

# BEACHES: Beamspace Channel Estimation for Multi-Antenna mmWave Systems and Beyond

Ramina Ghods, Alexandra Gallyas-Sanhueza, Seyed Hadi Mirfarshbafan, and Christoph Studer

*School of Electrical and Computer Engineering, Cornell University, Ithaca, NY; email: studer@cornell.edu*

**Abstract**—Massive multi-antenna millimeter wave (mmWave) and terahertz wireless systems promise high-bandwidth communication to multiple user equipments in the same time-frequency resource. The high path loss of wave propagation at such frequencies and the fine-grained nature of beamforming with massive antenna arrays necessitates accurate channel estimation to fully exploit the advantages of such systems. In this paper, we propose BEAmespace CHannel EStimation (BEACHES), a low-complexity channel estimation algorithm for multi-antenna mmWave systems and beyond. BEACHES leverages the fact that wave propagation at high frequencies is directional, which enables us to denoise the (approximately) sparse channel state information in the beamspace domain. To avoid tedious parameter selection, BEACHES includes a computationally-efficient tuning stage that provably minimizes the mean-square error of the channel estimate in the large-antenna limit. To demonstrate the efficacy of BEACHES, we provide simulation results for line-of-sight (LoS) and non-LoS mmWave channel models.

## I. INTRODUCTION

Massive multiuser (MU) multiple-input multiple-output (MIMO) [1] and millimeter-wave (mmWave) communication [2], [3] are among the key technologies of next-generation wireless systems. The high path loss of wave propagation at mmWave or terahertz (THz) frequencies and the fact that massive MU-MIMO enables fine-grained beamforming, requires the basestations (BSs) to acquire accurate channel state information (CSI) [4], [5]. In addition, the trend towards low-precision data converters in all-digital massive MU-MIMO BSs to reduce power, interconnect bandwidth, and costs [6] renders accurate channel estimation increasingly important.

At mmWave or terahertz frequencies, wave propagation is highly directional and real-world channels typically comprise only a small number of dominant propagation paths [2], [3]. These unique properties enable the deployment of channel estimation algorithms that effectively suppress noise [7]–[9]. As a consequence, compressive sensing (CS)-based methods have been proposed for mmWave channel estimation in [10], [11]. Most of such methods use a discretization procedure of the number of propagation paths that can be resolved in the *beamspace* domain [12], resulting in the well-known basis mismatch problem [13]. Methods that perform off-grid CS, such as atomic norm minimization (ANM) [14] or Newtonized orthogonal matching pursuit (NOMP) [15],

The work of RG was supported by the US National Science Foundation under grants ECCS-1408006, CCF-1535897, CCF-1652065, CNS-1717559, and ECCS-1824379. The work of AGS, SHM, and CS was supported in part by Xilinx, Inc. and by ComSenTer, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

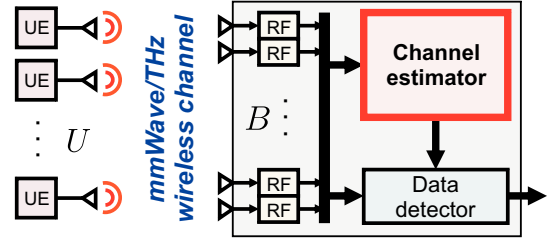


Fig. 1. Massive MU-MIMO mmWave uplink system:  $U$  UEs transmit pilots over a mmWave/THz wireless channel, which are used to estimate the channel vectors associated to each UE at the  $B$ -antenna base station.

avoid this basis mismatch problem. These methods, however, exhibit excessively high complexity, which renders the design of corresponding hardware designs challenging.

### A. Contributions

We propose a new channel estimation algorithm for massive MU-MIMO mmWave/terahertz communication systems that relies on Stein's unbiased risk estimator (SURE). Our algorithm, called BEAmespace CHannel EStimation (BEACHES), exploits sparsity of mmWave/terahertz channels in the beamspace domain and adaptively denoises the channel vectors at low complexity. We prove that BEACHES minimizes the mean square error (MSE) of the channel estimate in the large-antenna limit. We evaluate BEACHES for LoS and non-LoS mmWave channels and demonstrate that it performs on par with ANM and NOMP but at orders-of-magnitude lower complexity.

### B. Notation

Lowercase and uppercase boldface letters designate column vectors and matrices, respectively. For a vector  $\mathbf{a}$ , the  $k$ th entry is  $[\mathbf{a}]_k = a_k$ ; the real and imaginary parts are  $[\mathbf{a}]_{\mathcal{R}} = \mathbf{a}_{\mathcal{R}}$  and  $[\mathbf{a}]_{\mathcal{I}} = \mathbf{a}_{\mathcal{I}}$ , respectively. The transpose and conjugate transpose of matrix  $\mathbf{A}$  are  $\mathbf{A}^T$  and  $\mathbf{A}^H$ , respectively. A complex Gaussian vector  $\mathbf{a}$  with mean vector  $\mathbf{m}$  and covariance matrix  $\mathbf{K}$  is written as  $\mathbf{a} \sim \mathcal{CN}(\mathbf{m}, \mathbf{K})$  and its probability density function (PDF) as  $f^{\mathcal{CN}}(\mathbf{a}; \mathbf{m}, \mathbf{K})$ . A real Gaussian vector  $\mathbf{a}$  with mean vector  $\mathbf{m}$  and covariance matrix  $\mathbf{K}$  is written as  $\mathbf{a} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$  and its PDF as  $f^{\mathcal{N}}(\mathbf{a}; \mathbf{m}, \mathbf{K})$ . The expectation operator is  $\mathbb{E}[\cdot]$ .

## II. SYSTEM MODEL AND BEAMSPACE REPRESENTATION

### A. System Model

We consider an all-digital mmWave/THz massive MU-MIMO uplink system as illustrated in Fig. 1. The BS is equipped with a  $B$ -antenna uniform linear array (ULA) and

communicates with  $U$  single-antenna user equipments (UEs) in the same time-frequency resource. For simplicity, we focus on pilot-based channel estimation for flat-fading channels, where the BS estimates the  $B$ -dimensional complex channel vector  $\mathbf{h} \in \mathbb{C}^B$  for each UE. By assuming that (i) wave propagation is predominantly directional [4], [16], and (ii) the distance between UEs and BS is sufficiently large, the channel vectors in the antenna domain can be modeled as follows [17]:

$$\mathbf{h} = \sum_{\ell=0}^{L-1} \alpha_\ell \mathbf{a}(\Omega_\ell), \quad \mathbf{a}(\Omega) = [e^{j0\Omega}, e^{j1\Omega}, \dots, e^{j(B-1)\Omega}]^T. \quad (1)$$

Here,  $L$  refers to the total number of paths arriving at the antenna array (including a potential LoS path),  $\alpha_\ell \in \mathbb{C}$  is the complex-valued channel gain of the  $\ell$ th path, and  $\mathbf{a}(\Omega_\ell)$  represents a complex-valued sinusoid containing the relative phases between BS antennas, where  $\Omega_\ell \in [0, 2\pi)$  is determined by the incident angle of the  $\ell$ th path to the antenna array. We model the estimated channel vector in the antenna domain as  $\mathbf{y} = \mathbf{h} + \mathbf{e}$ , where  $\mathbf{e} \sim \mathcal{CN}(\mathbf{0}_{B \times 1}, E_0 \mathbf{I}_B)$  represents channel estimation error with variance  $E_0$  per complex entry.

### B. Beamspace Channel Vector Denoising

The channel vectors  $\mathbf{h}$  as modeled in (1) are a superposition of  $L$  complex-valued sinusoids. Hence, it is useful to transform the vector  $\mathbf{h}$  into the discrete Fourier transform (DFT) domain, also known as the *beamspace domain*,  $\hat{\mathbf{h}} = \mathbf{F}\mathbf{h}$ , where  $\mathbf{F}$  is the  $B \times B$  unitary DFT matrix. In the beamspace domain, each entry of  $\hat{\mathbf{h}}$  is associated to a specific incident angle with respect to the BS antenna array [12]. If the number of paths  $L$  is smaller than the number of BS antennas  $B$ , then the beamspace channel vector  $\hat{\mathbf{h}}$  will be (approximately) sparse [8]. This key property enables the use of denoising algorithms. More specifically, by transforming  $\mathbf{y}$  into the beamspace domain  $\hat{\mathbf{y}} = \mathbf{F}\mathbf{y} = \hat{\mathbf{h}} + \hat{\mathbf{e}}$ , where  $\hat{\mathbf{e}} = \mathbf{F}\mathbf{e}$  has the same statistics as  $\mathbf{e}$ , one can suppress noise while preserving the strong beamspace components. Prominent methods for beamspace denoising are ANM [14] and NOMP [15], which require high complexity.

## III. BEACHES: BEAMSPACE CHANNEL ESTIMATION

### A. Channel Vector Denoising via Soft-Thresholding

A widely-used sparsity-based denoising method is the least absolute shrinkage and selection operator (LASSO) [18], [19], which corresponds to the following optimization problem:

$$\hat{\mathbf{h}}^* = \arg \min_{\hat{\mathbf{h}}' \in \mathbb{C}^B} \frac{1}{2} \|\hat{\mathbf{y}} - \hat{\mathbf{h}}'\|_2^2 + \tau \|\hat{\mathbf{h}}'\|_1. \quad (2)$$

Here,  $\tau \in \mathbb{R}_+$  is a suitably-chosen denoising parameter. The solution to (2) in the complex case is the well-known *soft-thresholding operator*  $\eta(\hat{\mathbf{y}}, \tau)$  defined as [20, App. A]

$$[\eta(\hat{\mathbf{y}}, \tau)]_b = \frac{\hat{y}_b}{|\hat{y}_b|} \max\{|\hat{y}_b| - \tau, 0\}, \quad b = 1, \dots, B, \quad (3)$$

where we define  $y/|y| = 0$  for  $y = 0$ . For sparsity-based denoising via soft-thresholding, the performance strongly depends on the choice of the denoising parameter  $\tau$  [18], [21]. In wireless systems, it is particularly important to design robust methods to select this parameter, as many factors such

as the propagation conditions, the number of arriving paths, and the signal power can vary widely over time.

### B. Computing the Optimal Denoising Parameter

In what follows, we are interested in the optimal parameter  $\tau^*$  that minimizes the estimation MSE defined as

$$MSE = \frac{1}{B} \mathbb{E} [\|\hat{\mathbf{h}}^* - \hat{\mathbf{h}}\|_2^2], \quad (4)$$

where  $\hat{\mathbf{h}}^* = \eta(\hat{\mathbf{y}}, \tau^*)$  is the associated denoised beamspace vector. Determining the optimal parameter  $\tau^*$  requires knowledge of the noiseless beamspace vector  $\hat{\mathbf{h}}$ , which is unknown. To avoid the need for knowing the ground truth  $\hat{\mathbf{h}}$ , we propose to use Stein's unbiased risk estimate (SURE) [18] as a proxy for the MSE function. The following result provides SURE in the complex domain and shows that it is an unbiased estimator for the MSE. The proof is given in Appendix A.

**Theorem 1.** *Let  $\hat{\mathbf{h}} \in \mathbb{C}^B$  be an unknown vector and  $\hat{\mathbf{y}} \in \mathbb{C}^B$  a noisy observation vector distributed as  $\hat{\mathbf{y}} \sim \mathcal{CN}(\hat{\mathbf{h}}, E_0 \mathbf{I}_B)$ . Let  $\mu(\hat{\mathbf{y}})$  be an estimator of  $\hat{\mathbf{h}}$  from  $\hat{\mathbf{y}}$  that is weakly differentiable and operates element-wise on vectors. Then,*

$$SURE = \frac{1}{B} \|\mu(\hat{\mathbf{y}}) - \hat{\mathbf{y}}\|_2^2 + E_0 + \frac{1}{B} E_0 \sum_{b=1}^B \left( \frac{\partial [\mu_{\mathcal{R}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{R}}]_b} + \frac{\partial [\mu_{\mathcal{I}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{I}}]_b} - 2 \right) \quad (5)$$

is an unbiased estimate of the MSE, i.e.,  $\mathbb{E}[SURE] = MSE$ .

The following theorem shows that SURE for the soft-thresholding operator  $\eta(\hat{\mathbf{y}}, \tau)$  converges to the MSE in the large-antenna limit  $B \rightarrow \infty$ . The proof is given in Appendix B.

**Theorem 2.** *For the soft-thresholding function  $\mu(\hat{\mathbf{y}}) = \eta(\hat{\mathbf{y}}, \tau)$  in (3), SURE in (5) is given by<sup>1</sup>*

$$SURE_\tau = \frac{1}{B} \sum_{b: |\hat{y}_b| < \tau} |\hat{y}_b|^2 + \frac{1}{B} \sum_{b: |\hat{y}_b| > \tau} \tau^2 + E_0 - \frac{E_0}{B} \tau \sum_{b: |\hat{y}_b| > \tau} \frac{1}{|\hat{y}_b|} - 2 \frac{E_0}{B} \sum_{b: |\hat{y}_b| < \tau} 1, \quad (6)$$

which, in the limit  $B \rightarrow \infty$  converges to the MSE, i.e.,

$$\lim_{B \rightarrow \infty} SURE_\tau = MSE. \quad (7)$$

SURE in (6) is *independent* of the true beamspace channel vector  $\hat{\mathbf{h}}$ . The expression (6) only depends on the magnitudes of the observed beamspace channel vector  $\hat{\mathbf{y}}$ , the channel estimation error variance  $E_0$ , the number of BS antennas  $B$ , and the denoising parameter  $\tau$ . Thanks to (7) and the fact that  $B$  is large in massive MU-MIMO systems, we can use SURE as a surrogate to minimize the MSE and to determine the optimal denoising parameter. While no closed-form expression for the minimum of (6) is known, reference [21] proposes a bisection procedure to approximate the optimal value of a similar expression for sparse recovery. We next propose an efficient algorithm that computes the optimal parameter  $\tau^*$  using a deterministic procedure with complexity  $O(B \log(B))$ .

<sup>1</sup>As discussed in Appendix B, the value of SURE is undefined for  $\tau = \hat{y}_b$ ,  $b = 1, \dots, B$ , due to the non-differentiability of the function  $\eta$ .

---

**Algorithm 1** BEACHES: BEAmospace CHannel ESTimation

---

```

1: input  $\hat{\mathbf{y}} = \text{FFT}(\mathbf{y})$  and  $E_0$ 
2:  $S = 0$  and  $\text{SURE}_{\min} = \infty$ 
3:  $\hat{\mathbf{y}}^s = \text{sort}\{|\hat{\mathbf{y}}|, \text{'ascend'}\}$ 
4:  $V = \sum_{k=1}^B (|\hat{y}_k^s|)^{-1}$ ,  $\hat{y}_0^s = 0$ , and  $\hat{y}_{B+1}^s = \infty$ 
5: for  $k = 1, \dots, B+1$  do
6:    $\tau = \max\{\hat{y}_{k-1}^s, \min\{\hat{y}_k^s, \frac{E_0}{2(B-k+1)}V\}\}$ .
7:    $\text{SURE}_\tau = \frac{S}{B} + \frac{(B-k+1)}{B}\tau^2 + E_0 - \frac{E_0}{B}\tau V - 2\frac{E_0}{B}(k-1)$ 
8:   if  $\text{SURE}_\tau < \text{SURE}_{\min}$  then
9:      $\text{SURE}_{\min} = \text{SURE}_\tau$ 
10:     $\tau^* = \tau$ 
11:   end if
12:    $S = S + (\hat{y}_k^s)^2$  and  $V = V - (\hat{y}_k^s)^{-1}$ 
13: end for
14:  $[\hat{\mathbf{h}}^*]_k = \frac{\hat{y}_k}{|\hat{y}_k|} \max\{|\hat{y}_k| - \tau^*, 0\}$ ,  $k = 1, \dots, B$ 
15: return  $\mathbf{h}^* = \text{IFFT}(\hat{\mathbf{h}}^*)$ 

```

---

### C. The BEACHES Algorithm

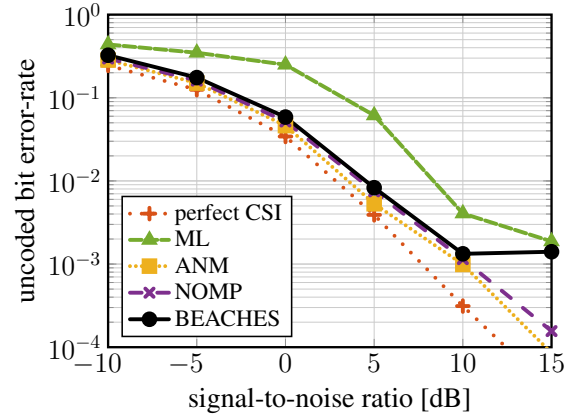
Reference [18] outlines an efficient procedure to minimize SURE for real-valued wavelet denoising. We propose a similar strategy to minimize (6) for the complex-valued case with soft-thresholding. Instead of continuously sweeping the denoising parameter  $\tau$  in the interval  $[0, \infty)$ , we first sort the absolute values of the vector  $\hat{\mathbf{y}}$  in ascending order which we call  $\hat{\mathbf{y}}^s$ . We then search for the optimal denoising parameter  $\tau$  only between each pair of consecutive elements of the sorted vector, i.e.,  $\tau \in (\hat{y}_{k-1}^s, \hat{y}_k^s)$  for  $k = 1, \dots, B+1$ . In each such interval, SURE is a quadratic function in  $\tau$  given by

$$\text{SURE}_\tau = \sum_{b=0}^{k-1} \frac{(\hat{y}_b^s)^2}{B} + \frac{(B-k+1)}{B}\tau^2 + E_0 - \frac{E_0}{B}\tau \sum_{b=k}^{B+1} (\hat{y}_b^s)^{-1} - 2\frac{E_0}{B}(k-1), \quad (8)$$

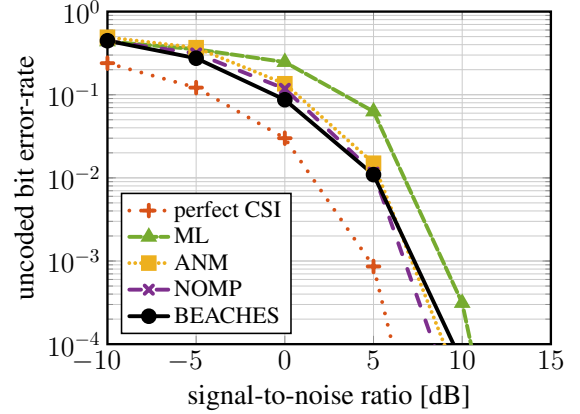
where we define  $\hat{y}_0^s = 0$  and  $\hat{y}_{B+1}^s = \infty$  to account for the first  $(0, \hat{y}_1^s)$  and last interval  $(\hat{y}_B^s, \infty)$ . For each  $k \in \{1, \dots, B+1\}$ , we compute the optimal value of  $\tau$  that minimizes SURE in each interval  $\tau \in (\hat{y}_{k-1}^s, \hat{y}_k^s)$ . Since there is a discontinuity in the SURE expression when progressing from one interval to the next, the minimal value in each interval is either the minimum of the quadratic function (8) or one of the boundaries of the interval  $\hat{y}_{k-1}^s$  and  $\hat{y}_k^s$ .<sup>2</sup> The minimum value of (8) is given by  $\tau_k^Q = \frac{E_0}{2(B-k+1)} \sum_{b=k}^{B+1} (\hat{y}_b^s)^{-1}$ . Since the function  $\text{SURE}_\tau$  is convex, we can identify the optimal parameter  $\tau$  in the interval  $(\hat{y}_{k-1}^s, \hat{y}_k^s)$  indexed by  $k$  by knowing the value of  $\tau_k^Q$  with respect to the interval boundaries  $\hat{y}_{k-1}^s$  and  $\hat{y}_k^s$ . Put simply, the optimal denoising parameter  $\tau_k^*$  in each interval  $k = 1, \dots, B+1$  is given by

$$\tau_k^* = \begin{cases} \tau_k^Q & \hat{y}_{k-1}^s < \tau_k^Q < \hat{y}_k^s \\ \hat{y}_{k-1}^s & \tau_k^Q < \hat{y}_{k-1}^s \\ \hat{y}_k^s & \tau_k^Q > \hat{y}_k^s, \end{cases}$$

<sup>2</sup>Note that SURE is not defined for  $\tau = \hat{y}_{k-1}^s$  and  $\tau = \hat{y}_k^s$ . Instead, we compute  $\text{SURE}_\tau$  for two values arbitrarily close to these boundaries, i.e.,  $\tau = \hat{y}_{k-1}^s + \epsilon$  and  $\tau = \hat{y}_k^s - \epsilon$ , where  $\epsilon > 0$  is small compared to  $\tau$ .



(a) LoS,  $B = 256$ ,  $U = 16$



(b) Non-LoS,  $B = 256$ ,  $U = 16$

Fig. 2. Unencoded bit error-rate (BER) of various channel denoising methods for LoS and non-LoS channels. We see that BEACHES performs on par with atomic norm minimization (ANM) and Newtonized OMP, and provides 2 dB to 3 dB SNR improvements over ML channel estimation at  $\text{BER} = 10^{-2}$ .

or simply  $\tau_k^* = \max\{\hat{y}_{k-1}^s, \min\{\hat{y}_k^s, \tau_k^Q\}\}$ . By knowing the optimal value of  $\tau$  in each interval, we only need to find the minimal value of  $\text{SURE}_{\tau_k^*}$  for  $k = 1, \dots, B+1$ . It is now key to realize that we do not need to recalculate SURE in (8) from scratch while searching through  $k = 1, \dots, B+1$ . Instead, we can sequentially update the quantities  $S = \sum_{b=0}^{k-1} (\hat{y}_b^s)^2$  and  $V = \sum_{b=k}^{B+1} (\hat{y}_b^s)^{-1}$ , thanks to sorting the magnitudes of the vector  $\hat{\mathbf{y}}$ . The resulting procedure, called BEACHES, is summarized in Algorithm 1. The computational complexity of BEACHES is only  $O(B \log(B))$ , which is caused by the FFT, sorting, and IFFT operations—the remaining computations in the for-loop (lines 5–13) are simple scalar operations.

## IV. PERFORMANCE AND RUNTIME OF BEACHES

### A. Bit Error-Rate Performance

To assess the performance of BEACHES, we consider an all-digital massive MU-MIMO system in which  $U = 16$  UEs communicate with a  $B = 256$  antenna BS. We focus on the situation in which the UEs first send orthogonal pilots, which are used to acquire maximum-likelihood channel estimates  $\mathbf{y}_u$  for each UE  $u = 1, \dots, U$ . The channel matrices are generated for both LoS and non-LoS conditions using the QuaDRiGa

TABLE I  
MATLAB RUNTIMES IN MILLISECONDS (AND NORMALIZED RUNTIMES).

Scenario	BEACHES	NOMP	ANM
LoS	1.64 (1×)	199.9 (120×)	47 968 (29 000×)
non-LoS	1.45 (1×)	2 204 (1 500×)	83 750 (58 000×)

mmMAGIC UMi model [22] at a carrier frequency of 60 GHz with a ULA using  $\lambda/2$  antenna spacing. The UEs are placed randomly within a  $120^\circ$  circular sector with minimum and maximum distance of 10 and 110 meters from the BS antenna array, respectively. We enforce UE separation of at least  $1^\circ$  (relative to the BS antenna array) and assume optimal UE power control. We then use different channel vector denoising methods, including (i) ANM-based denoising, where we use the debiased output of the code provided in [14], (ii) NOMP with a (manually tuned) false alarm rate of  $P_{fa} = 0.5$  using the code provided in [15], and (iii) “perfect CSI,” which is a baseline that uses the noiseless channel vectors. Finally, we transmit 16-QAM symbols and perform linear minimum MSE (L-MMSE) equalization with the denoised matrix to detect the transmitted bits. The resulting uncoded bit error-rate (BER) is used to assess the performance of various denoising methods.

Figure 2 shows that channel vector denoising in the beamspace domain provides 2 dB to 3 dB SNR performance improvements at  $BER = 10^{-2}$  compared to conventional ML channel estimation. The achieved performance gains are more pronounced under LoS conditions. Quite surprisingly, we observe that BEACHES performs on par to ANM and NOMP. This observation indicates that off-the-grid denoising methods, such as ANM and NOMP, do not provide a critical performance advantage over BEACHES (in terms of BER).

### B. Runtime Comparison

While the BER performance of BEACHES is comparable to ANM and NOMP, it exhibits significantly lower complexity. To support this claim, we measured their MATLAB runtimes in milliseconds on an Intel core i5-7400 CPU with 16 GB RAM at an SNR of 5 dB. Table I demonstrates that the runtime of BEACHES is orders of magnitude lower than that of NOMP (up to 1 500×) and ANM (up to 58 000×), while the speedup is more pronounced for the non-LoS scenario.

## V. CONCLUSIONS

We have proposed a new channel denoising algorithm for massive MU-MIMO mmWave and terahertz communication systems called BEAmespace CHannel ESTimation (BEACHES). BEACHES exploits sparsity of mmWave/terahertz channels in the beamspace domain to perform adaptive soft-thresholding via Stein’s unbiased risk estimate (SURE). We have shown that BEACHES minimizes the mean square error in the large-antenna limit and performs on par with sophisticated channel estimation algorithms for realistic LoS and non-LoS channel models but at orders-of-magnitude lower complexity. There are many avenues for future work. An extension of BEACHES to systems with low-precision quantizers and single-carrier transmission is an open research problem.

## APPENDIX A PROOF OF THEOREM 1

The MSE for  $\hat{\mathbf{h}}^* = \mu(\hat{\mathbf{y}})$  is defined as

$$MSE = \mathbb{E} \left[ \frac{1}{B} \|\hat{\mathbf{h}}^* - \hat{\mathbf{h}}\|_2^2 \right] = \mathbb{E} \left[ \frac{1}{B} \|\mu(\hat{\mathbf{y}}) - \hat{\mathbf{h}}\|_2^2 \right],$$

where we decompose the complex-valued vector  $\hat{\mathbf{y}}$  into the real part  $\hat{\mathbf{y}}_{\mathcal{R}} \sim \mathcal{N}(\mathbf{h}_{\mathcal{R}}, \frac{E_0}{2} \mathbf{I}_B)$  and imaginary part  $\hat{\mathbf{y}}_{\mathcal{I}} \sim \mathcal{N}(\mathbf{h}_{\mathcal{I}}, \frac{E_0}{2} \mathbf{I}_B)$ . Note that expectation is with respect to the noisy observation  $\hat{\mathbf{y}}$ . Define  $g(\hat{\mathbf{y}}) = \mu(\hat{\mathbf{y}}) - \hat{\mathbf{y}}$ . Hence,

$$\begin{aligned} MSE &= \frac{1}{B} \mathbb{E} \left[ \|g(\hat{\mathbf{y}}) + \hat{\mathbf{y}} - \hat{\mathbf{h}}\|_2^2 \right] \\ &= \mathbb{E} \left[ \frac{1}{B} \|g(\hat{\mathbf{y}})\|_2^2 \right] + \mathbb{E} \left[ \frac{1}{B} \|\hat{\mathbf{y}} - \hat{\mathbf{h}}\|_2^2 \right] \\ &\quad + \mathbb{E} \left[ \frac{2}{B} \left[ g(\hat{\mathbf{y}})^H (\hat{\mathbf{y}} - \hat{\mathbf{h}}) \right]_{\mathcal{R}} \right]. \end{aligned} \quad (9)$$

The last term can be simplified as

$$\begin{aligned} \frac{2}{B} \mathbb{E} \left[ \left[ g(\hat{\mathbf{y}})^H (\hat{\mathbf{y}} - \hat{\mathbf{h}}) \right]_{\mathcal{R}} \right] &= \frac{2}{B} \mathbb{E} \left[ g_{\mathcal{R}}(\hat{\mathbf{y}})^T (\hat{\mathbf{y}}_{\mathcal{R}} - \hat{\mathbf{h}}_{\mathcal{R}}) \right] \\ &\quad + \frac{2}{B} \mathbb{E} \left[ g_{\mathcal{I}}(\hat{\mathbf{y}})^T (\hat{\mathbf{y}}_{\mathcal{I}} - \hat{\mathbf{h}}_{\mathcal{I}}) \right]. \end{aligned}$$

We can now expand  $\frac{2}{B} \mathbb{E} \left[ g_{\mathcal{R}}(\hat{\mathbf{y}})^T (\hat{\mathbf{y}}_{\mathcal{R}} - \hat{\mathbf{h}}_{\mathcal{R}}) \right]$  which yields

$$\frac{2}{B} \mathbb{E} \left[ g_{\mathcal{R}}(\hat{\mathbf{y}})^T (\hat{\mathbf{y}}_{\mathcal{R}} - \hat{\mathbf{h}}_{\mathcal{R}}) \right] \quad (10)$$

$$\begin{aligned} &\stackrel{(a)}{=} \frac{2}{B} \int_{\hat{\mathbf{y}}_{\mathcal{I}}} f^{\mathcal{N}} \left( \hat{\mathbf{y}}_{\mathcal{I}}; \hat{\mathbf{h}}_{\mathcal{I}}, \frac{E_0}{2} \mathbf{I}_B \right) \sum_{b=1}^B \int_{\hat{\mathbf{y}}_{\mathcal{R}}} \frac{1}{(2\pi \frac{E_0}{2})^{B/2}} \times \\ &\quad \exp \left( -\frac{\|\hat{\mathbf{y}}_{\mathcal{R}} - \hat{\mathbf{h}}_{\mathcal{R}}\|^2}{2 \frac{E_0}{2}} \right) \frac{E_0}{2} \frac{\partial [g_{\mathcal{R}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{R}}]_b} d\hat{\mathbf{y}}_{\mathcal{R}} d\hat{\mathbf{y}}_{\mathcal{I}} \end{aligned} \quad (11)$$

$$= \frac{E_0}{B} \mathbb{E} \left[ \sum_{b=1}^B \left( \frac{\partial [\mu_{\mathcal{R}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{R}}]_b} - 1 \right) \right], \quad (12)$$

where (a) follows from integration by parts. Similarly, we have

$$\frac{2}{B} \mathbb{E} \left[ g_{\mathcal{I}}(\hat{\mathbf{y}})^T (\hat{\mathbf{y}}_{\mathcal{I}} - \hat{\mathbf{h}}_{\mathcal{I}}) \right] = \frac{E_0}{B} \mathbb{E} \left[ \sum_{b=1}^B \left( \frac{\partial [g_{\mathcal{I}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{I}}]_b} - 1 \right) \right]. \quad (13)$$

Recall that  $g(\hat{\mathbf{y}}) = \mu(\hat{\mathbf{y}}) - \hat{\mathbf{y}}$  and replace (12) and (13) in the original MSE expression in (9). This leads to

$$\begin{aligned} MSE &= \mathbb{E} \left[ \frac{1}{B} \|\mu(\hat{\mathbf{y}}) - \hat{\mathbf{y}}\|_2^2 \right] + \mathbb{E} \left[ \frac{1}{B} \|\hat{\mathbf{y}} - \hat{\mathbf{h}}\|_2^2 \right] \\ &\quad + \frac{E_0}{B} \mathbb{E} \left[ \sum_{b=1}^B \left( \frac{\partial [\mu_{\mathcal{R}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{R}}]_b} + \frac{\partial [\mu_{\mathcal{I}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{I}}]_b} - 2 \right) \right]. \end{aligned}$$

The second term in the MSE expression above equals  $E_0$ . For the first and third term we remove their expectations to arrive at the following SURE expression:

$$\begin{aligned} SURE &= \frac{1}{B} \|\mu(\hat{\mathbf{y}}) - \hat{\mathbf{y}}\|_2^2 + E_0 \\ &\quad + \frac{E_0}{B} \sum_{b=1}^B \left( \frac{\partial [\mu_{\mathcal{R}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{R}}]_b} + \frac{\partial [\mu_{\mathcal{I}}(\hat{\mathbf{y}})]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{I}}]_b} - 2 \right), \end{aligned}$$

which establishes the fact that  $\mathbb{E}[SURE] = MSE$ .

## APPENDIX B PROOF OF THEOREM 2

SURE in (5) for  $\mu(\hat{\mathbf{y}}) = \eta(\hat{\mathbf{y}}, \tau)$  is derived as follows. The only unknowns in the expression of SURE are its derivative of real and imaginary parts. For  $|\hat{y}_b| < \tau$ , we have

$$\frac{\partial [\eta_{\mathcal{R}}(\hat{\mathbf{y}}, \tau)]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{R}}]_b} = \frac{\partial [\eta_{\mathcal{I}}(\hat{\mathbf{y}}, \tau)]_b}{\partial [\hat{\mathbf{y}}_{\mathcal{I}}]_b} = 0.$$

For  $|\hat{y}_b| > \tau$ , we have

$$\begin{aligned}\frac{\partial[\eta_{\mathcal{R}}(\hat{\mathbf{y}}, \tau)]_b}{\partial[\hat{\mathbf{y}}_{\mathcal{R}}]_b} &= \frac{\partial}{\partial[\hat{\mathbf{y}}_{\mathcal{R}}]_b} \left( [\hat{\mathbf{y}}_{\mathcal{R}}]_b - \frac{[\hat{\mathbf{y}}_{\mathcal{R}}]_b \tau}{\sqrt{[\hat{\mathbf{y}}_{\mathcal{R}}]_b^2 + [\hat{\mathbf{y}}_{\mathcal{I}}]_b^2}} \right) \\ &= 1 - \tau \frac{[\hat{\mathbf{y}}_{\mathcal{R}}]_b^2}{([\hat{\mathbf{y}}_{\mathcal{R}}]_b^2 + [\hat{\mathbf{y}}_{\mathcal{I}}]_b^2)^{3/2}}\end{aligned}$$

and

$$\frac{\partial[\eta_{\mathcal{I}}(\hat{\mathbf{y}}, \tau)]_b}{\partial[\hat{\mathbf{y}}_{\mathcal{I}}]_b} = 1 - \tau \frac{[\hat{\mathbf{y}}_{\mathcal{I}}]_b^2}{([\hat{\mathbf{y}}_{\mathcal{R}}]_b^2 + [\hat{\mathbf{y}}_{\mathcal{I}}]_b^2)^{3/2}}.$$

Note that at  $|\hat{y}_b| = \tau$ , there is a discontinuity and thus the derivative and consequently SURE are not defined for this value. The complex-valued SURE expression reduces to

$$\begin{aligned}\text{SURE}_{\tau} &= \frac{1}{B} \sum_{b=1}^B \min\{|\hat{y}_b|, \tau\}^2 + E_0 \\ &\quad + \frac{E_0}{B} \sum_{b: |\hat{y}_b| > \tau} \left( 2 - \tau \frac{1}{\sqrt{[\hat{\mathbf{y}}_{\mathcal{R}}]_b^2 + [\hat{\mathbf{y}}_{\mathcal{I}}]_b^2}} - 2 \right) \\ &\quad + \frac{E_0}{B} \sum_{b: |\hat{y}_b| < \tau} (0 - 2).\end{aligned}$$

We now prove the convergence of SURE in (7). In [23, Lemma 4.3.], the authors prove convergence of SURE to MSE in the real domain for the soft-thresholding function. We follow the same procedure for the complex domain. Using [20, Thm. III.15 & III.16], we have that for any pseudo-Lipschitz function  $\gamma: \mathbb{C} \rightarrow \mathbb{R}$  the following equality holds:

$$\begin{aligned}\lim_{B \rightarrow \infty} \frac{1}{B} \sum_{b=1}^B \gamma(\eta(\hat{y}_b, \tau), \hat{h}_b) \\ = \mathbb{E}[\gamma(\eta(H + \sqrt{E_0}Z, \tau), H)].\end{aligned}\quad (14)$$

Here,  $Z \sim \mathcal{CN}(0, 1)$  and  $H$  is a random variable with the sparse distribution of the channel vector in the beamspace domain  $\hat{h}_b$ . Using (14), we have the following result

$$\lim_{B \rightarrow \infty} \frac{1}{B} \sum_{b=1}^B |\eta(\hat{y}_b, \tau) - \hat{y}_b|^2 = \mathbb{E}_{\hat{y}_b} [|\eta(\hat{y}_b, \tau) - \hat{y}_b|^2],$$

where,  $\hat{y}_b$  is any element of the random vector  $\hat{\mathbf{y}}$ . The expression above can be rewritten as

$$\lim_{B \rightarrow \infty} \frac{1}{B} \|\eta(\hat{\mathbf{y}}, \tau) - \hat{\mathbf{y}}\|_2^2 = \mathbb{E}_{\hat{\mathbf{y}}} \left[ \frac{1}{B} \|\eta(\hat{\mathbf{y}}, \tau) - \hat{\mathbf{y}}\|_2^2 \right]. \quad (15)$$

Now, since  $\frac{\partial[\eta_{\mathcal{R}}(\hat{\mathbf{y}}, \tau)]_b}{\partial[\hat{\mathbf{y}}_{\mathcal{R}}]_b} + \frac{\partial[\eta_{\mathcal{I}}(\hat{\mathbf{y}}, \tau)]_b}{\partial[\hat{\mathbf{y}}_{\mathcal{I}}]_b}$  is bounded, it is pseudo-Lipschitz and, hence, we can use (14) to obtain the following convergence result:

$$\begin{aligned}\lim_{B \rightarrow \infty} \frac{1}{B} \sum_{b=1}^B \left( \frac{\partial[\mu_{\mathcal{R}}(\hat{\mathbf{y}})]_b}{\partial[\hat{\mathbf{y}}_{\mathcal{R}}]_b} + \frac{\partial[\mu_{\mathcal{I}}(\hat{\mathbf{y}})]_b}{\partial[\hat{\mathbf{y}}_{\mathcal{I}}]_b} - 2 \right) \\ = \frac{1}{B} \mathbb{E} \left[ \sum_{b=1}^B \left( \frac{\partial[\mu_{\mathcal{R}}(\hat{\mathbf{y}})]_b}{\partial[\hat{\mathbf{y}}_{\mathcal{R}}]_b} + \frac{\partial[\mu_{\mathcal{I}}(\hat{\mathbf{y}})]_b}{\partial[\hat{\mathbf{y}}_{\mathcal{I}}]_b} - 2 \right) \right].\end{aligned}\quad (16)$$

By summing (15) and (16) and  $\mathbb{E} \left[ \frac{1}{B} \|\hat{\mathbf{y}} - \hat{\mathbf{h}}\|_2^2 \right] = E_0$ , we have established that  $\lim_{B \rightarrow \infty} \text{SURE}_{\tau} = \text{MSE}$ .

## REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [3] T. Rappaport, R. Heath, R. Daniels, and J. Murdock, *Millimeter Wave Wireless Communications*. Prentice Hall, 2015.
- [4] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [5] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
- [6] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.
- [7] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "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, Oct. 2014.
- [8] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proceedings of the 48th Asilomar Conference on Signals, Systems and Computers*, Nov. 2014, pp. 273–277.
- [9] J. Deng, O. Tirkkonen, and C. Studer, "mmWave channel estimation via atomic norm minimization for multi-user hybrid precoding," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2018, pp. 1–6.
- [10] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [11] J. Lee, G. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.
- [12] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.
- [13] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.
- [14] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.
- [15] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5066–5081, Oct. 2016.
- [16] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [17] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005.
- [18] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.
- [19] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [20] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4290–4308, Jul. 2013.
- [21] A. Mousavi, A. Maleki, and R. G. Baraniuk, "Parameterless optimal approximate message passing," Oct. 2013. [Online]. Available: <http://arxiv.org/abs/1311.0035>
- [22] S. Jaekel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, "QuaDRiGa - Quasi Deterministic Radio Channel Generator User Manual and Documentation," Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2.0.0, 2017.
- [23] A. Mousavi, A. Maleki, and R. G. Baraniuk, "Consistent parameter estimation for LASSO and approximate message passing," *arXiv preprint: 1511.01017 [math.ST]*, Nov. 2015.