Uplink-Downlink Channel Covariance Transformations and Precoding Design for FDD Massive MIMO

Mahdi Barzegar Khalilsarai, Yi Song, Tianyu Yang, Saeid Haghighatshoar, and Giuseppe Caire

Abstract—A large majority of cellular networks deployed today make use of Frequency Division Duplexing (FDD) where, in contrast with Time Division Duplexing (TDD), the channel reciprocity does not hold and explicit downlink (DL) probing and uplink (UL) feedback are required in order to achieve spatial multiplexing gain. In order to support massive MIMO, i.e., a very large number of antennas at the base station (BS) side, the overhead incurred by conventional DL probing and UL feedback schemes scales linearly with the number of BS antennas and, therefore, may be very large. In this paper, we present a new approach to achieve a very competitive tradeoff between spatial multiplexing gain and probing-feedback overhead in such systems. Our approach is based on two novel methods: (i) an efficient regularization technique based on Deep Neural Networks (DNN) that learns the Angular Spread Function (ASF) of users channels and permits to estimate the DL covariance matrix from the noisy i.i.d. channel observations obtained freely via UL pilots (UL-DL covariance transformation), (ii) a novel "sparsifying precoding" technique that uses the estimated DL covariance matrix from (i) and imposes a controlled sparsity on the DL channel such that given any assigned DL pilot dimension, it is able to find an optimal sparsity level and a corresponding sparsifying precoder for which the "effective" channel vectors after sparsification can be estimated at the BS with a low mean-square error. We compare our proposed DNN-based method in (i) with other methods in the literature via numerical simulations and show that it yields a very competitive performance. We also compare our sparsifying precoder in (ii) with the state-of-the-art statistical beamforming methods under the assumption that those methods also have access to the covariance knowledge in the DL and show that our method yields higher spectral efficiency since it uses in addition the instantaneous channel information after sparsification.

Index Terms—Massive MIMO, Sparse Scattering, Angular Spread Function (ASF), Uplink-Downlink Covariance Transformation (UDCT), Deep Neural Networks (DNNs), Sparisfying Precoder.

I. INTRODUCTION

Massive Multiple-Input Multiple-Output (MIMO) is a variation of conventional multi-user MIMO, where base station (BS) has a much larger number of antennas (or antenna ports) $M \gg 1$, and is considered to be a key technology for the next generation of wireless networks [1]. Large number of antennas permits to multiplex $K \gg 1$ data streams over the spatial domain to serve K users and guarantees significant advantages such as energy efficiency due to a large beamforming gain, reduced inter-cell interference, and simple user scheduling and rate adaptation due to the wellknown channel hardening phenomenon [1]. To achieve such benefits, especially in the a downlink (DL) scenario we are

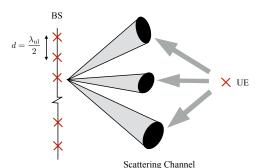


Fig. 1: Sparse scattering channel between a generic user and the BS. In this example, the channel consists of 3 large scatterers reflecting the power of the user to the BS array.

interested in this paper, the BS needs to learn the channel vectors of K users to M BS antennas in the DL. With *Time* Division Duplexing (TDD), due to uplink-downlink (UL-DL) channel reciprocity which holds under suitable calibration [2, 3], this can be done via transmitting mutually orthogonal pilots from the users only in the UL. Unfortunately, channel reciprocity does not hold in Frequency Division Duplexing (FDD) since UL and DL channels lie on disjoint and farseparated frequency bands. Consequently, the only way to learn the DL channel is to devote a fraction of resource elements (RE) [4] to estimate the DL channel of K users by transmitting DL pilots and then UL feedback. Conventional DL training consists of the transmission of an $M \times T_{dl}$ pilot matrix over $T_{\rm dl}$ REs, such that $T_{\rm dl} \ge M$ to permit each user to estimate its own DL channel vector. Then, the users feed their estimated channels back to the BS via the UL channel. Although this method works quite well for conventional MIMO systems with moderately small number of antennas M, it is quite inefficient in massive MIMO since $M \gg 1$ and full training of DL channel wastes at least $T_{dl} = M$ REs which may exhaust or be even larger than the whole REs available in the DL. This feedback bottleneck makes implementing massive MIMO in FDD quite challenging.

To overcome this bottleneck, several works have been proposed to reduce DL training and feedback overhead using the sparse structure of the channel in the angle-of-arrival (AoA) domain. This sparsity arises due to the fact that FDD systems in 5G will be mainly used for large cells (while TDD for smaller denser cells) with tower-mounted base stations [5], where the communication between the users and the BS occurs through a sparse cluster of scatterers with limited angular support (see, e.g., Fig. 1). As a result, the effective dimension of the channel s is much less than M. Building on this idea, these works proposed using compressed DL pilots and *Compressed Sensing* (CS)-based channel recovery techniques [6–8]. From standard results in CS, these method

The authors are with the Communications and Information Theory Group (CommIT), Technische Universität Berlin ({m.barzegarkhalilsarai, yi.song, tianyu.yang, saeid.haghighatshoar, caire}@tu-berlin.de).

require a pilot dimension of order $T_{dl} = O(s \log M)$, which may be much less than M for very sparse channels ($s \ll M$). Although the assumption $s \ll M$ may be met in very sparse Line-of-Sight (LoS) scattering scenarios, it is not fulfilled when the propagation occurs through diffuse clusters of moderately large AoA support size because in those cases the effective dimension s also scales like $s = \alpha M$ where α is proportional to the angular width of the scatterer. In those cases, CS methods incur a pilot dimension overhead of $T_{dl} = O(\alpha M \log M)$ which still grows proportionally to M. Overall, although the CS methods are able to exploit the sparsity of the channel they are still at the mercy of sparsity induced by the propagation environment since they are unable to "shape the channel sparsity" as desired.

A. Contribution

In this paper we propose an efficient scheme for realizing FDD massive MIMO. The fundamental assumption we make to achieve this goal is that, the channel is a Gaussian process with an *Angular Spread Function* (ASF) that remains the same for UL and DL. In particular, this implies that although the UL-DL reciprocity may not hold for the instantaneous channel vectors, some sort of statistical reciprocity still holds as the ASF does not vary between UL and DL. We can summarize our contributions in this paper as follows.

(i) A Novel UL-DL Covariance Transformation Using Deep Neural Networks. We use the reciprocity of the ASF in UL and DL to estimate the DL covariance matrix from the observation of noisy channel vectors freely available through the UL pilots. The ASF reciprocity was also assumed implicitly in several past works [9–11] by considering only ASFs with spike discrete components. Recent works [12–14] generalize these results to arbitrary ASFs consisting of both discrete spike and also continuous components, and show that such Uplink-Downlink covariance transformation (UDCT) is still feasible for sufficiently large number of BS antennas M although establishing the stability of UDCT is quite challenging when only a noisy estimate of the UL covariance matrix is available.

One of the aims of this paper is to improve the performance of UDCT by exploiting the natural group sparsity structure of the ASF. Unfortunately, designing good regularization methods that promote the group-sparsity of the ASF (see, e.g., Fig. 1) is not trivial. In particular, none of the structured group-sparse estimators widely adopted in sparse signal recovery literature are applicable here [15, 16]. In this paper, we bridge this gap by using Deep Neural Networks (DNNs). We train DNNs with suitable training data corresponding to group-sparse ASFs and let them learn/capture the notion of group-sparsity. Once the DNN is suitably trained, we use it as a black-box algorithm to estimate the ASF, thus, the DL covariance matrix, from the noisy channel vectors gathered in the UL. We illustrate via numerical simulations that our proposed DNN-based method is quite strong and recovers the DL covariance matrix much better than other methods.

(ii) A Novel Sparsifying Precoder. We use the estimated DL covariance matrix from (i) to design a *sparsifying precoder*, such that the following two criteria are met: (a) the reduceddim channel vectors of all the users after sparsification are sparse enough such that they are stably estimated with a very low error and (b) the dimension of the channel after sparsification is kept as large as possible in order to increase the rank of the effective channel matrix of the users after sparsification such that large number of users K can be served. We pose this as a Mixed Integer Linear Program (MILP) and solve it via off-the-shelf MILP solvers.

It is also worthwhile to mention that compared with CS methods, which are at the mercy of the sparsity of the channel due to the propagation, our proposed method is able to shape the sparsity of the channel completely flexibly depending on number of REs $T_{\rm dl}$ available for channel estimation. We illustrate via numerical simulations that the proposed method has an excellent performance much superior to other methods such as statistical beamforming, which also use the knowledge of the DL covariance matrix.

B. Notation

We denote vectors/matrices with small/large boldface letters (e.g., \mathbf{x}/\mathbf{X}), and sets with calligraphic letters (e.g., \mathcal{X}). We use the *i*-th element of a vector \mathbf{x} with $[\mathbf{x}]_i$ and the (i, j)-th element of a matrix \mathbf{X} with $[\mathbf{X}]_{i,j}$. For an integer k, we use the short-hand notation [k] for $\{1, \ldots, k\}$.

II. PROPOSED ASF ESTIMATION AND UPLINK-DOWNLINK COVARIANCE TRANSFORMATION

In this section, we explain our proposed method for UDCT using DNNs. For the sake of completeness, we first provide a summary of UDCT problem for the Uniform Linear Array (ULA) we consider in this paper (see, Fig. 1). Let us consider a generic use and let us denote the ASF of this user by $\gamma(\xi)$ where $\xi = \sin(\theta) \in [-1, 1]$ denotes a parametrization of the AoA $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and where $\gamma(\xi)$ is density of the received signal power at the AoA ξ . The covariance matrix of the channel vector of this user at UL/DL carrier frequency $f \in \{f_{ul}, f_{dl}\}$ is given by [12]

$$\boldsymbol{\Sigma}(f) = \int_{-1}^{1} \gamma(\xi) \mathbf{a}(\xi, f) \mathbf{a}(\xi, f)^{\mathsf{H}} d\xi, \qquad (1)$$

where $\mathbf{a}(\xi, f)$ denotes the array response vector at AoA ξ at frequency f given by

$$[\mathbf{a}(\xi, f)]_k = e^{j(k-1)\pi d\frac{f}{c_0}}, k \in [M],$$
(2)

where d denotes the antenna spacing and where c_0 is the speed of light. Note that in (1) we assumed implicitly that the ASF $\gamma(\xi)$ is the same in both UL and DL frequency range. As explained before, this provides some sort of statistical UL-DL channel reciprocity for FDD (in contrast with the instantaneous channel reciprocity which may not hold). With this notation, can pose the UDCT problem as follows.

UDCT Problem: Given the UL covariance matrix or an estimate thereof $\Sigma_{ul} := \Sigma(f_{ul})$, find the DL covariance matrix $\Sigma_{dl} := \Sigma(f_{dl})$.

For the ULA, we can gain a better understanding of UDCT by looking at the Fourier coefficients of the ASF $\gamma(\xi)$. We first assume that the array has the standard half wavelength spacing in the UL, namely, $d = \frac{\lambda_{\rm ul}}{2}$ where $\lambda_{\rm ul} = \frac{c_0}{f_{\rm ul}}$ denotes the wavelength at UL carrier frequency. Then, it is not difficult to show that $\Sigma_{\rm ul}$ is a Toeplitz matrix whose first column is given by $\sigma_{\rm ul}(\gamma) \in \mathbb{C}^M$ where

$$[\boldsymbol{\sigma}_{\mathrm{ul}}(\gamma)]_k = \int_{-1}^1 \gamma(\xi) e^{j(k-1)\pi\xi} d\xi, \quad k \in [M], \qquad (3)$$

denotes the (k-1)-th Fourier coefficient of $\gamma(\xi)$. Similarly, Σ_{dl} is also a Toeplitz matrix with first column $\sigma_{dl}(\gamma) \in \mathbb{C}^M$

$$[\boldsymbol{\sigma}_{\rm dl}(\gamma)]_k = \int_{-1}^1 \gamma(\xi) e^{j(k-1)\pi\beta\xi} d\xi, \quad k \in [M], \qquad (4)$$

where $\beta = \frac{f_{\rm dl}}{f_{\rm ul}}$ denotes the ratio between the DL and UL carrier frequencies. In current deployments of FDD systems typically $f_{\rm ul} > f_{\rm dl}$, thus, $\beta > 1$. Therefore, in the ideal case where $\Sigma_{\rm ul}$ is known exactly, one can pose UDCT as recovering the projections of the positive functions $\gamma(\xi)$ on the DL set of harmonic functions $\mathcal{H}_{\rm dl} = \{e^{j(k-1)\pi\beta\xi} : k \in [M]\}$ from the knowledge of its projections on the UL harmonic functions $\mathcal{H}_{\rm ul} = \{e^{j(k-1)\pi\xi} : k \in [M]\}$.

All the UDCT algorithms in the literature implicitly or explicitly do the following: (a) estimate a positive function $\hat{\gamma}(\xi)$ that has exactly (in the noiseless case) or approximately (in the noisy case) the same projections on the UL harmonic functions \mathcal{H}_{dl} as the original ASF $\gamma(\xi)$, (b) use the resulting estimate $\hat{\gamma}(\xi)$ to compute the projections onto the DL set \mathcal{H}_{dl} to recover an estimate of the DL covariance matrix. Note that even in the ideal noiseless case, the mapping from the ASF to the DL projections is a linear map from the infinitedim space of positive functions to the finite-dim space of Mprojections. Therefore, there are generally a large set of ASFs corresponding to a given UL projection $\sigma_{ul}(\gamma)$ produced by a generic ASF $\gamma(\xi)$:

$$\mathcal{A}(\gamma) := \left\{ \mu(\xi) : \int_{-1}^{1} \mu(\xi) e^{j(k-1)\pi\xi} d\xi = [\boldsymbol{\sigma}_{ul}(\gamma)]_k \right\}.$$
(5)

All the UDCT algorithms can be seen in one way or other as simply different strategies for selecting a specific candidate $ALG(\sigma_{ul}(\gamma)) \in \mathcal{A}(\gamma)$ according to specific criteria. One such criterion is to assume that the original ASF γ belongs to a specific subset \mathcal{A}_0 of structured ASFs. Also, the UDCT algorithm can written more generally as a method that produces $ALG(\sigma_{ul}(\gamma)) \in \mathcal{A}(\gamma) \cap \mathcal{A}_0$ when fed with the UL projections $\sigma_{ul}(\gamma)$.

A similar argument applies to the more general case where instead of Σ_{ul} one has access to a collection of N i.i.d. noisy UL channel vectors $\mathbf{y}(s) = \mathbf{h}_{ul}(s) + \mathbf{z}(s), s \in [N]$, gathered via UL pilot transmission where $\mathbf{h}_{ul}(s)$ and $\mathbf{z}(s)$ denote the UL channel vector and additive measurement noise over resource block $s \in [N]$. Then, one can compute the UL sample covariance matrix as

$$\widehat{\boldsymbol{\Sigma}}_{ul} = \frac{1}{N} \sum_{s \in [N]} \mathbf{y}(s) \mathbf{y}(s)^{\mathsf{H}}, \tag{6}$$

and design an algorithm ALG that produces a structured ASF in \mathcal{A}_0 whose UL covariance matrix matches $\widehat{\Sigma}_{ul}$ under suitable metric. For example, [12, 13] propose such algorithms using Non-Negative Least Squares (NNLS) and ℓ_2 -norm projections. However, none of these methods are able to capture the group-sparsity of the ASF in the AoA domain as illustrated in Fig. 1. For example, as we will illustrate via numerical simulations, the NNLS proposed in [12] is able to promote sparsity of ASF in the AoA domain but it does not yield necessarily group-sparse ASF. The ℓ_2 projection method [13], in contrast, is able to produce smooth ASFs but creates out-of-support components in the estimated ASF since ℓ_2 norm is inherently unable to promote sparsity.

Overall it is generally difficult to design suitable regularization methods that promote notion of group-sparsity of

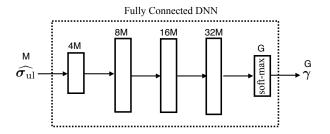


Fig. 2: The structure of DNN adopted for ASF estimation. DNN consists of 5 layers with 2M, 4M, 8M, 16M, and Gneurons, where M is the number of antennas and where G is the ASF quantization grid size. The last layer has a *soft-max* activation function and produces positive values for $\gamma \in \mathbb{R}^G_+$.

ASF we address here. In particular, none of the structured group-sparse estimators widely adopted in sparse signal recovery literature are applicable here [15, 16]. In this paper, we develop such a group-sparsity promoting regularization using DNNs as follows.

(a) Training Data. We first consider a class of group-sparse ASFs \mathcal{A}_0 that may potentially arise in practical propagation scenarios. To capture the notion of group sparsity, we assume that each ASF in \mathcal{A}_0 can be written as $\gamma(\xi) = \sum_{i=1}^{g} \kappa_i p_i(\xi)$ where g denotes the number of groups, where $p_i(\xi)$ is a normalized, i.e., $\int_{-1}^{1} p_i(\xi) d\xi = 1$, positive function with connected support in $[\xi_i, \xi_i + w_i]$ with w_i denoting the angular width of the ASF of the *i*-th group $p_i(\xi)$, and where $\kappa_i \in [0, 1]$ with $\sum_{i=1}^{g} \kappa_i = 1$ are the normalized weights corresponding to ASFs in g groups. For example, Fig. 1 corresponds to an ASF with g = 3 groups.

We use the ASFs in \mathcal{A}_0 to train a DNN as follows. Given a training sample size S, we select S ASFs from \mathcal{A}_0 completely randomly. For each specific ASF γ inside this training set, we compute the first row of the UL covariance matrix as (3) and produce N i.i.d. noisy UL channel vectors and their corresponding sample covariance $\widehat{\Sigma}_{ul}$ as in (6). Since the covariance matrices for ULA are Toeplitz, we toeplitizify $\widehat{\Sigma}_{ul}$ and define the first column of the resulting Toeplitz matrix as $\widehat{\sigma}_{ul}$ where

$$[\widehat{\boldsymbol{\sigma}_{ul}}]_k = \frac{\sum_{i=1}^{M-k+1} [\widehat{\boldsymbol{\Sigma}_{ul}}]_{i,i+k}}{M-k+1}.$$
(7)

We define a uniform quantization grid $\mathcal{G} := \{\xi_i : i \in [G]\}$ over the set of AoAs [-1,1] of size $G \gg M$ where $\xi_i = -1 + \frac{2(i-1)}{G}$ denotes the *i*-th quantization point. We also define the discrete quantization of the ASF $\gamma(\xi)$ over the grid \mathcal{G} as $\gamma = (\gamma(\xi_1), \dots, \gamma(\xi_G))^{\mathsf{T}} \in \mathbb{R}^G_+$. For simplicity, we always normalize γ to make sure that $\sum_{i=1}^G [\gamma]_i = 1$. Finally, we use $(\widehat{\sigma}_{ul}, \gamma)$ as input-output labeled pair for training the DNN illustrated in Fig. 2. By repeating this for all the *S* ASFs selected for training, we obtain a collection of *S* training samples for DNN.

(b) Supervised Learning. We use these S training samples to train a *Deep Neural Networks* (DNN). For training, we use the widely-adopted *Stochastic Gradient Descend* (SGD) with $\ell(\gamma, \hat{\gamma}) = \|\gamma - \hat{\gamma}\|_1$ as the loss function between the true γ and the estimate $\hat{\gamma}$ generated by the network. DNNs have recently been of tremendous importance in Machine Learning and also in many applications in wireless communications, such as signal detection, channel encoding [17], and decoding [18].

(c) Structure of DNN. One of the important factors affecting the performance of the ASF estimation using DNNs is the structure of the DNN consisting of the number of layers, the number of neurons in each layer, and the activation function of each layer. In this paper, we use a fully-connected network illustrated in Fig. 2, with 5 layers consisting of 2M, 4M, 8M, 16M, and G neurons, respectively, where M denotes the number of antennas, where the number of neurons G in the last layer corresponds to the grid size we are adopting for ASF quantization. The activation function of the 4 initial layers is the RelU function $x \mapsto \max\{x, 0\}$. For the last layer we use the *soft-max* activation function, which for an input vectors $(x_1, \ldots, x_G)^{\mathsf{T}}$ in the input produces the output as $(x_1, \ldots, x_G)^{\mathsf{T}} \mapsto \frac{(e^{x_1}, \ldots, e^{x_G})^{\mathsf{T}}}{\sum_{j=1}^{j} e^{x_j}}$. Note that the summation of the elements produced by soft-max layer is always 1, which produces a normalized γ , i.e., $\sum_{i=1}^{G} \gamma_i = 1$, as desired. (d) UDCT using DNN. Once the DNN was suitably trained

(d) UDCT using DNN. Once the DNN was suitably trained we use it as a black-box algorithm for ASF estimation and UDCT. More specifically, given N i.i.d. noisy channel vectors received in UL, we compute the UL sample covariance matrix as in (6) and $\hat{\sigma}_{ul}$ as in (6). Then, we feed $\hat{\sigma}_{ul}$ to the trained DNN and obtain an estimate of the quantized ASF $\hat{\gamma} \in \mathbb{R}^G_+$ where G is the grid size. Then, we build an estimate of the original continuous ASF $\hat{\gamma}(\xi)$ by triangular interpolation. Finally, we use $\hat{\gamma}(\xi)$ to compute the first column of the DL Toeplitz covariance matrix as in (4). Note that DNNs have the fundamental advantage that one does not need to run any iterative algorithm as in [12, 13], which may require many iteration to converge; instead one immediately computes the estimate by straightforward calculations through the network, which can be done even in parallel to obtain a tremendous speed-up.

III. SPARSIFICATION PRECODING

In this section, we provide an step-by-step method to design the sparsifying precoder for a collection of K users based on their estimated DL covariance matrix denoted by Σ_k , $k \in [M]$, where for simplicity we drop the label 'dl'. An essential requirement for the our method is the existence of the common eigen-vector property for the array and in particular for the set of K DL covariance matrices Σ_k , $k \in [K]$. In brief, this property is satisfied when the set of all the DL covariance matrices produced by the array is, at least asymptotically for large M, diagonalizable in the same basis. Fortunately, this assumption is met for Uniform Linear/Planar Array (ULA/UPA) widely adopted in wireless applications since the resulting covariance matrices have Toeplitz/Block-Toeplitz structure. And, it is well-known that, asymptotically for large M, all such matrices are diagonalizable with appropriate DFT (Discrete Fourier Transform) matrices, thus, common eigen-vector property.

In the following, we provide a detailed derivation of the this property for the ULA. We first calculate the circulant approximation of the estimated DL covariance matrices of all users, so that they all share the same set of eigen-vectors, or equivalently virtual beam-space representation.¹ The circulant approximation of large Toeplitz matrices imposes an small error as a result of the application of Szegö Theorem [19]. Let Σ_k , $k \in [K]$, be as before and define the diagonal

¹Recall that circulant matrices are diagonalizable with the orthogonal DFT matrix.

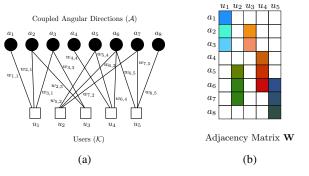


Fig. 3: (a) An example of a bipartite graph \mathcal{L} . (b) The corresponding weighted adjacency matrix \mathbf{W} .

matrices $\mathring{\Lambda}_k = \operatorname{diag}(\mathbf{F}^{\mathsf{H}} \Sigma_k \mathbf{F})$ for $k \in [K]$. There are several ways to define a circulant approximation [20], among which we choose $\mathring{\Sigma}_k = \mathbf{F}\mathring{\Lambda}_k\mathbf{F}^{\mathsf{H}}$. According to Szegö's theorem, for large M, $\mathring{\Lambda}_k$ converges to the diagonal eigenvalue matrix Λ_k of Σ_k , i.e. $\mathring{\Lambda}_k \to \Lambda_k$ as $M \to \infty$. This shows that, with a small error, we can find a set of common eigenvectors for all the estimated DL covariance matrices Σ_k , $k \in [K]$. As a consequence, the DL channel covariance of each user $k \in [K]$ is characterized simply by a vector of eigenvalues $\lambda_k \in \mathbb{R}^M$, with *m*-th element $\lambda_m^{(k)} = [\mathring{\Lambda}^{(k)}]_{m,m}$. In addition, the DFT matrix whose (m, n)-th entry is given by $[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{M}} e^{-j2\pi \frac{mn}{M}}$, $m, n \in [M]$, forms a unitary basis for (approximately) expressing any user channel vector via an (approximated) Karhunen-Loeve expansion. In particular, letting $\mathbf{f}_n := [\mathbf{F}]_{,n}$ denote the n^{th} column of \mathbf{F} , we can express the DL channel vector of user k as

$$\mathbf{h}^{(k)} \approx \sum_{m=0}^{M-1} g_m^{(k)} \sqrt{\lambda_m^{(k)}} \mathbf{f}_n \tag{8}$$

where $g_m^{(k)} \sim C\mathcal{N}(0, 1)$. Here, we wish to design the precoder **B** such that the support of the effective channels $\tilde{\mathbf{h}}_{\text{eff}}^{(k)} = \mathbf{B}\mathbf{h}^{(k)}$ is not larger than T_{dl} , such that all users have a chance of being served. Let $\mathbf{H} = \mathbf{L} \odot \mathbb{G} \in \mathbb{C}^{M \times K}$ denote the matrix of DL channel coefficients expressed in the DFT basis (8), in which each column of **H** represents the coefficients vector of a user, where **L** is a $M \times K$ matrix with elements $[\mathbf{L}]_{m,k} = \sqrt{\lambda_m^{(k)}}$, where $\mathbb{G} \in \mathbb{C}^{M \times K}$ has i.i.d. elements $[\mathbb{G}]_{m,k} = g_m^{(k)}$, and where \odot denotes the Hadamard (elementwise) product.

To design the sparsifying precoder, we first illustrate the joint user-beam association of all K users by a graphical model. Let $\mathbf{A} = [\mathbf{L}]$ denote a one-bit thresholded version of \mathbf{L} , such that $[\mathbf{A}]_{m,k} = 1$ if $\lambda_m^{(k)} > \epsilon$, where $\epsilon > 0$ is a suitable small threshold, used to identify the components that are significantly larger than 0 from the "almost zero" ones, and consider the $M \times K$ bipartite graph $\mathcal{L} = (\mathcal{A}, \mathcal{K}, \mathcal{E})$ with adjacency matrix \mathbf{A} and weights $w_{m,k} = \lambda_m^{(k)}$ on the edges $(m,k) \in \mathcal{E}$. An example of the bipartite graph \mathcal{L} and its corresponding weighted adjacency matrix \mathbf{W} is illustrated in Figs. 3a and 3b. Given a pilot dimension T_{dl} , our goal consists in selecting a subgraph $\mathcal{L}' = (\mathcal{A}', \mathcal{K}', \mathcal{E}')$ of \mathcal{L} in which each node on either side of the graph has a degree at least 1 and such that

- For all k ∈ K' we have deg_{L'}(k) ≤ T_{dl}, where deg_{L'} denotes the degree of a node in the selected subgraph.
- The sum of weights of the edges adjacent to any node k ∈ K' in the subgraph L' is greater than a threshold,

i.e. $\sum_{m \in \mathcal{N}_{\mathcal{L}'}(k)} w_{m,k} \ge P_0, \ \forall k \in \mathcal{K}'$, where $\mathcal{N}_{\mathcal{L}'}(k)$ denotes the set of neighbors in \mathcal{L}' of node k.

The channel matrix H_{A',K'} obtained from H by selecting a ∈ A' (referred to as "selected beam directions") and k ∈ K' (referred to as "selected users") has large rank.

The first criterion enables the stable estimation of the effective channel of any selected user with only $T_{\rm dl}$ common pilot dimensions and $T_{\rm dl}$ complex symbols of feedback per selected user. The second criterion makes sure that the effective channel strength of any selected user is greater than a certain desired threshold. The third criterion is motivated by the fact that the DL multiplexing gain is given by rank($\mathbf{H}_{\mathcal{A}',\mathcal{K}'}$) × max{0, 1 – $T_{\rm dl}/T$ }, where T denotes the whole number of REs, and it is obtained by serving a number of users equal to the rank of the effective channel matrix.

In [21] we show that the rank of **H** is given, with probability 1, by the size of the largest intersection submatrix whose associated bipartite graph contains a perfect matching.² With this observation, we can formulate the problem mentioned above as follows. Let $\mathcal{M}(\mathcal{A}', \mathcal{K}')$ denote a matching of the subgraph $\mathcal{L}'(\mathcal{A}', \mathcal{K}', \mathcal{E}')$ of the bipartite graph $\mathcal{L}(\mathcal{A}, \mathcal{K}, \mathcal{E})$. Find the solution of the following optimization problem:

$$\max_{\mathcal{A}' \subseteq \mathcal{A}, \mathcal{K}' \subseteq \mathcal{K}} \max \left| \mathcal{M} \left(\mathcal{A}', \mathcal{K}' \right) \right|$$
(9a)

subject to
$$\deg_{\mathcal{L}'}(k) \leq T_{\mathrm{dl}} \ \forall k \in \mathcal{K}',$$

$$\sum_{a \in \mathcal{N}_{\mathcal{L}'}(k)} w_{a,k} \ge P_0, \ \forall k \in \mathcal{K}'.$$
(9c)

(9b)

It turns out that, this problem can be transformed into the following equivalent mixed integer linear program (MILP), denoted by $\mathcal{P}_{\text{MILP}}$ [21]:

$$\underset{x_{m}, y_{k}, z_{m,k}}{\text{maximize}} \quad \sum_{m \in \mathcal{A}} \sum_{k \in \mathcal{K}} z_{m,k} + \epsilon \sum_{m \in \mathcal{A}} x_{m}$$
(10a)

subject to
$$z_{m,k} \leq [\mathbf{A}]_{m,k} \quad \forall m \in \mathcal{A}, k \in \mathcal{K},$$
 (10b)

$$\sum_{k \in \mathcal{K}} z_{m,k} \leqslant x_m \quad \forall m \in \mathcal{A}, \tag{10c}$$

$$\sum_{n \in \mathcal{A}} z_{m,k} \leqslant y_k \quad \forall k \in \mathcal{K},$$
(10d)

$$\sum_{m \in \mathcal{A}} [\mathbf{A}]_{m,k} x_m \leqslant T_{\mathrm{dl}} y_k + M(1 - y_k) \quad \forall k \in \mathcal{K},$$
(10e)

$$P_0 y_k \leqslant \sum [\mathbf{W}]_{m,k} x_m \quad \forall k \in \mathcal{K},$$
 (10f)

$$x_m \leqslant \sum_{k \in \mathcal{K}}^{m \in \mathcal{A}} [\mathbf{A}]_{m,k} y_k \quad \forall m \in \mathcal{A},$$
(10g)

$$x_m, y_k \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K},$$
(10h)

$$z_{m,k} \in [0,1] \quad \forall m \in \mathcal{A}, k \in \mathcal{K}, \tag{10i}$$

The binary variable x_m in (10) determines whether beam $m \in \mathcal{A}$ is selected or not, i.e., $x_m = 1$ if and only if the *m*-th beam is selected. Similarly, y_k controls the selection of user *k* for channel estimation (and eventually serving). For a given set of user DL covariance matrices, we denote by $\mathcal{B} = \{m : x_m^* = 1\} = \{m_1, m_2, \ldots, m_{M'}\}$ the set of selected beams directions of cardinality $|\mathcal{B}| = M'$ and by $\mathcal{K} = \{k : y_k^* = 1\}$ the set of selected users of cardinality $|\mathcal{K}| = K'$, where $\{x_m^*\}_{m=1}^M$ and $\{y_k^*\}_{k=1}^K$ are solutions to $\mathcal{P}_{\text{MILP}}$.

The desired sparsifying precoding matrix **B** is finally obtained as $\mathbf{B} = \mathbf{F}_{\mathcal{B}}^{\mathsf{H}}$, where $\mathbf{F}_{\mathcal{B}} = [\mathbf{f}_{m_1}, \dots, \mathbf{f}_{m_{M'}}]$ and

 \mathbf{f}_m denotes the *m*-th column of the $M \times M$ unitary DFT matrix **F**. Using (8), the effective DL channel vectors take on the form

$$\check{\mathbf{h}}_{\text{eff}}^{(k)} = \mathbf{B} \sum_{m \in \mathcal{S}_k} g_m^{(k)} \sqrt{\lambda_m^{(k)}} \mathbf{f}_m = \sum_{i:m_i \in \mathcal{B} \cap \mathcal{S}_k} \sqrt{\lambda_{m_i}^{(k)}} g_{m_i}^{(k)} \mathbf{u}_i,$$
(11)

where S_k is the support of $\mathbf{h}^{(k)}$ in the DFT domain and where \mathbf{u}_i denotes a $M' \times 1$ vector with all zero components but a single "1" in the *i*-th position. With this construction, the number of non-identically zero coefficients for each user k are $|\mathcal{B} \cap \mathcal{S}_k| \leqslant T_{\mathrm{dl}}$ and their positions (encoded in the vectors \mathbf{u}_i in (11)) are known to the BS. Hence, the effective channel vectors can be estimated from the $T_{\rm dl}$ dimensional DL pilot observation with an estimation MSE that vanishes as 1/SNR [21] by increasing the Signal-to-Noise Ration (SNR). With the above precoding, we have $\mathbf{BB}^{\mathsf{H}} = \mathbf{I}_{M'}$. Furthermore, we can choose the DL pilot matrix Ψ to be proportional to a random unitary matrix of dimension $T_{\rm dl} \times M'$, such that $\Psi \Psi^{\sf H} = P_{\rm dl} \mathbf{I}_{T_{\rm dl}}$. After DL pilot transmission, all users $k \in [K]$ send back their noisy observations $\mathbf{y}^{(k)}$ to the BS via analog unquantized feedback. The BS, having an estimate of the DL channel covariance, performs linear MMSE estimation to obtain an estimate of the DL channel.

We conclude by describing the channel precoding step. Let $\hat{\mathbf{H}}_{\text{eff}} = [\hat{\mathbf{h}}_{\text{eff}}^{(1)}, \dots, \hat{\mathbf{h}}_{\text{eff}}^{(K')}]$ be the matrix of the estimated effective DL channels for the selected users. We consider the ZF beamforming matrix V given by the column-normalized version of the Moore-Penrose pseudoinverse of the estimated channel matrix, i.e., $\mathbf{V} = \left(\hat{\mathbf{H}}_{eff}\right)^{\dagger} \mathbf{G}^{1/2}$, where $\left(\hat{\mathbf{H}}_{eff}\right)^{\dagger} = \hat{\mathbf{H}}_{eff} \left(\hat{\mathbf{H}}_{eff}^{\mathsf{H}} \hat{\mathbf{H}}_{eff}\right)^{-1}$ and \mathbf{G} is a diagonal matrix that makes the columns of \mathbf{V} to have unit norm. A channel use of the DL precoded data transmission phase at the k-th user receiver takes on the form $y^{(k)} = (\mathbf{h}^{(k)})^{\mathsf{H}} \mathbf{B}^{\mathsf{H}} \mathbf{V} \mathbf{P}^{1/2} \mathbf{d} + n^{(k)}$, where $\mathbf{d} \in \mathbb{C}^{K' \times 1}$ is a vector of unit-energy user data symbols and **P** is a diagonal matrix defining the power allocation to the DL data streams. The transmit power constraint is given by $tr(\mathbf{B}^{\mathsf{H}}\mathbf{V}\mathbf{P}\mathbf{V}^{\mathsf{H}}\mathbf{B}) = tr(\mathbf{V}^{\mathsf{H}}\mathbf{V}\mathbf{P}) = tr(\mathbf{P}) = P_{\mathrm{dl}}$, where we used $\mathbf{BB}^{\mathsf{H}} = \mathbf{I}_{M'}$ and the fact that $\mathbf{V}^{\mathsf{H}}\mathbf{V}$ has unit diagonal elements by construction. For simulation results, we use the simple uniform power allocation $P_k = P_{dl}/K'$ to each k-th user data stream. The received symbol at user k receiver is given by

$$y^{(k)} = b_{k,k}d_k + \sum_{k' \neq k} b_{k,k'}d_{k'} + n^{(k)}, \qquad (12)$$

where the coefficients $(b_{k,1}, \ldots, b_{k,K'})$ are given by the elements of the $1 \times K'$ row vector $(\mathbf{h}^{(k)})^{\mathsf{H}} \mathbf{B}^{\mathsf{H}} \mathbf{V} \mathbf{P}^{1/2}$. Of course, in the presence of an accurate channel estimation we expect that $b_{k,k} \approx \sqrt{G_k P_k}$ and $b_{k,k'} \approx 0$ for $k' \neq k$. To calculate achievable sum-rate, we assume that all coefficients $(b_{k,1}, \ldots, b_{k,K'})$ are known to the corresponding receiver k. Including the DL training overhead, this yields the rate expression

$$R_{\text{sum}} = \left(1 - \frac{T_{\text{dl}}}{T}\right) \sum_{k \in \mathcal{K}} \mathbb{E}\left[\log\left(1 + \frac{|b_{k,k}|^2}{1 + \sum_{k' \neq k} |b_{k,k'}|^2}\right)\right].$$
(13)

²A matching is a set of edges of a graph without common vertices.

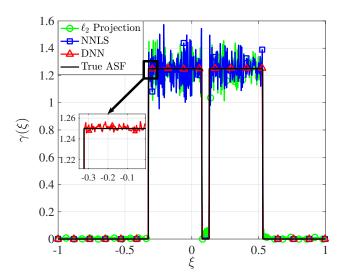


Fig. 4: Comparison of performance of ASF estimation for DNN, NNLS, and ℓ_2 Projection method.

IV. SIMULATION RESULTS

In this section, we perform numerical simulations to assess the performance of our proposed method. For simulations, we consider IMT-FDD band as in LTE standard [22] with a UL band [1920, 1980] MHz and a DL band [2110, 2170] MHz. Thus, we use $f_{\rm ul} = 1950$ and $f_{\rm dl} = 2140$, and set the UDCT parameter β in (4) to $\beta = \frac{f_{\rm dl}}{f_{\rm ul}} \approx 1.1 > 1$. We also set the number of BS antennas to M = 256.

A. ASF Estimation and UL-DL Covariance Transformation

In this section, we perform numerical simulations to compare a DNN trained with group-sparse ASFs with NNLS method in [12] and ℓ_2 projection method in [13].

ASF Estimation. We assume that each ASF in the set of feasible group-sparse ASFs consists of two groups g = 2 where the ASF corresponding to each group is a uniform distribution with a center randomly selected in the range of AoAs [-1,1] and with a random width of size at most $w_{\text{max}} = 0.4$. We consider a SNR of 20 dB for noisy UL channel vectors used for training and set the sampling ratio to $\frac{N}{M} = 2$ (see, e.g., (6)). We refer to Section II for a more detailed discussion of the data set used for training the DNN. We build a training data set of size S = 10000 where we use 80% of this data set for training and the remaining 20% for validation.

Fig. 4 illustrates the simulation results for ASF estimation. It is seen that DNN performs much better than NNLS and ℓ_2 projection. NNLS estimates the support very well but is quite spiky over the support. In contrast, ℓ_2 projection method is quite smooth over the support but produces considerable off-support elements.

UL-DL Covariance Transformation. In this part, we compare the UDCT performance of our proposed DNN with that of NNLS and ℓ_2 projection method for different sampling ratios $\frac{N}{M} \in \{1, \dots, 8\}$. As in the previous part, we assume a SNR of 20 dB for the noisy UL channel vectors. Recall that an UDCT algorithm takes N noisy UL channel vectors as the input and produces an estimate of the DL covariance matrix as the output. We use the following distortion metrics:

1) Normalized Frobenius-norm: defined as

$$E_{\rm NFD} = \mathbb{E}\left\{\frac{\|\boldsymbol{\Sigma}_{\rm dl} - \boldsymbol{\Sigma}_{\rm dl}^{\star}\|_{\rm F}}{\|\boldsymbol{\Sigma}_{\rm dl}\|_{\rm F}}\right\}$$

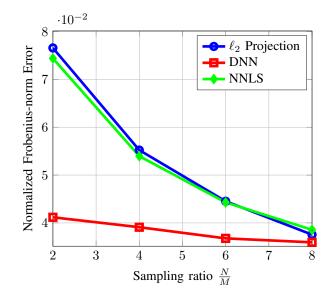


Fig. 5: Comparison of the performance of UDCT for DNN, NNLS, and ℓ_2 Projection method under Normalized Frobenius-norm distortion.

where $\|.\|_{\mathsf{F}}$ denotes the matrix Frobenius norm, where Σ_{dl} and Σ_{dl}^* denote the true and the estimated DL covariance matrices, and where the expectation is taken over random realizations of the UL channel vectors.

2) *Power-loss*: Suppose that the eigenvalue decomposition of the true and estimated DL covariance matrices are given as $\Sigma_{dl} = \mathbf{U}\mathbf{D}\mathbf{U}^{H}$ and $\Sigma_{dl}^{*} = \widetilde{\mathbf{U}}\widetilde{\mathbf{D}}\widetilde{\mathbf{U}}^{H}$, respectively. We define the efficiency metric of order $q \in [M]$ as

$$\eta_q = \frac{\mathsf{tr}\left(\widetilde{\mathbf{U}}_q^{\mathsf{H}} \mathbf{\Sigma}_{\mathrm{dl}} \widetilde{\mathbf{U}}_q\right)}{\mathsf{tr}\left(\mathbf{U}_q^{\mathsf{H}} \mathbf{\Sigma}_{\mathrm{dl}} \mathbf{U}_q\right)} \in [0, 1], \tag{14}$$

where tr(.) denotes the trace operator, and where U_q and \tilde{U}_q denotes $M \times q$ matrices consisting of the first qcolumns of U and \tilde{U} respectively. This metric indicates which fraction of the power lying in the dominant qdim subspace of the true DL covariance Σ_{dl} is captured by the estimated subspace with the same dimension in Σ_{dl}^* spanned by \tilde{U}_q [23, 24]. The closer the efficiency parameter η_q is to one, the better the power in the dominant q-dim subspace of Σ_{dl} is captured by the estimated covariance Σ_{dl}^* . As a worst-case distortion metric that works independent of the dimension of the subspace, we consider

$$E_{\mathsf{PLE}} = 1 - \mathbb{E}\{\min_{q \in [M]} \eta_q\}$$
(15)

where the expected value is taken with respect to the random realization of the noisy UL channel vectors.

Figs. 5 and 6 illustrate the simulation results. It is again seen that, as in the ASF estimation, our proposed method yields better performance in terms of UDCT under both distortion metrics, especially in the practically-relevant regime where the sampling ratio $\frac{N}{M}$ may be very small.

B. Sparsifying Precoder vs. Statistical Beamforming

In this paper, we designed the sparsification precoding from the estimated DL covariance matrices Σ_k , $k \in [K]$ (for simplicity, we dropped the label 'dl'). In the presence of channel sparsification imposed by this precoder, BS still

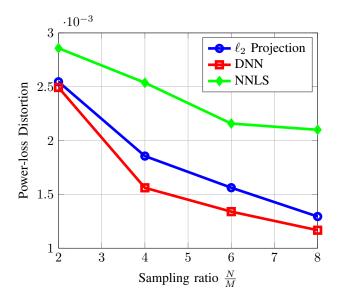


Fig. 6: Comparison of the performance of UDCT for DNN, NNLS, and ℓ_2 Projection method under Power-loss distortion.

needs to probe the reduced-dim channel in the DL and request the users to feedback their channel measurements during the probing to the BS, from which the BS is able to extract the instantaneous channel state of the users after sparsification.

As an alternative to sparsification precoding, one may use the statistical beamforming, which uses only the DL covariance knowledge to beamform to the users, and especially does not apply any DL probing and UL feedback to extract the instantaneous channel of the users [25, 26]. Consequently, the statistical beamforming does not suffer from the fractional rate loss factor $1 - \frac{T_{d1}}{T}$ in (13) caused because of devoting T_{d1} out of T REs to DL channel probing.

We compare the performance of the sparsification precoder with the statistical beamforming method proposed in [26]. Let $\{\Sigma_k\}_{k=1}^K$ denote the DL covariance matrices of the users, either exact or estimated via a UDCT method. In particular, in this section, we provide results for both the case in which the BS has access to exact DL covariances and the case where the DL covariances are estimated using DNN. Given $\{\Sigma_k\}_{k=1}^K$, the beamforming vector for the user k is given by

$$\mathbf{u}_{k} = \mathbf{u}_{\max} \left\{ \left(N_{0} \mathbf{I}_{M} + \sum_{\ell \neq k} \boldsymbol{\Sigma}_{\ell} \right)^{-1} \boldsymbol{\Sigma}_{k} \right\}, \qquad (16)$$

where $\mathbf{u}_{\max}(.)$ denotes the normalized eigenvector corresponding to the maximum eigenvalue of a matrix, and where N_0 denotes the normalized noise power. The statistical beamforming matrix is then given by $\mathbf{V} = [\mathbf{u}_1, \ldots, \mathbf{u}_K]$. We also assume uniform power allocation across the users, just as explained in Section III, and calculate the sum-rate according to (13).

We set the simulation parameters as follows: the BS has M = 256 antennas and serves K = 20 users, each having a randomly generated ASF as before. Fig. 7 illustrates the sumrate vs. pilot dimension results. The blue curves represent the results by assuming true DL covariances, and the red curves represent the results for estimated DL covariances via DNN. As we can see, the performance of the statistical beamforming (SBF) method is substantially inferior to the performance of the sparsification precoding for almost all range of pilot

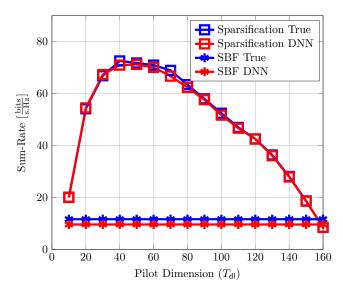


Fig. 7: Sum-rate comparison of the proposed sparsification precoder with statistical beamforming for SNR = 20 dB. We have M = 256 BS antennas and K = 20 served users.

dimensions. This implies that, it is always worthwhile to spend some of resource blocks to estimate instantaneous channels and design the beamformer upon them. This is true, even for small values of the pilot dimension, which implies that even a coarse estimation of the instantaneous channel using a few pilots yields a better performance than the that of statistical beamforming.

V. CONCLUSION

In this paper, we presented a new approach to achieve a very competitive tradeoff between spatial multiplexing gain and probing-feedback overhead in FDD massive MIMO systems. Our approach has two main ingradients: (i) an efficient regularization technique based on Deep Neural Networks (DNN) for structured ASF estimation and UL-DL covariance transformation, (ii) a novel "sparsifying precoding" technique that uses the estimated DL covariance matrices from (i) and shapes the channel sparsity depending on the pilot dimension such that the "effective" channel vectors after sparsification can be estimated at the BS with a low mean-square error. In particular, the proposed sparsifying precoder is not at the mercy of the channel sparsity induced by the propagation environment, which is a bottleneck for Compressed Sensing based channel estimation methods. We compared our proposed DNN-based method in (i) with other methods in the literature via numerical simulations and showed that it yields a very competitive performance. We also compared our sparsifying precoder in (ii) with the state-of-the-art statistical beamforming methods under the assumption that those methods also have access to the covariance knowledge in the DL and showed that our method yields higher spectral efficiency since it uses in addition the instantaneous channel information after sparsification.

REFERENCES

- E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [2] T. L. Marzetta, "How much training is required for multiuser MIMO?" in Fortieth Asilomar Conference on Signals, Systems

and Computers, 2006. ACSSC'06. IEEE, 2006, pp. 359-363.

- [3] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, "Achieving massive MIMO spectral efficiency with a not-solarge number of antennas," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, 2012.
- [4] X. Lin, J. Li, R. Baldemair, T. Cheng, S. Parkvall, D. Larsson, H. Koorapaty, M. Frenne, S. Falahati, A. Grövlen *et al.*, "5g new radio: Unveiling the essentials of the next generation wireless access technology," *arXiv preprint arXiv:1806.06898*, 2018.
- [5] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*. ACM, 2012, pp. 53–64.
- [6] X. Rao and V. K. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, 2014.
- [7] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6169–6183, 2015.
- [8] P.-H. Kuo, H. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Wireless Communications and Networking Conference (WCNC)*, 2012 IEEE. IEEE, 2012, pp. 492–497.
- [9] B. M. Hochwald and T. Maretta, "Adapting a downlink array from uplink measurements," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 642–653, 2001.
- [10] Y. Han, J. Ni, and G. Du, "The potential approaches to achieve channel reciprocity in FDD system with frequency correction algorithms," in *Communications and Networking in China* (CHINACOM), 2010 5th International ICST Conference on. IEEE, 2010, pp. 1–5.
- [11] A. Decurninge, M. Guillaud, and D. T. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *Globecom Workshops (GC Wkshps)*, 2015 *IEEE*. IEEE, 2015, pp. 1–6.
- [12] S. Haghighatshoar, M. B. Khalilsarai, and G. Caire, "Multiband covariance interpolation with applications in massive mimo," in 2018 IEEE International Symposium on Information Theory (ISIT). IEEE, 2018, pp. 386–390.
- [13] L. Miretti, R. L. Cavalcante, and S. Stanczak, "FDD massive MIMO channel spatial covariance conversion using projection methods," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 3609–3613.

- [14] R. L. Cavalcante, L. Miretti, and S. Stańczak, "Error bounds for FDD massive MIMO channel covariance conversion with set-theoretic methods," in 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018, pp. 1–7.
- [15] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on information theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [16] Y. C. Eldar and H. Bolcskei, "Block-sparsity: Coherence and efficient recovery," in 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2009, pp. 2885–2888.
- [17] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2018, pp. 2326–2330.
- [18] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Beery, "Deep learning methods for improved decoding of linear codes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 119–131, Feb 2018.
- [19] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: the large-scale array regime," *IEEE Trans. on Inform. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.
- [20] Z. Zhu and M. B. Wakin, "On the asymptotic equivalence of circulant and toeplitz matrices," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2975–2992, 2017.
- Information Theory, vol. 63, no. 5, pp. 2975–2992, 2017.
 [21] M. B. Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "Fdd massive mimo via ul/dl channel covariance extrapolation and active channel sparsification," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 121–135, 2018.
- [22] (2018) LTE frequency bands. [Online]. Available: https: //en.wikipedia.org/wiki/LTE-frequency-bands
- [23] S. Haghighatshoar and G. Caire, "Massive mimo channel subspace estimation from low-dimensional projections," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 303– 318, 2016.
- [24] —, "Low-complexity massive mimo subspace estimation and tracking from low-dimensional projections," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1832–1844, 2018.
- [25] I. Wajid, Y. C. Eldar, and A. Gershman, "Robust downlink beamforming using covariance channel state information," in 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2009, pp. 2285–2288.
- [26] C. Zhang, Z. Lu, Y. Huang, J. Zhang, and L. Yang, "Statistical beamforming for FDD massive mimo downlink systems," in *Communications in China (ICCC), 2015 IEEE/CIC International Conference on.* IEEE, 2015, pp. 1–6.