Compressed Sensing of Wireless Channels in Time, Frequency, and Space

Waheed U. Bajwa, Akbar Sayeed, and Robert Nowak
Dept. of Electrical and Computer Engineering, University of Wisconsin-Madison
bajwa@cae.wisc.edu, akbar@engr.wisc.edu, nowak@engr.wisc.edu

***Abstract*—Training-based channel estimation involves probing of the channel in time, frequency, and space by the transmitter with known signals, and estimation of channel parameters from the output signals at the receiver. Traditional training-based methods, often comprising of maximum likelihood estimators, are known to be optimal under the assumption of rich multipath channels. Numerous measurement campaigns have shown, however, that physical multipath channels exhibit a sparse structure in angle-delay-Doppler, especially at large signal space dimensions. In this paper, key ideas from the emerging theory of compressed sensing are leveraged to: (i) propose new methods for efficient estimation of sparse multi-antenna channels, and (ii) show that explicitly accounting for multipath sparsity in channel estimation can result in significant performance improvements when compared with existing training-based methods.**

I. Introduction

Coherent communication over multi-antenna (MIMO) channels requires knowledge of the channel state information (CSI) at the receiver. In practice, however, communication systems have seldom access to the CSI and the channel needs to be first learned at the receiver to reap the benefits of coherent communication. In this paper, we focus on learning *sparse* MIMO channels—channels with most of the multipath energy localized to relatively small regions within the angle-delay-Doppler spread of the channel.

One of the most popular and widely used approaches to learning a MIMO channel is to probe it with known signaling waveforms (referred to as training signals) and process the corresponding channel output to estimate the channel parameters. Almost all existing training-based channel estimation methods in the literature are based on the assumption of a *rich* underlying multipath environment; the number of degrees of freedom in the MIMO channel are assumed to scale linearly with the signal space dimensions. In contrast, physical MIMO channels encountered in practice tend to exhibit impulse responses dominated by a relatively small number of dominant resolvable paths, especially when operating at large bandwidths and signaling durations and/or with number of antennas [1], [2]. Traditional channel estimation schemes such as [3]–[6], however, lead to overutilization of the key communication resources of energy and bandwidth in sparse MIMO channels. In contrast, by leveraging key ideas from the theory of compressed sensing, we propose new training-based channel estimation methods in this paper that are provably more efficient than the traditional schemes. Our discussion focusses on the nature of the signals used by the transmitter for probing a sparse MIMO channel, the algorithms used at the receiver for learning the channel, and quantification of the mean squared-error in the resulting channel estimate.

The rest of this paper is organized as follows. In Section II, a modeling framework for MIMO channels is reviewed and the notion of sparse MIMO channels is formally described. Section III considers the problem of learning sparse narrowband MIMO channels. Finally, Section IV discusses extensions of the results of Section III to learn sparse wideband MIMO channels using multicarrier training signals.

II. Multipath Wireless Channel Modeling

Consider a MIMO channel corresponding to uniform linear arrays of N_T transmit antennas and N_R receive antennas. Throughout the paper, we implicitly consider signaling over this channel using packets of duration T and (two-sided) bandwidth W. In the absence of noise, the corresponding baseband transmitted and received signal are related as

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

where $\mathbf{x}(t)$ is the N_R-dimensional received signal, $\mathbf{S}(f)$ is the (element-wise) 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.

One of the most salient characteristics of multipath wireless channels is signal propagation over multiple spatially distributed paths. A MIMO channel can be accurately modeled in terms of these physical paths as

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

which represents signal propagation over N_p paths; here, β_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, 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 [7].[1] We assume that $\tau_n \in [0, \tau_{max}]$ and $\nu_n \in [-\frac{\nu_{max}}{2}, \frac{\nu_{max}}{2}]$, where τ_{max} denotes the delay spread and ν_{max} the (two-sided) Doppler spread of the channel. Further,

[1]The normalized angle variable 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.

the signaling parameters are chosen so that the channel is doubly-selective: $T\nu_{max} \geq 1$ (time-selective) and $W\tau_{max} \geq 1$ (frequency-selective), and maximum angular spreads are assumed at critical ($d = \lambda/2$) antenna spacing: $(\theta_{R,n}, \theta_{T,n}) \in [-1/2, 1/2] \times [-1/2, 1/2]$.

A. Virtual Representation of MIMO Channels

While the physical model (2) is highly accurate, it is difficult to analyze and learn owing to its *nonlinear* dependence on a potentially large number of physical parameters $\{(\beta_n, \theta_{R,n}, \theta_{T,n}, \tau_n, \nu_n)\}$. However, because of the finite (transmit and receive) array apertures, signaling duration, and bandwidth, the physical model can be well-approximated by a linear (in parameters) counterpart, known as a *virtual channel model*, with the aid of a Fourier series expansion [7], [8].

The key idea behind virtual channel modeling is to provide a low-dimensional approximation of (2) by uniformly sampling the multipath environment in the angle-delay-Doppler domain at a resolution commensurate with the signal space parameters: $(\Delta\theta_R, \Delta\theta_T, \Delta\tau, \Delta\nu) = (1/N_R, 1/N_T, 1/W, 1/T)$. That is,

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

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

where a phase and attenuation factor has been absorbed in the β_n's in (4). In (3), $N_R, N_T, L = \lceil W\tau_{max}\rceil + 1$, and $M = \lceil T\nu_{max}/2\rceil$ denote the maximum number of resolvable AoAs, AoDs, delays, and (one-sided) Doppler shifts within the channel angle-delay-Doppler spread, respectively. Due to the fixed angle-delay-Doppler sampling of (2), which defines the fixed basis functions in (3), the virtual representation is a *linear* channel representation completely characterized by the virtual channel coefficients $\{H_v(i,k,\ell,m)\}$. Further, the relation (4) 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 $(\hat{\theta}_{R,i}, \hat{\theta}_{T,k}, \hat{\tau}_\ell, \hat{\nu}_m) = (i/N_R, k/N_T, \ell/W, m/T)$ in the angle-delay-Doppler domain; we refer the reader to [8] for further details. In essence, the virtual representation (3) effectively approximates a physical MIMO channel in terms of a D_{max}-dimensional parameter comprising of the virtual channel coefficients $\{H_v(i,k,\ell,m)\}$, where $D_{max} = N_R N_T L(2M+1)$.

B. Sparse MIMO Channels

Channel measurement results dating as far back as 1987 [9] and as recent as 2007 [2] suggest that multipath components tend to arrive at the receiver in clusters. Based on the interspacings between different multipath clusters within the angle-delay-Doppler domain, MIMO channels can be characterized as either "rich" or "sparse". In a rich MIMO channel, the interspacings are smaller than $(\Delta\theta_R, \Delta\theta_T, \Delta\tau, \Delta\nu)$. Sparse MIMO channels, on the other hand, exhibit interspacings that are larger than $\Delta\theta_R, \Delta\theta_T, \Delta\tau$, and/or $\Delta\nu$. Not every angle-delay-Doppler bin of size $\Delta\theta_R \times \Delta\theta_T \times \Delta\tau \times \Delta\nu$ contains a physical path in this case. In particular, since a virtual coefficient consists of the sum of gains of all paths falling within its respective angle-delay-Doppler resolution bin, sparse MIMO channels tend to have far fewer than D_{max} nonzero channel coefficients at any fixed (but large enough) number of antennas, signaling duration, and bandwidth. We formalize this notion of multipath sparsity as follows.

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. We say that a MIMO channel is D-sparse if $D = |\mathcal{S}_D| \ll D_{max}$, where $D_{max} = N_R N_T L(2M+1)$ is the total number of resolvable paths (channel coefficients) within the angle-delay-Doppler spread of the channel. Further, the corresponding set of indices $\mathcal{S}_D$ is termed as the *channel sparsity pattern*.

III. Learning Sparse Narrowband MIMO Channels

In the case of a narrowband MIMO channel (corresponding to $W\tau_{max} \ll 1$), the physical channel model (2) and its virtual representation (3) reduce to

$$\mathbf{H} = \sum_n \beta_n \mathbf{a}_R(\theta_{R,n})\mathbf{a}_T^H(\theta_{T,n}) \approx \mathbf{A}_R\mathbf{H}_v\mathbf{A}_T^H \quad (5)$$

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 $D_{max} = N_R N_T$ virtual channel coefficients $\{H_v(i,k)\}$. We further assume that the channel is D-sparse in the angular domain ($D = |\mathcal{S}_D| \ll D_{max}$) and it remains constant over the packet signaling duration T (block-fading assumption corresponding to $T\nu_{max} \ll 1$).

To learn the $N_R \times N_T$ (antenna domain) matrix $\mathbf{H}$, training-based channel estimation methods dedicate part of the packet duration T to transmit known signals to the receiver. Assuming this training duration to be T_{tr}, many traditional training-based receivers stack the $M_{tr} = T_{tr}W$ received (vector-valued) training signals $\{\mathbf{x}(n), n = 1, \ldots, M_{tr}\}$ into an $M_{tr} \times N_R$ matrix $\mathbf{X}$ to yield the following system of equations

$$\mathbf{X} = \sqrt{\frac{\mathcal{E}}{M_{tr}}}\mathbf{S}\mathbf{H}^T + \mathbf{W} \quad (6)$$

where $\mathcal{E}$ is the total transmit energy budget available for training, $\mathbf{S}$ is the collection of M_{tr} training signal vectors $\{\mathbf{s}(n), n = 1, \ldots, M_{tr}\}$ stacked row-wise into an $M_{tr} \times N_T$ matrix with the constraint that $\|\mathbf{S}\|_F^2 = M_{tr}$, and $\mathbf{W}$ is an $M_{tr} \times N_R$ matrix of unit-variance additive white Gaussian noise. The goal then is to design the training matrix $\mathbf{S}$ using fewest number of training vectors M_{tr} and process the received signal matrix $\mathbf{X}$ to obtain an estimate $\widehat{\mathbf{H}}$ that is close to $\mathbf{H}$ in terms of the mean squared-error (MSE).

Conventionally, it is assumed that the number of training vectors $M_{tr} \geq N_T$ and (assuming that $\mathbf{S}$ has full column rank) linear reconstruction schemes such as maximum likelihood (ML) estimators are used to recover $\mathbf{H}$ from $\mathbf{X}$: $\widehat{\mathbf{H}}^T = \sqrt{M_{tr}/\mathcal{E}}\,(\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H\mathbf{X}$. Regardless of the form of $\mathbf{S}$, it can be shown in this case that the MSE in the channel estimate is lower bounded as [3], [4]

$$\mathbb{E}\left[\|\widehat{\mathbf{H}} - \mathbf{H}\|_F^2\right] \geq \frac{N_T\,(N_R N_T)}{\mathcal{E}} = \frac{N_T\,D_{max}}{\mathcal{E}}\ . \qquad (7)$$

Further, the requirement $M_{tr} \geq N_T$ means that traditional methods dedicate a total of $N_{tr} = N_R M_{tr} \geq D_{max}$ *receive* signal space dimensions for training. On the other hand, given that sparse MIMO channels are completely characterized by $D \ll D_{max}$ parameters, it is arguable whether $N_{tr} = O(D_{max})$ and (7) are really optimal. In this regard, we now propose two new training-based estimation schemes for sparse narrowband MIMO channels that leverage key ideas from the theory of compressed sensing to significantly reduce (i) the number of receive signal space dimensions needed for meaningful estimation, and/or (ii) the MSE of the resulting channel estimate. Before proceeding further, however, we briefly review some basic facts about compressed sensing.

A. Review of Compressed Sensing

Compressed sensing (CS) is a relatively new area of theoretical research that lies at the intersection of signal processing, statistics, and computational harmonic analysis. One of the central tenets of CS theory is that a relatively small number of (noisy) linear measurements of a sparse signal can capture most of its salient information. In addition, recent theoretical results have established that the signal can be near-ideally reconstructed from these measurements by solving a linear program, known as the Dantzig selector (DS) [10], provided the collection of measurement vectors satisfy the so-called *restricted isometry property* (RIP).

Definition 2 (Restricted Isometry Property): A matrix $\mathbf{\Psi}$ having unit ℓ_2-norm columns is said to satisfy RIP of order S with parameter $\delta_S \in (0,1)$ if for all $\mathbf{z} : \|\mathbf{z}\|_0 \leq S$

$$(1-\delta_S)\|\mathbf{z}\|_2^2 \leq \|\mathbf{\Psi}\mathbf{z}\|_2^2 \leq (1+\delta_S)\|\mathbf{z}\|_2^2\ . \qquad (8)$$

Our approach to training-based estimation of sparse MIMO channels is inspired by some of these recent advances in the CS theory. By leveraging some of the analytical insights provided in the related CS literature, such as in [10]–[12], we propose new channel estimation methods in this section that employ *nonlinear* DS-based reconstruction algorithms at the receiver, and which are provably more efficient than the traditional schemes. It is therefore instructive to state the reconstruction error performance of the DS. The following theorem is a slight variation on [10, Th. 1.1].

Theorem 1 (The Dantzig Selector): Let $\mathbf{y} = \mathbf{\Psi}\boldsymbol{\theta} + \boldsymbol{\eta} \in \mathbb{C}^n$ be a vector of noisy measurements of $\boldsymbol{\theta} \in \mathbb{C}^p : \|\boldsymbol{\theta}\|_0 \leq S$, where $\boldsymbol{\eta} \sim \mathcal{CN}(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$. Suppose that the $n \times p$ measurement matrix $\mathbf{\Psi}$ has unit ℓ_2-norm columns and it satisfies RIP of order $2S$ with $\delta_{2S} < \sqrt{2}-1$. Choose $\lambda = (2\sigma^2(1+a)\log p)^{1/2}$ for any $a \geq 0$. Then the estimate $\widehat{\boldsymbol{\theta}}$ obtained as a solution to the optimization program

$$\widehat{\boldsymbol{\theta}} = \underset{\tilde{\boldsymbol{\theta}} \in \mathbb{C}^p}{\arg\min}\ \|\tilde{\boldsymbol{\theta}}\|_1 \text{ subject to } \|\mathbf{\Psi}^H(\mathbf{y} - \mathbf{\Psi}\tilde{\boldsymbol{\theta}})\|_\infty \leq \lambda \quad \text{(DS)}$$

satisfies

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2 \leq c_1^2 \cdot \log p \cdot S \cdot \sigma^2 \qquad (9)$$

with probability at least $1 - 2(\pi(1+a)\log p \cdot p^{2a})^{-1/2}$. Here, the constant $c_1 = 4\sqrt{2(1+a)}/\left(1 - (\sqrt{2}+1)\delta_{2S}\right)$.

In the sequel, we will make use of the shorthand notation $\widehat{\boldsymbol{\theta}} = \text{DS}(\mathbf{\Psi}, \mathbf{y}, \lambda)$ to denote a solution of the (linear) program (DS) that takes as input $\mathbf{\Psi}, \mathbf{y}$, and λ.

B. Estimation Scheme I: Antenna Domain Processing

We are now ready to state the training structure and the associated DS-based reconstruction algorithm for our first proposed estimation scheme for D-sparse narrowband MIMO channels. The focus here is on receiver processing in the antenna domain, and we refer to this particular scheme as NBE-AP for *NarrowBand Estimation-Antenna Processing*.

NBE-AP Training: Let $\mathcal{S} = \{1,\ldots,N_R\} \times \{1,\ldots,N_T\}$ be the set of indices of elements within the antenna domain matrix $\mathbf{H}$. Further, let the number of receive signal space dimensions dedicated to training $N_{tr} \geq c_2 \cdot \log^5 D_{max} \cdot D$ for some constant $c_2 > 0$ and choose $\mathcal{S}_{tr}$ to be a set of N_{tr} ordered pairs sampled uniformly at random from $\mathcal{S}$. Define the number of (transmit) training vectors $M_{tr} = |\{k : (i,k) \in \mathcal{S}_{tr}\}|$. The NBE-AP training strategy corresponds to measuring N_{tr} elements of $\mathbf{H}$ at the receiver that are indexed by the set $\mathcal{S}_{tr}$: $\mathbf{x}_{tr} = \sqrt{\mathcal{E}/M_{tr}}\,\{H(i,k)\}_{\mathcal{S}_{tr}} + \mathbf{w}$. Here, $\mathbf{x}_{tr} \in \mathbb{C}^{N_{tr}}$ is the vector of received training data and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}_{N_{tr}}, \mathbf{I}_{N_{tr}})$.

NBE-AP Reconstruction: Let vec^{-1} denote the inverse of the vec operator (defined as the stacking of the columns of a matrix into a vector) and choose $\lambda = (2\tilde{\mathcal{E}}(1+a)\log D_{max})^{1/2}$ for any $a \geq 0$, where $\tilde{\mathcal{E}} = \mathcal{E}N_{tr}/D_{max}M_{tr}$. Further, let $\mathbf{a}_{T,i}^T$ and $\mathbf{a}_{R,i}^T$ denote the i-th row of $\mathbf{A}_T^*$ and $\mathbf{A}_R$, respectively, and consider an $N_{tr} \times D_{max}$ matrix $\mathbf{U}_{tr}$ that is comprised of $\{(\mathbf{a}_{T,k}^T \otimes \mathbf{a}_{R,i}^T) : (i,k) \in \mathcal{S}_{tr}\}$ as its rows. The NBE-AP estimate of $\mathbf{H}$ is obtained from $\mathbf{x}_{tr}$ as follows:

$$\widehat{\mathbf{H}} = \mathbf{A}_R\left(\text{vec}^{-1}\big(\text{DS}(\sqrt{\mathcal{E}/M_{tr}}\,\mathbf{U}_{tr}, \mathbf{x}_{tr}, \lambda)\big)\right)\mathbf{A}_T^H. \qquad (10)$$

Remark 1: As an illustrative example of NBE-AP training and reconstruction, consider the case of a 5×5 MIMO system with $N_{tr} = 3$ and let $\mathcal{S}_{tr} = \{(3,2),(1,4),(4,4)\}$. In this case, $M_{tr} = 2$ and the two training vectors are given by $\{\mathbf{s}(1) = \mathbf{e}_2, \mathbf{s}(2) = \mathbf{e}_4\}$, where $\mathbf{e}_i$ denotes the i-th standard basis element of $\mathbb{C}^{N_T}$. At the receiver, we first form $\mathbf{x}_{tr}$ by stacking the third element of $\mathbf{x}(1) = \sqrt{\mathcal{E}/2}\,\mathbf{H}\mathbf{s}(1) + \mathbf{w}(1)$ and first and fourth elements of $\mathbf{x}(2) = \sqrt{\mathcal{E}/2}\,\mathbf{H}\mathbf{s}(2) + \mathbf{w}(2)$ into a vector, and then reconstruct $\mathbf{H}$ using (10).

Finally, the following theorem states the MSE performance of NBE-AP for estimating narrowband MIMO channels.

Theorem 2 (NBE-AP MSE): The NBE-AP estimate (10) of a D-sparse narrowband MIMO channel satisfies

$$\|\widehat{\mathbf{H}} - \mathbf{H}\|_F^2 \leq c_3 \cdot \log D_{max} \cdot \left(\frac{N_T\,D}{\mathcal{E}}\right) \cdot \left(\frac{N_R M_{tr}}{N_{tr}}\right) \qquad (11)$$

with probability exceeding $1 - 2\max\left\{2(\pi(1+a)\log D_{max} \cdot D_{max}^{2a})^{-1/2}, c_4 D_{max}^{-c_5}\right\}$. Here, c_3, c_4 and c_5 are strictly positive constants that do not depend on $\mathcal{E}, N_R$ or N_T.

Proof Sketch: Define $\mathbf{h} = \text{vec}(\mathbf{H})$ and $\mathbf{h}_v = \text{vec}(\mathbf{H}_v)$. From (5), we have that $\mathbf{h} = (\mathbf{A}_T^* \otimes \mathbf{A}_R)\, \mathbf{h}_v$. It is then easy to see that NBE-AP training implies $\mathbf{x}_{tr} = \sqrt{\mathcal{E}/M_{tr}}\, \mathbf{U}_{tr}\mathbf{h}_v + \mathbf{w}$, where $\mathbf{U}_{tr}$ is as defined earlier. Next, define $\mathbf{U} = \mathbf{A}_T^* \otimes \mathbf{A}_R$, which is a unitary matrix, and note that by construction $\mathbf{U}_{tr}$ corresponds to randomly sampling N_{tr} rows of $\mathbf{U}$. Therefore, [12, Th. 3.3] implies that the matrix $\sqrt{N_R N_T/N_{tr}}\, \mathbf{U}_{tr}$ satisfies RIP of order $2D$ with $\delta_{2D} < \sqrt{2} - 1$ with probability at least $1 - c_4 D_{max}^{-c_5}$. The theorem follows by noting that (i) $\mathbf{H}$ and $\mathbf{H}_v$ are unitarily equivalent to each other, and (ii) $\|\mathbf{h}_v\|_0 = D$, and combining these observations with Theorem 1. ∎

A few remarks regarding the performance of the proposed scheme are in order now. First, NBE-AP training structure requires that the number of training vectors $M_{tr} \le N_T$ as opposed to $M_{tr} \ge N_T$ for traditional schemes. Second, MSE scaling of the NBE-AP channel estimate given in (11) is less than that for traditional methods given in (7). Finally, and perhaps most importantly, the number of receive signal space dimensions that NBE-AP dedicates to training is $N_{tr} \approx O(D)$ as opposed to $N_{tr} = O(D_{max})$ for conventional schemes. This is significant from the perspective of multiuser MIMO systems since these savings in the receive signal space dimensions can result in increased network spectral efficiency for appropriately designed cognitive/ad-hoc networks.

C. Estimation Scheme II: Beamspace Processing

We know from elementary estimation theory that at the very minimum D measurements are needed to reasonably estimate a D-dimensional parameter. Therefore, NBE-AP performs near optimally in terms of the number of receive signal space dimensions that it dedicates to training. However, the same cannot be said of NBE-AP for the number of training vectors M_{tr} and the MSE. In fact, by appealing to the classic "occupancy problem" [13], it can be shown that if $D \ge N_T$ then the NBE-AP training strategy results in $M_{tr} = N_T$ with high probability and hence, NBE-AP and traditional schemes perform near identically in terms of the MSE and M_{tr} in this case. We now present an alternative estimation scheme, referred to as NBE-BP, that circumvents this problem by focusing on receiver processing in the beamspace and, as a result, achieves near-optimal MSE of $O(N_T D/\mathcal{E})$.

NBE-BP Training: Let $\mathbf{h}_{v,i}^T$ denote the i-th row of $\mathbf{H}_v$ and define D_i to be the number of nonzero virtual coefficients in each row of $\mathbf{H}_v$; that is, $D_i = \|\mathbf{h}_{v,i}\|_0$. Choose the number of training vectors $M_{tr} \ge c_6 \cdot \log(N_T/\max_i D_i) \cdot \max_i D_i$ for some constant $c_6 > 0$ and let each training vector $\mathbf{s}(n), n = 1, \dots, M_{tr}$, be an i.i.d. vector of binary random variables taking values $+1/\sqrt{N_T}$ or $-1/\sqrt{N_T}$ with probability $1/2$ each. At the receiver, all the received training signals $\{\mathbf{x}(n),\, n = 1, \dots, M_{tr}\}$ are stacked into an $M_{tr} \times N_R$ matrix $\mathbf{X}$ (just like in the traditional training setup) to yield (6).

NBE-BP Reconstruction: First, define $\mathbf{X}_v = \mathbf{X}\mathbf{A}_R^*$ and let $\mathbf{x}_{v,i}$ denote the i-th column of $\mathbf{X}_v$. Next, fix some $a \ge 0$ and pick $\lambda = (2\mathcal{E}(1+a)(\log D_{max})/N_T)^{1/2}$. Finally, define $\widehat{\mathbf{h}}_{v,i} = \text{DS}\big(\sqrt{\mathcal{E}/M_{tr}}\, \mathbf{S}\mathbf{A}_T^*, \mathbf{x}_{v,i}, \lambda\big)$ for $i = 1, \dots, N_R$. The NBE-BP estimate of $\mathbf{H}$ is then given as follows:

$$\widehat{\mathbf{H}} = \mathbf{A}_R \begin{bmatrix} \widehat{\mathbf{h}}_{v,1} & \dots & \widehat{\mathbf{h}}_{v,N_R} \end{bmatrix}^T \mathbf{A}_T^H. \qquad (12)$$

The following theorem states the MSE performance of NBE-BP for estimating narrowband MIMO channels.

Theorem 3 (NBE-BP MSE): The NBE-BP estimate (12) of a D-sparse narrowband MIMO channel satisfies

$$\|\widehat{\mathbf{H}} - \mathbf{H}\|_F^2 \le c_7 \cdot \log D_{max} \cdot \left(\frac{N_T\, D}{\mathcal{E}}\right) \qquad (13)$$

with probability exceeding $1 - 4\max\big\{(\pi(1+a)\log D_{max} \cdot D_{max}^{2a})^{-1/2}, e^{-c_8 M_{tr}}\big\}$. Here, c_7 and c_8 are strictly positive constants that do not depend on $\mathcal{E}, N_R$ or N_T.

Proof Sketch: By definition $\mathbf{x}_{v,i} = \sqrt{\mathcal{E}/M_{tr}}\, \mathbf{S}\mathbf{A}_T^* \mathbf{h}_{v,i} + \mathbf{w}_{v,i}$, where $\|\mathbf{h}_{v,i}\|_0 = D_i$ and $\mathbf{w}_{v,i} \sim \mathcal{CN}(\mathbf{0}_{M_{tr}}, \mathbf{I}_{M_{tr}})$. Next, define $\bar{D} = \max_i D_i$ and note that [11, Th. 5.2] implies that $\sqrt{N_T/M_{tr}}\, \mathbf{S}\mathbf{A}_T^*$ satisfies RIP of order $2\bar{D}$ with $\delta_{2\bar{D}} < \sqrt{2} - 1$ with probability at least $1 - 2e^{-c_8 M_{tr}}$. It can then be shown through a slight variation of the proof of [10, Th. 1.1] that $\|\widehat{\mathbf{h}}_{v,i} - \mathbf{h}_{v,i}\|_2^2 = O(\log D_{max} \cdot (\frac{N_T D_i}{\mathcal{E}}))$ for *all* $i = 1, \dots, N_R$ with probability exceeding $1 - 4\max\big\{(\pi(1+a)\log D_{max} \cdot D_{max}^{2a})^{-1/2}, e^{-c_8 M_{tr}}\big\}$, and the proof of this theorem follows by noting the fact that $\sum_{i=1}^{N_R} D_i = D$. ∎

Remark 2: The aforementioned scheme remains unchanged if the random binary probing is carried out in the beamspace instead of in the antenna domain: $\mathbf{s}(n) = \mathbf{A}_T \mathbf{s}_v(n)$, where the $\mathbf{s}_v(n)$'s are now i.i.d. binary random vectors. The NBE-BP estimation can also be carried out using a completely different set of training vectors without significantly altering its efficacy as follows. Let $\mathcal{S}_{tr}$ be a (sorted) set of $M_{tr} = O(\log^5(N_T/\bar{D}) \cdot \bar{D})$ elements sampled uniformly at random from $\{1, \dots, N_T\}$ and define the M_{tr} training vectors to be $\{\mathbf{s}(n) = \mathbf{e}_i : i \in \mathcal{S}_{tr}\}$. Then the preceding analysis still follows through with the difference being that the second term in the $\max$ expression in Theorem 3 is changed to $O(N_T^{-O(1)})$.

The preceding discussion shows that NBE-BP achieves near-optimal MSE performance. However, the other two performance metrics of interest, namely, the number of training vectors (M_{tr}) and the number of receive signal space dimensions dedicated to training (N_{tr}), vary with the channel sparsity pattern in this scheme: $M_{tr} \approx O(\max_i D_i)$ and $N_{tr} \approx O(\max_i D_i N_R)$. However, if one assumes that the D nonzero virtual coefficients are uniformly distributed across the channel angular spread then we have $\max_i D_i \approx D/N_R$ and, in this case, NBE-BP 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 near-optimal performance of $N_{tr} \approx O(D)$ for NBE-BP versus $N_{tr} = O(D_{max})$ for ML-based estimators.

IV. Learning Sparse Wideband MIMO Channels

In this section, we extend the results of Section III to encompass sparse wideband MIMO channels (corresponding to $W\tau_{max} \ge 1$). Because of space constraints, we limit ourselves to block-fading channels (corresponding to $T\nu_{max} \ll 1$), and assume that the communication packet is comprised of $N_o \approx T/(T_f + \tau_{max}) \ge N_T$ orthogonal frequency division multiplexing (OFDM) vector-valued symbols and each OFDM symbol consists of $Q = T_f W \geq \lceil W\tau_{max}\rceil + 1$ tones. Here, $T_f < T$ denotes the OFDM symbol duration and extensions of this to the case when $N_o = 1$ ($T_f \approx T$), and to sparse doubly-selective MIMO channels will be reported in a journal version of this paper currently under preparation.

For block-fading wideband MIMO channels, the virtual channel representation (3) reduce to

$$\mathbf{H}(f) \approx \sum_{\ell=0}^{L-1} \widetilde{\mathbf{H}}(\ell) e^{-j2\pi \frac{\ell}{W} f} \tag{14}$$

where the (antenna domain matrices) $\widetilde{\mathbf{H}}(\ell)$'s are defined in terms of the beamspace matrices: $\widetilde{\mathbf{H}}(\ell) \approx \mathbf{A}_R \mathbf{H}_v(\ell) \mathbf{A}_T^H$. Here, the channel frequency response $\mathbf{H}(f)$ is completely characterized by the $D_{max} = N_R N_T L$ virtual channel coefficients $\{H_v(i,k,\ell)\}$, out of which only $D \ll D_{max}$ coefficients are assumed to be nonzero. To learn this D_{max}-dimensional channel, training-based methods dedicate M_{tr} of the $N_o Q$ OFDM tones as "pilot tones" and transmit known (vector-valued) training signals to the receiver over these tones. The transmitted and received training signals in this case are related to each other as

$$\mathbf{x}(n,q) = \sqrt{\frac{\mathcal{E}}{M_{tr}}} \mathbf{H}(q)\mathbf{s}(n,q) + \mathbf{w}(n,q)\ ,\ (n,q) \in \mathcal{P}_{tr} \tag{15}$$

where $\mathcal{E}$ is the transmit energy budget available for training, the matrix $\mathbf{H}(q) = \mathbf{H}(f)|_{f=q/T_f}$, $\mathbf{w}(n,q) \sim \mathcal{CN}(\mathbf{0}_{N_R}, \mathbf{I}_{N_R})$, $\mathcal{P}_{tr} \subset \{1,\dots,N_o\} \times \{0,\dots,Q-1\} : |\mathcal{P}_{tr}| = M_{tr}$, and the set of transmit training vectors $\{\mathbf{s}(n,q)\}_{\mathcal{P}_{tr}}$ is designed such that $\sum_{\mathcal{P}_{tr}} \|\mathbf{s}(n,q)\|_2^2 = M_{tr}$.

Traditional training-based methods often assume that the number of pilot tones $M_{tr} \geq N_T L$ and typically employ ML-based estimators at the receiver to recover $\{H_v(i,k,\ell)\}$ from the knowledge of $\{\mathbf{s}(n,q), \mathbf{x}(n,q)\}_{\mathcal{P}_{tr}}$.[2] Irrespective of the set of pilot tones $\mathcal{P}_{tr}$ and the exact nature of training vectors $\{\mathbf{s}(n,q)\}_{\mathcal{P}_{tr}}$, it can be shown in this case too that the MSE in the channel estimate is lower bounded as [5], [6]

$$\sum_{i,k,\ell} \mathbb{E}\left[|\widehat{H}_v(i,k,\ell) - H_v(i,k,\ell)|^2\right] \geq \frac{N_T\, D_{max}}{\mathcal{E}}\ . \tag{16}$$

In contrast, we now propose a new training-based estimation scheme that significantly improves upon the performance of traditional schemes, both in terms of the MSE and the number of pilot tones (resp. receive dimensions) dedicated to training. The proposed scheme can be thought of as an extension of the NBE-BP scheme to sparse wideband MIMO channels and is accordingly referred to as WBE-BP.[3]

<u>WBE-BP Training</u>: Let $\mathbf{h}_{v,i}^T(\ell)$ denote the i-th row of $\mathbf{H}_v(\ell)$ and define $\mathbf{H}_{v,i} = \begin{bmatrix}\mathbf{h}_{v,i}(0) & \dots & \mathbf{h}_{v,i}(L-1)\end{bmatrix}$. Further, define D_i to be the number of nonzero virtual coefficients in $\mathbf{H}_{v,i}$ ($D_i = \sum_{\ell=0}^{L-1} \|\mathbf{h}_{v,i}(\ell)\|_0$) and let the number of pilot tones $M_{tr} \geq c_9 \cdot \log(N_T Q/\max_i D_i) \cdot \max_i D_i$ for some constant $c_9 > 0$. Define the pilot tones $\mathcal{P}_{tr}$ to be a set of M_{tr} ordered pairs sampled uniformly at random from $\{1,\dots,N_T\} \times \{0,\dots,Q-1\}$, and the corresponding training vectors as $\{\mathbf{s}(n,q) = \mathbf{e}_n : (n,q) \in \mathcal{P}_{tr}\}$. At the receiver, the received signals $\{\mathbf{x}(n,q)\}_{\mathcal{P}_{tr}}$ are stacked row-wise to yield an $M_{tr} \times N_R$ matrix $\mathbf{X}$ consisting of $\mathbf{x}(n,q)$'s as its rows.

<u>WBE-BP Reconstruction</u>: First, define an $M_{tr} \times N_T L$ matrix $\mathbf{U}_{tr}$ that is comprised of $\{(\mathbf{u}_{f,q}^T \otimes \mathbf{a}_{T,n}^T) : (n,q) \in \mathcal{P}_{tr}\}$ as its rows; here, $\mathbf{u}_{f,q}^T = \begin{bmatrix} e^{-j2\pi \frac{q}{Q} 0} & \dots & e^{-j2\pi \frac{q}{Q}(L-1)}\end{bmatrix}$ and $\mathbf{a}_{T,n}^T$ denotes the n-th row of $\mathbf{A}_T^*$. Next, choose $\lambda = (2\mathcal{E}(1+a)(\log D_{max})/N_T)^{1/2}$ for any $a \geq 0$. Finally, define $\mathbf{X}_v = \mathbf{X}\mathbf{A}_R^*$ and let $\mathbf{x}_{v,i}$ denote the i-th column of $\mathbf{X}_v$. The WBE-BP estimate of $\mathbf{H}_{v,i}$'s is then given as follows:

$$\widehat{\mathbf{H}}_{v,i} = \text{vec}^{-1}\big(\text{DS}(\sqrt{\mathcal{E}/M_{tr}}\,\mathbf{U}_{tr}, \mathbf{x}_{v,i}, \lambda)\big)\ . \tag{17}$$

Theorem 4 (WBE-BP MSE): The WBE-BP estimate of the virtual channel coefficients of a D-sparse wideband MIMO channel (given by (17)) satisfies

$$\sum_{i,k,\ell} \mathbb{E}\left[|\widehat{H}_v(i,k,\ell) - H_v(i,k,\ell)|^2\right] = O\big(\log D_{max} \cdot \big(\tfrac{N_T D}{\mathcal{E}}\big)\big)$$

with probability exceeding $1 - 2\max\big\{2(\pi(1+a)\log D_{max} \cdot D_{max}^{2a})^{-1/2}, O(N_T L)^{-O(1)}\big\}$. Here, the scaling constants are independent of $\mathcal{E}, N_R, N_T$ or L.

We omit the proof of this theorem for the sake of brevity. Note that: (i) WBE-BP achieves near-optimal MSE scaling of $O(N_T D/\mathcal{E})$, and (ii) it requires that the number of pilot tones $M_{tr} \approx O(\max_i D_i)$ versus $M_{tr} \approx N_T L$ for traditional schemes—a significant improvement assuming uniformly distributed sparsity within the channel angle-delay spread.

[2]This means that traditional schemes dedicate a total of $N_{tr} = N_R M_{tr} \geq D_{max}$ receive dimensions for training in wideband MIMO channels.

[3]Likewise, an extension of NBE-AP to wideband channels also exists. Because of space constraints, however, we do not indulge in its details.

REFERENCES

[1] 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.

[2] N. Czink, X. Yin, H. Ozcelik, M. Herdin, E. Bonek, and B. Fleury, "Cluster characteristics in a MIMO indoor propagation environment," *IEEE Trans. Wireless Commun.*, Apr. 2007.

[3] T. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. Allerton Conf. 1999*.

[4] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, Apr. 2003.

[5] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Processing*, Jun. 2003.

[6] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, May 2006.

[7] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Processing*, Oct. 2002.

[8] ——, "A virtual representation for time- and frequency-selective correlated MIMO channels," in *Proc. IEEE ICASSP 2003*.

[9] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Select. Areas Commun.*, Feb. 1987.

[10] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, Dec. 2007.

[11] 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.

[12] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, Aug. 2008.

[13] R. Motwani and P. Raghavan, *Randomized Algorithms*. New York, NY: Cambridge University Press, 1995.