

## SPARSE MULTIPATH CHANNELS: MODELING AND ESTIMATION

Waheed U. Bajwa, Akbar Sayeed, and Robert Nowak

Electrical and Computer Engineering, University of Wisconsin-Madison  
 bajwa@cae.wisc.edu, akbar@engr.wisc.edu, nowak@engr.wisc.edu

### ABSTRACT

Multipath signal propagation is the defining characteristic of terrestrial wireless channels. Virtually all existing statistical models for wireless channels are implicitly based on the assumption of rich multipath, which can be traced back to the seminal works of Bello and Kennedy on the wide-sense stationary uncorrelated scattering model, and more recently to the i.i.d. model for multi-antenna channels proposed by Telatar, and Foschini and Gans. However, physical arguments and growing experimental evidence suggest that physical channels encountered in practice exhibit a sparse multipath structure that gets more 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 discuss applications of the emerging theory of compressed sensing for efficient estimation of sparse multipath channels.

### 1. INTRODUCTION

Multipath—signal propagation over multiple spatially distributed paths—is the most salient feature of wireless channels and necessitates statistical channel modeling due to the large number of propagation parameters involved. Multipath propagation is both a curse and a blessing from a communications viewpoint [1]. On the one hand, multipath propagation leads to signal *fading*—fluctuations in received signal strength—that severely impacts reliable communication. On the other hand, research in the last decade has shown that multipath is also a source of *diversity*—multiple statistically independent modes of communication—that can increase the rate and reliability of communication. Multipath diversity manifests itself in various forms, including delay, Doppler, spatial and multiuser. The impact of multipath fading versus diversity 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 diversity to combat fading. Further gains in capacity and reliability are possible if (even partial) CSI is available at the transmitter as well.

Statistical characteristics of a wireless channel depend on the interaction between the physical multipath propagation environment and the signal space of the wireless transceivers. For modern wideband, multi-antenna transceivers, this interaction happens in multiple dimensions of time, frequency and space. Technological advances in RF front-ends, including frequency- and bandwidth-agility and reconfigurable antenna arrays, are enabling sensing and exploitation of CSI at varying resolutions afforded by the spatio-temporal signal space. Accurate channel modeling and characterization in time, frequency and space, as a function of multipath and signal space characteristics, is thus critical for studying the impact and potential of such emerging agile wireless transceivers. In particular, while most existing models for wireless channels assume a rich multipath environment, there is growing experimental evidence

that physical channels exhibit a *sparse* structure even with a small number of antennas and especially at wide bandwidths [2, 3].

In this paper, we use a virtual representation of physical multipath channels that we have developed in the past several years to present a framework for modeling sparse wireless channels and to study the implications of multipath sparsity for channel estimation. The virtual channel representation, discussed in Sec. 2, samples the physical multipath in angle-delay-Doppler at a resolution commensurate with the signal space parameters, and the dominant non-vanishing virtual channel coefficients characterize the statistically independent degrees of freedom (DoF) in the channel. Sparse channels, discussed in Sec. 3, exhibit fewer DoF compared to channels induced by rich multipath. We also introduce the concept of *channel sparsity pattern* in Sec. 3 that captures the configuration of the sparse DoF in the angle-delay-Doppler domain and constitutes the most important element of CSI. In Sec. 4, we discuss estimation of sparse channels using the emerging theory of compressed sensing. For the sake of this exposition, we focus only on estimation of time- and frequency-selective single-antenna channels and block-fading narrowband multi-antenna channels. Our discussion focusses on the nature of the waveforms used by the transmitter for probing the channel, the algorithms used at the receiver for learning the sparse channel, and quantification of the mean-squared-error in the resulting channel estimate.

An important application of the modeling and estimation framework proposed in this paper is the emerging area of cognitive radio in which wireless transceivers sense and adapt to the wireless environment for enhanced spectral efficiency and interference management. In particular, the channel estimation strategies discussed in this paper can be leveraged for learning the network CSI—a critical element of cognitive radio. In a related work [4], we have also studied how accurate knowledge of the CSI of a sparse channel can be exploited by agile wireless transceivers for improved link performance.

### 2. VIRTUAL MODELING OF PHYSICAL MULTIPATH WIRELESS CHANNELS

Consider a time- and frequency-selective multi-antenna (MIMO) channel corresponding to a transmitter with  $N_T$  antennas and a receiver with  $N_R$  antennas. For simplicity, we assume uniform linear arrays (ULAs) of antennas and consider signaling over this channel using packets of duration  $T$  and (two-sided) bandwidth  $W$ . In the absence of noise, the transmitted and received signal are related as

$$\mathbf{x}(t) = \int_{-W/2}^{W/2} \mathbf{H}(t, f) \mathbf{S}(f) e^{j2\pi ft} df, \quad 0 \leq t \leq T \quad (1)$$

where  $\mathbf{x}(t)$  is the  $N_R$ -dimensional received signal,  $\mathbf{S}(f)$  is the Fourier transform of the  $N_T$ -dimensional transmitted signal  $\mathbf{s}(t)$ , and  $\mathbf{H}(t, f)$  is the  $N_R \times N_T$  time-varying frequency response matrix of the channel.

A physical multipath wireless channel can be accurately modeled as

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

which represents signal propagation over  $N_p$  paths; here,  $\beta_n$  denotes 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,  $\tau_n$  the (relative) delay, and  $\nu_n$  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  $\theta_T/\theta_R$  and are periodic in  $\theta$  with unit period [5].<sup>1</sup> We assume that  $\tau_n \in [0, \tau_{max}]$  and  $\nu_n \in [-\frac{\nu_{max}}{2}, \frac{\nu_{max}}{2}]$ , where  $\tau_{max}$  denotes the delay spread and  $\nu_{max}$  the (two-sided) Doppler spread of the channel. The signaling parameters are chosen so that the channel is doubly-selective:  $T\nu_{max} > 1$  (time-selective) and  $W\tau_{max} > 1$  (frequency-selective). We assume maximum angular spreads,  $(\theta_{R,n}, \theta_{T,n}) \in [-1/2, 1/2] \times [-1/2, 1/2]$ , at critical ( $d = \lambda/2$ ) antenna spacing. Finally, we assume that over the time-scales of interest, the physical path parameters  $\{\theta_{T,n}, \theta_{R,n}, \tau_n, \nu_n\}$  remain fixed; the only variation in the channel is due to variations in the amplitude and phases of  $\{\beta_n\}$ , which are assumed statistically independent across different paths.

While accurate (non-linear) estimation of AoAs, AoDs, delays and Doppler shifts is critical in certain applications, such as radar imaging, it is not critical in a communications context since the ultimate goal is to reliably communicate information over the channel. As such, studying the key communication-theoretic characteristics of doubly-selective MIMO channels is greatly facilitated by a *virtual representation* of the physical model (2) that we have developed in the past several years [5, 6]. The virtual representation of  $\mathbf{H}(t, f)$  is essentially a four-dimensional Fourier series imposed by the spatio-temporal signaling parameters ( $T$ ,  $W$ ,  $N_R$ , and  $N_T$ ) given by

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

$$H_v(i, k, \ell, m) = \frac{1}{N_R N_T T W} \int_0^T \int_{-\frac{W}{2}}^{\frac{W}{2}} \mathbf{a}_R^H \left( \frac{i}{N_R} \right) \mathbf{H}(t, f) \mathbf{a}_T \left( \frac{k}{N_T} \right) e^{-j2\pi \frac{\ell}{T} t} e^{j2\pi \frac{m}{W} f} dt df \quad (4)$$

where (3) represents the expansion of  $\mathbf{H}(t, f)$  in terms of the spatio-temporal Fourier basis functions, and the virtual representation is completely characterized by the coefficients  $\{H_v(i, k, \ell, m)\}$  that can be computed via (4). Comparing (2) and (3), we note that the virtual representation corresponds to sampling the physical multipath environment in the angle-delay-Doppler domain at uniformly spaced virtual AoAs, AoDs, delays and Doppler shifts at resolutions that are commensurate with the signal space parameters:

$$\Delta\theta_R = 1/N_R \quad , \quad \Delta\theta_T = 1/N_T \quad (5)$$

$$\Delta\tau = 1/W \quad , \quad \Delta\nu = 1/T \quad . \quad (6)$$

In (3),  $L = \lceil W\tau_{max} \rceil \geq 1$  denotes the maximum number of resolvable delays, and  $M = \lceil T\nu_{max}/2 \rceil \geq 1$  the maximum number

<sup>1</sup>The normalized angle variable  $\theta$  is related to the physical angle  $\phi$  (measured with respect to array broadside) as  $\theta = d \sin(\phi)/\lambda$  where  $d$  is the antenna spacing and  $\lambda$  is the wavelength of propagation.

of resolvable Doppler shifts within the channel spreads. For maximum angular spreads,  $N_T$  and  $N_R$  reflect the maximum number of resolvable AoDs and AoAs. Note that due to the fixed angle-delay-Doppler sampling, which defines the fixed basis functions in (3), the virtual representation is a *linear* channel representation characterized by the virtual channel coefficients  $\{H_v(i, k, \ell, m)\}$ . Statistical characterization of these virtual channel coefficients is critical from a communication-theoretic viewpoint and is discussed next.

## 2.1. Channel Statistics: Virtual Path Partitioning

An important and insightful property of the virtual representation is that its coefficients partition the physical propagation paths into approximately disjoint subsets. Specifically, define the following subsets of paths, associated with each coefficient  $H_v(i, k, \ell, m)$ , based on the resolution in angle, delay and Doppler

$$\begin{aligned} S_{R,i} &= \{n : \theta_{R,n} \in (i/N_R - 1/2N_R, i/N_R + 1/2N_R)\} \\ S_{T,k} &= \{n : \theta_{T,n} \in (k/N_T - 1/2N_T, k/N_T + 1/2N_T)\} \\ S_{\tau,\ell} &= \{n : \tau_n \in (\ell/W - 1/2W, \ell/W + 1/2W)\} \\ S_{\nu,m} &= \{n : \nu_n \in (m/T - 1/2T, m/T + 1/2T)\} \quad . \quad (7) \end{aligned}$$

For example,  $S_{R,i}$  denotes the set of paths whose *physical* AoAs lie within the resolution bin of size  $\Delta\theta_R$  centered around the  $i$ -th *virtual* receive angle  $\theta_R = i/N_R$  in (3). Then, it can be shown that [5, 6]

$$H_v(i, k, \ell, m) \approx \sum_{n \in S_{R,i} \cap S_{T,k} \cap S_{\tau,\ell} \cap S_{\nu,m}} \beta_n \quad (8)$$

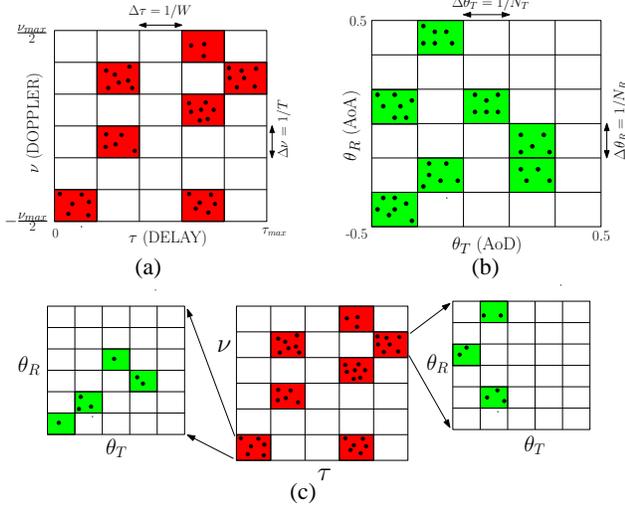
where a phase and attenuation factor has been absorbed in  $\beta_n$ . The relation (8) states that each  $H_v(i, k, \ell, m)$  is approximately equal to the sum of the complex gains of all physical paths whose angles, delays and Doppler shifts lie within an *angle-delay-Doppler resolution bin* of size  $\Delta\theta_R \times \Delta\theta_T \times \Delta\tau \times \Delta\nu$  centered around the virtual sample point  $(\theta_R, \theta_T, \tau, \nu) = (i/N_R, k/N_T, \ell/W, m/T)$  in the angle-delay-Doppler domain (see Fig. 1).

It then follows from (8) that *distinct*  $H_v(i, k, \ell, m)$ 's correspond to *approximately*<sup>2</sup> disjoint subsets of paths and, hence, the virtual channel coefficients are *approximately* statistically independent (due to independent path gains and phases). For simplicity, we assume that the virtual channel coefficients are perfectly independent. For Rayleigh fading,  $\{H_v(i, k, \ell, m)\}$  are zero-mean, statistically independent (complex) Gaussian random variables, and the channel statistics are characterized by the power in the virtual coefficients

$$\begin{aligned} \Psi(i, k, \ell, m) &= E[|H_v(i, k, \ell, m)|^2] \\ &\approx \sum_{n \in S_{R,i} \cap S_{T,k} \cap S_{\tau,\ell} \cap S_{\nu,m}} E[|\beta_n|^2] \quad (9) \end{aligned}$$

which is also a measure of the angle-delay-Doppler power spectrum. Thus, for Rayleigh fading,  $H_v(i, k, \ell, m) \sim \mathcal{CN}(0, \Psi(i, k, \ell, m))$ . We note that the Rayleigh fading assumption is justified by the central limit theorem if there are sufficiently many physical paths contributing to each  $H_v(i, k, \ell, m)$  in (8). However, the virtual representation (3) is not just limited to Rayleigh fading. The statistical independence of  $\{H_v(i, k, \ell, m)\}$ , due to path partitioning, facilitates a wide range of statistical channel models via appropriate modeling of the *marginal statistics* of each  $H_v(i, k, \ell, m)$ .

<sup>2</sup>The approximation is due to the sidelobes induced by the finite signaling parameters and gets more accurate with increasing  $T$ ,  $W$ ,  $N_R$  and  $N_T$ .



**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 coefficient. The total number of squares equals  $D_{max}$ . The shaded squares represent the SP,  $S_D$ , corresponding to the  $D < D_{max}$  dominant coefficients, and the dots represent the paths contributing to each dominant coefficient. (a) VCR and SP in delay-Doppler:  $\{H_v(\ell, m)\}_{S_D}$ . (b) VCR and SP in angle:  $\{H_v(i, k)\}_{S_D}$ . (c) VCR and SP in angle-delay-Doppler:  $\{H_v(i, k, \ell, m)\}_{S_D}$ . The paths contributing to a fixed dominant delay-Doppler coefficient,  $H_v(\ell_o, m_o)$ , are further resolved in angle to yield the conditional SP in angle:  $\{H_v(i, k, \ell_o, m_o)\}_{S_D(\ell_o, m_o)}$ .

## 2.2. Channel Degrees of Freedom

The number of dominant non-vanishing  $\{H_v(i, k, \ell, m)\}$  represent the statistically independent *degrees of freedom* (DoF),  $D$ , in the channel that govern its capacity and diversity.

**Definition 1** (Degrees of Freedom). Let  $D = |\{(i, k, \ell, m) : |\Psi(i, k, \ell, m)| > \epsilon\}|$  denote the number of dominant virtual channel coefficients for some appropriately chosen  $\epsilon > 0$ . Then  $D$ —the number of independent virtual coefficients that significantly contribute to channel power—reflects the number of statistically independent DoF in the channel.

The value of  $\epsilon$  in the above definition is nuanced and depends on the operating signal-to-noise ratio (SNR). An intuitive choice for  $\epsilon$  is the operating received SNR per channel dimension, meaning only channel coefficients with power above this threshold contribute to the DoF. Note that the maximum number of DoF in a channel is

$$D_{max} = N_R N_T (L + 1)(2M + 1) \approx \tau_{max} \nu_{max} N_R N_T T W \quad (10)$$

which corresponds to the *maximum number* of angle-delay-Doppler resolutions bins in the virtual representation, and reflects the maximum number of *resolvable* paths within the angular, delay and Doppler spreads (see Fig. 1). By virtue of (8), we have  $D_{max} \leq N_p$ ; also,  $D \leq D_{max}$  and  $D = D_{max}$  if there are at least  $N_p \geq D_{max}$  physical paths distributed in a way such that each angle-delay-Doppler resolution bin is populated by at least one path.

## 3. SPARSE MULTIPATH WIRELESS CHANNELS

Let  $N_c = N_R N_T T W$  denote the dimension of the spatio-temporal channel space, induced by the transmit signal space of dimension

$N_{s,t,x} = N_T T W$  and receive signal space of dimension  $N_{s,r,x} = N_R T W$ . We note from (10) that for underspread channels, corresponding to  $\tau_{max} \nu_{max} \ll 1$  [1],  $D_{max} = \tau_{max} \nu_{max} N_c \ll N_c$ . All existing models for doubly-selective MIMO channels are implicitly based on the assumption of wide-sense stationary uncorrelated scattering (WSSUS) [7, 8, 9, 10], which in turn implies a *rich scattering* environment in which there are sufficiently many propagation paths so that  $D = D_{max}$  and the channel DoF scale *linearly* with the channel dimension

$$D_{rich} = D_{max} = \tau_{max} \nu_{max} N_c = O(N_c). \quad (11)$$

In many physical channels encountered in practice, however, the number of paths may not be large enough to excite  $D_{max}$  DoF, especially as we increase the channel dimension by increasing the number of antennas, bandwidth, or signaling duration. This has been supported by experimental measurement campaigns, both for indoor MIMO channels (see, e.g., [3]), and ultrawideband single-antenna channels (see, e.g., [2]). When  $D \ll D_{max}$ , we refer to such channels as *sparse multipath channels*. We formalize this notion of sparsity in the following definition.

**Definition 2** (Sparse Multipath Channels). Let  $D$  denote the channel DoF. A sparse multipath channel satisfies  $D \ll D_{max}$  and, furthermore, its DoF scale sub-linearly with the channel dimension

$$D = o(N_c) \iff \lim_{N_c \rightarrow \infty} \frac{D}{N_c} = 0. \quad (12)$$

**Remark 1.** A variety of sub-linear scaling laws can be imposed on  $D$  to capture the sparsity in the channel. As an example, consider

$$\begin{aligned} D &= D_{S,max}^{\delta_1} D_{T,max}^{\delta_2} D_{W,max}^{\delta_3} \\ &= (N_R N_T)^{\delta_1} (T \nu_{max})^{\delta_2} (W \tau_{max})^{\delta_3} \end{aligned} \quad (13)$$

for some  $\delta_1, \delta_2, \delta_3 \in [0, 1]$ , where  $D_{S,max} = N_R N_T$ ,  $D_{T,max} = T \nu_{max}$  and  $D_{W,max} = W \tau_{max}$  denote the maximum number of DoF in the spatial, temporal and spectral dimension, respectively. Here, the extreme case  $\delta_1 = \delta_2 = \delta_3 = 1$  represents a rich multipath environment in which  $D$  scales linearly with  $N_c$ . On the other extreme,  $\delta_1 = \delta_2 = \delta_3 = 0$  represents a very sparse environment in which  $D$  remains constant regardless of  $N_c$ .

Sparse multipath channels represent a sparse distribution of resolvable paths in the angle-delay-Doppler domain. Sparsity in angle-delay-Doppler leads to *correlation* or *coherence* in space-frequency-time due to the Fourier relation between the angle-delay-Doppler and space-frequency-time domains. Furthermore, the *locations* of the  $D$  dominant virtual coefficients within the  $D_{max}$  angle-delay-Doppler resolution bins influence the nature of channel correlation in time, frequency and space. This information about the channel can be captured through the notion of the *channel sparsity pattern*.

**Definition 3** (Channel Sparsity Pattern). Let  $S_D = \{(i, k, \ell, m) : |\Psi(i, k, \ell, m)| > \epsilon\}$ , for some appropriately chosen  $\epsilon > 0$ , denote the channel sparsity pattern. That is,  $S_D$  is the set of indices of the  $D = |S_D|$  dominant virtual channel coefficients.

Essentially, the sparsity pattern  $S_D$  characterizes the  $D$ -dimensional subspace of the  $N_c$ -dimensional channel space that is excited by the  $D$  dominant and statistically independent virtual channel coefficients  $\{H_v(i, k, \ell, m)\}_{S_D}$  representing the stochastic DoF in the channel (see Fig. 1). This means that the statistical CSI of Rayleigh fading sparse channels is completely characterized by  $\{\Psi(i, k, \ell, m)\}_{S_D}$ , whereas their instantaneous CSI is characterized by the realizations of  $\{H_v(i, k, \ell, m)\}_{S_D}$ .

#### 4. ESTIMATION OF SPARSE MULTIPATH CHANNELS

One of the most popular and widely used approaches to learning a multipath wireless channel is to probe the channel with signaling waveforms that are known to the receiver (referred to as training waveforms) and process the corresponding channel output to estimate the channel parameters. The performance of such training-based channel estimation methods is completely characterized by the number of signal space dimensions,  $N_{tr}$ , occupied by the training signals and the mean-squared-error (MSE) associated with the estimation of channel parameters. As such, there are two salient aspects to training-based estimation schemes, namely, *sensing* and *reconstruction*. Sensing corresponds to the design of training waveforms used to probe a channel, while reconstruction is the problem of processing of the corresponding channel output at the receiver to recover the channel response.

Training-based methods that are designed under the assumption of rich multipath scattering typically utilize  $N_{tr} = O(D_{max})$  signal space dimensions for channel sensing and employ linear reconstruction strategies at the receiver for channel estimation—see, e.g., [11, 12, 13, 14, 15]. Traditional channel estimation schemes such as these, however, lead to overutilization of the key communication resources of energy and bandwidth in sparse multipath channels. Consequently, a number of alternative training-based methods have been proposed in recent years for estimating single- and multi-antenna sparse multipath channels—see, e.g., [16, 17, 18, 19]. However, these and similar investigations either lack a quantitative theoretical analysis of the performance of the proposed methods [16, 17, 18] or effectively assume perfect knowledge of the channel sparsity pattern [19]. In contrast, by leveraging key ideas from the theory of compressed sensing (CS), we propose new training-based channel estimation methods in this section that are provably more efficient than the traditional schemes and do not rely on knowledge of the sparsity pattern  $\mathcal{S}_D$ . Our proposed methods utilize  $N_{tr} \approx O(D)$  signal space dimensions for channel sensing, employ non-linear reconstruction algorithms based on convex/linear programming at the receiver, and come within a logarithmic factor of the performance of an ideal channel estimator that has perfect knowledge of the channel sparsity pattern.

##### 4.1. Estimation of Single-Antenna Channels: Sparsity in the Delay-Doppler Domain

In the case of a doubly-selective single-antenna channel, the physical channel model (2) and its virtual representation (3) reduce to

$$H(t, f) = \sum_n \beta_n e^{j2\pi\nu_n t} e^{-j2\pi\tau_n f} \approx \sum_{\ell=0}^L \sum_{m=-M}^M H_v(\ell, m) e^{j2\pi\frac{m}{T}t} e^{-j2\pi\frac{\ell}{W}f} \quad (14)$$

$$H_v(\ell, m) \approx \sum_{n \in \mathcal{S}_{\tau, \ell} \cap \mathcal{S}_{\nu, m}} \beta_n \quad (15)$$

and the (complex) baseband signal at the receiver is given by

$$x(t) \approx \sum_{\ell=0}^L \sum_{m=-M}^M H_v(\ell, m) e^{j2\pi\frac{m}{T}t} s(t - \ell/W) + w(t). \quad (16)$$

Here,  $s(t)$  represents the transmitted waveform and  $w(t)$  denotes the zero-mean, circularly symmetric, complex additive white Gaussian noise (AWGN) at the receiver. From (14), the maximum number of

DoF in a doubly-selective single-antenna channel is  $D_{max} = (L + 1)(2M + 1) \approx \tau_{max}\nu_{max}TW$ , and we assume that the channel is sparse in the delay-Doppler domain in the sense that  $D = |\mathcal{S}_D| \ll D_{max}$  for a given  $T$  and  $W$ .

##### 4.1.1. Orthogonal Short-Time Fourier Signaling

Signaling over orthogonal short-time Fourier (STF) (or Gabor) basis functions provides a very attractive approach for communication over sufficiently underspread doubly-selective channels since appropriately chosen STF basis functions serve as approximate eigenfunctions for such channels [20, 21]. A complete orthogonal STF basis for the  $TW$ -dimensional signal space is generated via time and frequency shifts of a fixed prototype pulse  $g(t)$ :  $\gamma_{i,k}(t) = g(t - iT_o)e^{j2\pi kW_o t}$ ,  $(i, k) \in \mathcal{S} = \{0, 1, \dots, N_t - 1\} \times \{0, 1, \dots, N_f - 1\}$ , where  $N_t = T/T_o$  and  $N_f = W/W_o$ . The prototype pulse is assumed to have unit energy,  $\int |g(t)|^2 dt = 1$ , and completeness of  $\{\gamma_{i,k}\}$  stems from the underlying assumption that  $T_o W_o = 1$ , which results in a total of  $N_t N_f = TW$  basis elements.<sup>3</sup> The transmitted signal in this case can be represented as

$$s(t) = \sum_{i=0}^{N_t-1} \sum_{k=0}^{N_f-1} s_{i,k} \gamma_{i,k}(t), \quad 0 \leq t \leq T \quad (17)$$

where  $\{s_{i,k}\}$  represent the  $TW$  symbols that are modulated onto the STF basis waveforms. At the receiver, assuming that the basis parameters  $T_o$  and  $W_o$  are matched to the channel parameters  $\tau_{max}$  and  $\nu_{max}$  so that  $\gamma_{i,k}$ 's serve as approximate eigenfunctions[21],<sup>4</sup> projecting the (noisy) received signal  $x(t)$  onto the STF basis waveforms yields

$$x_{i,k} = \langle s, \gamma_{i,k} \rangle \approx H_{i,k} s_{i,k} + w_{i,k}, \quad (i, k) \in \mathcal{S} \quad (18)$$

where  $\langle s, \gamma_{i,k} \rangle = \int s(t) \gamma_{i,k}^*(t) dt$  and  $\{w_{i,k}\}$  corresponds to an AWGN sequence. Here, the  $N_c = TW$  channel coefficients  $\{H_{i,k}\}$  represent the underlying channel in the STF domain and are related to the physical channel as  $H_{i,k} \approx H(t, f)|_{(t,f)=(iT_o, kW_o)}$  [21]. Using the virtual representation (14), we can further write the STF channel coefficients as

$$H_{i,k} = \sum_{\ell=0}^{L-1} \sum_{m=-M}^M H_v(\ell, m) e^{j2\pi\frac{m}{T}i} e^{-j2\pi\frac{\ell}{W}k} = \mathbf{u}_{f,k}^T \mathbf{H}_v \mathbf{u}_{t,i} = (\mathbf{u}_{t,i}^T \otimes \mathbf{u}_{f,k}^T) \text{vec}(\mathbf{H}_v) = (\mathbf{u}_{t,i}^T \otimes \mathbf{u}_{f,k}^T) \mathbf{h}_v \quad (19)$$

where  $\mathbf{H}_v$  is the  $(L + 1) \times (2M + 1)$  matrix of  $D_{max}$  virtual channel coefficients,  $\mathbf{u}_{f,k} = \begin{bmatrix} 1 & e^{-j2\pi\frac{1}{W}k} & \dots & e^{-j2\pi\frac{L}{W}k} \end{bmatrix}^T \in \mathbb{C}^{L+1}$ ,  $\mathbf{u}_{t,i} = \begin{bmatrix} e^{-j2\pi\frac{M}{T}i} & e^{-j2\pi\frac{(M-1)}{T}i} & \dots & e^{j2\pi\frac{M}{T}i} \end{bmatrix}^T \in \mathbb{C}^{2M+1}$ , and  $\mathbf{h}_v = \text{vec}(\mathbf{H}_v) \in \mathbb{C}^{D_{max}}$ .<sup>5</sup>

In training-based methods,  $N_{tr}$  of the  $N_c$  STF basis elements are dedicated as ‘‘pilot tones’’ for learning the channel. That is, the STF training waveform  $s_{tr}(t)$  takes the form

$$s_{tr}(t) = \sqrt{\frac{\mathcal{E}}{N_{tr}}} \sum_{(i,k) \in \mathcal{S}_{tr}} \gamma_{i,k}(t), \quad 0 \leq t \leq T \quad (20)$$

<sup>3</sup>Note that signaling over a complete orthogonal STF basis can be thought of as block orthogonal frequency division multiplexing (OFDM) signaling with OFDM symbol duration  $T_o$  and block length  $N_t = T/T_o$ .

<sup>4</sup>Two necessary matching conditions are: (i)  $\tau_{max} \leq T_o \leq 1/\nu_{max}$  and (ii)  $\nu_{max} \leq W_o \leq 1/\tau_{max}$ ; we refer the reader to [21] for further details.

<sup>5</sup>Under the assumption of STF basis parameters being matched to the channel parameters, we trivially have  $N_f \geq L + 1$  and  $N_t \geq 2M + 1$ .

where  $\mathcal{E}$  is the transmit energy budget available for training and  $\mathcal{S}_{tr}$  is the set of indices of  $N_{tr}$  pilot tones;  $\mathcal{S}_{tr} \subset \mathcal{S} : |\mathcal{S}_{tr}| = N_{tr}$ . At the receiver, we can stack the received training symbols  $\{x_{i,k}\}_{\mathcal{S}_{tr}}$  into an  $N_{tr}$ -dimensional vector  $\mathbf{x}$  to yield the following system of equations

$$\mathbf{x} = \sqrt{\frac{\mathcal{E}}{N_{tr}}} \mathbf{U}_{tr} \mathbf{h}_v + \mathbf{w} \quad (21)$$

where the  $N_{tr} \times D_{max}$  matrix  $\mathbf{U}_{tr}$  is comprised of  $\{(\mathbf{u}_{t,i}^T \otimes \mathbf{u}_{f,k}^T) : (i,k) \in \mathcal{S}_{tr}\}$  as its rows and the AWGN vector  $\mathbf{w}$  is distributed as  $\mathcal{CN}(\mathbf{0}_{N_{tr}}, \mathbf{I}_{N_{tr}})$ . The goal then is to choose a set of pilot tones  $\mathcal{S}_{tr}$  and process the corresponding channel output  $\mathbf{x}$  to obtain an estimate  $\hat{\mathbf{h}}_v$  that is close to  $\mathbf{h}_v$  in terms of the MSE.

In many traditional training-based receivers, it is assumed that the number of pilot tones  $N_{tr} \geq D_{max}$  and linear reconstruction schemes such as maximum likelihood (ML) estimators are used to recover  $\mathbf{h}_v$  from  $\mathbf{x}$ :  $\hat{\mathbf{h}}_v = \sqrt{N_{tr}/\mathcal{E}}(\mathbf{U}_{tr}^H \mathbf{U}_{tr})^{-1} \mathbf{U}_{tr}^H \mathbf{x}$ . It can be shown in this case that, regardless of the choice of  $\mathcal{S}_{tr}$ , the MSE in the channel estimate is lower bounded as [12, 22]

$$\mathbb{E} \left[ \|\hat{\mathbf{h}}_v - \mathbf{h}_v\|_2^2 \right] \geq \frac{D_{max}}{\mathcal{E}}. \quad (22)$$

On the other hand, consider an ideal (nonrealizable) channel estimator that has perfect knowledge of the sparsity pattern  $\mathcal{S}_D$  and assume for the sake of this exposition that the *non-dominant* virtual channel coefficients are identically zero:  $\Psi(\ell, m) = 0 \forall (\ell, m) \notin \mathcal{S}_D$ . Then, as long as the number of pilot tones  $N_{tr} \geq D$ , an ideal channel estimate  $\mathbf{h}_v^*$  can be obtained from  $\mathbf{x}$  by first forming a *restricted* ML estimate  $\mathbf{h}_{\mathcal{S}_D}^* = \sqrt{N_{tr}/\mathcal{E}}(\mathbf{U}_{tr, \mathcal{S}_D}^H \mathbf{U}_{tr, \mathcal{S}_D})^{-1} \mathbf{U}_{tr, \mathcal{S}_D}^H \mathbf{x}$ , where  $\mathbf{U}_{tr, \mathcal{S}_D}$  is the submatrix obtained by extracting the  $D$  columns of  $\mathbf{U}_{tr}$  corresponding to the indices in  $\mathcal{S}_D$ , and then setting  $\mathbf{h}_v^*$  equal to  $\mathbf{h}_{\mathcal{S}_D}^*$  on the indices in  $\mathcal{S}_D$  and zero on other locations. The MSE of this ideal channel estimate can be lower bounded as [22]

$$\mathbb{E} \left[ \|\mathbf{h}_v^* - \mathbf{h}_v\|_2^2 \right] \geq \frac{D}{\mathcal{E}}. \quad (23)$$

The preceding discussion suggests that it might be possible to improve upon the performance of traditional training-based channel learning methods by a factor of about  $O(D_{max}/D)$ , both in terms of the minimum number of pilot tones needed for meaningful estimation and the MSE of the resulting channel estimate. And while the ideal channel estimate  $\mathbf{h}_v^*$  is impossible to construct in practice, we now show that it is possible to obtain an estimate of  $\mathbf{h}_v$  that comes within a logarithmic factor of the performance of the ideal estimator.

**Theorem 1.** *Let the number of pilot tones  $N_{tr} \geq c_1 \cdot \log^5 N_c \cdot D$  and choose  $\mathcal{S}_{tr}$  to be a set of  $N_{tr}$  ordered pairs sampled uniformly at random from  $\mathcal{S}$ . Pick  $\lambda(\mathcal{E}, D_{max}) = \sqrt{2\mathcal{E}(1+a) \log D_{max}}$  for any  $a \geq 0$ . Then the channel estimate obtained as a solution to the convex program*

$$\hat{\mathbf{h}}_v = \arg \min_{\tilde{\mathbf{h}} \in \mathbb{C}^{D_{max}}} \|\tilde{\mathbf{h}}\|_1 \quad \text{subject to} \quad \left\| \sqrt{\frac{\mathcal{E}}{N_{tr}}} \mathbf{U}_{tr}^H \mathbf{r} \right\|_\infty \leq \lambda \quad (24)$$

satisfies

$$\|\hat{\mathbf{h}}_v - \mathbf{h}_v\|_2^2 \leq c_2 \cdot \log D_{max} \cdot \left( \frac{D}{\mathcal{E}} \right) \quad (25)$$

with probability  $\geq 1 - 2 \max \left\{ 2 \left( \sqrt{\pi(1+a) \log D_{max}} \cdot D_{max}^a \right)^{-1}, c_3 N_c^{-c_4} \right\}$ . Here,  $\mathbf{r}$  in (24) is the  $N_{tr}$ -dimensional vector of residuals:  $\mathbf{r} = \mathbf{x} - \sqrt{\mathcal{E}/N_{tr}} \mathbf{U}_{tr} \tilde{\mathbf{h}}$ , and  $c_1, c_2, c_3$  and  $c_4$  are strictly positive constants that do not depend on  $\mathcal{E}, D_{max}$  or  $N_c$ .

The proof of this theorem, which uses some of the key results in CS, is given in [22]. The convex program (24) goes by the name of Dantzig selector (DS) in the CS literature and is computationally tractable since it can be recast as a linear program [23]. Theorem 1 essentially states that our proposed DS-based channel estimator comes remarkably close to matching the performance of the ideal estimator and can potentially reduce both the number of pilots tones needed for channel estimation and the MSE in the resulting estimate by a factor of about  $O(D_{max}/D)$  when used as an alternative to existing methods for learning single-antenna channels.

## 4.2. Estimation of Narrowband Multi-Antenna Channels: Sparsity in the Angular Domain

For block-fading narrowband MIMO channels (corresponding to  $T\nu_{max}$  and  $W\tau_{max} \ll 1$ ), the physical channel model (2) and its virtual representation (3) reduce to

$$\mathbf{H} = \sum_n \beta_n \mathbf{a}_R(\theta_{R,n}) \mathbf{a}_T^H(\theta_{T,n}) \approx \mathbf{A}_R \mathbf{H}_v \mathbf{A}_T^H \quad (26)$$

where  $\mathbf{A}_R$  and  $\mathbf{A}_T$  are  $N_R \times N_R$  and  $N_T \times N_T$  unitary discrete Fourier transform (DFT) matrices, respectively. The  $N_R \times N_T$  beamspace matrix  $\mathbf{H}_v$  couples the virtual AoAs and AoDs, and its entries are given by the virtual channel coefficients

$$H_v(i, k) \approx \sum_{n \in \mathcal{S}_{R,i} \cap \mathcal{S}_{T,k}} \beta_n. \quad (27)$$

From (26), the maximum number of DoF in a block-fading narrowband MIMO channel is  $D_{max} = N_R N_T$ , and we assume that: (i) the channel is sparse in the angular domain in the sense that  $D = |\mathcal{S}_D| \ll D_{max}$  for a given  $N_R$  and  $N_T$ , and (ii) the non-dominant virtual channel coefficients are zero:  $\Psi(i, k) = 0 \forall (i, k) \notin \mathcal{S}_D$ .

### 4.2.1. Beamspace Signaling

Since  $\mathbf{H}$  and  $\mathbf{H}_v$  are unitarily equivalent to each other, we focus only on estimating  $\mathbf{H}_v$  and assume without loss of generality that signaling and reception take place in the beamspace:  $\mathbf{s} = \mathbf{A}_T \mathbf{s}_v$  and  $\mathbf{x}_v = \mathbf{A}_R^H \mathbf{x}$ , where  $\mathbf{s}$  and  $\mathbf{x}$  are the transmitted and received signals in the antenna domain, respectively. To learn the  $N_c = N_R N_T$ -dimensional matrix  $\mathbf{H}_v$ , training-based methods dedicate part of the packet duration  $T$  to transmit known signals to the receiver. Assuming this training duration to be  $T_{tr}$ , we can stack the  $M_{tr} = T_{tr} W$  received (vector-valued) training symbols into an  $M_{tr} \times N_R$  matrix  $\mathbf{X}_v$  to yield the following system of equations

$$\mathbf{X}_v = \sqrt{\frac{\mathcal{E}}{M_{tr}}} \mathbf{S}_v \mathbf{H}_v^T + \mathbf{W} \quad (28)$$

where  $\mathcal{E}$  is the transmit energy budget available for training,  $\mathbf{S}_v$  is the collection of  $M_{tr}$  (vector) training symbols stacked row-wise into an  $M_{tr} \times N_T$  matrix with the constraint that  $\|\mathbf{S}_v\|_F^2 = M_{tr}$ , and  $\mathbf{W}$  is an  $M_{tr} \times N_R$  matrix of unit-variance AWGN noise. The goal here is to design the training matrix  $\mathbf{S}_v$  using fewest number of training symbols  $M_{tr}$  and process the received signal matrix  $\mathbf{X}_v$  to obtain an estimate  $\hat{\mathbf{H}}_v$  that is close to  $\mathbf{H}_v$  in terms of the MSE.

To obtain a meaningful estimate of  $\mathbf{H}_v$ , traditional estimation schemes require the number of training symbols to satisfy  $M_{tr} \geq N_T$  and typically employ ML-based estimators at the receiver:  $\hat{\mathbf{H}}_v = \sqrt{M_{tr}/\mathcal{E}}(\mathbf{S}_v^H \mathbf{S}_v)^{-1} \mathbf{S}_v^H \mathbf{X}_v$ .<sup>6</sup> Regardless of the form of  $\mathbf{S}_v$ , it can be shown in this case that the MSE in the channel

<sup>6</sup>Note that the condition  $M_{tr} \geq N_T$  means that a total of  $N_{tr} \geq D_{max}$  receive signal space dimensions are dedicated to training.

estimate is lower bounded as [14, 15]

$$\mathbb{E} \left[ \|\hat{\mathbf{H}}_v - \mathbf{H}_v\|_F^2 \right] \geq \frac{N_T (N_R N_T)}{\mathcal{E}} = \frac{N_T D_{max}}{\mathcal{E}}. \quad (29)$$

On the other hand, given that there are only  $D$  unknowns (albeit at unknown locations) within  $\mathbf{H}_v$ , it is arguable whether (29) is really optimal. In fact, one could easily conceive estimators having perfect knowledge of the channel sparsity pattern  $\mathcal{S}_D$  that would yield  $\mathbb{E}[\|\hat{\mathbf{H}}_v - \mathbf{H}_v\|_F^2] = O(N_T D / \mathcal{E})$ . However, we now show that it is possible to get within a logarithmic factor of this ideal MSE scaling *without* any knowledge of the sparsity pattern.

**Theorem 2.** Let  $\mathbf{h}_{v,i}^T$  and  $\mathbf{x}_{v,i}$  denote the  $i$ -th row of  $\mathbf{H}_v$  and  $i$ -th column of  $\mathbf{X}_v$ , respectively. Further, let  $D_i$  denote the number of channel DoF contributed by each row of  $\mathbf{H}_v$ ; that is,  $D_i = \|\mathbf{h}_{v,i}\|_0$  (note that  $\sum_{i=1}^{N_R} D_i = D$ ). Choose the number of training symbols  $M_{tr} \geq c_5 \cdot \log(N_T / \max_i D_i) \cdot \max_i D_i$  and let  $\mathbf{S}_v$  be an i.i.d matrix of binary random variables taking values  $+1/\sqrt{N_T}$  or  $-1/\sqrt{N_T}$  with probability  $1/2$  each. Pick  $\lambda(\mathcal{E}, N_T) = \sqrt{2\mathcal{E}(1+a)(\log D_{max})/N_T}$  for any  $a \geq 0$  and let

$$\hat{\mathbf{h}}_{v,i} = \arg \min_{\hat{\mathbf{h}} \in \mathbb{C}^{N_T}} \|\hat{\mathbf{h}}\|_1 \quad \text{subject to} \quad \left\| \sqrt{\frac{\mathcal{E}}{M_{tr}}} \mathbf{S}_v^H \mathbf{r}_i \right\|_\infty \leq \lambda \quad (30)$$

where  $\mathbf{r}_i = \mathbf{x}_{v,i} - \sqrt{\mathcal{E}/M_{tr}} \mathbf{S}_v \hat{\mathbf{h}}$ ,  $i = 1, \dots, N_R$ . Then the channel estimate  $\hat{\mathbf{H}}_v = [\hat{\mathbf{h}}_{v,1} \ \dots \ \hat{\mathbf{h}}_{v,N_R}]^T$  satisfies

$$\|\hat{\mathbf{H}}_v - \mathbf{H}_v\|_F^2 \leq c_6 \cdot \log D_{max} \cdot \left( \frac{N_T D}{\mathcal{E}} \right) \quad (31)$$

with probability  $\geq 1 - 4 \max \left\{ \left( \sqrt{\pi(1+a) \log D_{max}} \cdot D_{max}^a \right)^{-1}, e^{-c_7 M_{tr}} \right\}$ . Here,  $c_5, c_6$  and  $c_7$  are strictly positive constants that do not depend on  $\mathcal{E}, N_R$  or  $N_T$ .

The proof of this theorem is given in [24]. It is clearly evident from (31) that our proposed DS-based (row-by-row) MIMO channel estimator (30) achieves near-optimal MSE performance. In addition, since  $\max_i D_i \leq N_T$ , it does so while potentially requiring considerably fewer number of training symbols as compared to traditional schemes. In fact, if one assumes that the  $D$  channel DoF are uniformly distributed across the angular spread then we have  $\max_i D_i \approx D/N_R$  and hence, the proposed estimator requires  $M_{tr} \approx O(D/N_R)$  as opposed to  $M_{tr} = O(N_T)$  for existing methods; in terms of the receive signal space dimensions dedicated to training, this translates into  $N_{tr} \approx O(D)$  for the proposed channel estimator versus  $N_{tr} = O(D_{max})$  for ML-based estimators.

## 5. REFERENCES

- [1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, U.K., 2005.
- [2] A. F. Molisch, "Ultrawideband propagation channels-Theory, measurement, and modeling," *IEEE Trans. Veh. Technol.*, pp. 1528–1545, Sept. 2005.
- [3] Z. Yan, M. Herdin, A. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Tech. Rep., University of Wisconsin-Madison, Feb. 2007.
- [4] A. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE J. Select. Topics Signal Processing*, pp. 156–166, June 2007.
- [5] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Processing*, pp. 2563–2579, Oct. 2002.
- [6] A. Sayeed, "A virtual representation for time- and frequency-selective correlated MIMO channels," in *Proc. ICASSP 2003*.
- [7] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun.*, pp. 360–393, Dec. 1963.
- [8] R. S. Kennedy, *Fading Dispersive Communication Channels*, Wiley-Interscience, 1969.
- [9] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, pp. 585–595, Nov. 1999.
- [10] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, pp. 311–335, Mar. 1998.
- [11] M.-A. Baissas and A. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers," *IEEE Trans. Signal Processing*, pp. 2037–2049, Aug. 2002.
- [12] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, pp. 1122–1128, Aug. 1998.
- [13] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, pp. 223–229, Sept. 2002.
- [14] T. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. 1999*.
- [15] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, pp. 951–963, Apr. 2003.
- [16] S. Cotter and B. Rao, "The adaptive matching pursuit algorithm for estimation and equalization of sparse time-varying channels," in *Proc. 34th Asilomar Conf. 2000*.
- [17] W. Dongming, H. Bing, Z. Junhui, G. Xiqi, and Y. Xiaohu, "Channel estimation algorithms for broadband MIMO-OFDM sparse channel," in *Proc. PIMRC 2002*.
- [18] G. Tauböck and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. ICASSP 2008*.
- [19] V. Raghavan, G. Hariharan, and A. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *IEEE J. Select. Topics Signal Processing*, pp. 357–371, Oct. 2007.
- [20] W. Kozek and A. F. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE J. Select. Areas Commun.*, pp. 1579–1589, Oct. 1998.
- [21] K. Liu, T. Kadous, and A. Sayeed, "Orthogonal time-frequency signaling over doubly dispersive channels," *IEEE Trans. Inform. Theory*, pp. 2583–2603, Nov. 2004.
- [22] W. Bajwa, A. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proc. 46th Annu. Allerton Conf. 2008* (also, Tech. Rep. ECE-08-02, Univ. of Wisconsin-Madison).
- [23] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ ," *Ann. Statist.*, pp. 2313–2351, Dec. 2007.
- [24] W. Bajwa, A. Sayeed, and R. Nowak, "Compressed sensing of wireless channels in time, frequency, and space," in *Proc. 42nd Asilomar Conf. 2008*.