Channel Estimation and Precoder Design for Millimeter-Wave Communications: The Sparse Way

Philip Schniter* and Akbar Sayeed†
*Dept. of ECE, The Ohio State University, Columbus, OH 43210. (Email: schniter@ece.osu.edu)
†Dept. of ECE, The University of Wisconsin, Madison WI 53706. (Email: akbar@engr.wisc.edu)

***Abstract*—We propose strategies for mmWave communications that exploit the inherent sparsity of mmWave channels in the angle and delay domains. In particular, we propose the use of aperture shaping to ensure a sparse virtual-domain MIMO channel representation; fast FFT-based modulation and demodulation schemes to expose the virtual-channel coefficients; a pilot design that facilitates fast LASSO-based sparse-channel estimation; and spectrally efficient precoding and decoding, via the Lanczos algorithm and waterfilling over both frequency and angle. Numerical experiments suggest that our approach comes close to achieving the perfect-CSI capacity of the mmWave channel.**

I. Introduction

The vast majority of today's wireless communications systems operate in the microwave spectrum (i.e., $<$3 GHz), which is by now a crowded, limited resource. Yet $200\times$ more bandwidth is available in the *millimeter-wave* (mmWave) spectrum from 30-300 GHz, offering the potential for huge increases in throughput [1]. However, communicating in the mmWave spectrum is challenging for several reasons. For one, mmWave signal propagation is impaired by severe path-loss and shadowing (e.g., recent urban experiments show that path losses are 40 dB worse at 28 GHz compared to 2.8 GHz [2]). Furthermore, Gb/sec throughput is fundamentally challenging from a signal-processing perspective, if billions of bits are to be processed and decoded each second.

A natural approach to counteracting this severe mmWave path-loss is to use many (e.g., $\geq$ 32) antennas at both the transmit and receive sides. With many antennas, the transmitter (and receiver) can exploit the angle domain, e.g., by focusing energy (and attention) on the dominant propagation paths. In this regard, an important property of mmWave channels is their extreme *sparsity* in both the angle and delay domains [2]–[5]. For example, measurement campaigns in dense-urban NLOS environments have revealed that mmWave channels typically exhibit only 3-4 scattering clusters, with relatively little delay/angle spreading within each cluster [5].

In this paper, we investigate whether angle and delay-domain sparsity can be exploited for the design of *computationally and spectrally efficient* many-antenna (i.e., "massive") multiple-input multiple-output (MIMO) systems. To this end, we propose the use of aperture shaping to ensure a sparse virtual-domain MIMO channel representation; fast FFT-based modulation and demodulation schemes to expose the virtual-channel coefficients; a pilot design that facilitates fast LASSO-based sparse-channel estimation; and spectrally efficient precoding and decoding, via the Lanczos algorithm and waterfilling over both frequency and angle. Numerical experiments suggest that our approach comes close to achieving the perfect-CSI capacity of the mmWave channel.

Our work is preliminary in that assumes the use of high-precision analog-to-digital converters (ADCs) at each receive-antenna output, which is impractical due to the high cost and power consumption of multi-Gb/sec ADCs. However, we believe that the sparsity-leveraging ideas proposed in this work can be extended (with minor modification) in two practical directions: continuous aperture phased (CAP) MIMO [6] and digital beamforming using one-bit ADCs [7]. For example, in [6], a discrete lens array is used to enable beamspace MIMO [8], whereas in our work FFTs are used for this purpose.

*Schniter acknowledges support from NSF grants CCF-1018368 and CCF-1218754 and Sayeed from NSF grants ECCS-1247583 and IIP-1444962.

II. System Model

We consider a communication system that uses N_{t} transmit antennas, N_{r} receive antennas, and cyclic-prefix (CP) block-transmission with N channel uses per block and a CP length of N_{d}. As an example, the 802.11ad 60 GHz standard [9] uses $N = 512$ and $N_{\mathsf{d}} = 128$.

Suppose that, after discarding the CP, the received complex-baseband samples for block $t \in \mathbb{Z}$ are collected into the matrix $\boldsymbol{Y}_t \in \mathbb{C}^{N\times N_{\mathsf{r}}}$, where the jth column represents the samples collected by the jth receive antenna. Assuming linear propagation, we can write

$$\boldsymbol{Y}_t = \sum_{d=0}^{N_{\mathsf{d}}-1} \boldsymbol{J}_d \boldsymbol{X}_t \boldsymbol{H}_d + \boldsymbol{W}_t, \tag{1}$$

where $\boldsymbol{J}_d$ is the d-circulant-delay matrix, $\boldsymbol{X}_t \in \mathbb{C}^{N\times N_{\mathsf{t}}}$ contains the transmitted samples of block t, $\boldsymbol{H}_d \in \mathbb{C}^{N_{\mathsf{t}}\times N_{\mathsf{r}}}$ is the MIMO channel response at delay d, and $\boldsymbol{W}_t$ is additive white Gaussian noise (AWGN) of variance ν_{w}.

The delay-d MIMO channel matrix can be written as

$$\boldsymbol{H}_d = \sum_{l=1}^{L} \beta_l\, p_{\mathsf{rc}}(d\,T_{\mathsf{s}} - \tau_l)\, \boldsymbol{f}_{N_{\mathsf{t}}}(\theta_{\mathsf{t},l})\, \boldsymbol{f}_{N_{\mathsf{r}}}(\theta_{\mathsf{r},l})^{\mathsf{H}}, \tag{2}$$

where L is the number of scattering clusters, $p_{\mathsf{rc}}(\tau)$ is the raised cosine (RC) pulse for T_{s}-spaced signaling evaluated at τ seconds, $\beta_l \in \mathbb{C}$ is the gain of the lth cluster, $\tau_l \in \mathbb{R}$ is the delay of the lth cluster, $\theta_{\mathsf{t},l} \in [0, 2\pi)$ and $\theta_{\mathsf{r},l} \in [0, 2\pi)$ are the transmit and receive angles associated with the lth cluster, and $\boldsymbol{f}_{N_{\mathsf{t}}}(\theta_{\mathsf{t},l}) \in \mathbb{C}^{N_{\mathsf{t}}\times 1}$ and $\boldsymbol{f}_{N_{\mathsf{r}}}(\theta_{\mathsf{r},l}) \in \mathbb{C}^{N_{\mathsf{r}}\times 1}$ are the transmit and receive array response vectors. With half-wavelength space uniform linear arrays (ULA), they become (using $\mathsf{j} \triangleq \sqrt{-1}$)

$$\boldsymbol{f}_{N_{\mathsf{t}}}(\theta_{\mathsf{t},l}) = \left[1 \;\; e^{\mathsf{j}\theta_{\mathsf{t},l}} \;\cdot\; e^{\mathsf{j}\theta_{\mathsf{t},l}(N_{\mathsf{t}}-1)}\right]^{\mathsf{H}} / \sqrt{N_{\mathsf{t}}} \tag{3}$$

$$\boldsymbol{f}_{N_{\mathsf{r}}}(\theta_{\mathsf{r},l}) = \left[1 \;\; e^{\mathsf{j}\theta_{\mathsf{r},l}} \;\cdot\; e^{\mathsf{j}\theta_{\mathsf{r},l}(N_{\mathsf{r}}-1)}\right]^{\mathsf{H}} / \sqrt{N_{\mathsf{r}}}. \tag{4}$$

Under the same ULA assumption, the corresponding virtual channel [8] matrix $\boldsymbol{G}_d \in \mathbb{C}^{N_{\mathsf{t}}\times N_{\mathsf{r}}}$ is defined as

$$\boldsymbol{G}_d \triangleq \boldsymbol{F}_{N_{\mathsf{t}}}^{\mathsf{H}} \boldsymbol{H}_d \boldsymbol{F}_{N_{\mathsf{r}}} \tag{5}$$

where $\boldsymbol{F}_{N_{\mathsf{t}}}$ and $\boldsymbol{F}_{N_{\mathsf{r}}}$ denote unitary DFT matrices of sizes N_{t} and N_{r}, respectively. We can then rewrite (1) in terms of the virtual channel

coefficients as

$$\boldsymbol{Y}_t = \sum_{d=0}^{N_{\mathsf{d}}-1} \boldsymbol{J}_d \boldsymbol{X}_t \boldsymbol{F}_{N_{\mathsf{t}}} \boldsymbol{G}_d \boldsymbol{F}_{N_{\mathsf{r}}}^{\mathsf{H}} + \boldsymbol{W}_t. \tag{6}$$

Note that, if the angles associated with the lth scattering cluster were of the form $(\theta_{\mathsf{t},l}, \theta_{\mathsf{r},l}) = (i_l \frac{2\pi}{N_{\mathsf{t}}}, j_l \frac{2\pi}{N_{\mathsf{r}}})$ with $i_l \in \{0, \ldots, N_{\mathsf{t}}-1\}$ and $j_l \in \{0, \ldots, N_{\mathsf{r}}-1\}$, then the contribution from this cluster would manifest exclusively in the single virtual channel coefficient $[\boldsymbol{G}_d]_{i_l,j_l}$. This can be seen from the fact that

$$\boldsymbol{G}_d = \sum_{l=1}^{L} \beta_l \, p_{\mathsf{rc}}(d\,T_{\mathsf{s}} - \tau_l) \boldsymbol{F}_{N_{\mathsf{t}}}^{\mathsf{H}} \boldsymbol{f}_{N_{\mathsf{t}}}(\theta_{\mathsf{t},l}) \boldsymbol{f}_{N_{\mathsf{r}}}(\theta_{\mathsf{r},l})^{\mathsf{H}} \boldsymbol{F}_{N_{\mathsf{r}}}, \tag{7}$$

which implies

$$\begin{aligned}[\boldsymbol{G}_d]_{i,j} = \sum_{l=1}^{L} & \frac{e^{\mathsf{j}\frac{N_{\mathsf{t}}-1}{2}(\theta_{\mathsf{t},l} - \frac{2\pi}{N_{\mathsf{t}}} i)} \sin(\frac{N_{\mathsf{t}}}{2}(\theta_{\mathsf{t},l} - \frac{2\pi}{N_{\mathsf{t}}} i))}{N_{\mathsf{t}} \sin(\frac{1}{2}(\theta_{\mathsf{t},l} - \frac{2\pi}{N_{\mathsf{t}}} i))} \\ & \times \frac{e^{\mathsf{j}\frac{N_{\mathsf{r}}-1}{2}(\theta_{\mathsf{r},l} - \frac{2\pi}{N_{\mathsf{r}}} j)} \sin(\frac{N_{\mathsf{r}}}{2}(\theta_{\mathsf{r},l} - \frac{2\pi}{N_{\mathsf{r}}} j))}{N_{\mathsf{r}} \sin(\frac{1}{2}(\theta_{\mathsf{r},l} - \frac{2\pi}{N_{\mathsf{r}}} j))} \beta_l \, p_{\mathsf{rc}}(d\,T_{\mathsf{s}} - \tau_l)\end{aligned} \tag{8}$$

Furthermore, if all L scattering clusters had this property, then the matrix $\boldsymbol{G}_d$ would be *sparse* with $\leq L$ non-zero coefficients.

In practice, however, the scattering angles do not lie on these uniform grids. Rather, they have the form $(\theta_{\mathsf{t},l}, \theta_{\mathsf{r},l})$ for non-integer $\theta_{\mathsf{t},l} N_{\mathsf{t}}/(2\pi)$ and/or $\theta_{\mathsf{r},l} N_{\mathsf{r}}/(2\pi)$. In this case, while the scattering energy is most concentrated in the element $[\boldsymbol{G}_d]_{i_l,j_l}$, where $i_l = \lfloor \theta_{\mathsf{t},l} N_{\mathsf{t}}/(2\pi) \rceil$ and $j_l = \lfloor \theta_{\mathsf{r},l} N_{\mathsf{r}}/(2\pi) \rceil$, it "leaks" out to all other elements in $\boldsymbol{G}_d$. As a result, the virtual channel matrix $\boldsymbol{G}_d$ is generally non-sparse. A similar leakage phenomenon is well-known in the context of spectral analysis [10, §10.2].

With these considerations in mind, the model (6) can be interpreted as performing a *critical* sampling of the transmit and receive angle spaces (uniformly over N_{t} and N_{r} locations, respectively). In addition, (6) explicitly accounts for delay spreading. In contrast, the recent mmWave beamforming papers [6], [11], [12] used oversampling in the angular domain and/or ignored delay-spread.

III. Aperture Shaping

Consider the previously described MIMO system, but where the transmitter applies a fixed gain of $\alpha_{\mathsf{t},i}$ to the ith antenna and the receiver applies a fixed gain of $\alpha_{\mathsf{r},j}$ to the jth antenna. In the sequel, we will refer to this practice as *aperture shaping*. The system model (1) then generalizes to

$$\boldsymbol{Y}_t = \left(\sum_{d=0}^{N_{\mathsf{d}}-1} \boldsymbol{J}_d \boldsymbol{X}_t \, \boldsymbol{D}(\boldsymbol{\alpha}_{\mathsf{t}}) \boldsymbol{H}_d + \boldsymbol{W}_t \right) \boldsymbol{D}(\boldsymbol{\alpha}_{\mathsf{r}}), \tag{9}$$

where $\boldsymbol{D}(\boldsymbol{w})$ is the matrix obtained by diagonalizing the vector $\boldsymbol{w}$. If we define the *shaped virtual MIMO channel matrix* as

$$\overline{\boldsymbol{G}}_d \triangleq \boldsymbol{F}_{N_{\mathsf{t}}}^{\mathsf{H}} \boldsymbol{D}(\boldsymbol{\alpha}_{\mathsf{t}}) \boldsymbol{H}_d \, \boldsymbol{D}(\boldsymbol{\alpha}_{\mathsf{r}}) \boldsymbol{F}_{N_{\mathsf{r}}}, \tag{10}$$

then we can rewrite (9) as

$$\boldsymbol{Y}_t = \sum_{d=0}^{N_{\mathsf{d}}-1} \boldsymbol{J}_d \boldsymbol{X}_t \boldsymbol{F}_{N_{\mathsf{t}}} \overline{\boldsymbol{G}}_d \boldsymbol{F}_{N_{\mathsf{r}}}^{\mathsf{H}} + \boldsymbol{W}_t \, \boldsymbol{D}(\boldsymbol{\alpha}_{\mathsf{r}}). \tag{11}$$

In [13], it was proposed to design $\boldsymbol{\alpha}_{\mathsf{t}}$ and $\boldsymbol{\alpha}_{\mathsf{r}}$ so that the sparsity in $\overline{\boldsymbol{G}}_d$ is enhanced. We now briefly summarize this approach. For the lth scattering cluster, recall that the dominant angle bins are $i_l \triangleq \lfloor \theta_{\mathsf{t},l} N_{\mathsf{t}}/(2\pi) \rceil$ and $j_l \triangleq \lfloor \theta_{\mathsf{r},l} N_{\mathsf{r}}/(2\pi) \rceil$. We are thus interested in suppressing the contributions of the lth cluster to the bin pairs *outside* of (i_l, j_l). For this purpose, let us define the "signal" energy

$$\mathcal{E}_{\mathsf{s}}(\boldsymbol{\alpha}_{\mathsf{t}}, \boldsymbol{\alpha}_{\mathsf{r}}) \triangleq \mathrm{E} \left\{ \left| \sum_{l=1}^{L} \sum_{d=0}^{N_{\mathsf{d}}-1} [\overline{\boldsymbol{G}}_d]_{i_l,j_l} \right|^2 \right\} \tag{12}$$

and the "interference" energy

$$\mathcal{E}_{\mathsf{i}}(\boldsymbol{\alpha}_{\mathsf{t}}, \boldsymbol{\alpha}_{\mathsf{r}}) \triangleq \mathrm{E} \left\{ \left| \sum_{l=1}^{L} \sum_{i \notin \mathcal{I}_l} \sum_{j \notin \mathcal{J}_l} \sum_{d=0}^{N_{\mathsf{d}}-1} [\overline{\boldsymbol{G}}_d]_{i,j} \right|^2 \right\}, \tag{13}$$

where $\mathcal{I}_l$ is a small neighborhood about i_l, $\mathcal{J}_l$ is a small neighborhood about j_l, and the expectations are with respect to the scattering parameters $\beta_l, \theta_{\mathsf{t},l}, \theta_{\mathsf{r},l}$ and τ_l. In [13], it was proposed to maximize the signal-to-interference ratio (SIR), i.e., to design the shaping coefficients as

$$\arg\max_{\boldsymbol{\alpha}_{\mathsf{t}}, \boldsymbol{\alpha}_{\mathsf{r}}} \frac{\mathcal{E}_{\mathsf{s}}(\boldsymbol{\alpha}_{\mathsf{t}}, \boldsymbol{\alpha}_{\mathsf{r}})}{\mathcal{E}_{\mathsf{i}}(\boldsymbol{\alpha}_{\mathsf{t}}, \boldsymbol{\alpha}_{\mathsf{r}})} \tag{14}$$

under the assumption of i.i.d. scattering with angles $\theta_{\mathsf{t},l}$ and $\theta_{\mathsf{r},l}$ uniformly distributed over $[0, 2\pi)$ and delays τ_l sufficiently far from the endpoints of the interval $[0, N_{\mathsf{d}} T_{\mathsf{s}})$ to prevent edge effects. In the latter case, the effects of delay spreading can be ignored in shaping design. The optimal windows $\boldsymbol{\alpha}_{\mathsf{t}}, \boldsymbol{\alpha}_{\mathsf{r}}$ can then be computed as the solutions to a generalized eigenvalue problem. Due to space limitations, we refer the reader to [13] for additional details.

With aperture shaping, the lth scattering cluster is expected to contribute to only a small neighborhood of transmit-angle bins $\mathcal{I}_l$ and receive-angle bins $\mathcal{J}_l$. (See Fig. 1 for an example.) Similarly, due to RC temporal shaping, each scattering cluster is expected to contribute to only a small neighborhood of delay bins. For example, if the neighborhood size was 3 along each dimension, then each scattering cluster would contribute significantly to only $3 \times 3 \times 3 = 27$ bins out of a total of $N_{\mathsf{d}} N_{\mathsf{t}} N_{\mathsf{r}}$ bins in $\{\overline{\boldsymbol{G}}_d\}_{d=0}^{N_{\mathsf{d}}-1}$. In this case, the sparsity rate of the shaped virtual channel would be at most $27L/(N_{\mathsf{d}} N_{\mathsf{t}} N_{\mathsf{r}})$. So, if $L = 4$ (as suggested by [5]), $N_{\mathsf{d}} = 128$ (as in 802.11ad), and $N_{\mathsf{t}} = N_{\mathsf{r}} = 64$, then the sparsity rate would be < 0.00021.

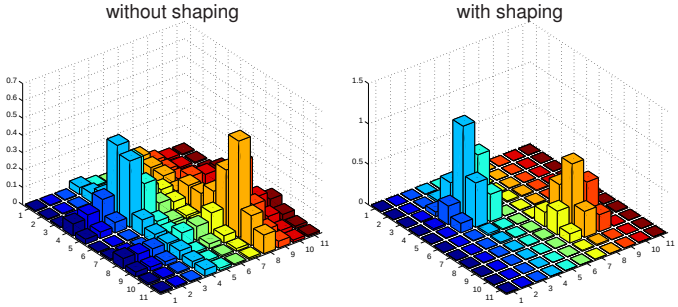

Fig. 1. Virtual-channel coefficient matrix $\overline{\boldsymbol{G}}_d$ without aperture shaping (left) and with SIR-optimal aperture-domain windowing (right), for $L = 2$ scattering paths, $N_{\mathsf{t}} = N_{\mathsf{r}} = 11$ antenna uniform linear arrays without delay spread.

IV. Modulation and Demodulation

In the sequel, we propose channel-estimation and precoding strategies that are both based on the following FFT-based approach to modulation and demodulation. For modulation, we propose to structure the transmitted matrix as

$$\boldsymbol{X}_t = \boldsymbol{F}_N^{\mathsf{H}} \boldsymbol{S}_t \boldsymbol{F}_{N_{\mathsf{t}}}^{\mathsf{H}}, \tag{15}$$

where the structure of $\boldsymbol{S}_t$ will be specified in the sequel. Here, the temporal IFFT $\boldsymbol{F}_N^{\mathsf{H}}$ amounts to OFDM precoding, while the transmit-aperture FFT $\boldsymbol{F}_{N_{\mathsf{t}}}^{\mathsf{H}}$ is used to expose the virtual channel coefficients,

as will be seen. For demodulation, we propose to FFT-process the received samples $\boldsymbol{Y}_t$ both temporally and across the receive-aperture. According to (11), this yields

$$\begin{aligned}\overline{\boldsymbol{Y}}_t &\triangleq \boldsymbol{F}_N \boldsymbol{Y}_t \boldsymbol{F}_{N_\mathsf{r}} \qquad (16)\\ &= \sum_{d=0}^{N_\mathsf{d}-1} \boldsymbol{F}_N \boldsymbol{J}_d \boldsymbol{F}_N^\mathsf{H} \boldsymbol{S}_t \overbrace{\boldsymbol{F}_{N_\mathsf{t}}^\mathsf{H} \boldsymbol{D}(\boldsymbol{\alpha}_\mathsf{t}) \boldsymbol{H}_d \boldsymbol{D}(\boldsymbol{\alpha}_\mathsf{r}) \boldsymbol{F}_{N_\mathsf{r}}}^{\overline{\boldsymbol{G}}_d}\\ &\quad + \boldsymbol{F}_N \boldsymbol{W}_t \boldsymbol{D}(\boldsymbol{\alpha}_\mathsf{r}) \boldsymbol{F}_{N_\mathsf{r}}. \qquad (17)\end{aligned}$$

We now show that (17) can be simplified. Denoting the circulant matrix with first column $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$ by $\boldsymbol{C}(\boldsymbol{x})$, we know [10]

$$\boldsymbol{C}(\boldsymbol{x}) = \boldsymbol{F}_N^\mathsf{H} \operatorname{diag}(\sqrt{N}\boldsymbol{F}_N \boldsymbol{x}) \boldsymbol{F}_N. \qquad (18)$$

Thus, since $\boldsymbol{J}_d = \boldsymbol{C}(\boldsymbol{\delta}_d)$, where $\boldsymbol{\delta}_d$ denotes the dth column of the identity matrix, we have that

$$\begin{aligned}\boldsymbol{F}_N \boldsymbol{J}_d \boldsymbol{F}_N^\mathsf{H} &= \boldsymbol{F}_N \boldsymbol{F}_N^\mathsf{H} \operatorname{diag}(\sqrt{N}\boldsymbol{F}_N \boldsymbol{\delta}_d) \boldsymbol{F}_N \boldsymbol{F}_N^\mathsf{H} \qquad (19)\\ &= \sqrt{N} \operatorname{diag}(\boldsymbol{f}_{N,d}), \qquad (20)\end{aligned}$$

where $\boldsymbol{f}_{N,d} \triangleq \boldsymbol{F}_N \boldsymbol{\delta}_d$ is the dth column of $\boldsymbol{F}_N$. Furthermore, we can write $\boldsymbol{S}_t \overline{\boldsymbol{G}}_d = \sum_{i=0}^{N_\mathsf{t}-1} \boldsymbol{s}_{t,:,i} \overline{\boldsymbol{g}}_{d,i}^\mathsf{T}$, where $\boldsymbol{s}_{t,:,i}$ is the ith column of $\boldsymbol{S}_t$ and $\overline{\boldsymbol{g}}_{d,i}^\mathsf{T} \in \mathbb{C}^{N_\mathsf{r}}$ is the ith row of $\overline{\boldsymbol{G}}_d$, implying

$$\begin{aligned}&\textstyle\sum_{d=0}^{N_\mathsf{d}-1} \operatorname{diag}(\boldsymbol{f}_{N,d}) \boldsymbol{S}_t \overline{\boldsymbol{G}}_d\\ &= \textstyle\sum_{d=0}^{N_\mathsf{d}-1} \operatorname{diag}(\boldsymbol{f}_{N,d}) \sum_{i=0}^{N_\mathsf{t}=1} \boldsymbol{s}_{t,:,i}\, \overline{\boldsymbol{g}}_{d,i}^\mathsf{T} \qquad (21)\\ &= \textstyle\sum_{i=0}^{N_\mathsf{t}-1} \operatorname{diag}(\boldsymbol{s}_{t,:,i}) \sum_{d=0}^{N_\mathsf{d}-1} \boldsymbol{f}_{N,d}\, \overline{\boldsymbol{g}}_{d,i}^\mathsf{T} \qquad (22)\\ &= \left[\operatorname{diag}(\boldsymbol{s}_{t,:,0}) \overline{\boldsymbol{F}}_N \cdots \operatorname{diag}(\boldsymbol{s}_{t,:,N_\mathsf{t}-1}) \overline{\boldsymbol{F}}_N\right] \overline{\boldsymbol{G}} \qquad (23)\end{aligned}$$

where $\overline{\boldsymbol{F}}_N$ contains the first N_d columns of $\boldsymbol{F}_N$ and

$$\overline{\boldsymbol{G}} \triangleq \begin{bmatrix} \overline{\boldsymbol{g}}_{0,0}^\mathsf{T} \\ \vdots \\ \overline{\boldsymbol{g}}_{N_\mathsf{d}-1,0}^\mathsf{T} \\ \hline \vdots \\ \hline \overline{\boldsymbol{g}}_{0,N_\mathsf{t}-1}^\mathsf{T} \\ \vdots \\ \overline{\boldsymbol{g}}_{N_\mathsf{d}-1,N_\mathsf{t}-1}^\mathsf{T} \end{bmatrix} \in \mathbb{C}^{N_\mathsf{d}N_\mathsf{t}\times N_\mathsf{r}}. \qquad (24)$$

Applying these facts to (17), we find

$$\overline{\boldsymbol{Y}}_t = \underbrace{\sqrt{N}\left[\operatorname{diag}(\boldsymbol{s}_{t,:,0})\overline{\boldsymbol{F}}_N \cdots \operatorname{diag}(\boldsymbol{s}_{t,:,N_\mathsf{t}-1})\overline{\boldsymbol{F}}_N\right]}_{\triangleq \boldsymbol{A}_t} \overline{\boldsymbol{G}} + \overline{\boldsymbol{W}}_t \qquad (25)$$

where $\overline{\boldsymbol{W}}_t \triangleq \boldsymbol{F}_N \boldsymbol{W}_t \boldsymbol{D}(\boldsymbol{\alpha}_\mathsf{r}) \boldsymbol{F}_{N_\mathsf{r}}$ contains Gaussian noise that is spectrally white but possibly correlated across receive-angle bins due to $\boldsymbol{\alpha}_\mathsf{r}$. In particular, each column of $\overline{\boldsymbol{W}}_t$ is AWGN with variance $\nu_{\overline{\mathsf{w}}} \triangleq \nu_\mathsf{w}\|\boldsymbol{\alpha}_\mathsf{r}\|^2/N_\mathsf{r}$, while each row of $\overline{\boldsymbol{W}}_t$ is zero-mean Gaussian with covariance

$$\boldsymbol{R}_{\overline{\mathsf{w}}} \triangleq \nu_\mathsf{w} \boldsymbol{F}_{N_\mathsf{r}}^\mathsf{H} \boldsymbol{D}(|\boldsymbol{\alpha}_\mathsf{r}|^2) \boldsymbol{F}_{N_\mathsf{r}}, \qquad (26)$$

where $|\boldsymbol{\alpha}_\mathsf{r}|^2$ is the component-wise magnitude-square of $\boldsymbol{\alpha}_\mathsf{r}$.

V. Channel Estimation

We now propose a training-based channel estimation scheme that leverages the sparsity of the virtual MIMO channel $\overline{\boldsymbol{G}}$. Assuming the channel is time-invariant over blocks $t \in \{1,\dots,T\}$, we construct $\overline{\boldsymbol{Y}} \triangleq [\overline{\boldsymbol{Y}}_1^\mathsf{T},\dots,\overline{\boldsymbol{Y}}_T^\mathsf{T}]^\mathsf{T}$, $\boldsymbol{A} \triangleq [\boldsymbol{A}_1^\mathsf{T},\dots,\boldsymbol{A}_T^\mathsf{T}]^\mathsf{T} \in \mathbb{C}^{NT\times N_\mathsf{d}N_\mathsf{t}}$, and $\overline{\boldsymbol{W}} \triangleq [\overline{\boldsymbol{W}}_1^\mathsf{T},\dots,\overline{\boldsymbol{W}}_T^\mathsf{T}]^\mathsf{T}$, so that

$$\overline{\boldsymbol{Y}} = \boldsymbol{A}\overline{\boldsymbol{G}} + \overline{\boldsymbol{W}}. \qquad (27)$$

In (27), the ratio of observations to unknowns is $NT/(N_\mathsf{d}N_\mathsf{t})$, so that when $T < N_\mathsf{d}N_\mathsf{t}/N$ the problem becomes one of *compressed channel sensing* [14]. Ignoring the possible correlation across the columns of $\overline{\boldsymbol{W}}$, we separately estimate $\overline{\boldsymbol{g}}_j$, the jth column of $\overline{\boldsymbol{G}}$, from $\overline{\boldsymbol{y}}_j$, the jth column of $\overline{\boldsymbol{Y}}$, noting that

$$\overline{\boldsymbol{y}}_j = \boldsymbol{A}\overline{\boldsymbol{g}}_j + \overline{\boldsymbol{w}}_j, \qquad (28)$$

where $\overline{\boldsymbol{w}}_j$ is $\nu_{\overline{\mathsf{w}}}$-variance AWGN. Leveraging the sparsity in $\overline{\boldsymbol{g}}_j$, we propose to use LASSO [15] for estimation:

$$\widehat{\boldsymbol{g}}_j = \arg\min_{\overline{\boldsymbol{g}}_j} \|\overline{\boldsymbol{g}}_j\|_1 \text{ s.t. } \tfrac{1}{NT}\|\overline{\boldsymbol{y}}_j - \boldsymbol{A}\overline{\boldsymbol{g}}_j\|^2 \le \nu_{\overline{\mathsf{w}}}. \qquad (29)$$

For good estimation performance, we desire that the linear map $\boldsymbol{A} : \mathbb{C}^{N_\mathsf{d}N_\mathsf{t}} \to \mathbb{C}^{NT}$ is approximately isometric, which occurs (with high probability) when $\boldsymbol{A}$ is randomly constructed with pilots $\{s_{t,n,i}\}$ drawn i.i.d. uniform from $\{1, \mathsf{j}, -1, -\mathsf{j}\}$. Low estimation complexity can be achieved through the use of a "first-order" LASSO solver such as SPGL1 [16], noting that $\boldsymbol{A}\overline{\boldsymbol{g}}_j$ can be computed using only $O\big(N_\mathsf{t}N\log N\big)$ multiplies via FFTs when N is a power of 2.

As a baseline, we also consider the performance of linear minimum mean-squared error (LMMSE) estimation, i.e.,

$$\widehat{\boldsymbol{g}}_j = \boldsymbol{A}^\mathsf{H}\left(\boldsymbol{A}\boldsymbol{A}^\mathsf{H} + \frac{\nu_{\overline{\mathsf{w}}}}{\nu_{\overline{\mathsf{g}}_j}}\boldsymbol{I}_{NT}\right)^{-1}\overline{\boldsymbol{y}}_j \qquad (30)$$

for which the elements in $\overline{\boldsymbol{g}}_j$ were treated as if they were uncorrelated with mean zero and variance $\nu_{\overline{\mathsf{g}}_j} = \frac{1}{N_\mathsf{d}N_\mathsf{t}}\operatorname{E}\{\|\overline{\boldsymbol{g}}_j\|^2\}$.

VI. Precoding

Since the noise in $\overline{\boldsymbol{W}}_t$ is independent across rows, we can decouple the decoding across subcarriers without loss of optimality. Note that the nth row (or subcarrier) of $\overline{\boldsymbol{Y}}_t$ is

$$\begin{aligned}\overline{\boldsymbol{y}}_{t,n}^\mathsf{T} &= \sqrt{N}\left[s_{t,n,0}\overline{\boldsymbol{f}}_{N,n}^\mathsf{T} \;\cdots\; s_{t,n,N_\mathsf{t}-1}\overline{\boldsymbol{f}}_{N,n}^\mathsf{T}\right]\overline{\boldsymbol{G}} + \overline{\boldsymbol{w}}_{t,n}^\mathsf{T} \qquad (31)\\ &= \underbrace{\left[s_{t,n,0},\dots,s_{t,n,N_\mathsf{t}-1}\right]}_{\triangleq\, \boldsymbol{s}_{t,n,:}^\mathsf{T}} \underbrace{\begin{bmatrix} \sum_{d=0}^{N_\mathsf{d}-1} \overline{\boldsymbol{g}}_{d,0}^\mathsf{T} e^{-\mathsf{j}\frac{2\pi}{N}dn} \\ \vdots \\ \sum_{d=0}^{N_\mathsf{d}-1} \overline{\boldsymbol{g}}_{d,N_\mathsf{t}-1}^\mathsf{T} e^{-\mathsf{j}\frac{2\pi}{N}dn} \end{bmatrix}}_{\triangleq\, \overline{\boldsymbol{G}}_n^\mathsf{T}} + \overline{\boldsymbol{w}}_{t,n}^\mathsf{T}, \qquad (32)\end{aligned}$$

where (with a slight abuse of notation) $\overline{\boldsymbol{G}}_n \in \mathbb{C}^{N_\mathsf{r}\times N_\mathsf{t}}$ denotes the virtual MIMO channel matrix associated with the nth subcarrier.

We consider linear precoding over $N_\mathsf{s} \le \min(N_\mathsf{t}, N_\mathsf{r})$ streams. In this case, the signal modulated on the nth subcarrier is

$$\boldsymbol{s}_{t,n,:} = \boldsymbol{B}_n \boldsymbol{c}_{t,n} \in \mathbb{C}^{N_\mathsf{t}}, \qquad (33)$$

where each element in $\boldsymbol{c}_{t,n} \in \mathbb{C}^{N_\mathsf{s}}$ contains an independently coded datastream (presumably coded over $T \gg 1$ blocks t) and $\boldsymbol{B}_n \in \mathbb{C}^{N_\mathsf{t}\times N_\mathsf{s}}$ is a suitable precoding matrix. Without loss of generality, we assume that $\boldsymbol{c}_{t,n}$ has an identity covariance matrix, after which the power constraint $\sum_{n=0}^{N-1}\operatorname{E}\{\|\boldsymbol{s}_{t,n,:}\|^2\} \le P_\mathsf{s}\ \forall t$ translates to $\sum_{n=0}^{N-1}\|\boldsymbol{B}_n\|_F^2 \le P_\mathsf{s}$.

In the slow-fading case, the capacity of the system (conditioned on the true channel $\overline{\boldsymbol{G}}$) for a given N_s is [17]

$$\max_{\{\boldsymbol{B}_n\}:\sum_n\|\boldsymbol{B}_n\|_F^2\le P_\mathsf{s}} \sum_{n=0}^{N-1} \log_2\det\left(\boldsymbol{I}_{N_\mathsf{s}} + \boldsymbol{B}_n^\mathsf{H}\overline{\boldsymbol{G}}_n^\mathsf{H}\boldsymbol{R}_{\overline{\mathsf{w}}}^{-1}\overline{\boldsymbol{G}}_n\boldsymbol{B}_n\right). \qquad (34)$$

Furthermore, with the eigenvalue decomposition $\boldsymbol{V}_n\boldsymbol{\Lambda}_n\boldsymbol{V}_n^\mathsf{H} \triangleq \overline{\boldsymbol{G}}_n^\mathsf{H}\boldsymbol{R}_{\overline{\mathsf{w}}}^{-1}\overline{\boldsymbol{G}}_n$, the optimal precoders take the form

$$\boldsymbol{B}_n = \boldsymbol{V}_n \boldsymbol{D}(\sqrt{\boldsymbol{p}_n}), \qquad (35)$$

with powers $\boldsymbol{p}_n = [p_{n,0}, \ldots, p_{n,N_\mathsf{s}-1}]^\mathsf{T}$ that solve the well-known waterfilling problem

$$\max_{\{p_{n,i}\}} \sum_{n,i} \log_2\left(1+\lambda_{n,i}p_{n,i}\right) \text{ s.t. } p_{n,i} \geq 0, \sum_{n,i} p_{n,i} \leq P_\mathsf{s}. \quad (36)$$

The numerical experiments in Section VIII suggest that $N_\mathsf{s} = L$ is sufficient to achieve capacity, where typically $L \ll \min(N_\mathsf{t}, N_\mathsf{r})$.

VII. Decoding under Imperfect CSI

In practice, the true channel $\{\overline{\boldsymbol{G}}_n\}$ is unknown at the transmitter and receiver. Thus, it is natural to wonder to what extent channel estimation error degrades throughput.

In our preliminary investigation of this question, we assume that the channel estimate is available to both the receiver and transmitter. This scenario might arise, e.g., with time-division-duplex (TDD) operation and a reciprocal channel. We also assume that the transmitter uses the channel estimate $\widehat{\boldsymbol{G}}_n$ as if it was correct, i.e., that the precoder $\boldsymbol{B}_n$ is designed as in Section VI but with $\widehat{\boldsymbol{G}}_n$ used in place of $\overline{\boldsymbol{G}}_n$. Our assumptions about the receiver are described next.

Without loss of generality, we assume that the receiver pre-processes the observations $\overline{\boldsymbol{y}}_{t,n}$ from (32) using the invertible matrix $\widehat{\boldsymbol{U}}_n^\mathsf{H}\boldsymbol{R}_{\overline{\mathsf{w}}}^{-1/2}$, where $\boldsymbol{R}_{\overline{\mathsf{w}}}^{-1/2}$ accomplishes noise-whitening and $\widehat{\boldsymbol{U}}_n$ is defined via the singular value decomposition

$$\widehat{\boldsymbol{U}}_n\widehat{\boldsymbol{\Sigma}}_n\widehat{\boldsymbol{V}}_n^\mathsf{H} \triangleq \boldsymbol{R}_{\overline{\mathsf{w}}}^{-1/2}\widehat{\boldsymbol{G}}_n \quad (37)$$

$$\boldsymbol{U}_n\boldsymbol{\Sigma}_n\boldsymbol{V}_n^\mathsf{H} \triangleq \boldsymbol{R}_{\overline{\mathsf{w}}}^{-1/2}\overline{\boldsymbol{G}}_n. \quad (38)$$

We note that, when N_s is small, the dominant singular values/vectors can be efficiently computed via the Lanczos algorithm [18], which reduces to the simple "power method" in the case that $N_\mathsf{s} = 1$. In any case, the pre-processed observations take the form

$$\boldsymbol{z}_{t,n} \triangleq \widehat{\boldsymbol{U}}_n^\mathsf{H}\boldsymbol{R}_{\overline{\mathsf{w}}}^{-1/2}\overline{\boldsymbol{y}}_{t,n} \quad (39)$$

$$= \widehat{\boldsymbol{U}}_n^\mathsf{H}\boldsymbol{R}_{\overline{\mathsf{w}}}^{-1/2}\overline{\boldsymbol{G}}_n\boldsymbol{B}_n\boldsymbol{c}_{t,n} + \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{N_\mathsf{r}}) \quad (40)$$

$$= \underbrace{\widehat{\boldsymbol{U}}_n^\mathsf{H}\boldsymbol{U}_n\boldsymbol{\Sigma}_n\boldsymbol{V}_n^\mathsf{H}\widehat{\boldsymbol{V}}_n\boldsymbol{D}(\sqrt{\widehat{\boldsymbol{p}}_n})}_{\triangleq\, \boldsymbol{\Gamma}_n}\boldsymbol{c}_{t,n} + \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{N_\mathsf{r}}). \quad (41)$$

Under perfect channel state information (CSI), $\boldsymbol{\Gamma}_n$ is diagonal, and so the kth stream of $\boldsymbol{c}_{t,n}$ can be independently decoded from the kth component of $\boldsymbol{z}_{t,n}$ without loss of optimality. In the general case, $\boldsymbol{\Gamma}_n$ is non-diagonal, and so inter-stream interference exists. But $\boldsymbol{\Gamma}_n$ is unknown, which complicates the application of standard multi-user detection strategies like MMSE-SIC [17]. Thus, we assume that independent stream decoding is used even with imperfect CSI. In this case, (41) implies that the kth component of $\boldsymbol{z}_{t,n}$ can be written as

$$z_{t,n,k} = \gamma_{n,k,k}c_{t,n,k} + e_{t,n,k}, \quad (42)$$

where (conditioned on $\overline{\boldsymbol{G}}_n$ and $\widehat{\boldsymbol{G}}_n$) the interference $e_{t,n,k} \triangleq \sum_{l\neq k}\gamma_{n,k,l}c_{t,n,l} + \mathcal{N}(0,1)$ has variance $1+\sum_{l\neq k}|\gamma_{n,k,l}|^2$ and is statistically independent of $\gamma_{n,k,k}c_{t,n,k}$. From the mutual-information perspective, the worst-case distribution for the interference $e_{t,n,k}$ is Gaussian [17]. Under Gaussian interference, the channel (42) supports a throughput of $\log_2(1+|\gamma_{n,k,k}|^2/(1+\sum_{l\neq k}|\gamma_{n,k,l}|^2))$ bits/sec/Hz. Summing over all subcarriers n and streams k, a lower bound on the total throughput is

$$I(\overline{\boldsymbol{G}}, \widehat{\boldsymbol{G}}) \triangleq \sum_{n=0}^{N-1}\sum_{k=0}^{N_\mathsf{s}-1}\log_2\left(1+\frac{|\gamma_{n,k,k}|^2}{1+\sum_{l\neq k}|\gamma_{n,k,l}|^2}\right). \quad (43)$$

The average value of (43) is numerically investigated below.

VIII. Numerical Experiments

We now investigate the average performance of the proposed system using the channel model (2) with raised-cosine parameter 1 and $L = 4$ i.i.d. scattering centers, each with delay τ_l, transmit angle $\theta_{\mathsf{t},l}$, and receive angle $\theta_{\mathsf{r},l}$ chosen uniformly at random, and with Rayleigh fading gain β_l. In particular, we investigate the performance of *aperture shaping* (9) (relative to no shaping) and *sparse channel estimation* (29) (relative to LMMSE estimation (30)) in terms of both normalized mean-squared error (NMSE) and system throughput, where for the latter we use the bound (43).

For our numerical study, we used block length $N = 512$ and cyclic prefix length $N_\mathsf{d} = 128$ (as in the 802.11ad 60 GHz standard [9]), along with $N_\mathsf{t} = 64$ transmit antennas and $N_\mathsf{r} = 64$ receive antennas. For channel estimation, we considered both the use of $T = 1$ training block, which leads to a sub-Nyquist sampling rate of $NT/(N_\mathsf{d}N_\mathsf{t}) = 1/16$ (i.e., compressed channel sensing), and $T = 16$ training blocks, which leads to a Nyquist sampling rate of $NT/(N_\mathsf{d}N_\mathsf{t}) = 1$. We considered $\mathsf{SNR} \triangleq \mathrm{E}\{\|\boldsymbol{A}\overline{\boldsymbol{G}}\|_F^2\}/\mathrm{E}\{\|\overline{\boldsymbol{W}}\|_F^2\}$ in the interval $[-15, 0]$ dB since, with the beamforming gain provided by $N_\mathsf{r} = 64$ receive antennas, this translates to subcarrier SNRs in the interval $[3, 18]$ dB.

Figure 2 shows the average channel-estimation NMSE for both LASSO (blue) and LMMSE (green) estimators. The NMSE performance of LASSO is far better than that of LMMSE, essentially because LASSO is able to exploit the sparsity of the channel. For example, the $T = 1$ traces (solid lines) show that LASSO is able to accomplish successful *compressed* channel sensing, whereas LMMSE is not. But even in the non-compressed case ($T = 16$, dashed lines), LASSO exhibits a huge gain over LMMSE. Comparing the left and right subplots in Fig. 2, we see that aperture shaping improves LASSO by ≈ 1 dB when $\mathsf{SNR} = -15$ dB and ≈ 6 dB when $\mathsf{SNR} = 0$ dB, essentially because shaping makes the channel sparser (recall Fig. 1) and thus easier to estimate.

Figure 3 shows the average throughput (43) under LASSO (blue) and LMMSE (green) channel estimators as well as perfect CSI (black). There we see that the throughput under LASSO channel estimates is close to that under perfect-CSI and much higher than that under LMMSE channel estimates. More precisely, under Nyquist-rate training ($T = 16$, dashed lines), the throughput achieved under LASSO estimates is 96% *without* shaping but 98% optimal *with* shaping, while the throughput achieved under LMMSE estimates is only 32% optimal. Meanwhile, with compressed training ($T = 1$, solid lines), the throughput achieved under LASSO estimates is 82% optimal *without* shaping but 95% optimal *with* shaping, whereas the throughput achieved under LMMSE estimates is only 2% optimal. Thus, we conclude that aperture shaping brings a noticeable increase in throughput due to its improvement of the underlying channel estimates, especially when the number of training blocks T is few.

We note that, in practice, the use of few training blocks T will lead to direct benefits in throughput *beyond* that which is visible in Fig. 3, because the "slow-fading" metric (43) used for Fig. 3 does not take the training overhead into account. A more precise understanding of the relation between throughput and training overhead requires assumptions on the training interval (and thus the channel coherence time), which we postpone to future work.

Finally, we notice from Fig. 3 that the throughput attained via beamforming (i.e., $N_\mathsf{s} = 1$, "+" markers) is close to capacity, i.e., that attained via full spatial multiplexing ($N_\mathsf{s} = \min(N_\mathsf{t}, N_\mathsf{r}) = 64$, "□" markers), at very low SNR. Moreover, the throughput attained

with two streams (i.e., $N_{\mathsf{s}} = 2$, "$\times$" markers) is reasonably close to the capacity over the full SNR range considered in Fig. 3. Finally, although it is not directly visible from Fig. 3, at most 4 streams were activated through waterfilling power allocation in the $N_{\mathsf{s}} = 64$ case. Thus, for this statistical channel model, it suffices to use $N_{\mathsf{s}} = 4$ streams throughout the SNR range shown in Fig. 3. Given that there were $L = 4$ scattering clusters, the optimality of $N_{\mathsf{s}} = 4$ is not surprising.

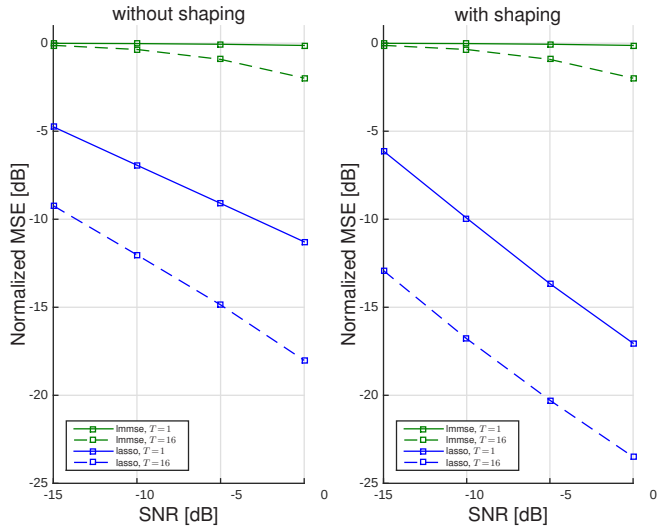

Fig. 2. Average NMSE of LASSO and LMMSE channel estimates versus SNR for $T \in \{1, 16\}$ training blocks, with and without shaping.

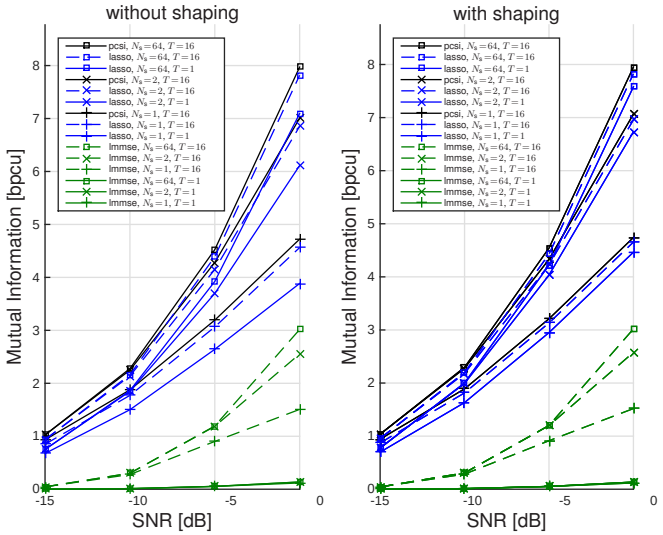

Fig. 3. Average throughput bound (43) versus SNR for LASSO, LMMSE, and perfect channel estimates; for number of streams $N_{\mathsf{s}} \in \{1, 2, 64\}$; and for number of training blocks $T \in \{1, 16\}$; with and without shaping.

IX. Conclusions

In this work we proposed techniques for computationally and spectrally efficient massive MIMO that exploit the inherent delay- and angle-domain sparsity exhibited by mmWave channels. Our approaches are based on the use of FFTs, across both time and aperture, for low-complexity modulation, demodulation, and training-based sparse-channel estimation. In addition, low-complexity Lanczos methods were proposed for precoder and decoder design, waterfilling for power allocation, and aperture shaping for sparsification of the channel representation. Numerical experiments suggested that our fast, pilot-aided designs come close to matching the spectral efficiency of the perfect-CSI capacity-optimal system. In future work, we plan to extend the approach presented here to the CAP-MIMO framework from [6] and the one-bit digital beamforming framework from [7], making it practically realizable.

References

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, pp. 101–107, June 2011.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges,," *Proceedings of the IEEE*, vol. 102, pp. 366–385, Mar. 2014.

[3] H. Zhang, S. Venkateswaran, and U. Madhow, "Channel modeling and MIMO capacity for outdoor millimeter wave links," in *Proc. IEEE Wireless Communication and Networking Conf.*, pp. 1–6, 2010.

[4] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, 2013.

[5] M. Akdeniz, Y. Liu, S. Sun, S. Rangan, T. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1164–1179, June 2014.

[6] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. on Antennas and Propagation*, vol. 61, pp. 3814–3827, July 2013.

[7] J. Mo, P. Schniter, N. González-Prelcic, and R. W. Heath, Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2014.

[8] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. on Signal Processing*, pp. 2563–2579, Oct. 2002.

[9] "IEEE 802.11ad standard draft D0.1." [Online], 2012. Available: www.ieee802.org/11/Reports/tgad_update.htm.

[10] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 3 ed., 2009.

[11] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays,," *Proc. Information Theory and Applications Workshop*, pp. 234–239, Feb. 2012.

[12] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[13] P. Schniter and A. M. Sayeed, "A sparseness-preserving virtual MIMO channel model," in *Proc. Conf. on Information Science and Systems*, Mar. 2004.

[14] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, pp. 1058–1076, June 2010.

[15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[16] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[17] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge University Press, 2005.

[18] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins University Press, 3rd ed., 1996.