Preprocessor Design for Hybrid Preprocessing with Selection in Massive MISO Systems

Vishnu V. Ratnam*, Ozgun Y. Bursalioglu†, Haralabos C. Papadopoulos† and Andreas F. Molisch*
*Ming Hsieh Dept. of Electrical Engineering, University of Southern California
Los Angeles, California, USA. Email: {ratnam, molisch}@usc.edu
†Docomo Innovations, Palo Alto, CA, 94304 USA.
Email: {obursalioglu, hpapadopoulos}@docomoinnovations.com

***Abstract*—Hybrid preprocessing, where an analogue preprocessing matrix is used to feed a large antenna array to fewer up/down-conversion chains, helps reduce hardware cost of massive Multiple-Input-Multiple-Output systems. Here, we analyze a variant of hybrid preprocessing, namely hybrid preprocessing with selection (HPwS), as an attractive solution to reduce this hardware cost while retaining good performance. In HPwS, the preprocessing matrix, built from radio frequency (RF) hardware, has a larger number of input ports L than the number of up/down-conversion chains K. A bank of RF switches connects the instantaneously best K input ports to the up/down-conversion chains. The preprocessor is designed based on average channel statistics and therefore needs to be updated only infrequently. This allows for a higher diversity-order and/or simpler RF hardware than some conventional hybrid preprocessing systems. In this paper, we propose a generic architecture of HPwS with a possibly non-unitary rectangular preprocessing matrix. A novel preprocessor is designed that maximizes a capacity lower bound for channels with isotropic scattering. In addition, we study how L, the number of users, and the use of low complexity RF hardware, such as discrete phase-shifters, impact system performance. We also present a method to extend the preprocessor design to anisotropic channels.**

I. Introduction

Multiple-input-multiple-output (MIMO) systems are viewed as a key enabler towards meeting the rising throughput demands in cellular systems, due to their ability to boost spectral efficiency by increasing the spatial degrees of freedom and/or providing beamforming gains. Massive MIMO (enabled by using a massive antenna array at the base-station) takes these benefits to the extreme, while requiring only simplified transmission methods [1]. Therefore, massive MIMO is one of the main areas of focus for 5G and millimeter (mm) wave systems [2]. However, this general trend towards larger number of antennas is leading to an increase in the hardware cost of MIMO transceivers. Though the antenna elements are themselves cheap, the associated up/down-conversion chains having analog-to-digital converters, mixers and filters are expensive and energy consuming, especially for the wide bandwidths encountered at mm-wave frequencies. In light of this fact, hybrid preprocessing was proposed in [3], [4] and further investigated for mm-wave frequencies in [5]–[7], wherein the N antenna elements are connected to K up/down-conversion chains, with $K \ll N$. This is achieved via a radio frequency (RF) preprocessing matrix built from RF hardware such as RF phase shifters and RF amplifiers, that connects the K up/down-conversion chains to the N antenna elements.

This work is partially supported by the National Science Foundation.

In conventional single-user hybrid preprocessing, the RF preprocessing matrix forms beams into the channel subspace spanned by the singular vectors corresponding to the K largest singular values of either the channel matrix or the channel spatial correlation matrix. The former approach, referred to as hybrid preprocessing based on instantaneous channel state information (HPiCSI) [3], [5], leads to the highest possible signal-to-noise ratio (SNR) but requires instantaneous channel state information (iCSI) across the whole channel dimension. Since typically $N \gg K$, this imposes a significant channel estimation overhead [8]. Further, the RF preprocessing matrix needs to be updated frequently (in each coherence time interval) which puts a very stringent constraint on the design of the RF hardware. In the latter approach, referred to as hybrid preprocessing based on average channel state information (HPaCSI) [4], [6], the preprocessing matrix needs to be updated infrequently since the spatial correlation matrix changes relatively slowly and can be learned with significantly smaller overhead. Furthermore, iCSI is required only in the channel subspace corresponding to the K largest singular values of the spatial correlation matrix, thereby significantly reducing the channel estimation overhead. Despite these advantages, HPaCSI leads to a low SNR since the preprocessor beams do not adapt to iCSI.

Even with a preprocessor designed by the spatial correlation matrix, it is possible to adapt the beams to iCSI through use of additional RF hardware and selection techniques [4]. In this work we explore a generalization of this approach, which we call hybrid preprocessing with selection (HPwS), as an attractive alternate hybrid preprocessing technique, to achieve performance comparable to HPiCSI while retaining the benefits of HPaCSI. The block diagram of a general HPwS system implemented at the transmitter is depicted in Fig. 1. Here, the RF preprocessing matrix is again designed based on average channel statistics like the spatial correlation matrix, but it has L ports, with $L > K$. A bank of RF switches is used to connect K out-of-the L ports to the up/down-conversion chains based on iCSI. The premise for this design is that, unlike other RF hardware, RF switches can be easily designed to switch quickly within a coherence time interval. Since the dimension of the preprocessing matrix is $L > K$ and switches adapt the effective beams to iCSI, a higher diversity order and beamforming gain can be achieved in comparison to HPaCSI, thereby, bridging the performance gap between HPaCSI and HPiCSI. On the downside, depending on the design of the preprocessing matrix, the channel estimation overhead for HPwS may be higher than for HPaCSI. However, an analysis of the channel estimation overhead is beyond the scope of this paper. See [8], [9] for some preliminary results. Also, since the preprocessor has size $N \times L$, as opposed to $N \times K$ for HPaCSI, a larger number of RF components are required. Varying the number of ports L, thereby allows a trade-off between performance and hardware cost.

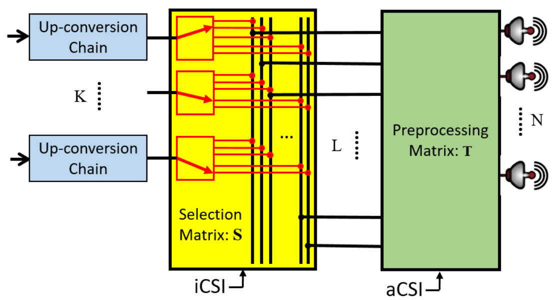


Fig. 1. Block diagram of a transmitter with HPwS.[1]

Several forms of HPwS have been explored in the literature, e.g., antenna selection [9], [10] and beam selection [3], where the RF preprocessor is either absent or fixed. An exception is [4], where a preprocessor that adapts to the average channel statistics was proposed. However in most of the prior work, the preprocessing matrix is unitary, where the number of ports (L) equals the number of transmit antennas (N) and a single user scenario is considered. These designs are capacity sub-optimal, especially in spatially sparse channels. In this paper, we find more a generic design for the preprocessor, that provides additional beamforming gains. Furthermore, due to superior beam-shaping capabilities, it is expected that this improved preprocessor design will yield higher gains in multi-user systems than in systems with a single user. As a result, although we optimize the preprocessor for a single-user scenario here, we also present a preliminary evaluation of its performance in a multi-user setting via simulations.

The contributions of this work are as follows:

1) We propose a generic architecture of HPwS for low complexity Multiple-input-single-output (MISO) transceivers, wherein, the preprocessing matrix may be a non-unitary rectangular matrix i.e., $L \neq N$.
2) We formulate the problem of finding the ergodic capacity maximizing preprocessor design for channels with isotropic scattering and show that it is an optimization problem over the Grassmannian manifold.
3) A lower bound on the ergodic capacity is derived and shown to be maximized by Grassmannian line-packing the columns of the preprocessor.
4) We study the impact of low complexity RF hardware like discrete phase shifters on the system performance.
5) We study the performance of designed preprocessor in a multi-user scenario.
6) We present a method to extend the preprocessor design to more generic channels with non-isotropic scattering.

[1]For $L = K$, this reduces to the conventional hybrid scheme - HPaCSI.

The organization of this paper is as follows: The general assumptions and the channel model are discussed in Section II. The expression for the system signal-to-noise ratio is derived and the problem for finding the optimal preprocessing matrix is formulated in Section III. The preprocessor design problem is solved in Section IV. The simulation results for impact of number of columns L and impact of reduced complexity RF hardware and performance in multi-user setting are presented in Section V. The extension of the design to anisotropic channels is discussed in Section VI and the conclusions are summarized in Section VII.

Notation used in this work is as follows: scalars are represented by light-case letters; vectors by bold-case letters; matrices by capitalized bold-case letters and sets, subspaces are represented by calligraphic letters. Additionally, $\mathbf{a}_i$ represents the i-th element of a vector $\mathbf{a}$, $|\mathbf{a}|$ represents the L_2 norm of a vector $\mathbf{a}$, $\mathbf{A}_{i,j}$ represents the (i,j)-th element of a matrix $\mathbf{A}$, $[\mathbf{A}]_{c\{i\}}$ and $[\mathbf{A}]_{r\{i\}}$ represent the i-th column and row vectors of matrix $\mathbf{A}$ respectively, $\mathbf{A}^\dagger$ is the conjugate transpose of a matrix $\mathbf{A}$ and $|\mathcal{A}|$ represents the cardinality of a set $\mathcal{A}$ or dimension of a vector space $\mathcal{A}$. Also, $\overset{d}{=}$ is equivalence in distribution, $\mathbb{E}(\cdot)$ represents the expectation operator, $\mathbb{P}(\cdot)$ is the probability operator, $\mathbb{I}_i$ and $\mathbb{O}_{i,j}$ are the $i \times i$ and $i \times j$ identity and zero matrices respectively, and $\mathbb{R}$ and $\mathbb{C}$ represent the fields of real and complex numbers, respectively.

II. General Assumptions and Channel Model

We consider a single transmitter-receiver pair where the transmitter has N antennas and the receiver has a single antenna. The transmitter has $K < N$ up-conversion chains and implements HPwS where the preprocessing matrix has $L > K$ ports.[2] The channel has a transfer function that can be multiplicatively decomposed into path-loss, a shadow fading component and a small-scale fading component. Furthermore, the channel is assumed to have Rayleigh amplitude statistics and frequency flat and block fading i.e., the channel remains constant for a coherence time interval and then changes to a different realization. Under these assumptions, the received signal during any coherence time interval can be expressed as:

$$y = \sqrt{\rho}\mathbf{h}\mathbf{T}\mathbf{S}\mathbf{G}\mathbf{u} + n \tag{1}$$

where ρ is the average channel SNR[3], $\mathbf{h} \sim \mathcal{CN}(\mathbb{O}_{1\times N}, \mathbf{R}_{\text{tx}})$ is the $1 \times N$ small-scale fading channel matrix, $\mathbf{T}$ is the $N \times L$ RF-preprocessing matrix implemented using RF hardware, $\mathbf{S}$ is a $L \times K$ sub-matrix of the identity matrix $\mathbb{I}_L$ that represents the switch positions, $\mathbf{Gu}$ is the $K \times 1$ data vector from the up-conversion chains, where $\mathbf{G}$ is a $K \times K$ matrix that orthonormalizes columns of $\mathbf{TS}$, i.e., $\mathbf{G}^\dagger\mathbf{S}^\dagger\mathbf{T}^\dagger\mathbf{TSG} = \mathbb{I}_K$ and

[2]The presented results can also be easily extended to a single-input-multiple-output scenario having multiple antennas and HPwS at the receiver.

[3]The pathloss, shadow fading component and the transmit power term are assumed to be included in ρ.

$n \sim \mathcal{CN}(0,1)$ is the normalized additive white Gaussian noise. Here the selection matrix $\mathbf{S}$ is composed of those K columns of $\mathbb{I}_L$ that correspond to the K input ports of $\mathbf{T}$ that are connected to the up-conversion chains via the switch bank. Let us define the set of all selection matrices as $\mathcal{S} = \{\mathbf{S}_1, ..., \mathbf{S}_{|\mathcal{S}|}\}$. This set, referred to as the switch position set, depends on the hardware implementation of the switch bank. For any selection matrix $\mathbf{S}_i \in \mathcal{S}$, the transmit power constraint is given by:

$$\text{tr}\{\mathbf{T}\mathbf{S}_i\mathbf{G}_i\mathbb{E}_u\{\mathbf{u}\mathbf{u}^\dagger\}\mathbf{G}_i^\dagger\mathbf{S}_i^\dagger\mathbf{T}^\dagger\} \leq 1$$
$$\Rightarrow \mathbb{E}_u\{\mathbf{u}\mathbf{u}^\dagger\} \leq 1 \quad (2)$$

where $\mathbf{G}_i$ orthonormalizes the columns of $\mathbf{T}\mathbf{S}_i$, i.e., $\mathbf{G}_i^\dagger\mathbf{S}_i^\dagger\mathbf{T}^\dagger\mathbf{T}\mathbf{S}_i\mathbf{G}_i = \mathbb{I}_K$. Though $\mathbf{G}_i$ is also a function of $\mathbf{T}$, this dependence is not shown explicitly for ease of representation. We assume the transmitter has knowledge of iCSI. The selection matrix $\mathbf{S} \in \mathcal{S}$ is chosen based on knowledge of iCSI but $\mathbf{T}$ is designed only based on the knowledge of the spatial correlation matrix $\mathbf{R}_{\text{tx}} = \mathbb{E}_{\mathbf{h}}\{\mathbf{h}^\dagger\mathbf{h}\}$.

In this work we consider both the cases of $L \leq \text{rank}\{\mathbf{R}_{\text{tx}}\}$ and $L > \text{rank}\{\mathbf{R}_{\text{tx}}\}$. While having $L > \text{rank}\{\mathbf{R}_{\text{tx}}\}$ does not provide any additional diversity benefit, it allows selection from a larger range of transmit beams, thereby providing additional beamforming gains. These additional beam choices are also expected to aid user separation in a multi-user scenario, as explored via simulations in Section V. For convenience, we first consider the case of $\mathbf{R}_{\text{tx}} = \mathbb{I}_N$. Extension to more generic channels is considered later in Section VI. As $\mathbf{R}_{\text{tx}} = \mathbb{I}_N$, $\mathbf{T}$ is a fixed matrix for majority of this paper, except Section VI.

III. Problem Formulation

From (1)–(2) and assuming the use of the capacity optimal maximal ratio transmission [11], the instantaneous receiver SNR and the mean capacity can then be expressed as:

$$\gamma(\mathbf{T},\mathbf{h}) = \max_{1\leq i\leq|\mathcal{S}|}\left[\rho\mathbf{h}\mathbf{T}\mathbf{S}_i\mathbf{G}_i\mathbf{G}_i^\dagger\mathbf{S}_i^\dagger\mathbf{T}^\dagger\mathbf{h}^\dagger\right] \quad (3)$$
$$C(\mathbf{T}) = \mathbb{E}_{\mathbf{h}}\{\log(1+\gamma(\mathbf{T},\mathbf{h}))\}. \quad (4)$$

The main aim of this paper is to find the preprocessing matrix design, that maximizes the mean capacity, i.e. to find:

$$\mathbf{T}_{\text{opt}} = \text{argmax}_{\mathbf{T}\in\mathbb{C}^{N\times L}}\{C(\mathbf{T})\}. \quad (5)$$

Here, with slight abuse of notation, by argmax{} we refer to any one (of the possibly many) maximizing arguments. Note that the search space in (5) is $\mathbb{C}^{N\times L}$ - the set of all $N\times L$ complex matrices, which is unbounded. However, as shall be shown in the next theorem, many of these matrices are equivalent and therefore the search space can be reduced to a bounded region.

Theorem 1 (Bounding the search space)**.** *For any* $\mathbf{T} \in \mathbb{C}^{N\times L}$*, both* $\mathbf{T}$ *and* $\mathbf{T}\mathbf{\Lambda}_\theta$ *attain the same mean capacity* (4)*, where* $\mathbf{\Lambda}_\theta$ *is any arbitrary* $L\times L$ *complex diagonal matrix.*

Proof. Let $\mathbf{T}$ and $\mathbf{T}_\theta = \mathbf{T}\mathbf{\Lambda}_\theta$ be two $N\times L$ preprocessing matrices where $\mathbf{\Lambda}_\theta$ is any non-singular $L\times L$ complex diagonal matrix. For any selection matrix $\mathbf{S}_i$, let $\mathbf{G}_i$ be the orthonormalization matrix for $\mathbf{T}\mathbf{S}_i$. Then by defining a corresponding matrix $\mathbf{G}_{i\theta} \triangleq \mathbf{S}_i^\dagger\mathbf{\Lambda}_\theta^{-1}\mathbf{S}_i\mathbf{G}_i$, we have:

$$\mathbf{G}_{i\theta}^\dagger\mathbf{S}_i^\dagger\mathbf{T}_\theta^\dagger\mathbf{T}_\theta\mathbf{S}_i\mathbf{G}_{i\theta}$$
$$= \mathbf{G}_i^\dagger\mathbf{S}_i^\dagger[\mathbf{\Lambda}_\theta^{-1}]^\dagger\mathbf{S}_i\mathbf{S}_i^\dagger\mathbf{\Lambda}_\theta^\dagger\mathbf{T}^\dagger\mathbf{T}\mathbf{\Lambda}_\theta\mathbf{S}_i\mathbf{S}_i^\dagger\mathbf{\Lambda}_\theta^{-1}\mathbf{S}_i\mathbf{G}_i \quad (6a)$$
$$= \mathbf{G}_i^\dagger\mathbf{S}_i^\dagger\mathbf{S}_i\mathbf{S}_i^\dagger\mathbf{T}^\dagger\mathbf{T}\mathbf{S}_i\mathbf{S}_i^\dagger\mathbf{S}_i\mathbf{G}_i \quad (6b)$$
$$= \mathbf{G}_i^\dagger\mathbf{S}_i^\dagger\mathbf{T}^\dagger\mathbf{T}\mathbf{S}_i\mathbf{G}_i \quad (6c)$$
$$= \mathbb{I}_K$$

where (6b) follows from the fact that $\mathbf{S}_i\mathbf{S}_i^\dagger$ is diagonal and hence commutes with $\mathbf{\Lambda}_\theta$ and (6c) uses the fact that $\mathbf{S}_i^\dagger\mathbf{S}_i = \mathbb{I}_K$. This proves that $\mathbf{G}_{i\theta}$ ortho-normalizes $\mathbf{T}_\theta\mathbf{S}_i$. Therefore from (3), following similar steps as above, we have:

$$\gamma(\mathbf{T}_\theta,\mathbf{h}) = \max_{1\leq i\leq|\mathcal{S}|}\left[\rho\mathbf{h}\mathbf{T}_\theta\mathbf{S}_i\mathbf{G}_{i\theta}\mathbf{G}_{i\theta}^\dagger\mathbf{S}_i^\dagger\mathbf{T}_\theta^\dagger\mathbf{h}^\dagger\right]$$
$$= \max_{1\leq i\leq|\mathcal{S}|}\left[\rho\mathbf{h}\mathbf{T}_\theta\mathbf{\Lambda}_\theta^{-1}\mathbf{S}_i\mathbf{S}_i^\dagger\mathbf{S}_i\mathbf{G}_i\mathbf{G}_i^\dagger\mathbf{S}_i^\dagger\mathbf{S}_i\mathbf{S}_i^\dagger\mathbf{\Lambda}_\theta^{-1}\mathbf{T}_\theta^\dagger\mathbf{h}^\dagger\right]$$
$$= \gamma(\mathbf{T},\mathbf{h}).$$

This proves the statement of the theorem. □

Now from theorem 1, using $\mathbf{T}_\theta = \mathbf{T}\mathbf{\Lambda}_\theta$ where:

$$[\Lambda_\theta]_{\ell,\ell} = \frac{[\mathbf{T}_{1,\ell}]^\dagger}{\left|[\mathbf{T}]_{c\{\ell\}}\right|\left|\mathbf{T}_{1,\ell}\right|} \quad \forall 1\leq\ell\leq L$$

instead of $\mathbf{T}$ in (5), the optimal preprocessor design problem can be reduced to:

$$\mathbf{T}_{\text{opt}} = \text{argmax}_{\mathbf{T}\in\mathcal{T}_{\mathcal{G}}}\{C(\mathbf{T})\} \quad \text{where}, \quad (7)$$
$$\mathcal{T}_{\mathcal{G}} = \left\{\mathbf{T}\in\mathbb{C}^{N\times L}\middle|\left|[\mathbf{T}]_{c\{\ell\}}\right| = 1, \text{Im}\{\mathbf{T}_{1,\ell}\} = 0 \ \forall\ell = 1,...,L\right\}$$

where, Im{} represents the imaginary component. Notice that since from the theorem, $C(\mathbf{T})$ is invariant to scaling of the columns of $\mathbf{T}$, each column $[\mathbf{T}]_{c\{\ell\}}$ can be considered to represent a 1-dimensional linear subspace in $\mathbb{C}^{N\times 1}$. Therefore (7) is actually an optimization problem over the complex Grassmannian manifold $\mathcal{G}(N,1)$.[4]

IV. Preprocessor Design

Notice that in (3) each column of $\mathbf{T}$ occurs in multiple terms of the maximization step, which makes finding the optimal solution to (7) difficult. In this section we shall therefore consider an analytically tractable lower bound to $\gamma(\mathbf{T},\mathbf{h})$.

For any selection matrix $\mathbf{S}_i$, let $\{\mu_1^i, .., \mu_K^i\}$ represent the set of ports, and consequently the columns, of the preprocessor matrix connected to the up-conversion chains. Then for any preprocessor $\mathbf{T}\in\mathcal{T}_{\mathcal{G}}$ and any $k\in\{1,..,K\}$, we have:

$$\mathbf{h}\mathbf{T}\mathbf{S}_i\mathbf{G}_i\mathbf{G}_i^\dagger\mathbf{S}_i^\dagger\mathbf{T}^\dagger\mathbf{h}^\dagger$$
$$\geq \mathbf{h}[\mathbf{T}]_{c\{\mu_k^i\}}[\mathbf{T}]_{c\{\mu_k^i\}}^\dagger\mathbf{T}\mathbf{S}_i\mathbf{G}_i\mathbf{G}_i^\dagger\mathbf{S}_i^\dagger\mathbf{T}^\dagger[\mathbf{T}]_{c\{\mu_k^i\}}[\mathbf{T}]_{c\{\mu_k^i\}}^\dagger\mathbf{h}^\dagger$$
$$= \mathbf{h}[\mathbf{T}]_{c\{\mu_k^i\}}[\mathbf{T}]_{c\{\mu_k^i\}}^\dagger\mathbf{h}^\dagger$$

[4]The complex Grassmannian manifold $\mathcal{G}(a,b)$ is the set of all b-dimensional linear subspaces over a vector space of dimension a with field $\mathbb{C}$.

where the last step follows from the fact that $[\mathbf{T}]_{c\{\mu_k^i\}}$ belongs to the column space of the semi-unitary matrix[5] $\mathbf{T}\mathbf{S}_i\mathbf{G}_i$. Therefore, from (3), we can lower bound the instantaneous SNR as:

$$\begin{aligned}\gamma(\mathbf{T},\mathbf{h}) &\geq \max_{1\leq i\leq|\mathcal{S}|}\max_{1\leq k\leq K}\left[\rho\mathbf{h}[\mathbf{T}]_{c\{\mu_k^i\}}[\mathbf{T}]_{c\{\mu_k^i\}}^\dagger\mathbf{h}^\dagger\right]\\ &= \max_{1\leq \ell\leq L}\left[\rho\mathbf{h}[\mathbf{T}]_{c\{\ell\}}[\mathbf{T}]_{c\{\ell\}}^\dagger\mathbf{h}^\dagger\right] \qquad (8)\end{aligned}$$

where the last step follows from the fact that each input port, and therefore each column of $\mathbf{T}$, is picked by at least one selection matrix for any reasonable switch position set $\mathcal{S}$. Unlike (3), in (8) each column of $\mathbf{T}$ occurs only in one term, thereby making the analysis easier. For any unit vectors $\mathbf{a}$ and $\mathbf{b}$, let us define the chordal distance between them [12] as:

$$d_{\text{chord}}(\mathbf{a},\mathbf{b}) \triangleq \sqrt{1-|\mathbf{a}^\dagger\mathbf{b}|^2}.$$

Now following a somewhat similar approach to [13], we have the following theorem.

Theorem 2 (Line-packing lower bound)**.** *$C(\mathbf{T}) \geq C_{\text{LB}}(\mathbf{T})$, where:*

$$C_{\text{LB}}(\mathbf{T}) = L\left(\frac{\delta}{2}\right)^{2N-2}\mathbb{E}_{|\mathbf{h}|}\left\{\log\left(1+\rho|\mathbf{h}|^2(1-\delta^2/4)\right)\right\} \quad (9)$$

and $\delta = \min_{a\neq b} d_{\text{chord}}([\mathbf{T}]_{c\{a\}},[\mathbf{T}]_{c\{b\}})$ for some $\mathbf{T}\in\mathcal{T}_{\mathcal{G}}$. Furthermore, if $N\geq 2$ and $\rho N\geq 2$, we have:

$$\text{argmax}_{\mathbf{T}\in\mathcal{T}_{\mathcal{G}}}\{C_{\text{LB}}(\mathbf{T})\} \equiv \text{argmax}_{\mathbf{T}\in\mathcal{T}_{\mathcal{G}}}\{f_{\text{chord}}(\mathbf{T})\} \quad (10)$$

where, $f_{\text{chord}}(\mathbf{T}) = \min_{a\neq b} d_{\text{chord}}([\mathbf{T}]_{c\{a\}},[\mathbf{T}]_{c\{b\}})$.

Proof. Since $\mathbf{R}_{\text{tx}} = \mathbb{I}_N$, $\mathbf{h}$ has complex Gaussian i.i.d. components and $\bar{\mathbf{h}} \triangleq \mathbf{h}/|\mathbf{h}|$, $|\mathbf{h}|$ are independently distributed. Further, $\bar{\mathbf{h}}$ is isotropically distributed i.e., $\bar{\mathbf{h}} \overset{d}{=} \bar{\mathbf{h}}\mathbf{U}$ for any $N\times N$ unitary matrix $\mathbf{U}$ and $|\mathbf{h}|$ is Chi-square distributed with $2N$ degrees of freedom.

Since $\log()$ is a non-decreasing function, from (8) we have:

$$C(\mathbf{T}) \geq \mathbb{E}_{|\mathbf{h}|,\bar{\mathbf{h}}}\left\{\log\left(1+\max_{1\leq\ell\leq L}\left[\rho|\mathbf{h}|^2\bar{\mathbf{h}}[\mathbf{T}]_{c\{\ell\}}[\mathbf{T}]_{c\{\ell\}}^\dagger\bar{\mathbf{h}}^\dagger\right]\right)\right\}. \quad (11)$$

Let us define $\delta = \min_{a\neq b} d_{\text{chord}}([\mathbf{T}]_{c\{a\}},[\mathbf{T}]_{c\{b\}})$ for $\mathbf{T}\in\mathcal{T}_{\mathcal{G}}$. Then the selection regions given by:

$$\begin{aligned}\mathcal{H}_\ell &= \left\{\bar{\mathbf{h}}\middle| d_{\text{chord}}(\bar{\mathbf{h}}^\dagger,[\mathbf{T}]_{c\{\ell\}}) < \delta/2\right\}\\ &= \left\{\bar{\mathbf{h}}\middle|\bar{\mathbf{h}}[\mathbf{T}]_{c\{\ell\}}[\mathbf{T}]_{c\{\ell\}}^\dagger\bar{\mathbf{h}}^\dagger > 1-\delta^2/4\right\}\end{aligned}$$

are all disjoint [13]. Since $0\leq d_{\text{chord}}(\mathbf{a},\mathbf{b})\leq 1$ for any unit vectors $\mathbf{a},\mathbf{b}$, from (11) we can lower bound $C(\mathbf{T})$ as:

$$C(\mathbf{T}) \geq \sum_{\ell=1}^{L}\mathbb{P}\left(\bar{\mathbf{h}}\in\mathcal{H}_\ell\right)\mathbb{E}_{|\mathbf{h}|}\left\{\log\left(1+\rho|\mathbf{h}|^2(1-\frac{\delta^2}{4})\right)\right\}. \quad (12)$$

Since $\bar{\mathbf{h}}$ is isotropically distributed, from [13], [14] we have:

$$\sum_{\ell=1}^{L}\mathbb{P}\left(\bar{\mathbf{h}}\in\mathcal{H}_\ell\right) = L\left(\frac{\delta}{2}\right)^{2N-2}. \quad (13)$$

[5]Here, an $a\times b$ rectangular matrix $\mathbf{A}$ as semi-unitary if $\mathbf{A}^\dagger\mathbf{A} = \mathbb{I}_b$

Using (12) and (13) we arrive at (9). Note that $\mathbf{T}$ affects $C_{\text{LB}}(\mathbf{T})$ only via the term δ (for a fixed L). Therefore, if the partial derivative of $C_{\text{LB}}(\mathbf{T})$ with respect to δ is non-negative, then maximizing δ is equivalent to maximizing $C_{\text{LB}}(\mathbf{T})$. The required condition is $\frac{\partial C_{\text{LB}}(\mathbf{T})}{\partial\delta}\geq 0$ i.e.,

$$\begin{aligned}(N-1)\mathbb{E}_{|\mathbf{h}|}&\left\{\log\left(1+\rho|\mathbf{h}|^2(1-\delta^2/4)\right)\right\}\\ &\geq \mathbb{E}_{|\mathbf{h}|}\left[\frac{\rho|\mathbf{h}|^2\delta^2/4}{1+\rho|\mathbf{h}|^2(1-\delta^2/4)}\right]. \qquad (14)\end{aligned}$$

By upper bounding the right hand side and lower bounding the left hand sides of (14), we get a sufficient condition as:

$$\mathbb{P}(|\mathbf{h}|^2\geq N)\log\left(1+\frac{3\rho N}{4}\right)\geq\frac{1}{3N-3}.$$

Using the distribution of $|\mathbf{h}|^2$, it can be verified that this holds if $N\geq 2$ and $\rho N\geq 2$. □

The derived bound $C_{\text{LB}}(\mathbf{T})$ is tight only when L is of the order of 2^N [13] and further it is independent of the switch position set $\mathcal{S}$. However, this bound is easier to analyze than $C(\mathbf{T})$. Hence, for analytical tractability, we consider the preprocessor design approach that maximizes this capacity lower bound, i.e.,

$$\mathbf{T}_{\text{LP}} = \text{argmax}_{\mathbf{T}\in\mathcal{T}_{\mathcal{G}}}\{f_{\text{chord}}(\mathbf{T})\}. \quad (15)$$

Here we implicitly assume the conditions of Theorem 2 hold. A more stringent design for the preprocessor, that depends on the switch position set, has been considered in our extended journal version [15].

Note that (15) is equivalent to the problem of Grassmannian line packing with chordal distance metric. Several algorithms have been proposed in literature to find good line-packing solutions [16], [17]. These algorithms can be leveraged to design good preprocessing matrices under various hardware constraints as illustrated in Section V. Note that for $L\leq N$, any $N\times L$ semi-unitary[5] matrix $\mathbf{T}$ is an optimal solution for line-packing (15). Therefore, a generalization of the preprocessor design in [4]: $\mathbf{T}_{\text{sud}} = [\mathbf{E}_{\text{tx}}]_{c\{1:L\}}$, (where $\mathbf{E}_{\text{tx}}$ is the eigen-vector matrix of $\mathbf{R}_{\text{tx}}$) is optimal for (15), for $L\leq N$. This is not true when $\mathbf{R}_{\text{tx}}\neq\mathbb{I}_N$ as explored in Section VI.

V. Simulation Results

We first consider a single transmitter-receiver pair where the transmitter implements HPwS and the receiver has a single antenna with a mean SNR of $\rho = 10$ dB. We consider a reduced complexity switch bank where each up-conversion chain has an exclusive set of input ports to connect to, i.e.,

$$\mathcal{S} = \left\{[\mathbb{I}_L]_{c\{\ell_1,..,\ell_K\}}\middle|(k-1)\left\lfloor\frac{L}{K}\right\rfloor<\ell_k\leq k\left\lfloor\frac{L}{K}\right\rfloor, k\in\{1,..,K\}\right\}.$$

Since the mean capacity (4) is not known in closed form, in this section we study the performance of the preprocessing matrix using Monte-Carlo simulations. Here, we perform a brute-force search among $\mathcal{S}$ to pick the best selection matrix $\mathbf{S}$ for each channel realization. The design of low-complexity algorithms for selecting $\mathbf{S}$, for a given $\mathbf{h}$, is beyond the scope of this paper (see [10] and references therein). The performance of HPaCSI and HPiCSI can be obtained by replacing $\mathcal{S} = \{\mathbb{I}_K\}$ (no selection) and $\mathbf{T}_{\text{HPaCSI}} = [\mathbb{I}_N]_{c\{1:K\}}$ or $\mathbf{T}_{\text{HPiCSI}} = [\mathbf{E_h}]_{c\{1:K\}}$ in (4), where $\mathbf{E_h}$ is the $N \times N$ eigen-vector matrix of $\mathbf{h}^\dagger\mathbf{h}$ with eigenvalues sorted in the descending order of magnitude. However, note that HPiCSI can actually attain this performance using a single RF chain.

For now, we assume that all the elements of $\mathbf{T}$ can have arbitrary magnitude and phase. Various hardware constraints on the preprocessing matrix are discussed later in this section. Note that this design can be implemented by using a variable RF phase-shifter and RF amplifier for each element. For such an unconstrained preprocessor, designed by (15), the mean capacity as a function of the number of columns of the preprocessor (L) is studied in Fig. 2. As expected, we observe

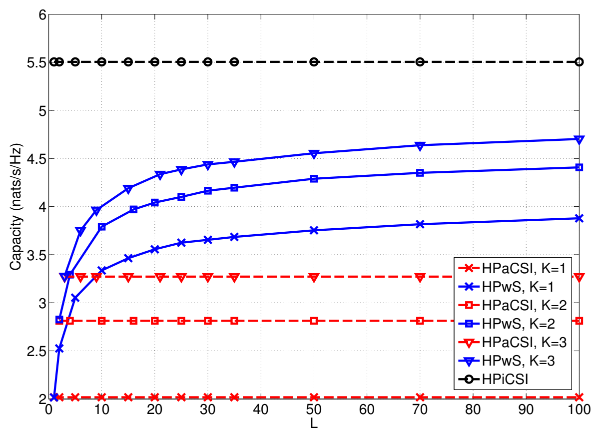


Fig. 2. Mean capacity versus the number of columns of the preprocessing matrix L (simulation parameters: $N = 25, \rho = 10$, $\mathbf{T}$ is generated from the MATLAB implementation [18] of the Grassmannian line-packing algorithm in [16])

that HPwS bridges the performance gap between HPaCSI and HPiCSI without requiring iCSI based adaptation of the preprocessing matrix. Intuitively, we use redundancy in RF hardware to effectively make the beams more channel adaptive. We also observe that there is a diminishing improvement in performance as L increases and therefore for a good trade-off between performance and hardware cost, the number of columns L should be of the order of N.

We next consider the case where elements of $\mathbf{T}$ have a fixed magnitude and the phase belongs to a set of discrete phases, i.e., $[\mathbf{T}]_{a,b} = \frac{e^{j\phi_{a,b}}}{\sqrt{N}}$ where $\phi_{a,b} \in \{0, \frac{2\pi}{B}, ..., \frac{2\pi(B-1)}{B}\}$. Such a preprocessor is easier to implement since it requires only a discrete phase-shifter for each element of $\mathbf{T}$. For this case, $\mathbf{T}$ that maximizes (15) can be designed using fixed alphabet line-packing algorithms, such as in [16].[6] The performance of HPwS with such a restricted preprocessor is studied in Fig. 3 as a function of the number of possible phase shifts B. As observed from the results, even for $B = 4$ the performance is almost as good as with the unrestricted preprocessor design,

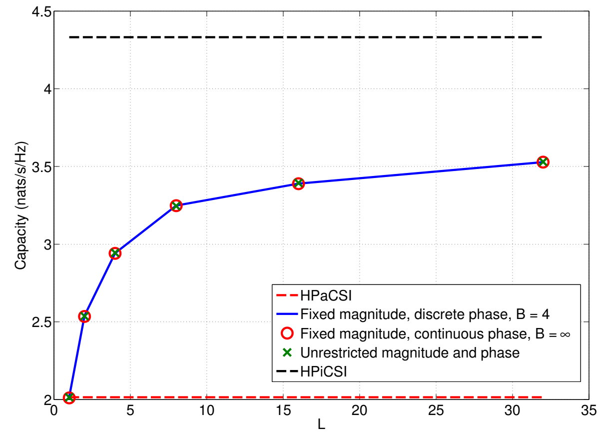


Fig. 3. Mean capacity versus the number of discrete phases (B) of the preprocessing matrix. For $B = 4, \infty$ and $L > N$, we use the line packed preprocessors for QPSK alphabet and constant modulus from [16], [18]. For $L < N$, we use a $N \times L$ submatrix of the $N \times N$ real Hadamard matrix (simulation parameters: $N = 8, K = 1, \rho = 10$)

suggesting that a good preprocessor can be implemented using a simple discrete phase-shifter array.

While the preprocessor in (15) is optimized for a single receiver, here we also study its performance in a multi-user scenario. The received signal vector at U single-antenna receivers, each having independent Rayleigh fading channels with same transmit correlation matrix[7], can be expressed as:

$$\mathbf{y}_U = \sqrt{\rho}\mathbf{H}_U\mathbf{T}\mathbf{S}_i\mathbf{G}_i\mathbf{u} + \mathbf{n}_U$$

for any $\mathbf{S}_i \in \mathcal{S}$. Assuming the transmitter performs zero-forcing, by pseudo-inverting the *effective channel* $\mathbf{H}_U\mathbf{T}\mathbf{S}_i\mathbf{G}_i$, the impact of L on performance is studied in Fig. 4. We observe that for $L > N$, the slope of the curves, and hence fractional performance gain, increases with U. This suggests that the additional beam choices with $L > N$, may help improve user separability.

VI. Extension to Anisotropic Channels

The isotropic scattering assumption i.e., $\mathbf{R}_{\text{tx}} = \mathbb{I}_N$ in Section II was only for convenience and is rarely met in practice. Practical wireless channels are usually spatially sparse i.e., most of the channel power is concentrated within a sub-space, referred to as the *dominant channel sub-space*. Such a channel can be modeled by assuming that $\mathbf{R}_{\text{tx}}$ has a rank D, where $K < D \leq N$. Under these conditions the small scale fading matrix in (1) can be re-written as:

$$\mathbf{h} = \mathbf{g}[\mathbf{\Lambda}_{\text{tx}}^D]^{1/2}[\mathbf{E}_{\text{tx}}^D]^\dagger \quad (16)$$

where $\mathbf{g} \sim \mathcal{CN}(\mathbb{O}_{1\times D}, \mathbb{I}_D)$ is a $1 \times D$ complex Gaussian vector with independent and identically distributed (i.i.d.) entries and $\mathbf{\Lambda}_{\text{tx}}^D, \mathbf{E}_{\text{tx}}^D$ are the $D \times D$, $N \times D$ matrices of eigenvalues and eigen-vectors of $\mathbf{R}_{\text{tx}}$, respectively, corresponding to the D non-zero eigenvalues. Note that even if $\mathbf{R}_{\text{tx}}$ has a full rank, it may be beneficial to approximate $D < N$ to reduce channel

[6] Other approaches for finding a constrained precoder, such as in [4], [6] may also be used.

[7] This is a common assumption for a co-located group of users, see for example [19].

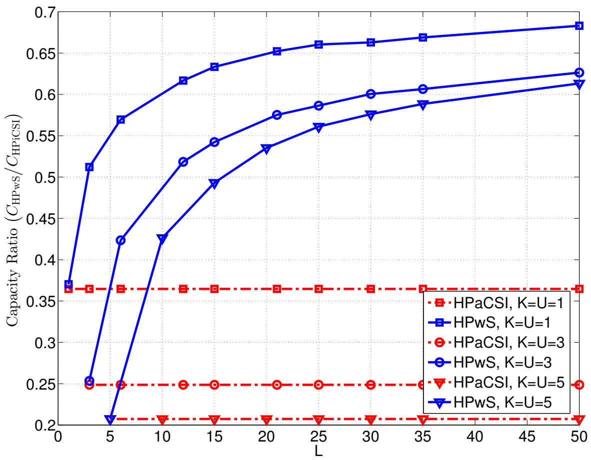


Fig. 4. Mean sum-rate (normalized with respect to mean sum-rate of HPiCSI) versus L for a multi-user scenario (simulation parameters: $N = 25, \mathbf{R}_{\text{tx}} = \mathbb{I}_N, \rho = 10$, $\mathbf{T}$ is generated from the MATLAB implementation of the Grassmannian line-packing algorithm [18])

estimation overhead [8]. Given (16), it can be shown that the optimal preprocessing matrix in (5) should be of the form $\mathbf{T}_{\text{opt}} = \mathbf{E}_{\text{tx}}^D \hat{\mathbf{T}}_{\text{opt}}$, where $\hat{\mathbf{T}}_{\text{opt}}$ is a $D \times L$ matrix. This design ensures that the preprocessor transmits power only into the dominant channel sub-space, and a formal proof of the above is included in our extended journal version [15]. Therefore the preprocessor design problem in (5) can be reduced to finding:

$$\hat{\mathbf{T}}_{\text{opt}} = \text{argmax}_{\hat{\mathbf{T}} \in \mathbb{C}^{D \times L}} \{\mathbb{E}_{\mathbf{g}}\{\log(1 + \hat{\gamma}(\hat{\mathbf{T}}, \mathbf{g}))\} \quad (17)$$

where the instantaneous SNR is simplified to:

$$\hat{\gamma}(\hat{\mathbf{T}}, \mathbf{g}) = \max_{1 \le i \le |\mathcal{S}|} \left[\rho \mathbf{g} [\mathbf{\Lambda}_{\text{tx}}^D]^{1/2} \hat{\mathbf{T}} \mathbf{S}_i \mathbf{G}_i \mathbf{G}_i^\dagger \mathbf{S}_i^\dagger \hat{\mathbf{T}}^\dagger [\mathbf{\Lambda}_{\text{tx}}^D]^{1/2} \mathbf{g}^\dagger \right] \quad (18)$$

A. Isotropic in dominant subspace

If $\mathbf{\Lambda}_{\text{tx}}^D = \mathbb{I}_D$, we observe that (3),(5) are very similar to (17),(18), respectively. Therefore, the previous analysis for $\mathbf{R}_{\text{tx}} = \mathbb{I}_N$ in Sections III and IV, is directly applicable to the design of $\hat{\mathbf{T}}$ by replacing N by D. Therefore, from theorem 2, assuming $D \ge 2$ and $\rho D \ge 2$, a good preprocessor design that minimizes a lower bound on capacity can be obtained as:

$$\mathbf{T}_{\text{iso}} = \mathbf{E}_{\text{tx}}^D \hat{\mathbf{T}}_{\text{LP}} \quad (19)$$

where the $D \times L$ matrix $\hat{\mathbf{T}}_{\text{LP}}$ is designed by line-packing, similar to (15). Furthermore, the simulation results and observations in Section V are all directly applicable for this case by replacing N with D. This justifies our use of small values of $N (\le 25)$ in the results in Section V. While usually in a massive MIMO scenario N may be much larger, these values are reasonable for D in a realistic channel [20]. The only exception is Fig. 3 which characterizes the impact of discrete components on $\hat{\mathbf{T}}_{\text{LP}}$ and not on $\mathbf{E}_{\text{tx}}^D$. Approaches to discretize $\mathbf{E}_{\text{tx}}^D$ are discussed, for example, in [4], [6].

B. Anisotropic in dominant subspace

In the more general case of $\mathbf{\Lambda}_{\text{tx}}^D \ne \mathbb{I}_D$, let us assume the eigenvalues in $\mathbf{\Lambda}_{\text{tx}}^D$ are arranged in descending order of magnitude. Then, similar to the approach in [21], we use the companding trick to design the preprocessor as:

$$[\mathbf{T}_{\text{ani}}]_{c\{\ell\}} = \frac{\mathbf{E}_{\text{tx}}^D \mathbf{\Lambda}_{\text{tx}}^D [\hat{\mathbf{T}}_{\text{LP}}]_{c\{\ell\}}}{\sqrt{[\hat{\mathbf{T}}_{\text{LP}}]_{c\{\ell\}}^\dagger [\mathbf{\Lambda}_{\text{tx}}^D]^2 [\hat{\mathbf{T}}_{\text{LP}}]_{c\{\ell\}}}} \quad 1 \le \ell \le L \quad (20)$$

where $\hat{\mathbf{T}}_{\text{LP}}$ is the $D \times L$ line-packed preprocessor, as in Section VI-A. Intuitively, (20) skews the columns of $\mathbf{T}_{\text{ani}}$, and therefore also the precorders $\mathbf{T}_{\text{ani}} \mathbf{S}_i \mathbf{G}_i$, to be more densely packed near the eigen-vectors corresponding to the larger eigenvalues of $\mathbf{R}_{\text{tx}}$. Note that for $L \le D$, any $D \times L$ semi-unitary[5] matrix $\hat{\mathbf{T}}$ is an optimal solution for Grassmannian line-packing. However, not all such $\hat{\mathbf{T}}$ yield good performance after skewing by (20). This problem typically does not arise when $L > D$. Therefore for $L \le D$, we suggest the use of $\hat{\mathbf{T}}_{\text{LP}} = \left[\, \mathbf{W} \quad \mathbb{O}_{(D-L) \times L} \,\right]^\dagger$ in (20), where $\mathbf{W}$ is the $L \times L$ DFT matrix i.e., $[\mathbf{W}]_{a,b} = e^{\frac{j 2\pi ab}{L}} / \sqrt{L}$. This choice ensures that $\mathbf{T}_{\text{iso}}$ (and $\mathbf{T}_{\text{ani}}$) span the "best" channel subspace and also ensures proper skewing in (20).

For simulations, we assume the transmitter has a $\lambda/2$-spaced uniform linear array, with $N = 100$ antenna elements. The transmit power angle spectrum is given by:

$$\text{PAS}(\theta) = e^{(-\eta|\theta + \frac{\pi}{3}|)} \Pi(\theta + \frac{\pi}{3}) + e^{(-\eta|\theta|)} \Pi(\theta) + e^{(-\eta|\theta - \frac{\pi}{6}|)} \Pi(\theta - \frac{\pi}{6}) \quad (21)$$

$$\text{where: } \Pi(\theta) = \begin{cases} 1 & \text{for } |\theta| \le \pi/36 \\ 0 & \text{otherwise} \end{cases}$$

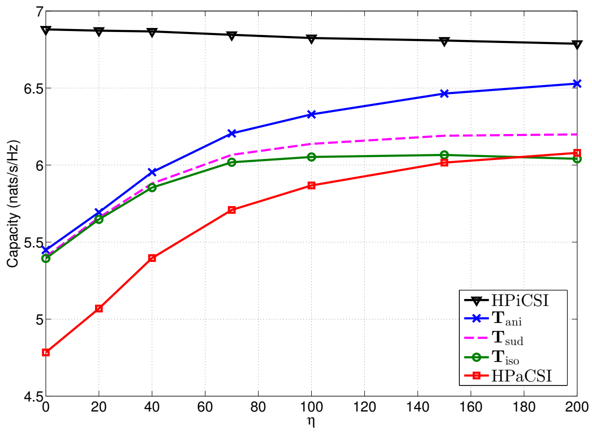
and η is a factor that controls the anisotropy of the channel. The transmit correlation matrix is then computed as:

$$[\mathbf{R}_{\text{tx}}]_{ab} = \int_{-\pi}^{\pi} \text{PAS}(\theta) e^{j\pi(a-b)\sin(\theta)} d\theta \Big/ \int_{-\pi}^{\pi} \text{PAS}(\phi) d\phi$$

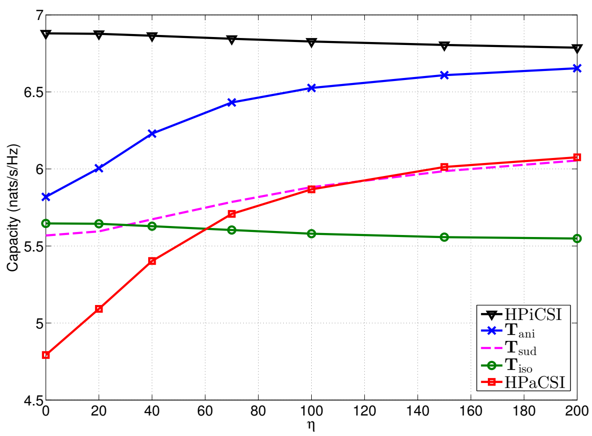
It can be readily verified (using simulations) from (21) that, for all values of η, $\mathbf{R}_{\text{tx}}$ has at-most $D = 25$ dominant eigenvalues with negligible power outside. We denote the $N \times N$ eigen-vector matrix of the $\mathbf{R}_{\text{tx}}$ as $\mathbf{E}_{\text{tx}}$ and its dominant $N \times D$ sub-matrix is denoted by $\mathbf{E}_{\text{tx}}^D$. For this channel, the performance of the skewed preprocessing matrix $\mathbf{T}_{\text{ani}}$ compared to $\mathbf{T}_{\text{iso}}$ is studied in Fig. 5 as a function of the channel anisotropy for the both cases of $L \le D$ and $L > D$. We also plot the performance of $\mathbf{T}_{\text{sud}} = [\mathbf{E}_{\text{tx}}]_{c\{1:L\}}$, a generalization of the preprocessor in [4]. We observe from the results that the capacity gap between HPiCSI and HPaCSI reduces as the channel anisotropy increases. Also, unlike with $\mathbf{T}_{\text{iso}}$, the performance of $\mathbf{T}_{\text{ani}}$ does not degrade with increasing anisotropy. Further, we observe that $\mathbf{T}_{\text{ani}}$ outperforms $\mathbf{T}_{\text{sud}}$ over the whole range of η for both $D \le L$ and $D > L$. Therefore the designed preprocessor is not only more generic (can be extended to the case of $L > N$) but also leads to better performance in comparison to other designs.

VII. Conclusions and Future Directions

In this work, we explore the use of HPwS as a solution to reduce the hardware cost of massive MIMO systems, without degrading the performance significantly. We formulate the problem of preprocessor design and show that a lower bound to mean capacity can be maximized using Grassmannian line-packing on the columns of the preprocessor. The simulation



(a) $\{D, L, K\} = \{25, 8, 2\}$



(b) $\{D, L, K\} = \{25, 50, 2\}$

Fig. 5. Mean capacity for the skewed preprocessor $\mathbf{T}_{\text{ani}}$ as a function of channel anisotropy. (a) $\hat{\mathbf{T}}_{\text{iso}} = \left[\mathbf{W} \; \mathbb{O}_{(D-L)\times L}\right]^{\dagger}$, where $\mathbf{W}$ is the $L \times L$ DFT matrix (b) $\hat{\mathbf{T}}_{\text{iso}}$ is generated from the line packing algorithm in [18] (simulation parameters: $N = 100$ and $\rho = 10$).

results suggest that significant capacity gains can be achieved using HPwS in comparison to HPaCSI. Similarly, the mean capacity is only slightly inferior to HPiCSI, with significantly simpler RF hardware. The results also suggest that for a good trade-off between performance and hardware cost, the number of columns of the preprocessor L should be of the order of rank$\{\mathbf{R}_{\text{tx}}\}$. However, for multi-receiver scenarios, it may be practical to use larger values of L. Furthermore, results show that building the preprocessor with discrete phase shifters having just 4 possible phases, is sufficient to achieve good performance. Results for anisotropic channels show that skewing the line-packed preprocessor yields better performance when compared to alternate designs.

However, the designed preprocessor maximizes a lower bound that is independent of the switch position set $\mathcal{S}$ and therefore maybe sub-optimal. Furthermore it does not give an intuition into designing good switch position sets. An analysis which deals with the above issues is considered in [15].

As is the case with other such diversity techniques [10], the performance gain of HPwS over HPaCSI decreases with frequency selective fading. This is because in the limit of a large transmission bandwidth, frequency diversity makes all the L selection ports equivalent. Hence, HPwS is more suited for narrow-to-medium bandwidth systems, where channels are only moderately frequency selective.

References

[1] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, November 2010.

[2] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, pp. 74–80, February 2014.

[3] X. Zhang, A. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Transactions on Signal Processing*, vol. 53, pp. 4091–4103, Nov 2005.

[4] P. Sudarshan, N. Mehta, A. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 3501–3511, December 2006.

[5] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1499–1513, March 2014.

[6] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Information Theory and Applications Workshop (ITA)*, pp. 1–5, Feb 2013.

[7] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO - A survey," *CoRR*, vol. abs/1609.05078, 2016.

[8] V. V. Ratnam, A. F. Molisch, N. Rabeah, F. Alawwad, and H. Behairy, "Diversity versus training overhead trade-off for low complexity switched transceivers," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec 2016.

[9] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, 2016.

[10] A. Molisch, "MIMO systems with antenna selection - an overview," in *Radio and Wireless Conference, 2003. RAWCON '03. Proceedings*, pp. 167–170, Aug 2003.

[11] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Transactions on Communications*, vol. 47, pp. 1458–1461, Oct 1999.

[12] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the grassmann manifold," *IEEE Transactions on Information Theory*, vol. 48, pp. 2450–2454, Sep 2002.

[13] D. Love, R. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, pp. 2735–2747, Oct 2003.

[14] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 49, pp. 2562–2579, Oct 2003.

[15] V. V. Ratnam, O. Y. Bursalioglu, H. C. Papadopoulos, and A. F. Molisch, "Hybrid preprocessing with selection in multi-user massive MIMO systems." in preparation, 2017.

[16] A. Medra and T. N. Davidson, "Flexible codebook design for limited feedback systems via sequential smooth optimization on the grassmannian manifold," *IEEE Transactions on Signal Processing*, vol. 62, pp. 1305–1318, March 2014.

[17] I. S. Dhillon, R. W. Heath, T. Strohmer, and J. A. Tropp, "Constructing packings in grassmannian manifolds via alternating projection," *Experiment. Math.*, vol. 17, no. 1, pp. 9–35, 2008.

[18] A. Medra and T. N. Davidson, "Flexible codebook design for limited feedback systems [online]," 2015. http://www.ece.mcmaster.ca/ davidson/pubs/Flexible_codebook_design.html.

[19] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing – the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, pp. 6441–6463, Oct 2013.

[20] 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, pp. 1164–1179, June 2014.

[21] D. Love and R. Heath, "Grassmannian beamforming on correlated MIMO channels," in *IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 1, pp. 106–110 Vol.1, Nov 2004.