Beam Selection for Performance-Complexity Optimization in High-Dimensional MIMO Systems

John Hogan and Akbar Sayeed
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
Madison, WI 53706

***Abstract*—Millimeter-wave (mm-wave) communications systems offer a promising solution to meeting the increasing data demands on wireless networks. Not only do mm-wave systems allow orders of magnitude larger bandwidths, they also create a high-dimensional spatial signal space due to the small wavelengths, which can be exploited for beamforming and multiplexing gains. However, the complexity of digitally processing the entire high-dimensional signal is prohibitive. By exploiting the inherent channel sparsity in beamspace due to highly directional propagation at mm-wave, it is possible to design near-optimal transceivers with dramatically lower complexity. In such beamspace MIMO systems, it is first necessary to determine the set of beams which define the low-dimensional communication subspace. In this paper, we address this beam selection problem and introduce a simple power-based classifier for determining the beamspace sparsity pattern that characterizes the communication subspace. We first introduce a physical model for a small cell which will serve as the setting for our analysis. We then develop a classifier for the physical model, and show its optimality for a class of ideal signals. Finally, we present illustrative numerical results and show the feasibility of the classifier in mobile settings.**

***Index Terms*—3D Beamforming, 2D Arrays, multiuser MIMO**

I. Introduction

The data demands on wireless networks have been increasing exponentially in recent years, driven by the rapid increase in data hungry wireless devices such as smartphones. There are multiple strategies being pursued to meet the spectrum crunch, including small cells and multiple input multiple output (MIMO) technology [1]. Even so, there is a growing consensus that moving to the millimeter-wave (mm-wave) spectrum, ranging from 30-300 GHz, will be feasible, as well as necessary, to provide the multi-gigabit speeds that will be needed in next-generation networks [2].

Mm-wave spectrum has multiple advantages. First, moving to these higher frequencies makes available wider ranges of spectrum, on the order of GHz [3]. Furthermore, mm-wave systems enable high dimensional MIMO (multiple input multiple output) operation, due to the small wavelengths. The large number of MIMO degrees of freedom can be exploited for a number of critical capabilities, including higher beamforming gain for power efficiency; higher spatial multiplexing gain for spectral efficiency; and highly directional communication for reduced interference [4]–[6]. Due to the predominance of line-of-sight (LoS) and sparse multipath components, communication can be performed in a low-dimensional subspace of the overall signal space, enabling complexity reduction.

We build on previous work on hybrid analog-digital transceiver architectures exploiting the concept of beamspace MIMO (B-MIMO) communication – multiplexing data onto orthogonal spatial beams – for near-optimal performance with reduced transceiver complexity [4], [5], [7], [8] . In order to take advantage of the performance-complexity optimization afforded by B-MIMO, it is first necessary to determine the optimal subset of orthogonal beams, which we term *beam selection* (BS). As practical array sizes scale up [6], efficient BS algorithms will become necessary. Previous approaches to reducing complexity, such as antenna selection, suffer from significant performance degradation [9]. Recent optimal approaches to BS have used strategies such as brute force optimization [10], sparse reconstruction [11], and iterative precoding using a hierarchical codebook [12], [13]. Although [10]–[13] present effective classifiers, they suffer from high computational complexity and training overhead.

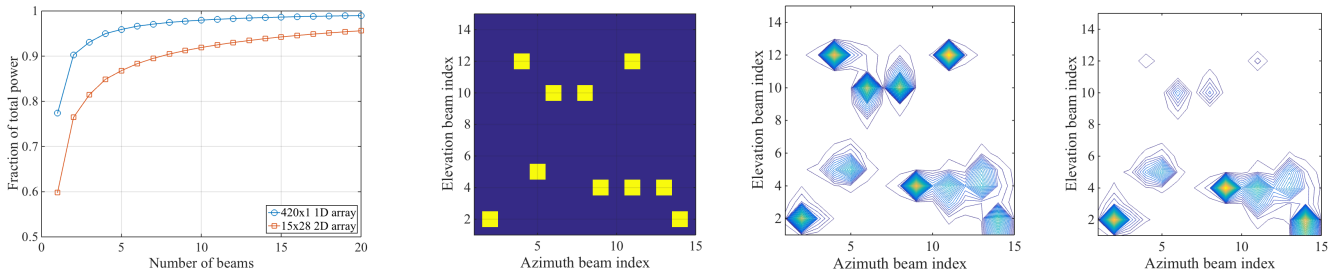
In this paper, we propose a simple power-focused criterion for beam selection, and examine its performance in a small cell access setting. In Sec. II, we describe our system model and typical beamspace patterns resulting from 1D and 2D antenna arrays. In Sec. III, we model the geometry of the small cell, introduce the BS criterion, and show it is the optimal ML classifier under an idealized model. In Sec. IV, we describe our simulation setup and present numerical results which illustrate that with BS, we can achieve near optimal performance with a low-complexity system. Finally, in Sec. V, we examine the overhead consumed by our BS estimation procedure in a small cell setting with user mobility.

II. System Model

We consider a narrowband model for the uplink multiuser-MIMO channel consisting of an Access Point (AP) with an n-element antenna array, and K single-antenna Mobile Stations (MS). At the AP, the received signal vector $\mathbf{r}$ is given by

$$\mathbf{r} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w} \ , \ \mathbf{H} = [\mathbf{h}_1, \cdots \mathbf{h}_K] \tag{1}$$

This work is partly supported by the NSF grants 1247583 and 1444962, and the Wisconsin Alumni Research Foundation.



(a) (b) (c) (d)

Fig. 1. a): Average fractional power capture of p strongest beams in 420-element arrays. (b): For one realization of $\mathbf{H}_b$, $K = 10$: mask used to generate the matrix. (c): Contour plot of $\sum_{i=1}^{K} \frac{|\mathbf{h}_{b,i}|^2}{\|\mathbf{h}_{b,i}\|^2}$ (d): Contour plot of $\sum_{i=1}^{K} |\mathbf{h}_{b,i}|^2$.

where ρ denotes transmit signal to noise ratio (SNR), $\mathbf{x}$ is the K-dimensional transmitted signal vector with the power constraint $E[|\mathbf{x}(i)|^2] = 1$, $\mathbf{H}$ is the $n \times K$ antenna domain channel matrix, $\mathbf{h}_k$ is the channel vector of the k^{th} user, and $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I})$ is complex additive white Gaussian noise.

A. Channel Model

The channel can be modeled as a superposition of multipath components. For a critically sampled n-dimensional ULA each multipath component is modeled using a steering vector

$$\mathbf{a}_n(\theta) = [e^{-j2\pi\theta i}]_{i\in\mathcal{I}(n)} \ , \ \theta = 0.5\sin(\phi) \tag{2}$$

where $\mathcal{I}(n) = \{\ell - (n-1)/2 : \ell = 0, 1, \cdots, n-1\}$ is a symmetric set of indices, $\phi \in [-\pi/2, \pi/2]$ is the angle of arrival of the multipath component, and θ is the spatial frequency induced by ϕ. Due to the highly directional propagation of mm-waves, multipath is expected to consist of a sparse set of single-bounce components 5-10 dB below the LoS path [2]. The k^{th} column of $\mathbf{H}$ can be modeled as

$$\mathbf{h}_k = \sum_{i=0}^{N_p} \beta_{k,i}\mathbf{a}_n(\theta_{k,i}) \tag{3}$$

where $\beta_{k,i}$ and $\theta_{k,i}$ represent the complex path gain and angle of arrival for the i^{th} multipath component of the k^{th} MS. In this paper, we focus on the LoS path, i.e. $\beta_{k,i} = 0$ for $i \neq 0$, for all k. We denote $\theta_{k,0} = \theta_k$ and $\beta_{k,0} = \beta_k$. We will model the LoS path magnitude as $|\beta_k| = \frac{\lambda}{4\pi\sqrt{R_k^2+h^2}}$, where λ is the operating wavelength, R_k is the distance of the k^{th} MS from the AP, and h is the AP height. This is the Friis transmission formula with isotropic radiators ($G_T = G_R = 1$). We model the phase $\angle\beta_k$ as uniformly distributed over $[-\pi, \pi]$.

We focus on a rectangular critically spaced uniform planar array (UPA) with dimension $n_{az} \times n_{el}$, $n = n_{az}n_{el}$. The 2D steering vector can be written as [8]

$$\mathbf{a}_n(\theta^{az}, \theta^{el}) = \mathbf{a}_{n_{az}}(\theta^{az}) \otimes \mathbf{a}_{n_{el}}(\theta^{el})$$

where $\otimes$ is the Kronecker product, θ^{az} and θ^{el} represent the azimuth and elevation spatial frequencies, corresponding to ϕ^{az} and ϕ^{el} with $-\pi/2 \leq \phi^{az} \leq \pi/2$ and $0 \leq \phi^{el} \leq \pi/2$. The channel matrix $\mathbf{H}$ is an $n \times K$ matrix, with the array indices arranged in Kronecker order. The k^{th} column of $\mathbf{H}$ representing the k^{th} user is modeled as

$$\mathbf{h}_k = \beta_k\mathbf{a}_n(\theta_k^{az}, \theta_k^{el}). \tag{4}$$

B. Beamspace MIMO Representation

The beamspace MIMO representation is obtained via the beamforming matrix, $\mathbf{U}_n$, whose columns form an orthonormal basis for the n-dimensional signal space. Each column of $\mathbf{U}_n$ is a steering/response vector at a different spatial frequency representing an orthogonal beam. For orthogonality, the spatial frequencies are spaced by $\Delta\theta = \frac{1}{n}$:

$$\mathbf{U}_n = \frac{1}{\sqrt{n}}\left[\mathbf{a}_n\left(i\Delta\theta\right)\right]_{i\in\mathcal{I}(n)} \ . \tag{5}$$

$\mathbf{U}_n$ is a unitary discrete Fourier transform (DFT) matrix: $\mathbf{U}_n^H\mathbf{U}_n = \mathbf{U}_n\mathbf{U}_n^H = \mathbf{I}$. The beamspace system equation is

$$\mathbf{r}_b = \mathbf{U}_n^H\mathbf{r} = \mathbf{H}_b\mathbf{x} + \mathbf{w}_b \ , \ \mathbf{H}_b = \mathbf{U}_n^H\mathbf{H} = [\mathbf{h}_{b,1}, \cdots, \mathbf{h}_{b,K}]$$

The statistics of $\mathbf{w}_b = \mathbf{U}_n^H\mathbf{w}$ are equivalent to the statistics of $\mathbf{w}$ because $\mathbf{U}_n$ is unitary. For a 2D UPA, the beamforming matrix is given by $\mathbf{U}_n = \mathbf{U}_{n_{az}} \otimes \mathbf{U}_{n_{el}}$ [8].

The elements of the n-dimensional beamspace channel vector for k^{th} user are given by:

$$\mathbf{h}_{b,k}(\ell, m) = \frac{\beta_k}{\sqrt{n}} f_{n_{az}}(\theta_k^{az} - \ell/n) f_{n_{el}}(\theta_k^{el} - m/n). \tag{6}$$

Each $\mathbf{h}_{b,k} = \beta_k\mathbf{U}_{n_{az}}^H\mathbf{a}_{n_{az}}(\theta_k^{az}) \otimes \mathbf{U}_{n_{el}}^H\mathbf{a}_{n_{el}}(\theta_k^{el})$ is the Kronecker product of two Dirichlet sinc functions (denoted by $f_n(\theta) = \sin(n\pi\theta)/\sin(\pi\theta)$) in azimuth and elevation. In a 1D ULA, $\mathbf{h}_{b,k}(i) = \beta_k f_n(\theta_k - i/n)/\sqrt{n}$, $i \in \mathcal{I}(n)$.

C. Power Leakage in Beam Sidelobes and SNR Loss

When the user is exactly at the center of the beam, all the power is concentrated in that beam. In general, the dirichlet sincs in (6) result in a user's power bleeding into beams in the neighborhood of the dominant beam. This results in SNR loss for each user as well as interference between different users. We define *average SNR loss* as $\delta = \mathbb{E}_\theta[\max_i |\mathbf{h}_{b,k}(i)|^2]/n$ where $\theta \sim \text{unif}[-1/2, 1/2]$, which measures the drop in SNR from the maximum gain (n), and hence the amount of power bleeding into adjacent beams. Fig. 1(a) plots the fraction of channel power captured by the p largest beams, $1 \leq p \leq$

TABLE I
AVERAGE POWER BLEEDING IN 4 STRONGEST BEAMS

	Dominant	2^{nd}	3^{rd}	4^{th}
420×1 Array	-1.11 dB	-8.89 dB	-15.49 dB	-17.25 dB
15×28 Array	-2.23 dB	-7.79 dB	-12.97 dB	-14.70 dB

20, and Table I lists the average loss in power for the four largest beams relative to the maximum. Observe that a 2D array exhibits more bleeding than a 1D array with the same number of elements.

III. BS Classifiers for a Small Cell Setting

A. Physical Array Model

In this section, we develop a signal model for a physical 2D UPA and the BS criterion. We describe the dominant beams with a 2D sparsity mask: $\mathcal{M} = \{(\ell_k, m_k) : \ell_k \in \mathcal{I}(n_{az}), m_k \in \mathcal{I}(n_{el}), k = 1, \cdots, K\}$. The indices in $\mathcal{M}$ represent the distinct orthogonal beams in which the K users are located (see Fig. 1(b)). Each index (ℓ_k, m_k) in the sparsity mask corresponds to a disjoint interval of spatial frequencies θ_k^{az} and θ_k^{el}, which we refer to as a *spatial bin*:

$$\theta_k^{az} \in (\ell_k \Delta\theta^{az} \pm \Delta\theta^{az}/2) \; ; \; \theta_k^{el} \in (m_k \Delta\theta^{el} \pm \Delta\theta^{el}/2)$$

where $\Delta\theta^{el} = 1/n_{el}$ and $\Delta\theta^{az} = 1/n_{az}$. The k^{th} user is associated with the beam (ℓ_k, m_k) corresponding with its spatial bin, defined by the array resolution. Furthermore, there is a one-to-one mapping between a spatial frequency pair $(\theta^{az}, \theta^{el})$ and a point on the ground described by the polar coordinates (R, ϕ^{az}). Therefore each frequency bin is associated with a corresponding area inside the small cell which we call a beam footprint (see Fig. 2(b)).

B. Small Cell Geometry

In this paper we focus on an AP with a small cell coverage area of $5\text{m} \le R \le 100\text{m}$, $-60^\circ \le \phi^{az} \le 60^\circ$. See Fig. 2(a) a for a visual depiction of the AP geometry. For simplicity we assume the AP is pointed parallel to the ground ($\psi = 0$). Refer to [8] for equations which generalize to nonzero ψ. With $\psi = 0$, $\eta = \phi^{az}$, and the $(\theta^{az}, \theta^{el}) \to (R, \phi^{az})$ mapping reduces to:

$$\theta^{az} = \frac{1}{2}\sin(\phi^{az}) \; ; \; \theta^{el} = \frac{1}{2}\frac{h}{\sqrt{(R\cos(\phi^{az}))^2 + h^2)}}. \quad (7)$$

Note that in our case $x = R\cos(\phi^{az})$ and θ^{el} depends only on x and h. Fig. 2(b) shows the beam footprints for our small cell and AP described above with $n_{az} = 15$, $n_{el} = 28$ ($n = 420$). Observe that only a subset of the beams actually map to the coverage area, whose indices we include in $\tilde{\mathcal{I}}(n_{az}, n_{el})$. We denote the quantity of these beams as $\tilde{n}$. For our cell, $\tilde{n} = 158$, and these beams are contained within the rectangle ($2 \le \ell \le 14$, $1 \le m \le 13$). This is advantageous because it reduces our search area for beam selection without compromising on beamforming gain.

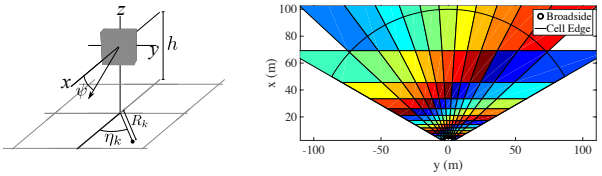


Fig. 2. (a) Coordinate geometry for an AP. (b) Beam footprint for our 15×28 small cell with an AP 10m high, $\psi = 0$.

C. Beam Selection Criterion and Sparsity Mask

The columns of $\mathbf{H}_b$ are sparse and most of the power in $\mathbf{h}_{b,k}$ is concentrated in a few dominant entries in the neighborhood of the k^{th} MS's LoS angle of arrival θ_k (see Fig. 1(c)-(d)). Let $\mathcal{SM}^{(p)}(\mathbf{x})$ be the set function that outputs the indices of the $p \ge 1$ components of $\mathbf{x}$ with largest power:

$$\mathcal{SM}^{(p)}(\mathbf{x}) = \{\{i_1, \cdots, i_p\} : i_\ell \in \mathcal{I}(n), \\ |x(i_1)|^2 \ge |x(i_2)|^2 \cdots \ge |x(i_p)|^2 \cdots \ge |x(i_n)|^2\} \quad (8)$$

We define the p-beam sparsity mask for the AP as

$$\mathcal{M}^{(p)} = \mathcal{SM}^{(p)}\left(\sqrt{\sum_{i=1}^{K} \frac{|\mathbf{h}_{b,i}|^2}{\|\mathbf{h}_{b,i}\|^2}} \circ \mathbb{1}_{\tilde{\mathcal{I}}(n_{az}, n_{el})}\right) \quad (9)$$

where $|\cdot|$, and $\sqrt{\cdot}$ are element-wise operations, $\circ$ is the Hadamard product, and $\mathbb{1}$ is the indicator function on the set of beams in the coverage area. The sparsity mask defines the number of beams simultaneously used by the AP. This selection criteria selects a fixed number of beams, which is a desirable property in physical systems where we have a fixed budget of RF chains. Larger fraction of the channel power is captured by a larger p (Fig. 1(a)), and by capturing a sufficiently large fraction, the resulting low-complexity system can deliver near-optimal performance [7], [8]. The amount of power captured depends on the beam leakage profile, discussed in Sec. II-C. As we will see, $p = 2K$ is sufficient for 1D arrays and 2D arrays. Typically $p \ll n$ and thus operating on $p = |\mathcal{M}|$ beams leads to significant complexity reduction.

When receive SNR varies across different MSs, the normalization term in (9) ensures that each MS is treated fairly, similar to the criterion developed in [7]. The normalization reduces the likelihood that an MS with low path loss will be allocated multiple beams before the high path loss MS is allocated its first and most critical beam (see Fig. 1(c)).

D. Ideal Model Analysis

In this section, we consider BS under an idealized channel model to develop intuition and to motivate the proposed BS criterion. We make a few assumptions. First, we assume that each $\mathbf{h}_{b,k}$ is 1-sparse; that is, all the power for each user is concentrated in a single beam. Second, we assume that different MS's occupy distinct beams. Finally, we assume that the number of users K is known and they are transmitting

simultaneously with equal power. With these assumptions, we can write the following simplified model

$$\mathbf{r}_b = \sqrt{\rho}\mathbf{\Phi}\mathbf{h}_b + \mathbf{w} \ , \ \mathbf{w} \sim \mathcal{CN}(0, \mathbf{I}) \tag{10}$$

where ρ represents SNR for each user, $\mathbf{h}_b = \sum_{k=1}^{K} \mathbf{h}_{b,k}$, and $\mathbf{\Phi} = \text{diag}(e^{j\varphi_1}, \cdots, e^{j\varphi_K})$ is a diagonal matrix of random phases. In this simple model, $\mathbf{h}_b$ is an n-dimensional K-sparse binary vector associated with a K-element mask $\mathcal{M}^{(K)}$:

$$\mathbf{h}_b \leftrightarrow \mathcal{M}^{(K)} \ ; \ \mathbf{h}_b(i) = \begin{cases} 1 \ , & i \in \mathcal{M}^{(K)} \\ 0 \ , & i \notin \mathcal{M}^{(K)} \end{cases} , \ i \in \mathcal{I}(n)$$

Let $\mathcal{S}_K$ denote the set of K-sparse binary vectors. We model $\mathbf{h}_b$ as uniformly distributed over such $\binom{n}{K}$ arrangements:

$$p(\mathbf{h}_b) = \begin{cases} \frac{1}{\binom{n}{K}} \ , & \mathbf{h}_b \in \mathcal{S}_K \\ 0 \ , & \text{otherwise} \end{cases} . \tag{11}$$

The sparsity mask classifier yields an estimate $\hat{\mathbf{h}}_b$ of $\mathbf{h}_b$ based on $|\mathbf{r}_b|^2$, the vector of element-wise magnitude-squared entries of $\mathbf{r}_b$. We choose $\hat{\mathbf{h}}_b$ to minimize the probability of error, P_e, defined as the event when there is at least one error in $\hat{\mathbf{h}}_b$:

$$P_e = P(E) \ ; \ E = \{\hat{\mathbf{h}}_b \neq \mathbf{h}_b\} = \bigcup_{i=1}^{n} E_i \tag{12}$$

where $E_i = \{\hat{\mathbf{h}}_b(i) \neq \mathbf{h}_b(i)\}$ represents an individual error associated with the i-th beam. The minimum probability of error classifier is the maximum likelihood classifier:

$$\hat{\mathbf{h}}_{b,ML} = \underset{\mathbf{h}_b \in \mathcal{S}_K}{\text{argmax}} f(|\mathbf{r}_b|^2 | \mathbf{h}_b) \ . \tag{13}$$

Because $\mathbf{w}$ is i.i.d., we can write the density as a product of the densities for each $\mathbf{r}_b(i)$. Introducing the scaling factor of two onto $|\mathbf{r}_b|^2$ aids our analysis because $f(2|\mathbf{r}_b(i)|^2 | \ \mathbf{h}_b(i) = 0)$ is chi-squared [14].

$$\begin{aligned} f(2|\mathbf{r}_b|^2 | \ \mathbf{h}_b) &= \prod_{i \in \mathcal{I}: \mathbf{h}_b(i)=0} f_0(2|\mathbf{r}_b(i)|^2) \prod_{i \in \mathcal{I}: \mathbf{h}_b(i)=1} f_1(2|\mathbf{r}_b(i)|^2) \\ &= \prod_{i \in \mathcal{I}} f_0(2|\mathbf{r}_b(i)|^2) \prod_{i \in \mathcal{I}: \mathbf{h}_b(i)=1} \frac{f_1(2|\mathbf{r}_b(i)|^2)}{f_0(2|\mathbf{r}_b(i)|^2)} \end{aligned}$$

where $f_j(2|\mathbf{r}_b(i)|^2) = f(2|\mathbf{r}_b(i)|^2 | \mathbf{h}_b(i) = j) \ , \ j = \{0, 1\}$. $f_0(\cdot) \sim \chi_2$ is chi-squared with two degrees of freedom (DoF) and $f_1 \sim \chi_{2,2\rho}$ is non-central chi-squared with two DoF and noncentrality parameter 2ρ. We plug $f(2|\mathbf{r}_b|^2 | \ \mathbf{h}_b)$ into (13), omitting the first product since it has no dependence on $\mathbf{h}_b$:

$$\hat{\mathbf{h}}_{b,ML} = \underset{\mathbf{h}_b \in \mathcal{S}_K}{\text{argmax}} \prod_{i \in \mathcal{I}: \mathbf{h}_b(i)=1} \frac{f_1(2|\mathbf{r}_b(i)|^2)}{f_0(2|\mathbf{r}_b(i)|^2)} \ . \tag{14}$$

The ratio of $f_1(\cdot)$ to $f_0(\cdot)$ is the likelihood ratio of a noncentral chi-squared to central chi-squared, and is strictly increasing with $|\mathbf{r}_b(i)|^2$. Since our sparsity constraint dictates that we can only choose K elements of $\mathbf{h}_b$ to be one, the argument is maximized by selecting these at the elements with the largest magnitudes. We have proven that when $\mathbf{h}_{b,k}$ vectors are 1-sparse and in distinct beams, the largest magnitude classifier is optimal. We will apply this classifier to the general class of $\mathbf{h}_{b,k}$ vectors, which are difficult to evaluate analytically:

$$\hat{\mathbf{h}}_b \leftrightarrow \hat{\mathcal{M}} \ ; \ \hat{\mathcal{M}} = \mathcal{SM}^{(K)}(\mathbf{r}_b) \tag{15}$$

where $\mathcal{SM}^{(K)}(\cdot)$ is defined in (8). We use the normalized version in (9) to approximately achieve equal receive power and to treat users fairly, and we only search beams in the coverage area.

IV. Numerical Results

A. Simulation Setup

Simulations are run over 100,000 channel realizations (10,000 for the full-dimensional system). For each trial, we distribute users as follows. Each orthogonal beam is associated with a θ-interval (spatial bin) and (x, y)-interval (beam footprint) where it is the highest power beam. First, we uniformly select a subset of K beams to distribute our users within, as shown in Fig. 1(a), ensuring that no two users are in the same frequency bin. Next, for each user, we distribute each $(\theta^{az}, \theta^{el})$ uniformly within its bin, with the constraint that the (x, y) coordinates fall within the coverage area.

Our noisy estimate of $\mathbf{H}_b$ is modeled as:

$$\hat{\mathbf{H}}_b = \sqrt{\rho}\mathbf{H}_b + \mathbf{W} \ , \ \mathbf{W}_{i,j} \sim \mathcal{CN}(0, 1) \tag{16}$$

where $\mathbf{W}_{i,j}$ are independent, ρ represents transmit SNR, and the columns of $\mathbf{H}_b$ are generated according to (6). To compute the path loss coefficients, we use $\lambda = 10.7\text{mm}$ (28 GHz frequency). In our cell, $-75.4 \text{ dB} \leq |\beta_k|^2 \leq -101.4 \text{ dB}$. Next, we apply our BS classifier from (9) to both $\hat{\mathbf{H}}_b$ and $\mathbf{H}_b$ to obtain our noisy BS estimate as well as the ground truth:

$$\hat{\mathcal{M}}^{(p)} = \mathcal{SM}^{(p)}(\hat{\mathbf{H}}_b) \ , \ \mathcal{M}^{(p)} = \mathcal{SM}^{(p)}(\mathbf{H}_b).$$

We use the sparsity mask to reduce our system to the near-optimal, low-dimensional communication subspace, described by the beamspace channel matrix $\tilde{\mathbf{H}}_b$. We have four variants of $\tilde{\mathbf{H}}_b$, originating from the permutations of: perfect channel state information (CSI), or noisy observations in the BS and channel estimation (CE) phases:

$$\begin{aligned} &\tilde{\mathbf{H}}_b = [\mathbf{h}_{b,i}]_{i \in \mathcal{M}^{(p)}} \ , \ \tilde{\mathbf{H}}_b^{(BS)} = [\mathbf{h}_{b,i}]_{i \in \hat{\mathcal{M}}^{(p)}} \\ &\tilde{\mathbf{H}}_b^{(CE)} = [\hat{\mathbf{h}}_{b,i}]_{i \in \mathcal{M}^{(p)}} \ , \ \tilde{\mathbf{H}}_b^{(BS,CE)} = [\hat{\mathbf{h}}_{b,i}]_{i \in \hat{\mathcal{M}}^{(p)}}. \end{aligned} \tag{17}$$

B. Performance Metrics

We will assess the performance of the beam selection algorithm by measuring the sum-rate capacity:

$$C(\rho | \mathbf{H}_b) = \sum_{i=1}^{K} \log_2(1 + \text{SINR}_i(\mathbf{H}_b)) \text{ b/s/Hz} \tag{18}$$

$$\text{SINR}_i(\mathbf{H}_b) = \frac{\rho \left| \mathbf{f}_i^H \mathbf{h}_{b,i} \right|^2}{\sum\limits_{i,j \in \mathcal{M}, i \neq j} \rho \left| \mathbf{f}_i^H \mathbf{h}_{b,j} \right|^2 + \|\mathbf{f}_i\|^2} \tag{19}$$

where $\mathbf{f}_i$ and $\mathbf{h}_{b,i}$ denote the i^{th} columns of the spatial equalization matrix $\mathbf{F}$ and channel matrix $\mathbf{H}_b$, respectively. Recall the noise is unit variance, and so $\rho \|\mathbf{h}_{b,k}\|^2$ reflects the

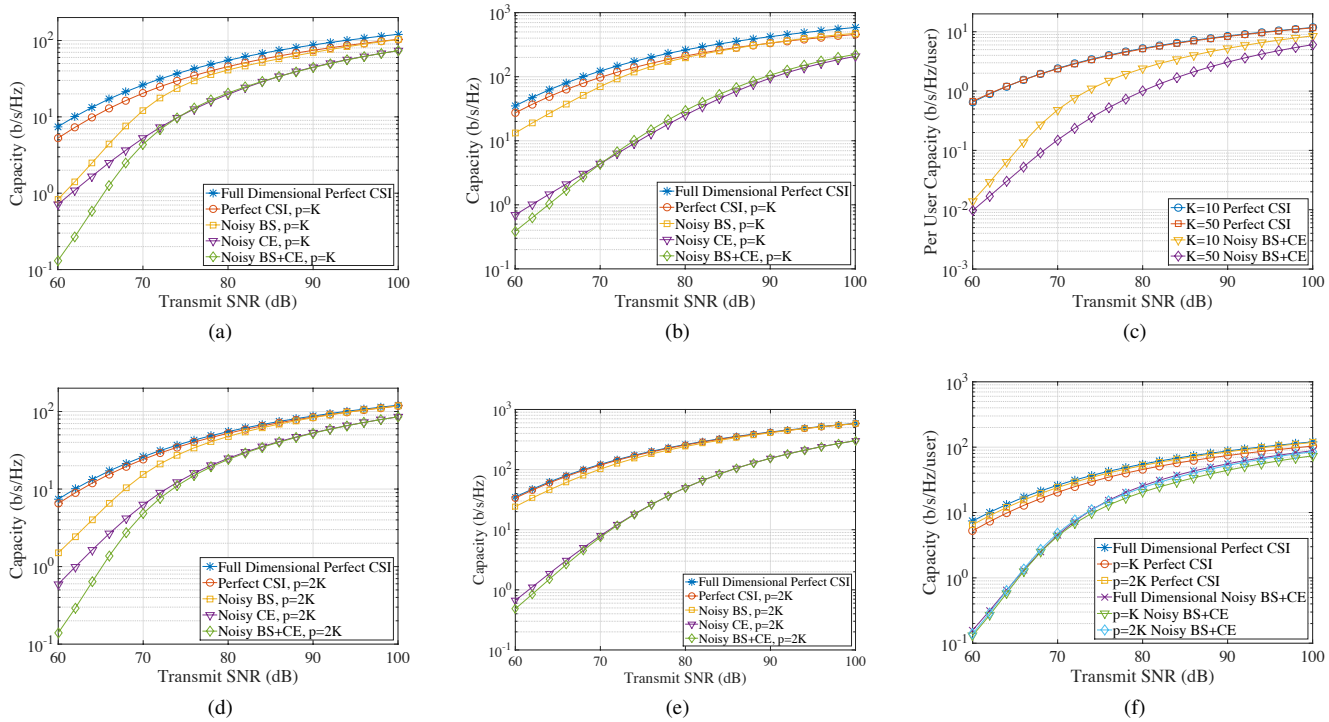


Fig. 3. Simulation Results. (a): Capacity plots in bits/s/Hz for $p = K = 10$. (b): Capacity for $p = K = 50$. (c) Per user capacity (b/s/Hz/user) for $K = 10, 50,\ p = 2K$ with perfect CSI and noisy estimates. (d) Capacity plots for $K = 10$, $p = 2K$. (e) Capacity plots for $K = 50$, $p = 2K$. (f) Capacity plots of the full-dimensional system vs $p = K$ and $p = 2K$, K=10, for perfect CSI and noisy BS and CE phases.

SNR of each user. For our simulations, we use the MMSE filter for $\mathbf{F}$:

$$\mathbf{F}(\tilde{\mathbf{H}}_b) = \tilde{\mathbf{H}}_b^H(\rho\tilde{\mathbf{H}}_b\tilde{\mathbf{H}}_b^H + \mathbf{I}). \tag{20}$$

We provide one of our four variants of $\tilde{\mathbf{H}}_b$ as input to the capacity formula to simulate the system operating with either perfect CSI or with a noisy estimate. We also simulate the full dimensional system; that is, the system which uses all $\tilde{n}$ beams which fall inside the coverage area, rather than all $p \leq \tilde{n}$ beams .

C. Simulation Results

Fig. 3(a)-(b),(d)-(e) plot the sum-rate capacity across a range of SNR values for different values of K and p. Each figure displays the results for the full-dimensional $\tilde{n} \times K$ system, as well as for $\tilde{\mathbf{H}}_b$, $\tilde{\mathbf{H}}_b^{(BS)}$, $\tilde{\mathbf{H}}_b^{(CE)}$, and $\tilde{\mathbf{H}}_b^{(BS,CE)}$. Results show that in the presence of noise, BS converges much more quickly than CE. Increasing number of beams from $p = K$ to $p = 2K$ yields a significant performance advantage, but adding more beams yields diminishing returns.

Fig. 3(c) plots the per user capacity for $K = 10$ and $K = 50$. The results show that with perfect channel estimates, sum-rate capacity scales linearly with K. However, with noisy CSI there is some loss in per user capacity with increasing K.

Fig. 3(f) plots the capacity curves for the full dimensional system and low complexity systems with $p = K$, and $p = 2K$. The results show that with just two beams per user (20 beams), we are able to achieve nearly the same capacity as the full dimensional system, which uses 158 beams. The capacity for the low-complexity, noisy CSI system are also nearly as high as the full dimensional, noisy CSI system.

V. Mobility and Dynamics

One key concern when designing the BS and channel estimation protocols is the overhead they will consume in a communications data frame. We introduce the concept of beam coherence time, which, analogous to channel coherence time, captures the timescale of variation in the sparsity mask.

A. Beam Selection Overhead

1) Beam Mask Coherence Time: We express the beam coherence time as the time taken by an MS crossing from one beam footprint to the next. Thus, we need to re-estimate our sparsity mask at the same rate as an MS traveling at the maximum supported velocity transitions from beam to beam. So, the beam coherence time is expressed as

$$T_{b,coh} = \frac{d_{min}}{v_{max}} \tag{21}$$

where d_{min} is the minimum distance across any beam, and v_{max} is the maximum supported MS speed. In general, we can calculate the widths of each beam footprint in the x-dimension and y-dimension by mapping the θ values demarcating the boundary between footprints into the geometric coordinates (x, y). Then we can find the minimum width. However, for square or nearly square arrays, the footprints are usually narrower in the azimuth direction (see Fig. 2(b)).

For an AP with $\psi = 0$, the narrowest beam in the azimuth dimension is the broadside beam. By the small angle approximation, we can approximate this width as $\frac{2R_{min}}{n_{az}}$. For now, we will assume the footprints are narrower in y-direction, and say $d_{min} = \frac{2R_{min}}{n_{az}}$. Therefore, we have

$$T_{b,coh} \approx \frac{2R_{min}}{n_{az} v_{max}} \tag{22}$$

2) BS Estimation Time: Suppose we transmit a beam selection training sequence of length N_{bs}. Thus, the training time required by the BS protocol is $T_{bs} = N_{bs}T$, where $T = \frac{1}{W}$ is the symbol period and W is the bandwidth. The length of the training sequence, N_{bs} controls the tradeoff between BS accuracy and length. We assume a hardware structure similar to [5] where our RF chain budget allows us to sample p beams simultaneously, and switching beams takes T_{sw} seconds.

3) Brute force BS Overhead: Suppose that we employ a brute force training sequence where each MS takes turns transmitting constant modulus symbols for as long as it takes to collect N_s samples at each beam. N_s is an integer that tunes the accuracy-overhead tradeoff. We can obtain more accurate estimates at the cost of longer training sequences. Using this protocol,

$$N_{bs} = K(N_s + N_{sw}) \cdot \lceil \tilde{n}/p \rceil \tag{23}$$

where $N_{sw} = \lceil \frac{T_{sw}}{T} \rceil$ is the number of symbols periods required to switch.

The BS overhead α can be calculated as $\alpha = T_{bs}/T_{b,coh}$. Plugging in the N_{bs} value from the brute force training sequence, we obtain the overhead expression:

$$\alpha = T \frac{K(N_s + N_{sw}) n_{az} v_{max}}{2R_{min}} \cdot \lceil \tilde{n}/p \rceil \tag{24}$$

B. Channel Estimation overhead

Overhead depends on the complexity induced by the BS procedure ($|\mathcal{M}|$), the Doppler spread, and the delay spread. The coherence time, T_{coh} is the time period over which the channel is correlated, and the maximum time interval between channel estimation frames. The coherence time is approximated by [15]:

$$T_{coh} \approx \frac{1}{\Delta f} = \frac{c}{f_c v_{max}},$$

where Δf is the doppler spread, c is the speed of light, f_c is the carrier frequency, and v_{max} is the maximum velocity of a user relative to the AP. Assuming the BS classifier does not exceed our RF budget, the time required by the channel estimation procedure is $T_{ce} = K\tau$. This allows us to estimate the temporal channel for each pairwise combination of transmit and receive beam. The channel estimation overhead, β, is the fraction of the channel coherence time needed to obtain the channel estimate.

$$\beta = \frac{T_{ce}}{T_{coh}} = \frac{K\tau f_c v_{max}}{c}$$

In the Table II, we list BS and CE overhead values α and β as a function of different parameters. Here, we fix $r_{min} = 5\text{m}$, $r_{max} = 100\text{m}$, $v_{max} = 20$ m/s, $N_s = 8$, $T_{sw} = 20$ ns, $T = 1$ ns, $f_c = 28$ GHz, and $\tau = 100$ ns. Although the BS overhead equation is sensitive to dimension, even quite large array dimensions yield small overheads due to beam mask coherence times on the order of a millisecond.

TABLE II
BS AND CE OVERHEAD VALUES

n_{az}	n_{el}	p	K	$\tilde{n}$	α	β
15	28	20	10	158	6.72×10^{-5}	9.33×10^{-4}
15	28	100	50	158	8.40×10^{-5}	0.0047
26	26	100	100	270	4.37×10^{-4}	0.0187
40	40	200	200	612	0.0018	0.0373
60	60	400	400	1296	0.0054	0.0747

VI. CONCLUSION

We have proposed a simple low-complexity approach to solving the important problem of BS in sparse high-dimensional MIMO systems. Motivated by a power-based criterion, we propose a simple classifier for BS. We use an idealized model for a multiuser MIMO to show its optimality, and extend it to a physical model based on 2D UPAs and small cells. Numerical results illustrate the promising performance of the classifiers, and mobility discussions show that BS estimation occupies minimal overhead even for large arrays.

REFERENCES

[1] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Comm. Mag*, vol. 46, no. 9, pp. 59–67, Sep. 2008.

[2] Rappaport, T.S. et al., "Millimeter Wave Mobile Communications for 5g Cellular: It Will Work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[3] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Comm. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[4] A. Sayeed and N. Behdad, "Continuous aperture phased MIMO: Basic theory and applications," in *2010 Proc. Allerton Conference*, Sep. 2010, pp. 1196–1203.

[5] J. Brady, N. Behdad, and A. 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.

[6] Rusek, F. et al., "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Sig. Proc. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[7] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *2013 IEEE GLOBECOM*, Dec. 2013, pp. 3679–3684.

[8] J. Brady and A. Sayeed, "Beamspace MU-MIMO for high-density gigabit small cell access at millimeter-wave frequencies," in *2014 IEEE (SPAWC)*, Jun. 2014, pp. 80–84.

[9] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

[10] P. Amadori and C. Masouros, "Low RF-Complexity Millimeter-Wave Beamspace-MIMO Systems by Beam Selection," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2212–2223, Jun. 2015.

[11] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Trans. on Wireless Comm.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[12] Song Noh, Michael Zoltowski, David Love, "Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment," *Purdue Dept. ECE Technical Reports*, 2015.

[13] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Sig. Proc.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[14] J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, Jun. 2006.

[15] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2001.