \arg\min_{\mathbf{x}_{tr}(t)} \mathbb{E}\left[\Delta\left(\mathbf{H}_v^{\mathsf{est}}\right)\right] \tag{9}$$

where we have used the shorthand notation $\Delta(\mathbf{H}) = \|\mathbf{H}_v - \mathbf{H}\|_F^2$ in the above equation. As a measure of its spectral efficiency, the resulting training signal also has associated with it the concepts of *temporal training dimensions*, defined as $M_{tr} = \#\{\text{temporal signal space dimensions occupied by } \mathbf{x}_{tr}^{\mathsf{opt}}(t)\}$, and *receive training dimensions*, defined as $N_{tr} = M_{tr} \times N_R$. Since every receive signal space dimension utilized for training means one less dimension available for communication, the effectiveness of a particular training-based method for a fixed training SNR $\mathcal{E}$ is measured in terms of both the receive training dimensions, N_{tr}, dedicated to $\mathbf{x}_{tr}^{\mathsf{opt}}(t)$ and the ensuing reconstruction error $\mathbb{E}[\Delta(\mathbf{H}_v^{\mathsf{est}}(\mathbf{x}_{tr}^{\mathsf{opt}}))]$.

Traditional training-based methods, such as those in [5]–[12], have been developed under the implicit assumption that the number of DoF, d, in $\boldsymbol{\mathcal{H}}$ 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 a few exceptions such as [17]–[22], nearly all training-based methods proposed in the past make use of the minimum least squares (LS) error criterion—or its Bayesian counterpart, the minimum mean squared error criterion, for a Bayesian formulation of $\boldsymbol{\mathcal{H}}$—to obtain an estimate of $\mathbf{H}_v$ from $\mathbf{Y}$

$$\mathbf{H}_v^{\mathsf{LS}} = \arg\min_{\mathbf{H}} \left\|\mathbf{Y} - \sqrt{\frac{\mathcal{E}}{N_T}}\,\mathbf{X}\mathbf{H}\right\|_F^2. \tag{10}$$

This is a well-known problem in the statistical estimation literature [37] and its closed-form solution is given by $\mathbf{H}_v^{\mathsf{LS}} = \sqrt{N_T/\mathcal{E}}\,\mathbf{X}^\dagger\mathbf{Y}$, 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 $\mathbf{H}_v^{\mathsf{LS}}$ equals $\mathbf{H}_v$ in the noiseless setting—reconstruction methods based on the LS error criterion further require that the sensing matrix $\mathbf{X}$ has at least as many rows as D/N_R, resulting in the following form for $\mathbf{H}_v^{\mathsf{LS}}$

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

where it is assumed that the training signal $\mathbf{x}_{tr}(t)$ is such that $\mathbf{X}$ has full column rank. It is easy to show in this case that the accompanying reconstruction error of a LS-based channel estimation method is given by

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

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., [38, Th. 1])

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

[4]Recall that $\mathbf{H}_v^{\mathsf{est}}$ depends on the training signal $\mathbf{x}_{tr}(t)$ through $\mathbf{X}$.

where the equality in (a) holds if and only if $\mathbf{X}$ has orthonormal columns, while (b) follows from the fact that $\mathsf{trace}(\mathbf{X}^{\mathsf{H}}\mathbf{X}) = \|\mathbf{X}\|_F^2 = D/N_R$. Consequently, an optimal training signal $\mathbf{x}_{tr}^{\mathsf{opt}}(t)$ for LS-based estimation methods is the one that leads to $\mathbf{X}^{\mathsf{H}}\mathbf{X} = \mathbf{I}_{N_T L(2M+1)}$, and much of the emphasis in the previously proposed 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. 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 II, depending upon the type of signaling waveforms used for training and the channel class to which $\boldsymbol{\mathcal{H}}$ belongs, this requirement often translates into the condition that the number of receive training dimensions dedicated to $\mathbf{x}_{tr}(t)$ must be at least as large as the maximum number of DoF in $\boldsymbol{\mathcal{H}}$: $N_{tr} = \Omega(D)$;[5] 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 $\mathbb{E}[\Delta(\mathbf{H}_v^{\mathsf{LS}})] = \Omega(D \cdot (N_T/\mathcal{E}))$.

In the light of the above observations, a natural question to ask is: *how good is the performance of traditional training-based methods?* In fact, if one assumes that $\boldsymbol{\mathcal{H}}$ is not sparse (in other words, $d = D$) then it is easy to argue the optimality of these methods [37]: (i) $\mathbf{H}_v^{\mathsf{LS}}$ in this case is also the maximum-likelihood estimate of $\mathbf{H}_v$, and (ii) 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 $\boldsymbol{\mathcal{H}}$ is d-sparse. In particular, note that 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 $\mathbb{E}[\Delta(\mathbf{H}_v^{\mathsf{est}})] = \Omega(d \cdot (N_T/\mathcal{E}))$ and, for signaling and channel configurations that require $N_{tr} = \Omega(D)$ in the case of LS-based estimation methods, $N_{tr} = \Omega(d)$ are the actual fundamental limits in sparse-channel estimation.

In the sequel, we present new training-based estimation methods for six particular signaling and channel configurations (see Table II) and show that our intuition is indeed correct (modulo polylogarithmic factors). In particular, a key feature of our approach to estimating sparse multipath channels—originally proposed in [24] for frequency-selective single-antenna channels and later generalized in [25], [26] 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 [23]. This makes our approach—termed as compressed channel sensing (CCS)—fundamentally different from the traditional training-based 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

[5] 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 $g_n = O(f_n)$.

TABLE II

SUMMARY AND COMPARISON OF CCS RESULTS FOR THE SIGNALING AND CHANNEL CONFIGURATIONS STUDIED IN THE PAPER[a]

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

[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 last two conditions in the second column are for the case when the conditional sparsity of each AoA equals the average AoA sparsity.

proposed various training-based methods for sparse multipath channels that are based on nonlinear reconstruction techniques [17]–[22]. The thing that distinguishes CCS from the prior work is that the CCS framework is highly amenable to analysis. Specifically, in order to give a summary of the results to come, define the *conditional* sparsity pattern associated with the i-th 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 the sequel that:

R.1 The performance of CCS in terms of the reconstruction error is provably better than the traditional methods. The training signals and reconstruction procedures specified by CCS for the signaling and channel configurations studied in the paper ensure that $\Delta(\mathbf{H}_v^{\textsf{ccs}}) = O(d \cdot (N_T/\mathcal{E}) \cdot \log D)$ with high probability.

R.2 CCS is provably more spectrally efficient than the traditional 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 traditional methods require that $N_{tr} = \Omega(D)$ for certain signaling and channel configurations, CCS only requires that $N_{tr} = \Omega(d \times \text{polylog factor})$ for the same configurations.

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

Table II provides a compact summary of the CCS results as they pertain to the six signaling and channel configurations studied in the paper and compares them to the corresponding results for traditional training-based 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 traditional 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 methods implicitly require that the temporal signal space dimensions be as large as possible: $N_o \longrightarrow \infty$. As such, the CCS condition is in fact a relaxation of this implicit requirement for traditional methods.

As is evident from the preceding discussion and analysis, the performance of CCS is a significant improvement over that of traditional training-based 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 II. 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 [39] 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

$$r_i = \boldsymbol{\psi}_i^{\mathsf{T}} \boldsymbol{\theta} + \eta_i \,, \ i = 1, \dots, n \tag{14}$$

where $(\cdot)^{\mathsf{T}}$ denotes the transpose operation, $\boldsymbol{\psi}_i \in \mathbb{C}^p$ 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 comprised 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 convex optimization programs or efficient greedy algorithms; see [40] for the references of 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) [41].

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

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

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 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 check 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 RIP with high probability. For example, let the $n \times p$ matrix $\mathbf{\Psi}$ be such that either (i) its entries are drawn independently from a $\mathcal{N}(0, \frac{1}{n})$ distribution, or (ii) 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 $1/\sqrt{p}$ and then scaled by a factor of $\sqrt{p/n}$. Then, for every $\delta_S \in (0,1)$, it is known that $\mathbf{\Psi} \in RIP(S, \delta_S)$ with probability exceeding $1 - e^{-O(n)}$ in the former case if $n = \Omega(S \cdot \log \frac{p}{S})$ [42], while $\mathbf{\Psi} \in RIP(S, \delta_S)$ with probability exceeding $1 - p^{-O(1)}$ in the latter case if $n = \Omega(S \cdot \log^5 p)$ [43].[6]

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 sequel—goes by the name of *Dantzig selector* (DS) [27]. 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 CS reconstruction procedures that perform near-optimally vis-à-vis stochastic noise (the others being the risk minimization method of Haupt and Nowak [44] and the lasso [28], which also goes by the name of basis pursuit denoising [45]). Second, unlike the method of [44], 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 the lasso also enjoys many of the useful properties of the DS, including the reconstruction error bounds that are very similar to that of the DS [46], [47]. 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 [48] and SpaRSA [49], 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 variation on [27, Th. 1.1], states the reconstruction error performance of the DS for the case when $\boldsymbol{\theta}$ is sparse.[7]

Theorem 1 (The Dantzig Selector [27]): Let $\mathbf{\Psi}\boldsymbol{\theta} + \boldsymbol{\eta} = \mathbf{r} \in \mathbb{C}^n$ be a vector of noisy measurements of $\boldsymbol{\theta} \in \mathbb{C}^p : \|\boldsymbol{\theta}\|_0 \leq S$, where the $n \times p$ matrix $\mathbf{\Psi}$ 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 $\mathbf{\Psi} \in RIP(2S, \delta_{2S})$ for some $\delta_{2S} < \sqrt{2} - 1$ and choose $\lambda = (2\sigma^2(1+a)\log p)^{1/2}$ for any $a \geq 0$. Then the estimate $\boldsymbol{\theta}^{\mathsf{DS}}$ obtained as a solution to the optimization program

$$\boldsymbol{\theta}^{\mathsf{DS}} = \operatorname*{arg\,min}_{\tilde{\boldsymbol{\theta}} \in \mathbb{C}^p} \|\tilde{\boldsymbol{\theta}}\|_1 \text{ subject to } \|\mathbf{\Psi}^{\mathsf{H}}(\mathbf{r} - \mathbf{\Psi}\tilde{\boldsymbol{\theta}})\|_\infty \leq \lambda \qquad \text{(DS)}$$

satisfies

$$\|\boldsymbol{\theta}^{\mathsf{DS}} - \boldsymbol{\theta}\|_2^2 \leq c_1^2 \cdot S \cdot \log p \cdot \sigma^2 \qquad (16)$$

with probability at least $1 - 2(\pi(1+a)\log p \cdot p^{2a})^{-1/2}$. Here, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the ℓ_1- and ℓ_∞-norm of a vector, respectively, and the constant $c_1 = 4\sqrt{2(1+a)}/(1 - (\sqrt{2}+1)\delta_{2S})$.

[6]In fact, the actual requirement in the latter case is $n = \Omega(S \log^2(p) \log(S \log p) \log^2(S))$ [43]; for the sake of compactness, however, we use the lax requirement $n = \Omega(S \cdot \log^5 p)$ in the paper.

[7]The variation is primarily due to the complex-valued setup in the paper as opposed to the real-valued one in [27, Th. 1.1] and noticing the fact that $\theta_{S,2S} < \sqrt{2}\delta_{2S}$; we refer the reader to [27] for further details.

In the sequel, 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. Finally, note that the statement of Theorem 1 assumes that $\boldsymbol{\Psi} \in RIP(2S, \delta_{2S})$ almost surely. However, if this is not true then in this case the only thing that changes in the statement is that $\boldsymbol{\theta}^{\mathsf{DS}}$ satisfies (16) with probability at least $1 - 2\max\big\{2(\pi(1+a)\log p \cdot p^{2a})^{-1/2}, \Pr\{\boldsymbol{\Psi} \in RIP(2S, \delta_{2S})\}\big\}$. We are now ready to discuss the specifics of CCS for sparse multipath channels.

V. Compressed Channel Sensing: Single-Antenna Channels

A. Estimating Sparse Frequency-Selective Channels

For a single-antenna channel that is frequency-selective (also, see Example 1), 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 given by [cf. (7)]

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

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 (multi-carrier) *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 $x_{tr}(t)$ can be represented as

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

where $g(t)$ is the *chip waveform* having unit energy $(\int |g(t)|^2 dt = 1)$, $T_c \approx 1/W$ is the chip duration, and $\{x_n\}$ is the N_o-dimensional spreading code associated with the training signal also having unit energy $(\sum_n |x_n|^2 = 1)$. In this case, chip-matched filtering the received training signal (18) yields the discrete-time representation [24]

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

where $*$ denotes discrete-time convolution, $\mathbf{h}_v \in \mathbb{C}^L$ is the vector of channel coefficients $\{H_v(\ell)\}$, and $\mathbf{x} \in \mathbb{C}^{N_o}$ is comprised 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 is $\begin{bmatrix} x_0 & \mathbf{0}_{L-1}^{\mathsf{T}} \end{bmatrix}$ and first column is $\begin{bmatrix} \mathbf{x}^{\mathsf{T}} & \mathbf{0}_{L-1}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$.

Note that (20) is the single-antenna version of the standard form (8). Therefore, from (13), the reconstruction error 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, and show that for d-sparse channels it leads to an improvement of a factor of about L/d (modulo a logarithmic factor).

CCS -1 – SS Training and Reconstruction

Training: Pick the spreading code $\{x_n\}_{n=0}^{N_o-1}$ associated with $x_{tr}(t)$ 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 $a \geq 0$ and pick $\lambda = (2\mathcal{E}(1+a)\log L)^{1/2}$. 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.

Theorem 2: Suppose that the number of temporal signal space dimensions $N_o \, (= TW) = \Omega(d^2 \cdot \log L)$. Then the CCS estimate of $\mathbf{h}_v$ satisfies

$$\Delta\left(\mathbf{h}_v^{\mathsf{ccs}}\right) \leq c_1^2 \cdot \frac{d}{\mathcal{E}} \cdot \log L \tag{21}$$

with probability at least $1 - 2\max\left\{2(\pi(1+a)\log L \cdot L^{2a})^{-1/2}, e^{-O(N_o/d^2)}\right\}$. Here, the constant c_1 is the same as in Theorem 1 (with δ_{2d} in place of δ_{2S}).

Note that the frequency-selective channel being d-sparse simply means that $\|\mathbf{h}_v\|_0 \leq d \ll L$. Therefore, the proof of this theorem essentially follows from Theorem 1 and [24, Th. 2] (also, see [38, Th. 6]), where it was shown that $\Pr\{\mathbf{X} \notin RIP(2d, \delta_{2d})\} \leq e^{-O(N_o/d^2)}$ for any $\delta_{2d} \in (0,1)$, provided $N_o = \Omega(d^2 \cdot \log L)$.

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

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

where $g(t)$ is a prototype pulse having unit energy, $\mathcal{S}_{tr} \subset \mathcal{S} = \{0, 1, \dots, N_o - 1\}$ is the set of indices of *pilot tones* used for training, and N_{tr}—the number of receive training dimensions—denotes the total number of pilot tones in this case, $N_{tr} = |\mathcal{S}_{tr}|$, and is a measure of the spectral efficiency of $x_{tr}(t)$. Finally, matched filtering the received training signal (18) with the OFDM basis waveforms $\{g(t)e^{j2\pi \frac{n}{T} t}\}_{\mathcal{S}_{tr}}$ 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 $N_{tr} \times L$ sensing matrix that is comprised of $\left\{\frac{1}{\sqrt{N_{tr}}}\begin{bmatrix}1 & \omega_{N_o}^{n \cdot 1} & \dots & \omega_{N_o}^{n \cdot (L-1)}\end{bmatrix} : n \in \mathcal{S}_{tr}\right\}$ as its rows, where $\omega_{N_o} = e^{-j\frac{2\pi}{N_o}}$, and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_{N_{tr}}, \mathbf{I}_{N_{tr}})$.[8]

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

CCS -2 – OFDM Training and Reconstruction

Training: Choose the number of pilot tones $N_{tr} = \Omega(d \cdot \log^5 N_o)$ and pick $\mathcal{S}_{tr}$ to be a set of N_{tr} indices sampled uniformly at random (without replacement) from $\mathcal{S} = \{0, 1, \dots, N_o - 1\}$.

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

[8]An assumption often made in the case of OFDM signaling is that $T \gg \tau_{max}$ [1], which implies $N_o \gg L$.

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

Theorem 3: Under the assumption that the frequency-selective channel is d-sparse, the reconstruction error of $\mathbf{h}_v^{\mathsf{ccs}}$ satisfies (21) with probability at least $1 - 2\max\left\{2(\pi(1+a)\log L \cdot L^{2a})^{-1/2}, N_o^{-O(1)}\right\}$.

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 the discussion following the definition, Pr$\{\mathbf{X} \notin RIP(2d, \delta_{2d})\} \leq N_o^{-O(1)}$ for any $\delta_{2d} \in (0,1)$, provided $N_{tr} = \Omega(d \cdot \log^5 N_o)$ [43].

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 corresponding received training signal can be written as

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

Signaling waveforms that are often used to communicate over a doubly-selective channel can be broadly categorized as (single-carrier) SS waveforms and (multi-carrier) *short-time Fourier* (STF) waveforms, which are a generalization of OFDM waveforms for doubly-selective channels [50], [51]. 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 $x_{tr}(t)$ in the case of a doubly-selective channel has exactly the same form as given 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 [25]

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

Nevertheless, it has been shown in [25, § III-A] that this received training data can be represented into the standard form (8) 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

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{-M} & \dots & \mathbf{X}_0 & \dots & \mathbf{X}_M \end{bmatrix}. \tag{26}$$

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 $\mathbf{W}_m = \text{diag}(\omega_{\tilde{N}_o}^{-m\cdot 0}, \omega_{\tilde{N}_o}^{-m\cdot 1}, \dots, \omega_{\tilde{N}_o}^{-m\cdot(\tilde{N}_o-1)})$ and $\mathbf{T}$ is a Toeplitz matrix of dimensions $\tilde{N}_o \times L$ having $\begin{bmatrix} x_0 & \mathbf{0}_{L-1}^{\mathsf{T}} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{x}^{\mathsf{T}} & \mathbf{0}_{L-1}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$ as its first row and first column, respectively.

Note that under the assumption that the doubly-selective channel is underspread ($\tau_{max}\nu_{max} \ll 1$), we have that $TW \gg \tau_{max}\nu_{max}TW \implies \tilde{N}_o > L(2M+1)$. 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, which is markedly similar to CCS -1, 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 in the case of CCS -1.

Reconstruction: Fix any $a \geq 0$ and pick $\lambda = (2\mathcal{E}(1+a)\log L(2M+1))^{1/2}$. 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: Suppose that the number of temporal signal space dimensions $N_o = \Omega(d^2 \cdot \log L(2M+1))$. Then the CCS estimate of $\mathbf{h}_v$ satisfies

$$\Delta\left(\mathbf{h}_v^{\mathsf{ccs}}\right) \leq c_1^2 \cdot \frac{d}{\mathcal{E}} \cdot \log L(2M+1) \tag{27}$$

with probability at least $1 - 2\max\left\{2(\pi(1+a)\log L(2M+1) \cdot (L(2M+1))^{2a})^{-1/2}, e^{-O(N_o/d^2)}\right\}$.

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 [25, Th. 2] in which it was established that $\Pr\{\mathbf{X} \notin RIP(2d, \delta_{2d})\} \leq e^{-O(N_o/d^2)}$ for any $\delta_{2d} \in (0,1)$, provided $N_o = \Omega(d^2 \cdot \log L(2M+1))$.

*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 [50], [51], the training signal $x_{tr}(t)$ is of the form

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

where $g(t)$ is again a prototype pulse having unit energy, $\mathcal{S}_{tr} \subset \mathcal{S} = \{0, 1, \dots, N_t - 1\} \times \{0, 1, \dots, N_f - 1\}$ is the set of indices of STF pilot tones used for training, and N_{tr}—a measure of the spectral efficiency of $x_{tr}(t)$—denotes the total number of pilot tones: $N_{tr} = |\mathcal{S}_{tr}|$. 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 $T_o W_o = 1$ [51]. 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 $N_f = W/W_o$.

Matched filtering the received training signal (24) in this case with $\{g(t - nT_o)e^{j2\pi m W_o t}\}_{\mathcal{S}_{tr}}$ yields [51]

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

where the STF channel coefficients $H_{n,m} \approx \tilde{H}(t,f)\big|_{(t,f)=(nT_o, mW_o)}$. As shown in [25, § IV-A], collection of this matched-filtered output into a vector $\mathbf{y} \in \mathbb{C}^{N_{tr}}$ followed by simple manipulations yields the standard form $\mathbf{y} = \sqrt{\mathcal{E}}\,\mathbf{X}\mathbf{h}_v + \mathbf{z}$, where $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_{N_{tr}}, \mathbf{I}_{N_{tr}})$ and the $N_{tr} \times L(2M+1)$ matrix $\mathbf{X}$ is comprised of

$$\left\{\frac{1}{\sqrt{N_{tr}}}\begin{bmatrix}\omega_{N_t}^{n\cdot M} & \omega_{N_t}^{n\cdot(M-1)} & \dots & \omega_{N_t}^{-n\cdot M}\end{bmatrix} \otimes \begin{bmatrix}1 & \omega_{N_f}^{m\cdot 1} & \dots & \omega_{N_f}^{m\cdot(L-1)}\end{bmatrix} : (n,m) \in \mathcal{S}_{tr}\right\}$$

as its rows.[9] Consequently, traditional methods impose the condition $N_{tr} = \Omega(L(2M+1))$ in order to satisfy the

[9]Here, $\otimes$ denotes the Kronecker product; also, $T_o \in [\tau_{max}, 1/\nu_{max}]$ and $W_o \in [\nu_{max}, 1/\tau_{max}] \implies N_t \geq 2M+1$ and $N_f \geq L$.

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 than the LS-based approach but is also spectrally more efficient.

CCS - 4 – STF Training and Reconstruction

Training: Choose the number of pilot tones $N_{tr} = \Omega(d \cdot \log^5 N_o)$ and pick $\mathcal{S}_{tr}$ to be a set of N_{tr} ordered pairs sampled uniformly at random (without replacement) from $\mathcal{S} = \{0, 1, \dots, N_t - 1\} \times \{0, 1, \dots, N_f - 1\}$.

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

Theorem 5: Under the assumption that the doubly-selective channel is d-sparse, the reconstruction error of $\mathbf{h}_v^{\mathsf{ccs}}$ satisfies (27) with probability at least $1 - 2\max\{2(\pi(1+a)\log L(2M+1) \cdot (L(2M+1))^{2a})^{-1/2}, N_o^{-O(1)}\}$.

Note that [25, Th. 3] specifies 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 II 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-antenna (MIMO) channel is of the form [cf. (3)]

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

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

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

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

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

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

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 [26, § III], pre-multiplying the $\tilde{\mathbf{y}}_n$'s with $\mathbf{A}_R^{\mathsf{H}}$ and row-wise stacking the resulting vectors into an $M_{tr} \times N_R$ matrix $\mathbf{Y}$ yields the standard linear form (8): $\mathbf{Y} = \sqrt{\frac{\mathcal{E}}{N_T}} \mathbf{X}\mathbf{H}_v + \mathbf{Z}$, where the entries of $\mathbf{Z}$ are independently distributed as $\mathcal{CN}(0,1)$. Here, $\mathbf{X}$ is an $M_{tr} \times N_T$ matrix of the form

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \dots & \mathbf{x}_{M_{tr}-1} \end{bmatrix}^{\mathsf{T}} \mathbf{A}_T^* \tag{33}$$

where $(\cdot)^*$ denotes the conjugation operation. In order to estimate nonselective MIMO channels, traditional training-based methods such as those in [9], [10] therefore require that $M_{tr} = \Omega(N_T)$ so as to ensure that $\mathbf{X}$ has full column rank and produce an estimate that satisfies $\mathbb{E}[\Delta(\mathbf{H}_v^{\mathsf{LS}})] = \Omega(N_R N_T^2/\mathcal{E})$. In particular, note that the condition $M_{tr} = \Omega(N_T)$ means that traditional methods in this case require the number of receive training dimensions to satisfy $N_{tr} = M_{tr} N_R = \Omega(N_R N_T)$. 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 and receive training dimensions. Before proceeding further, however, recall that the conditional sparsity pattern associated with the i-th 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: Choose the number of training time slots $M_{tr} = \Omega(\bar{d} \cdot \log N_T/\bar{d})$ and pick $\{\mathbf{x}_n, n = 0, \dots, M_{tr} - 1\}$ to be a sequence of i.i.d. Rademacher vectors in which each entry independently takes the value $+1/\sqrt{M_{tr}}$ or $-1/\sqrt{M_{tr}}$ with probability $1/2$ each.

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

Theorem 6: Assuming that the nonselective MIMO channel is d-sparse, the CCS estimate of $\mathbf{H}_v$ satisfies

$$\Delta\left(\mathbf{H}_v^{\mathsf{CCS}}\right) \leq c_1^2 \cdot \frac{d \cdot N_T}{\mathcal{E}} \cdot \log N_R N_T \tag{34}$$

with probability at least $1 - 4\max\left\{(\pi(1+a)\log N_R N_T \cdot (N_R N_T)^{2a})^{-1/2}, e^{-O(M_{tr})}\right\}$.

The proof of this theorem is sketched in [26, Th. 3], and is based on [42, Th. 5.2] and a slight modification of the proof of Theorem 1 in [27]. 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 $N_{tr} = M_{tr} N_R = \Omega(\bar{d}\, N_R \cdot \log N_T/\bar{d})$ for CCS, which—modulo the logarithmic factor—is always better than $N_{tr} = \Omega(N_R N_T)$ for traditional 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 $N_{tr} = \Omega(d \cdot \log N_R N_T/d)$.

B. Estimating Sparse Frequency-Selective Channels

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

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

where the matrices $\mathbf{A}_R$ and $\mathbf{A}_T$ are as given in (30), and $\mathbf{H}_v(\ell) = \Big[\mathbf{h}_{v,1}(\ell) \quad \dots \quad \mathbf{h}_{v,N_R}(\ell)\Big]$ is an $N_T \times N_R$ matrix in which the i-th 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, § IV-B] and [12, § 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

$$\mathbf{x}_{tr}(t) = \sqrt{\frac{\mathcal{E}}{N_T}} \sum_{(n,m)\in\mathcal{S}_{tr}} \mathbf{x}_{n,m}\, g(t - nT_o)e^{j2\pi m W_o t}, \ 0 \leq t \leq T \tag{36}$$

where $\mathcal{S}_{tr} \subset \mathcal{S} = \{0, 1, \dots, N_t - 1\} \times \{0, 1, \dots, N_f - 1\}$ 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 $\sum_{\mathcal{S}_{tr}} \|\mathbf{x}_{n,m}\|_2^2 = N_T$. 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): $M_{tr} = |\mathcal{S}_{tr}|$.[10]

From [51], matched filtering $\mathbf{y}_{tr}(t) = \boldsymbol{\mathcal{H}}(\mathbf{x}_{tr}(t)) + \mathbf{z}_{tr}(t)$ in this case with $\{g(t - nT_o)e^{j2\pi m W_o t}\}_{\mathcal{S}_{tr}}$ yields

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

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 $\mathbf{H}_m \approx \tilde{\mathbf{H}}(f)\big|_{f=mW_o}$. As in the case of nonselective MIMO channels, we can pre-multiply the received training vectors $\mathbf{y}_{n,m}$'s with $\mathbf{A}_R^{\mathsf{H}}$ and row-wise stack the resulting vectors $\mathbf{A}_R^{\mathsf{H}}\mathbf{y}_{n,m}$ to yield an $M_{tr} \times N_R$ matrix $\mathbf{Y}$. Further, as in [26, § IV], the right-hand side of (37) can be manipulated to express this matrix into the standard form (8): $\mathbf{Y} = \sqrt{\frac{\mathcal{E}}{N_T}}\,\mathbf{X}\mathbf{H}_v + \mathbf{Z}$. Here, $\mathbf{H}_v = \begin{bmatrix}\mathbf{h}_{v,1} & \dots & \mathbf{h}_{v,N_R}\end{bmatrix}$ is the $N_T L \times N_R$ channel matrix in which the i-th column consists of the coefficients $\{H_v(i,k,\ell)\}$ associated with the i-th resolvable AoA, while $\mathbf{X}$ is an $M_{tr} \times N_T L$ matrix comprising of $\left\{\begin{bmatrix}1 & \omega_{N_f}^{m\cdot 1} & \dots & \omega_{N_f}^{m\cdot(L-1)}\end{bmatrix} \otimes \mathbf{x}_{n,m}^{\mathsf{T}}\mathbf{A}_T^{*} : (n,m) \in \mathcal{S}_{tr}\right\}$ as its rows.

Once again, the form of the sensing matrix $\mathbf{X}$ here dictates that $M_{tr} = \Omega(N_T L)$ for the traditional methods such as those in [11], [12] to obtain a meaningful estimate of $\mathbf{H}_v$, and we have from (13) that $\mathbb{E}[\Delta(\mathbf{H}_v^{\mathsf{LS}})] = \Omega(N_R N_T^2 L/\mathcal{E})$ in that case. Note that in terms of the receive training dimensions, this implies that traditional methods require $N_{tr} = \Omega(N_R N_T L)$ 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 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: Choose the number of pilot tones $M_{tr} = \Omega(\bar{d} \cdot \log^5 N_T N_f)$. Further, pick $\mathcal{S}_{tr}$ to be a set of M_{tr} ordered pairs sampled uniformly at random (without replacement) from $\{0, 1, \dots, N_T - 1\} \times \{0, 1, \dots, N_f - 1\}$ and define the corresponding sequence of training vectors as $\{\mathbf{x}_{n,m} = \sqrt{N_T/M_{tr}}\,\mathbf{e}_{n+1} : (n,m) \in \mathcal{S}_{tr}\}$, where $\mathbf{e}_i$ denotes the i-th standard basis element of $\mathbb{C}^{N_T}$.

Reconstruction: Fix any $a \geq 0$ and pick $\lambda = (2\mathcal{E}(1+a)(\log N_R N_T L)/N_T)^{1/2}$. The CCS estimate of $\mathbf{H}_v$ is then given as follows: $\mathbf{H}_v^{\mathsf{ccs}} = \begin{bmatrix} DS(\sqrt{\mathcal{E}/N_T}\,\mathbf{X}, \mathbf{y}_1, \lambda) & \dots & DS(\sqrt{\mathcal{E}/N_T}\,\mathbf{X}, \mathbf{y}_{N_R}, \lambda)\end{bmatrix}$.

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

Theorem 7: For a frequency-selective MIMO channel that is d-sparse, the CCS estimate of $\mathbf{H}_v$ satisfies

$$\Delta\left(\mathbf{H}_v^{\text{ccs}}\right) \leq c_1^2 \cdot \frac{d \cdot N_T}{\mathcal{E}} \cdot \log N_R N_T L \tag{38}$$

with probability at least $1 - 2\max\left\{2(\pi(1+a)\log N_R N_T L \cdot (N_R N_T L)^{2a})^{-1/2}, (N_T N_f)^{-O(1)}\right\}$.

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 [25, Th. 3] and then essentially follows along the lines of the proof of Theorem 6 in [26]. One key observation from the description of the training signal above is that $N_{tr} = \Omega(\bar{d} N_R \cdot \log^5 N_T N_f)$ 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 $N_{tr} = \Omega(d \cdot \log^5 N_o)$ in this setting as opposed to $N_{tr} = \Omega(N_R N_T L)$ for traditional 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 estimating methods. Since—by definition—the 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 [52], 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. Our approach is based on some of the recent advances in the theory of compressed sensing and is accordingly termed as compressed channel sensing (CCS). Ignoring polylogarithmic factors, two distinct features of CCS are: (i) it has a reconstruction error that scales like $O(d)$ as opposed to $\Omega(D)$ for traditional methods, and (ii) it requires the number of receive training dimensions, N_{tr}, to scale like $N_{tr} = \Omega(d)$ for certain signaling and channel configurations as opposed to $N_{tr} = \Omega(D)$ for traditional methods. 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 optimality of CCS in this paper, it has been shown in [24], [25] that its 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 (also, see the accompanying numerical simulations in [24], [25]). Somewhat similar arguments can be made to argue the near-optimal nature of CCS for multiple-antenna sparse channels also. Second, the main ideas underlying the theory of CCS can be easily generalized to channel representations that make use of a basis other than the Fourier one and to other application areas such as high-resolution radar imaging. Third, one expects the representation of real-world multipath channels in certain bases to be often approximately sparse because of the so-called *leakage* effect. While our primary focus in this paper has been on characterizing the performance of CCS for exactly sparse channels, it works equally well for approximately sparse channels thanks to the near-optimal nature of the Dantzig selector; see, e.g., [38, Th. 4]. 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.

REFERENCES

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

[2] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. 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 Processing 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. 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 Processing*, 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 Processing*, 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.* 37*th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 1999.

[10] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. 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 Processing*, 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] A. F. Molisch, "Ultrawideband propagation channels-Theory, measurement, and modeling," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1528–1545, Sep. 2005.

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

[15] 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.

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

[17] 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.

[18] W. Dongming, H. Bing, Z. Junhui, G. Xiqi, and Y. Xiaohu, "Channel estimation algorithms for broadband MIMO-OFDM sparse channel," in *Proc. 14th IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun. (PIMRC '02)*, vol. 2, Beijing, China, Sep. 2003, pp. 1929–1933.

[19] 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.

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

[21] 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 and Signal Processing (ICASSP '08)*, Las Vegas, NV, Apr. 2008, pp. 2885–2888.

[22] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing (ICASSP '08)*, Las Vegas, NV, Apr. 2008, pp. 3469–3472.

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

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

[25] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proc. 45th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2008.

[26] ——, "Compressed sensing of wireless channels in time, frequency, and space," in *Proc. 42nd Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2008.

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

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

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

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

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

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

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

[34] A. M. Sayeed, "A virtual representation for time- and frequency-selective correlated MIMO channels," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing (ICASSP '03)*, vol. 4, Hong Kong, Apr. 2003, pp. 648–651.

[35] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Select. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[36] 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.

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

[38] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inform. Theory*, submitted. [Online]. Available: http://www.ece.wisc.edu/~nowak/sub08_toep.pdf

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

[40] (2009) The compressive sensing resources webpage. Rice University. [Online]. Available: http://www-dsp.rice.edu/cs

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

[42] 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, 2008.

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

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

[45] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Comput.*, vol. 20, no. 1, pp. 33–61, Jan. 1998.

[46] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis of lasso and Dantzig selector," *Ann. Statist.*, in press. [Online]. Available: arXiv:0801.1095v1

[47] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Near-oracle performance of basis pursuit under random noise," *IEEE Trans. Signal Processing*, submitted. [Online]. Available: arXiv:0903.4579v1

[48] 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. Select. Topics Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[49] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Processing*, in press. [Online]. Available: http://www.lx.it.pt/~mtf/SpaRSA/

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

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

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