

# FDD Massive MIMO via UL/DL Channel Covariance Extrapolation and Active Channel Sparsification

Mahdi Barzegar Khalilsarai<sup>\*</sup>, Saeid Haghighatshoar<sup>\*</sup>, Xinpeng Yi<sup>†</sup>, and Giuseppe Caire<sup>\*</sup>

## Abstract

We propose a novel method for massive Multiple-Input Multiple-Output (massive MIMO) in Frequency Division Duplexing (FDD) systems. Due to the large frequency separation between Uplink (UL) and Downlink (DL), in FDD systems channel reciprocity does not hold. Hence, in order to provide DL channel state information to the Base Station (BS), closed-loop DL channel probing and *Channel State Information* (CSI) feedback is needed. In massive MIMO this incurs typically a large training overhead. For example, in a typical configuration with  $M \simeq 200$  BS antennas and fading coherence block of  $T \simeq 200$  symbols, the resulting rate penalty factor due to the DL training overhead, given by  $\max\{0, 1 - M/T\}$ , is close to 0. To reduce this overhead, we build upon the well-known fact that the *Angular Scattering Function* (ASF) of the user channels is *invariant* over frequency intervals whose size is small with respect to the carrier frequency (as in current FDD cellular standards). This allows to estimate the users' DL channel covariance matrix from UL pilots without additional overhead. Based on this covariance information, we propose a novel *sparsifying precoder* in order to maximize the rank of the effective sparsified channel matrix subject to the condition that each effective user channel has sparsity not larger than some desired DL pilot dimension  $T_{dl}$ , resulting in the DL training overhead factor  $\max\{0, 1 - T_{dl}/T\}$  and CSI feedback cost of  $T_{dl}$  pilot measurements. The optimization of the sparsifying precoder is formulated as a *Mixed Integer Linear Program*, that can be efficiently solved. Extensive simulation results demonstrate the superiority of the proposed approach with respect to concurrent state-of-the-art schemes based on compressed sensing or UL/DL dictionary learning.

## Index Terms

FDD massive MIMO, downlink covariance estimation, active channel sparsification.

<sup>\*</sup> Communications and Information Theory Group, Technische Universität Berlin ({m.barzegarkhalilsarai, saeid.haghighatshoar, caire}@tu-berlin.de).

<sup>†</sup> Department of Electrical Engineering and Electronics, University of Liverpool (xinpeng.yi@liverpool.ac.uk).

## I. INTRODUCTION

Multuser *Multiple-Input Multiple-Output* (MIMO) consists of exploiting multiple antennas at the Base Station (BS) side, in order to multiplex over the spatial domain multiple data streams to multiple users sharing the same time-frequency transmission resource (channel bandwidth and time slots). For a block-fading channel with spatially independent fading and coherence block of  $T$  symbols,<sup>1</sup> the high-SNR sum-capacity behaves as  $C(\text{SNR}) = M^*(1 - M^*/T) \log \text{SNR} + O(1)$ , where  $M^* = \min\{M, K, T/2\}$ ,  $M$  denotes the number of BS antennas, and  $K$  denotes the number of single-antenna users [2–4]. When  $M$  and the number of users are potentially very large, the system *pre-log factor*<sup>2</sup> is maximized by serving  $K = T/2$  data streams (users). While any number  $M \geq K$  of BS antennas yields the same (optimal) *pre-log factor*, a key observation made in [6] is that, when training a very large number of antennas comes at no additional overhead cost, it is indeed convenient to use  $M \gg K$  antennas at the BS. In this way, at the cost of some additional *hardware complexity*, very significant benefits at the system level can be achieved. These include: i) energy efficiency (due to the large beamforming gain); ii) inter-cell interference reduction; iii) a dramatic simplification of user scheduling and rate adaptation, due to the inherent large-dimensional channel hardening [7]. Systems for which the number of BS antennas  $M$  is much larger than the number of DL data streams  $K$  are generally referred to as *massive MIMO* (see [6–8] and references therein). Massive MIMO has been the object of intense research investigation and development and is expected to be a cornerstone of the forthcoming 5th generation of wireless/cellular systems [9].

In order to achieve the benefits of massive MIMO, the BS must learn the downlink channel coefficients for  $K$  users and  $M \gg K$  BS antennas. For Time Division Duplexing (TDD) systems, due to the inherent *Uplink-Downlink* (UL-DL) channel reciprocity [3], this can be obtained from  $K$  mutually orthogonal UL pilots transmitted by the users. Unfortunately, the UL-DL channel reciprocity does not hold for *Frequency Division Duplexing* (FDD) systems, since the UL and

<sup>1</sup>This is the number of signal dimensions over which the fading channel coefficients can be considered constant over time and frequency [1].

<sup>2</sup> With this term we indicate the the number of spatial-domain data streams supported by the system, such that each stream has spectral efficiency that behaves as an interference-free Gaussian channel, i.e.,  $\log \text{SNR} + O(1)$ . In practice, although the system may be interference limited (e.g., due to inter-cell interference in multicell cellular systems), a well-design system would exhibit a regime of practically relevant SNR for which its sum-rate behaves as an affine function of  $\log \text{SNR}$  [5].

DL channels are separated in frequency by much more than the channel coherence bandwidth [1]. Hence, unlike TDD systems, in FDD the BS must actively probe the DL channel by sending a common DL pilot signal, and request the users to feed their channel state back.

In order to obtain a “fresh” channel estimate for each coherence block,  $T_{\text{dl}}$  out of  $T$  symbols per coherence block must be dedicated to the DL common pilot. Assuming (for simplicity of exposition) a delay-free channel state feedback, the resulting DL *pre-log factor* is given by  $K \times \max\{0, 1 - T_{\text{dl}}/T\}$ , where  $K$  is the number of served users, and  $\max\{0, 1 - T_{\text{dl}}/T\}$  is the penalty factor incurred by DL channel training. Conventional DL training consists of sending orthogonal pilot signals from each BS antenna. Thus, in order to train  $M$  antennas, the minimum required training dimension is  $T_{\text{dl}} = M$ . Hence, with such scheme, the number of BS antennas  $M$  cannot be made arbitrarily large. For example, consider a typical case taken from the LTE system [10], where groups of users are scheduled over resource blocks spanning 14 OFDM symbols  $\times$  12 subcarriers, for a total dimension of  $T = 168$  symbols in the time-frequency plane. Consider a typical massive MIMO configuration serving  $K \sim 20$  users with  $M \geq 200$  antennas (e.g., see [11]). In this case, the entire resource block dimension would be consumed by the DL pilot, leaving no room for data communication. Furthermore, feeding back the  $M$ -dimensional measurements (or estimated/quantized channel vectors) represents also a significant feedback overhead for the UL [12–16].

While the argument above is kept informal on purpose, it can be made information-theoretically rigorous. The central issue is that, if one insists to estimate the  $K \times M$  channel matrix in an “agnostic” way, i.e., without exploiting the channel fine structure, a hard dimensionality bottleneck kicks-in and fundamentally limits the number of data streams that can be supported in the DL by FDD systems. It follows that gathering “massive MIMO gains” in FDD systems is a challenging problem. On the other hand, current wireless networks are mostly based on FDD. Such systems are easier to operate and more effective than TDD systems in situations with symmetric traffic and delay-sensitive applications [17–19]. In addition, converting current FDD systems to TDD would represent a non-trivial cost for wireless operators. With these motivations in mind, a significant effort has been recently devoted in order to reduce the common DL training dimension and feedback overhead in order to materialize significant massive MIMO gains also for FDD systems.

### A. Related works: compressed DL pilots

Several works have proposed to reduce both the DL training and UL feedback overheads by exploiting the sparse structure of the massive MIMO channel. In particular, these works assume that propagation between the BS array and the user antenna occurs through a limited number of scattering clusters, with limited support<sup>3</sup> in the *Angle-of-Arrival/Angle-of-Departure* (AoA-AoD) domain.<sup>4</sup> Hence, by decomposing the angle domain into discrete “virtual beam” directions, the  $M$ -dimensional user channel vectors admit a sparse representation in the beam-space domain (e.g., see [20, 21]). Building on this idea, a large number of works (e.g., see [19, 22–28]) proposed to use “compressed pilots”, i.e., a reduced DL pilot dimension  $T_{dl} < M$ , in order to estimate the channel vectors using *Compressed Sensing* (CS) techniques [29, 30]. In [21] sparse representation of channel multipath components in angle, delay and Doppler domains was exploited to propose CS methods for channel estimation using far fewer measurements than required by conventional least-squares (LS) methods. For example, in [24], the authors noticed that the angles of the multipath channel components are common among all the subcarriers in the OFDM signaling and exploited the common sparsity to further reduce the number of required channel measurements. This gives rise to a so-called Multiple Measurement Vector (MMV) setting, arising when multiple snapshot of a random vector with common sparse support can be acquired and jointly processed (e.g., see [31, 32]). This was adapted to FDD in massive MIMO regime were introduced next, where the frequent idea is to probe the channel using compressed pilots in the downlink, receiving the measurements at the BS via feedback and performing channel estimation there. A recent work based on this approach was presented in [19], starting with the observation that, as shown in many experimental studies [33–36], the propagation between the BS antenna array and the users occurs along given scattering clusters, that may be common to multiple users, since they all belong to the same scattering environment. In turns, this yields that the channel sparse representations (in the angle/beam-space domain) share a common part of their support. Hence, [19] considers a scheme where the users feed back their noisy DL pilot measurements to the BS and the latter runs a *joint recovery* algorithm, coined as *Joint Orthogonal Matching Pursuit* (J-OMP), able to take advantage of the common

<sup>3</sup>Throughout the paper the term “support” indicates a set of intervals/indices over which a function/vector has non-zero value.

<sup>4</sup>From the BS perspective, AoD for the DL and AoA for the UL indicate the same domain. Hence, we shall simply refer to this as the “angle domain”, while the meaning of departure (DL) or arrival (UL) is clear from the context.

sparsity. It follows that in the presence of common sparsity, J-OMP improves upon the basic CS schemes that estimate each user channel separately.

More recent CS-based methods, in addition, make use of the *angular reciprocity* between the UL and the DL channels in FDD systems to improve channel estimation. Namely, this refers to the fact that the directions (angles) of propagation for the UL and DL channel are invariant over the frequency range spanning the UL and DL bands, which is generally very small with respect to the carrier frequency (e.g., UL/DL separation of the order of 100MHz, for carrier frequencies ranging between 2 and 6 GHz) [37–39]. In [28] the sparse set of AoAs is estimated from a preamble transmission phase in the UL, and this information is used for user grouping and channel estimation in the DL according to the well-known JSMD paradigm [4, 40]. In [25] the authors proposed a dictionary learning-based approach. First, in a preliminary learning phase a pair of UL-DL dictionaries able to sparsely representing the channel are obtained. Then, these dictionaries are used for a joint sparse estimation of instantaneous UL-DL channels. An issue with this method is that the dictionary learning phase requires off-line training and must be re-run if the propagation environment around the BS changes (e.g., due to large moving objects such as truck and buses, or new building). In addition, the computation involved in the instantaneous channel estimation is prohibitively demanding for real-time operations with a large number of antennas ( $M > 100$ ). In [27] the authors propose estimating the DL channel using a sparse Bayesian learning framework aiming at joint maximum a posteriori (MAP) estimation of the off-grid AoAs and multipath component strength by observing instantaneous UL channel measurements. This method has the drawback that it fundamentally assumes discrete and separable (in the AoA domain) multipath components and assumes that the order of the channel (number of AoA components) is a priori known. Hence, the method simply cannot be applied in the case of continuous (diffuse) scattering, where the scattering power is distributed over a continuous interval of in the angle domain.

### B. Contribution

The focus of this paper is an efficient scheme for massive MIMO in FDD systems. Our goal is to be able to serve as many users as possible even with very small number of DL pilots, compared to the inherent channel dimension. Similar to previous works [19, 25, 27], we consider a scheme where each user sends back its  $T_{dl}$  noisy pilot observations per slot, using unquantized analog

feedback (see [12, 13]). Hence, achieving a small  $T_{dl}$  yields both a reduction of DL training and UL feedback overhead. We summarize the major contributions of our work as follows:

- **DL covariance estimation:** the first problem addressed in this paper is how to estimate DL channel covariance from UL pilot symbols, which are sent anyway in order to enable a coherent multiuser MIMO reception in the UL (see Section III). The covariance matrix can be expressed as an integral transform of the channel *Angular Scattering Function* (ASF), which encodes the signal power distribution over the angle domain. Because of the already mentioned UL/DL angle reciprocity, the channel ASF is invariant with respect to frequency over frequency intervals that are small with respect to the carrier frequency. Stemming from the ASF reciprocity, the idea of UL to DL covariance estimation/transformation is studied in several previous works, including [41–45]. Our approach consists of estimating the channel ASF of each user from UL pilots, and using it to “extrapolate” the covariance matrix from UL to DL. As shown in our recent work [46], this extrapolation problem is non-trivial and must be posed in a robust min-max sense. In [46] we also show that robust covariance reconstruction can be obtained as long as one ensures that the estimated channel ASF is a real, positive function and that its generated UL antenna correlation is consistent with the true UL antenna correlation. Unlike most of the works in the literature, including the ones mentioned above, our covariance extrapolation technique does not rely on any regularity assumption on the ASF. That is to say, we do not assume the ASF to be discrete or sparse, and the estimation method works for a generic ASF. In contrast, it exploits the Toeplitz (resp., block-Toeplitz) structure of the channel covariance matrix resulting from Uniform Linear Arrays (ULA) (resp., Uniform Planar Arrays (UPA)).

- **Active channel sparsification:** the second problem addressed in this paper is how to effectively and *artificially* reduce each user channel dimension, such that a single common DL pilot of assigned dimension  $T_{dl}$  is sufficient to estimate a large number of user channels (see Section IV). In the CS-based works reviewed above, the pilot dimension depends on the channel sparsity level  $s$  (number of non-zero components in the angle/beam-space domain). In fact, standard CS theory states that stable sparse signal reconstruction is possible using  $T_{dl} = O(s \log M)$  measurements.<sup>5</sup> In a rich scattering situation,  $s$  is large or may in fact vary from user to user or in different cell locations. Even if the channel support is known, one needs at least  $s$  measurements for a stable

<sup>5</sup>As commonly defined in the CS literature, we say that a reconstruction method is stable if the resulting MSE vanishes as  $1/\text{SNR}$ , where SNR denotes the Signal-to-Noise Ratio of the measurements.

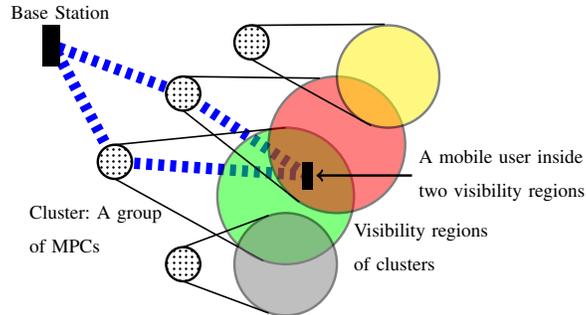


Fig. 1: A sketch of the clusters and visibility regions in the COST 2100 model.

channel estimation. Hence, these CS-based methods (including the ones having access to support information) may or may not work well, depending on the propagation environment. In order to allow channel estimation with an *assigned* pilot dimension  $T_{dl}$ , we use the DL covariance information in order to design an optimal *sparsifying precoder*. This is a linear transformation that depends only on the channel second order statistics (estimated DL covariances) that imposes that the effective channel matrix (including the precoder) has large rank and yet each column has sparsity not larger than  $T_{dl}$ . In this way, our method is not at the mercy of nature, i.e. it is flexible with respect to various types of environments and channel sparsity orders. We cast the optimization of the sparsifying precoder as a Mixed Integer Linear Program (MILP), which can be efficiently solved using standard off-the-shelf solvers.

## II. SYSTEM SETUP

We consider a directional channel propagation model formed by multiple multipath components (MPCs), each corresponding to a scattering cluster characterized by a certain angle width and AoA direction. In addition, as in [19], we consider the possibility that different users have partially overlapped multipath components. An example of such spatially consistent scattering model is provided by the COST 2100 channel model [47], where each MPC is associated to a visibility region, and users inside its visibility region are coupled with the BS array through the corresponding scattering cluster (see Fig. 1).

This model implies that the scattering geometry of the channel between the BS antenna array and the UE antenna remains constant over time intervals corresponding to the UE remaining in the same intersection of visibility regions. Since moving across the regions occurs at a time

scale much larger than moving across one wavelength, it is safe to assume that the channel scattering geometry is locally stationary over intervals much longer than the time scale of the transmission of channel codewords. Such fixed geometry yields the so-called *Wide Sense Stationary Uncorrelated Scattering* (WSSUS) channel model, for which the channel vectors evolve in time according to a WSS processes. Also, we use the ubiquitous block-fading approximation, and assume that the channel random process can be approximated as locally piecewise constant over blocks of  $T$  time-frequency symbols, where  $T \approx W_c T_c$ ,  $W_c$  denoting the channel coherence bandwidth and  $T_c$  denoting the channel coherence time [1]. We consider a BS equipped with an ULA with  $M \gg 1$  antennas and single-antenna UEs.<sup>6</sup> In an FDD system, communication takes place over two disjoint frequency bands. The UEs transmit to the BS over the frequency interval  $[f_{\text{ul}} - \frac{W_{\text{ul}}}{2}, f_{\text{ul}} + \frac{W_{\text{ul}}}{2}]$ , where  $f_{\text{ul}}$  is the UL carrier frequency and  $W_{\text{ul}}$  is the UL bandwidth. Likewise, the BS transmits to the UEs over the frequency band  $[f_{\text{dl}} - \frac{W_{\text{dl}}}{2}, f_{\text{dl}} + \frac{W_{\text{dl}}}{2}]$  where  $f_{\text{dl}}$  is the DL carrier frequency and  $W_{\text{dl}}$  is the DL bandwidth. The channel bandwidth is always much less than the carrier frequency, i.e.  $\frac{W_{\text{ul}}}{f_{\text{ul}}} \ll 1$ ,  $\frac{W_{\text{dl}}}{f_{\text{dl}}} \ll 1$ . Let  $\alpha = \frac{f_{\text{dl}}}{f_{\text{ul}}}$  denote the ratio between the DL and the UL carrier frequencies. Notice that in FDD systems in operation today, we always have  $\alpha > 1$  (e.g., see [48]). A general form for the above WSSUS channel model in the time-frequency-antenna domain is given by

$$\mathbf{h}(t, f) = \int_{\Theta} \rho(t, d\theta) \mathbf{a}(\theta, f) \in \mathbb{C}^M, \quad (1)$$

where  $\Theta := [-\theta_{\text{max}}, \theta_{\text{max}})$  is the angular range scanned by the ULA, the vector  $\mathbf{a}(\theta, f) \in \mathbb{C}^M$  is the *array response* at frequency  $f$  and angle  $\theta$ , with  $m$ -th element given by

$$[\mathbf{a}(\theta, f)]_m = e^{j2\pi \frac{f}{c_0} m d \sin \theta}, \quad (2)$$

where  $c_0$  denotes the speed of light and  $d$  the distance between two consecutive antennas, and  $\rho(t, d\theta)$  is a random gain dependent on the time  $t$  and the angle range  $[\theta, \theta + d\theta]$ . We model  $\rho(t, d\theta)$  to be a *zero-mean* Gaussian stochastic process with independent increments respect to  $\theta$  (uncorrelated scattering) and WSS with respect to  $t$ . The angular autocorrelation function is given by

$$\mathbb{E} [\rho(t, d\theta) \rho(t, d\theta')] = \gamma(d\theta) \delta(\theta - \theta'), \quad (3)$$

<sup>6</sup> The approach of this paper can be immediately generalized to UPAs for 3-dim beamforming. Here we restrict to a planar geometry for the sake of simplicity.

where  $\gamma(d\theta)$  is the channel ASF, modeling the power received from scatterers located at any angular interval. It is convenient to assume that  $\gamma(d\theta)$  is a normalized density function, such that  $\int_{\Theta} \gamma(d\theta) = 1$ . Based on the narrow-band assumption we consider the array response to be a constant function of frequency over each of the UL and DL bands separately and write  $\mathbf{a}_{\text{ul}}(\theta) := \mathbf{a}(\theta, f_{\text{ul}})$  and  $\mathbf{a}_{\text{dl}}(\theta) := \mathbf{a}(\theta, f_{\text{dl}})$ . We let  $d = \kappa \frac{\lambda_{\text{ul}}}{2 \sin(\theta_{\text{max}})}$ , where  $\lambda_{\text{ul}} = \frac{f_{\text{ul}}}{c_0}$  is the UL carrier wavelength and  $\kappa$  is the spatial oversampling factor, usually (including here) set to  $\kappa = 1$ . With this definition we have that  $[\mathbf{a}_{\text{ul}}(\theta)]_m = e^{jm\pi \frac{\sin(\theta)}{\sin(\theta_{\text{max}})}}$  and  $[\mathbf{a}_{\text{dl}}(\theta)]_m = e^{jm\pi \alpha \frac{\sin(\theta)}{\sin(\theta_{\text{max}})}}$ . Notice that the exponents of the array response elements for UL and DL differ by the factor  $\alpha$ , which is typically slightly larger than 1 (e.g., for the LTE-IMT bands we have  $\alpha = \frac{2140}{1950} \approx 1.1$  [48]).

The channel vector covariance matrix is thereby given as follows

$$\mathbf{C}_{\mathbf{h}}(f) = \mathbb{E} [\mathbf{h}(t, f)\mathbf{h}(t, f)^{\text{H}}] = \int_{\Theta} \gamma(d\theta) \mathbf{a}(\theta, f) \mathbf{a}(\theta, f)^{\text{H}}, \quad (4)$$

which is time-invariant due to stationarity. The dependence of the covariance matrix on frequency is due to the fact that, as discussed before, the array response vector is a function of frequency. The covariance matrix is Toeplitz positive semidefinite Hermitian and hence can be described by its first column  $\mathbf{c}(f)$  as  $\mathbf{C}_{\mathbf{h}}(f) = \mathcal{T}(\mathbf{c}(f))$ ,<sup>7</sup> where the first column is given by  $\mathbf{c}(f) = \int_{\Theta} \gamma(d\theta) \mathbf{a}(\theta, f)$ . We denote UL and DL covariance matrices by  $\mathbf{C}_{\text{ul}} := \mathbf{C}_{\mathbf{h}}(f_{\text{ul}})$  and  $\mathbf{C}_{\text{dl}} := \mathbf{C}_{\mathbf{h}}(f_{\text{dl}})$ , respectively.

### III. DL COVARIANCE ESTIMATION FROM UL PILOTS

Our proposed DL covariance estimation method exploits the assumption that the channel ASF is the same for UL and DL (angular reciprocity) [37–39]. Unlike previous works, we do not assume the ASF to be sparse, or to consist of only “discrete” components. In fact, as we have shown in a companion paper [46], any estimate of the ASF that is real, positive and consistent with the UL covariance, regardless of being sparse, is good enough for the purpose of DL covariance estimation.

<sup>7</sup> For  $\mathbf{x} \in \mathbb{C}^M$ , we let  $\mathcal{T}(\mathbf{x})$  denote the Toeplitz Hermitian matrix with first column  $\mathbf{x}$ , i.e., with  $(i, j)$ -th element  $[\mathcal{T}(\mathbf{x})]_{i,j} = x_{i-j}$  for  $i \geq j$  and  $[\mathcal{T}(\mathbf{x})]_{i,j} = x_{|i-j|}^*$  for  $i < j$ . If  $\mathbf{x}$  is a sampled autocorrelation function, then  $\mathcal{T}(\mathbf{x})$  is positive semidefinite.

### A. Uplink covariance estimation

Since the user channel vectors are mutually independent, and we assume Additive White Gaussian Noise (AWGN), the estimation of each user channel covariance in the UL is decoupled and we can focus on the estimation of a generic user. The received UL pilot observation during the  $i$ -th UL coherence block, after projecting over the orthogonal pilot sequence of the given generic user, is given by  $\mathbf{y}[i] = \mathbf{h}_{\text{ul}}[i] + \mathbf{n}[i]$  (see [6]), where  $\mathbf{h}_{\text{ul}}[i]$  denotes the generic user channel vector during the  $i$ -th coherence block and where  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$  is the measurement noise vector. Collecting a window of  $N_{\text{ul}}$  UL measurements and assuming the noise variance  $\sigma^2$  to be known we estimate the UL covariance as follows. We first calculate the sample covariance matrix as  $\tilde{\mathbf{C}}_{\text{ul}} = \frac{1}{N_{\text{ul}}} \sum_{i=1}^{N_{\text{ul}}} \mathbf{y}[i] \mathbf{y}[i]^H$ . The sample covariance is not necessarily Toeplitz and therefore, to improve the estimate, we project it to the Toeplitz, positive semidefinite cone using the following convex program as suggested in [45],

$$\hat{\mathbf{C}}_{\text{ul}} = \arg \min_{\mathbf{X} \in \mathbf{T}_+^M} \|\mathbf{X} - (\tilde{\mathbf{C}}_{\text{ul}} - \sigma^2 \mathbf{I}_M)\|_F, \quad (5)$$

where  $\mathbf{T}_+^M$  is the cone of Toeplitz, Hermitian, positive semidefinite  $M \times M$  matrices and  $\|\cdot\|_F$  is the Frobenius norm. Being a Toeplitz Hermitian matrix,  $\hat{\mathbf{C}}_{\text{ul}}$  can be fully described by its first column which is denoted by  $\hat{\mathbf{c}}_{\text{ul}}$  hereafter.

### B. Estimation of the channel ASF

Define  $\mathcal{G}$  as a uniform grid consisting of  $G \gg M$  discrete angular points  $\{\theta_i\}_{i=1}^G$ , where each point is given by  $\theta_i = \sin^{-1} \left( (-1 + \frac{2(i-1)}{G}) \sin(\theta_{\max}) \right) \in \Theta$ , and define  $\mathbf{G} \in \mathbb{C}^{M \times G}$  to be a matrix whose  $i^{\text{th}}$  column is given by  $\frac{1}{\sqrt{M}} \mathbf{a}_{\text{ul}}(\theta_i)$ ,  $i \in [G]$ . A discrete approximation of the ASF  $\gamma$  on the grid  $\mathcal{G}$  can be written as  $\gamma(d\theta) \approx \sum_{i=1}^G [\mathbf{z}]_i \delta(\theta - \theta_i)$  for some vector  $\mathbf{z} \in \mathbb{R}_+^G$ . We find  $\mathbf{z}$  by solving a non-negative least squares (NNLS) convex optimization program [46]:

$$\mathbf{z}^* = \arg \min_{\mathbf{z} \in \mathbb{R}_+^G} \|\mathbf{G}\mathbf{z} - \hat{\mathbf{c}}_{\text{ul}}\|. \quad (6)$$

The particularly desirable property of NNLS is that, it yields a real, positive approximation of the ASF and, minimizes the  $\ell_2$  distance of its generated UL covariance samples  $\mathbf{G}\mathbf{z}$  and the estimated UL covariance samples  $\hat{\mathbf{c}}_{\text{ul}}$  to satisfy a data consistency constraint. In fact, as we show in [46], positivity and data consistency are the only two requirements needed for guaranteeing a stable DL covariance estimation. Furthermore, the NNLS solution can be efficiently computed via several

convex optimization techniques [49]. By solving (6), the estimated discretized approximation of the ASF is simply given as  $\hat{\gamma}(d\theta) = \sum_{i=1}^G [\mathbf{z}^*]_i \delta(\theta - \theta_i)$ .

### C. Covariance extrapolation via Fourier transform resampling

Building on the theory developed in our companion paper [46], the problem of extrapolating the estimated UL covariance matrix to the DL frequency can be seen as the resampling of the Fourier transform of the channel ASF. To see this, notice that the  $m$ -th components of the first column  $\mathbf{c}_{\text{ul}}$  of  $\mathbf{C}_{\text{ul}}$  are given by

$$[\mathbf{c}_{\text{ul}}]_m = \int_{\Theta} \gamma(d\theta) e^{jm\pi \frac{\sin \theta}{\sin \theta_{\max}}} = \int_{-1}^1 \gamma(d\xi) e^{jm\pi \xi}, \quad m \in [M], \quad (7)$$

where we introduce the change of variable  $\xi = \frac{\sin \theta}{\sin \theta_{\max}}$ . Define the continuous Fourier transform of the positive measure  $\gamma(d\xi)$  as  $\check{\gamma}(x) = \int_{-1}^1 \gamma(d\xi) e^{jx\pi \xi}$ . Then it is clear from (7) that  $[\mathbf{c}_{\text{ul}}]_m = \check{\gamma}(m)$ ,  $m \in [M]$ . In words, the first column of the UL covariance matrix is simply a sampling of the Fourier transform of the positive measure  $\gamma(d\xi)$  at points  $m = 0, \dots, M-1$ . Taking similar steps, one can show that the components of the first column of the DL covariance matrix are given by  $[\mathbf{c}_{\text{dl}}]_m = \int_{-1}^1 \gamma(d\xi) e^{j\alpha m \pi \xi}$ ,  $m \in [M]$  and hence  $[\mathbf{c}_{\text{dl}}]_m = \check{\gamma}(\alpha m)$ ,  $m \in [M]$ . Estimating the DL covariance from the UL covariance is equivalent to resampling  $\check{\gamma}(\cdot)$  over a grid  $\{0, \alpha, 2\alpha, \dots, (M-1)\alpha\}$ , knowing its samples at the integer grid  $\{0, 1, 2, \dots, M-1\}$ . Summarizing, the proposed DL covariance estimation method consists of the following steps:

- 1) Estimate a discrete approximation of the positive measure  $\gamma(d\theta)$  using the the UL sample covariance estimator and solving (6). The samples of the Fourier transform of this measure on the grid  $\{0, \dots, M-1\}$  asymptotically converge to those generated from the true angular scattering function [50] for large sample size  $N_{\text{ul}}$ .

- 2) Calculate the Fourier transform of the estimated measure on the grid  $\{\alpha m\}_{m=0}^{M-1}$  to obtain the estimated DL antenna autocorrelation function

$$[\hat{\mathbf{c}}_{\text{dl}}]_m = \sum_{i=1}^G \hat{\gamma}(\theta_i) e^{j\alpha(m-1)\pi \frac{\sin \theta_i}{\sin \theta_{\max}}}, \quad m \in [M]. \quad (8)$$

The resulting DL covariance matrix is given by the Toeplitz completion  $\hat{\mathbf{C}}_{\text{dl}} = \mathcal{T}(\hat{\mathbf{c}}_{\text{dl}})$ . As a final remark in this section, notice that the above DL covariance estimation method does not rely on particular features of the channel ASF. For example, it does not require that the ASF has a sparse or discrete support, as needed in other ad-hoc methods (e.g., see [28, 39, 44]).

#### D. Circulant approximation of the DL covariance matrices

The DL covariance estimation from UL pilot signals is performed for all the users  $k \in [K]$  at the BS. These covariance matrices are Toeplitz by construction, due to the structure of the ULA as described before. In Section IV we will introduce the novel idea of active channel sparsification where, for a given DL pilot dimension, the BS selects a set of angular directions to transmit data to the users, such that the number of DL data streams that the system can support is maximized. A necessary step before performing sparsification is that all of the estimated DL covariance matrices share a common set of eigenvectors, namely, the same virtual beam-space representation. In the massive MIMO regime where  $M \gg 1$ , this is possible by considering the circulant approximation of Toeplitz matrices that follows as an application of Szegő Theorem (see details in [4] and references therein). Let  $\mathbf{C}_k$  denote the estimated DL channel covariance of user  $k$  for  $k \in [K]$ , where from now on we shall drop the subscript “dl” since it is clear from the context, as we consider only DL multiuser MIMO transmission. Define the diagonal matrices  $\mathring{\Lambda}_k$ ,  $k \in [K]$  for which  $[\mathring{\Lambda}_k]_{m,m} = [\mathbf{F}^H \mathbf{C}_k \mathbf{F}]_{m,m}$ , where  $\mathbf{F}$  is the  $M \times M$  DFT matrix, whose  $(m, n)$ -th entry is given by  $[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{M}} e^{-j2\pi \frac{mn}{M}}$ ,  $m, n \in [M]$ . There are several ways to define a circulant approximation [51], among which we choose the following:

$$\mathring{\mathbf{C}}_k = \mathbf{F} \mathring{\Lambda}_k \mathbf{F}^H. \quad (9)$$

According to Szegő’s theorem, for large  $M$ ,  $\mathring{\Lambda}_k$  converges to the diagonal eigenvalue matrix  $\Lambda_k$  of  $\mathbf{C}_k$ , i.e.  $\mathring{\Lambda}_k \rightarrow \Lambda_k$  as  $M \rightarrow \infty$ . Hence, within a small error for large  $M$ , the sought set of (approximate) common eigenvectors for all the users is provided by the columns of the  $M \times M$  DFT matrix. As a consequence, the DL channel covariance of user  $k$  is characterized simply via a vector of eigenvalues  $\boldsymbol{\lambda}^{(k)} \in \mathbb{R}^M$ , with  $m$ -th element  $[\boldsymbol{\lambda}^{(k)}]_m = [\mathring{\Lambda}^{(k)}]_{m,m}$ . In addition, the DFT matrix forms a unitary basis for (approximately) expressing any user channel vector via an (approximated) Karhunen-Loeve expansion. In particular, let  $\mathbf{f}_m := [\mathbf{F}]_{:,m}$  denote the  $m$ -th column of  $\mathbf{F}$ . We can express the DL channel vector of user  $k$  as

$$\mathbf{h}^{(k)} \approx \sum_{m=0}^{M-1} g_m^{(k)} \sqrt{[\boldsymbol{\lambda}^{(k)}]_m} \mathbf{f}_m, \quad (10)$$

where  $g_m^{(k)} \sim \mathcal{CN}(0, 1)$  are i.i.d. random variables. The columns of  $\mathbf{F}$  are very similar to array response vectors and in fact, recalling equation (2), we have that  $\mathbf{f}_m = \frac{1}{\sqrt{M}} \mathbf{a}_{\text{dl}}(\sin^{-1}(\frac{\lambda_{\text{dl}}}{d} \frac{m}{M}))$ .

Hence, each column with index  $m \in [M]$  of the DFT matrix can be seen as the array response to an angular direction and  $[\boldsymbol{\lambda}^{(k)}]_m$  can be seen as the power of the channel vector associated with user  $k$  along that direction. Due to the limited number of local scatterers as seen at the BS and the large number of antennas of the array, only a few entries of  $\boldsymbol{\lambda}^{(k)}$  are significantly large, implying that the DL channel vector  $\mathbf{h}^{(k)}$  is sparse in the Fourier basis. This sparsity in the beam-space domain is precisely what has been exploited in the CS-based works discussed in Section I-A, in order to reduce the DL pilot dimension  $T_{\text{dl}}$ . It is also evident that this channel representation combined with the geometrically consistent model reviewed in Section II yields the common sparsity across users, as exploited by J-OMP in [19]. As seen in the next section, our proposed approach does not rely on any intrinsic channel sparsity assumption, but adopts a novel artificial sparsification technique.

#### IV. ACTIVE CHANNEL SPARSIFICATION AND DL CHANNEL PROBING

In this section we consider the estimation of the *instantaneous realization* of the DL user channel vectors. As in [4], we consider the concatenation of the physical channel with a fixed precoder, i.e., a linear transformation that may depend on the user channel statistics (notably, on their covariance matrices estimated as explained in Section III), but is independent of the instantaneous channel realizations, which in fact must be estimated via the closed-loop DL probing and channel state feedback mechanism as discussed in Section I.

The BS transmits a training space-time matrix  $\boldsymbol{\Psi}$  of dimension  $T_{\text{dl}} \times M'$ , such that each row  $\boldsymbol{\Psi}_{i,\cdot}$  is transmitted simultaneously from the  $M' \leq M$  inputs of a precoding matrix  $\mathbf{B}$  of dimension  $M' \times M$ , and where  $M'$  is a suitable intermediate dimension that will be determined later. The precoded DL training length (in time-frequency symbols) spans therefore  $T_{\text{dl}}$  dimensions, and the DL training phase is repeated at each DL slot of dimension  $T$ . Stacking the  $T_{\text{dl}}$  DL training symbols in a column vector, the corresponding observation at the UE  $k$  receiver is given by

$$\mathbf{y}^{(k)} = \boldsymbol{\Psi} \mathbf{B} \mathbf{h}^{(k)} + \mathbf{n}^{(k)} = \boldsymbol{\Psi} \tilde{\mathbf{h}}_{\text{eff}}^{(k)} + \mathbf{n}^{(k)}, \quad (11)$$

where  $\mathbf{B}$  is the precoding matrix,  $\mathbf{h}^{(k)}$  is the channel vector of user  $k$ , and we define  $\tilde{\mathbf{h}}_{\text{eff}}^{(k)} := \mathbf{B} \mathbf{h}^{(k)}$  as the effective channel vector, formed by the concatenation of the actual DL channel (antenna-to-antenna) with the precoder  $\mathbf{B}$ . The measurement noise is AWGN with distribution

$\mathbf{n}^{(k)} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_{T_{\text{dl}}})$ . The training matrix and precoding matrix are normalized such that

$$\text{tr}(\Psi \mathbf{B} \mathbf{B}^H \Psi^H) = T_{\text{dl}} P_{\text{dl}}, \quad (12)$$

where  $P_{\text{dl}}$  denotes the total BS transmit power and we define the DL SNR as  $\text{SNR} = P_{\text{dl}}/N_0$ . Notice that most works on channel estimation focus on the estimation of the actual channels  $\{\mathbf{h}^{(k)}\}$ . This is recovered in our setting by letting  $\mathbf{B} = \mathbf{I}_M$ . However, our goal here is to design a *sparsifying precoder*  $\mathbf{B}$  such that each user effective channel has low dimension (in the beam-space representation) and yet the collection of effective channels for  $k \in [K]$  form a high-rank matrix. In this way, each user channel can be estimated using a small pilot overhead  $T_{\text{dl}}$ , but the BS is still able to serve many data streams using spatial multiplexing in the DL (in fact, as many as the rank of the effective matrix).

#### A. Necessity and implication of stable channel estimation

For simplicity of exposition, in this section we assume that the channel representation (10) holds exactly and that the eigenvalue vectors  $\boldsymbol{\lambda}^{(k)}$  have support  $\mathcal{S}_k = \{m : [\boldsymbol{\lambda}^{(k)}]_m \neq 0\}$  with sparsity level  $s_k = |\mathcal{S}_k|$ . We hasten to say that the above are convenient *design assumptions*, made in order to obtain a tractable problem, and that the precoder designed according to our simplifying assumption is applied to the actual physical channels. Under these assumptions, the following lemma yields necessary and sufficient conditions of stable estimation of the channel vectors  $\mathbf{h}^{(k)}$ .

**Lemma 1:** Consider the sparse Gaussian vector  $\mathbf{h}^{(k)}$  with support set  $\mathcal{S}_k$  given by the RHS of (10). Let  $\hat{\mathbf{h}}^{(k)}$  denote any estimator for  $\mathbf{h}^{(k)}$  based on the observation<sup>8</sup>  $\mathbf{y}^{(k)} = \Psi \mathbf{h}^{(k)} + \mathbf{n}^{(k)}$ , and let  $\mathbf{R}_e = \mathbb{E}[(\mathbf{h}^{(k)} - \hat{\mathbf{h}}^{(k)})(\mathbf{h}^{(k)} - \hat{\mathbf{h}}^{(k)})^H]$  denote the corresponding estimation error covariance matrix. If  $T_{\text{dl}} \geq s_k$  there exist pilot matrices  $\Psi \in \mathbb{C}^{T_{\text{dl}} \times M}$  for which  $\lim_{N_0 \downarrow 0} \text{tr}(\mathbf{R}_e) = 0$  for all support sets  $\mathcal{S}_k : |\mathcal{S}_k| = s_k$ . Conversely, for any support set  $\mathcal{S}_k : |\mathcal{S}_k| = s_k$  any pilot matrix  $\Psi \in \mathbb{C}^{T_{\text{dl}} \times M}$  with  $T_{\text{dl}} < s_k$  yields  $\lim_{N_0 \downarrow 0} \text{tr}(\mathbf{R}_e) > 0$ .  $\square$

*Proof:* See appendix VII-A.  $\blacksquare$

As a direct consequence of Lemma 1, we have that any scheme relying on intrinsic channel sparsity cannot yield stable estimation if  $T_{\text{dl}} < s_k$  for some users. Furthermore, we need to impose that the effective channel sparsity (after the introduction of the sparsifying precoder

<sup>8</sup>Note that this coincides with (11) with  $\mathbf{B} = \mathbf{I}_M$ , i.e., without the sparsifying precoder.

B) is less or equal to the desired DL pilot dimension  $T_{\text{dl}}$ . It is important to note that the requirement of estimation stability is *essential* in order to achieve high spectral efficiency in high SNR conditions, irrespectively of the DL precoding scheme. In fact, if the estimation MSE of the user channels does not vanish as  $N_0 \downarrow 0$ , the system self-interference due to the imperfect channel knowledge grows proportionally to the signal power, yielding a Signal-to-Interference plus Noise Ratio (SINR) that saturates to a constant when SNR becomes large. Hence, for sufficiently high SNR, the best strategy would consist of transmitting just a single data stream, since any form of multiuser precoding would inevitably lead to an interference limited regime, where the sum rate remains bounded while  $\text{SNR} \rightarrow \infty$  [52]. In contrast, it is also well-known that when the channel estimation error vanishes as  $O(N_0)$  for  $N_0 \downarrow 0$ , the high-SNR sum rate behaves as if the channel was perfectly known and can be achieved by very simple linear precoding [12]. A possible solution to this problem consists of serving only the users whose channel support  $s_k$  is not larger than  $T_{\text{dl}}$ . This is assumed *implicitly* in all CS-based schemes (see Section I-A), and represents a major intrinsic limitation of the CS-based approaches. In contrast, by artificially sparsifying the user channels, we manage to serve all users given a fixed DL pilot dimension  $T_{\text{dl}}$ .

### B. Sparsifying precoder optimization

Before proceeding in this section, we introduce some graph-theoretic terms [53]. A bipartite graph is a graph whose vertices (nodes) can be divided into two sets  $\mathcal{V}_1$  and  $\mathcal{V}_2$ , such that every edge in the set of graph edges  $\mathcal{E}$  connects a vertex in  $\mathcal{V}_1$  to one in  $\mathcal{V}_2$ . One can denote such a graph by  $\mathcal{L} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ . A subgraph of  $\mathcal{L}$  is a graph  $\mathcal{L}' = (\mathcal{V}'_1, \mathcal{V}'_2, \mathcal{E}')$  such that  $\mathcal{V}'_1 \subseteq \mathcal{V}_1$ ,  $\mathcal{V}'_2 \subseteq \mathcal{V}_2$  and  $\mathcal{E}' \subseteq \mathcal{E}$ . With regards to  $\mathcal{L}$ , the following terms shall be defined and later used.

- **Degree of a vertex:** for a vertex  $x \in \mathcal{V}_1 \cup \mathcal{V}_2$ , the degree of  $x$  refers to the number of edges in  $\mathcal{E}$  incident to  $x$  and is denoted by  $\text{deg}_{\mathcal{L}}(x)$ .
- **Neighbors of a vertex:** the neighbors of a vertex  $x \in \mathcal{V}_1 \cup \mathcal{V}_2$  are the set of vertices  $y \in \mathcal{V}_1 \cup \mathcal{V}_2$  connected to  $x$ . This set is denoted by  $\mathcal{N}_{\mathcal{L}}(x)$ .
- **Matching:** a matching in  $\mathcal{L}$  is a subset of edges in  $\mathcal{E}$  without common vertices.
- **Maximal matching:** a maximal matching  $\mathcal{M}$  of  $\mathcal{L}$  is a matching with the property that if any edge outside  $\mathcal{M}$  and in  $\mathcal{E}$  is added to it, it is no longer a matching.
- **Perfect matching:** a perfect matching in  $\mathcal{L}$  is a matching that covers all vertices of  $\mathcal{L}$ .

We propose to design the sparsifying precoder using a graphical model, where a bipartite graph is formed by a set of vertices representing users on one side and another set of vertices representing beams on the other side. An edge of the bipartite graph between a beam and a user represents the presence of that beam in the user angular profile, with its weight denoting the user channel power along that beam. Now, we wish to design the precoder  $\mathbf{B}$  such that the support of the effective channels  $\check{\mathbf{h}}_{\text{eff}}^{(k)} = \mathbf{B}\mathbf{h}^{(k)}$  is not larger than  $T_{\text{dl}}$  for all  $k$ , such that all users have a chance of being served. Let  $\check{\mathbf{H}} = \mathbf{L} \odot \mathbb{G} \in \mathbb{C}^{M \times K}$  denote the matrix of DL channel coefficients expressed in the DFT basis (10), in which each column of  $\check{\mathbf{H}}$  represents the coefficients vector of a user, where  $\mathbf{L}$  is a  $M \times K$  matrix with elements  $[\mathbf{L}]_{m,k} = \sqrt{[\boldsymbol{\lambda}^{(k)}]_m}$ , where  $\mathbb{G} \in \mathbb{C}^{M \times K}$  has i.i.d. elements  $[\mathbb{G}]_{m,k} = g_m^{(k)} \sim \mathcal{CN}(0, 1)$ , and where  $\odot$  denotes the Hadamard (elementwise) product. Let  $\mathbf{A}$  denote a one-bit thresholded version of  $\mathbf{L}$ , such that  $[\mathbf{A}]_{m,k} = 1$  if  $[\boldsymbol{\lambda}^{(k)}]_m > \text{th}$ , where  $\text{th} > 0$  is a suitable small threshold, used to identify the significant components, and consider the  $M \times K$  bipartite graph  $\mathcal{L} = (\mathcal{A}, \mathcal{K}, \mathcal{E})$  with adjacency matrix  $\mathbf{A}$  and weights  $w_{m,k} = [\boldsymbol{\lambda}^{(k)}]_m$  on the edges  $(m, k) \in \mathcal{E}$ .

Given a pilot dimension  $T_{\text{dl}}$ , 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:

- 1) For all  $k \in \mathcal{K}'$  we have  $\text{deg}_{\mathcal{L}'}(k) \leq T_{\text{dl}}$ , where  $\text{deg}_{\mathcal{L}'}$  denotes the degree of a node in the selected subgraph.
- 2) The sum of weights of the edges incident to any node  $k \in \mathcal{K}'$  in the subgraph  $\mathcal{L}'$  is greater than a threshold, i.e.  $\sum_{m \in \mathcal{N}_{\mathcal{L}'}(k)} w_{m,k} \geq P_0, \forall k \in \mathcal{K}'$ .
- 3) The channel matrix  $\check{\mathbf{H}}_{\mathcal{A}', \mathcal{K}'}$  obtained from  $\check{\mathbf{H}}$  by selecting  $a \in \mathcal{A}'$  (referred to as “selected beam directions”) and  $k \in \mathcal{K}'$  (referred to as “selected users”) has large rank.

The first criterion enables stable estimation of the effective channel of any selected user with only  $T_{\text{dl}}$  common pilot dimensions and  $T_{\text{dl}}$  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 desired threshold, since we do not want to spend resources on probing and serving users with weak effective channels (where “weak” is quantitatively determined by the value of  $P_0$ ). Therefore  $P_0$  is a parameter that serves to obtain a tradeoff between the rank of the effective matrix (which ultimately determines the number of spatially multiplexed DL data streams) and the beamforming gain (i.e., the power effectively conveyed along each selected user effective

channel). The third criterion is motivated by the fact that the DL pre-log factor is given by  $\text{rank}(\check{\mathbf{H}}_{\mathcal{A}', \mathcal{K}'}) \times \max\{0, 1 - T_{\text{dl}}/T\}$ , and it is obtained by serving a number of users equal to the rank of the effective channel matrix. The following lemmas relate the rank of the effective channel matrix to a graph-theoretic quantity, namely, the size of the maximal matching.

**Lemma 2:** [Skeleton or ‘‘CUR’’ decomposition [54]] Consider  $\check{\mathbf{H}} \in \mathbb{C}^{M \times K}$ , of rank  $r$ . Let  $\mathbf{Q}$  be an  $r \times r$  non-singular intersection submatrix obtained by selecting  $r$  rows and  $r$  columns of  $\check{\mathbf{H}}$ . Then, we have  $\check{\mathbf{H}} = \mathbf{C}\mathbf{U}\mathbf{R}$ , where  $\mathbf{C} \in \mathbb{C}^{M \times r}$  and  $\mathbf{R} \in \mathbb{C}^{r \times K}$  are the matrices of the selected columns and rows forming the intersection  $\mathbf{Q}$  and  $\mathbf{U} = \mathbf{Q}^{-1}$ .  $\square$

**Lemma 3:** [Rank and perfect matchings] Let  $\mathbf{Q}$  denote an  $r \times r$  matrix with some elements identically zero, and the non-identically zero elements independently drawn from a continuous distribution. Consider the associated bipartite graph with adjacency matrix  $\mathbf{A}$  such that  $\mathbf{A}_{i,j} = 1$  if  $\mathbf{Q}_{i,j}$  is not identically zero, and  $\mathbf{A}_{i,j} = 0$  otherwise. Then,  $\mathbf{Q}$  has rank  $r$  with probability 1 if and only if the associated bipartite graph contains a perfect matching.  $\square$

*Proof:* See appendix VII-B.  $\blacksquare$

A similar theorem can be found in [55], but we provide a direct proof in Appendix VII-B for the sake of completeness. Lemmas 2 and 3 result in the following corollary, which is an original contribution of this work.

**Corollary 1:** The rank  $r$  of a random matrix  $\check{\mathbf{H}} \in \mathbb{C}^{M \times K}$  with either identically zero elements or elements independently drawn from a continuous distribution is given, with probability 1, by the size of the largest intersection submatrix whose associated bipartite graph (defined as in Lemma 3) contains a perfect matching.  $\square$

Obviously this corollary holds in our case where the non-zero elements of  $\check{\mathbf{H}}$  are drawn from the complex Gaussian distribution. Using Corollary 1 this problem can be formulated as:

**Problem 1:** Let  $T_{\text{dl}}$  denote the available DL pilot dimension and 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:

$$\underset{\mathcal{A}' \subseteq \mathcal{A}, \mathcal{K}' \subseteq \mathcal{K}}{\text{maximize}} \quad |\mathcal{M}(\mathcal{A}', \mathcal{K}')| \quad (13a)$$

$$\text{subject to} \quad \text{deg}_{\mathcal{L}'}(k) \leq T_{\text{dl}} \quad \forall k \in \mathcal{K}', \quad (13b)$$

$$\sum_{a \in \mathcal{N}_{\mathcal{L}'}(k)} w_{a,k} \geq P_0, \quad \forall k \in \mathcal{K}'. \quad (13c)$$

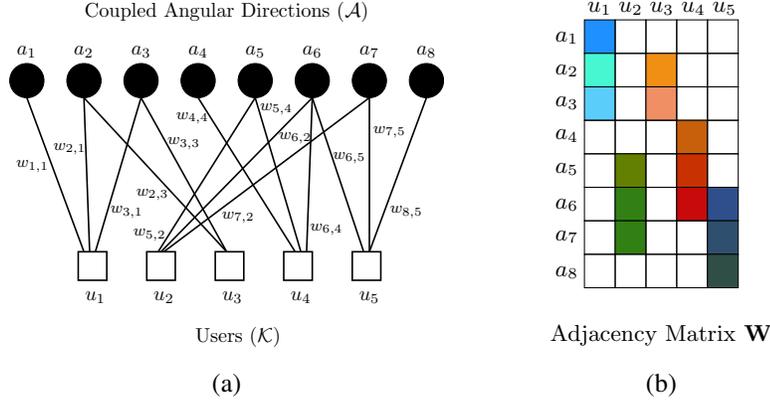


Fig. 2: (a) An example of a bipartite graph  $\mathcal{L}$ . (b) The corresponding weighted adjacency matrix  $\mathbf{W}$ .

◇

The following theorem shows that Problem 1 can be solved in a tractable way.

**Theorem 1:** The optimization problem in (13) is equivalent to the mixed integer linear program (MILP) below:

$$\mathcal{P}_{\text{MILP}} : \underset{x_m, y_k, z_{m,k}}{\text{maximize}} \quad \sum_{m \in \mathcal{A}} \sum_{k \in \mathcal{K}} z_{m,k} \quad (14a)$$

$$\text{subject to} \quad z_{m,k} \leq [\mathbf{A}]_{m,k} \quad \forall m \in \mathcal{A}, k \in \mathcal{K}, \quad (14b)$$

$$\sum_{k \in \mathcal{K}} z_{m,k} \leq x_m \quad \forall m \in \mathcal{A}, \quad (14c)$$

$$\sum_{m \in \mathcal{A}} z_{m,k} \leq y_k \quad \forall k \in \mathcal{K}, \quad (14d)$$

$$\sum_{m \in \mathcal{A}} [\mathbf{A}]_{m,k} x_m \leq T_{\text{dl}} y_k + M(1 - y_k) \quad \forall k \in \mathcal{K}, \quad (14e)$$

$$P_0 y_k \leq \sum_{m \in \mathcal{A}} [\mathbf{W}]_{m,k} x_m \quad \forall k \in \mathcal{K}, \quad (14f)$$

$$x_m \leq \sum_{k \in \mathcal{K}} [\mathbf{A}]_{m,k} y_k \quad \forall m \in \mathcal{A}, \quad (14g)$$

$$x_m, y_k \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \quad (14h)$$

$$z_{m,k} \in [0, 1] \quad \forall m \in \mathcal{A}, k \in \mathcal{K}, \quad (14i)$$

where  $\mathbf{W}$  is the  $|\mathcal{A}| \times |\mathcal{K}|$  weighted adjacency matrix in which  $[\mathbf{W}]_{m,k} = w_{m,k}$  (see the example in Fig. 2a and Fig. 2b). The solution sub-graph is given by the set of nodes  $\mathcal{A}' = \{m : x_m^* = 1\}$

and  $\mathcal{K}' = \{k : y_k^* = 1\}$ , with  $\{x_m^*\}_{m=1}^M$  and  $\{y_k^*\}_{k=1}^K$  being a solution of (14).  $\square$

*Proof:* See Appendix VII-C.  $\blacksquare$

The introduced MILP can be efficiently solved using an off-the-shelf optimization toolbox. The solution to this optimization, however, is not necessarily unique, i.e. there may exist several sub-graphs with the same (maximum) matching size. In order to limit the solution set we introduce a regularization term to the objective of (14) to favor solutions containing more “active” beams. The regularized form of (14) is given as

$$\begin{aligned} \mathcal{P}_{\text{MILP}} : \quad & \underset{x_m, y_k, z_{m,k}}{\text{maximize}} \quad \sum_{m \in \mathcal{A}} \sum_{k \in \mathcal{K}} z_{m,k} + \epsilon \sum_{m \in \mathcal{A}} x_m \\ & \text{subject to} \quad \{x_m, y_k, z_{m,k}\}_{m \in \mathcal{A}, k \in \mathcal{K}} \in \mathcal{S}_{\text{feasible}}, \end{aligned} \quad (15)$$

where the feasibility set  $\mathcal{S}_{\text{feasible}}$  encodes the constraints (14a)-(14i). Here the regularization factor  $\epsilon$  is chosen to be a small positive value such that it does not effect the matching size of the solution sub-graph. In fact choosing  $\epsilon < \frac{1}{M}$  ensures this, since then  $\epsilon \sum_{m \in \mathcal{A}} x_m < 1$  and a solution to (15) must have the same matching size as a solution to (14), otherwise the objective of (15) can be improved by choosing a solution with a larger matching size.

### C. Channel estimation and multiuser precoding

For a given set of user DL covariance matrices, let  $\{x_m^*\}_{m=1}^M$  and  $\{y_k^*\}_{k=1}^K$  denote the MILP solution and denote by  $\mathcal{B} = \{m : x_m^* = 1\} = \{m_1, m_2, \dots, m_{M'}\}$  the set of selected beam directions of cardinality  $|\mathcal{B}| = M'$  and by  $\mathcal{K} = \{k : y_k^* = 1\}$  the set of selected users of cardinality  $|\mathcal{K}| = K'$ . The resulting sparsifying precoding matrix  $\mathbf{B}$  in (11) is simply obtained as  $\mathbf{B} = \mathbf{F}_{\mathcal{B}}^H$ , 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  $\mathbf{F}$ . Given a DFT column  $\mathbf{f}_m$ , we have

$$\mathbf{B}\mathbf{f}_m = \begin{cases} \mathbf{0} & \text{if } m \notin \mathcal{B} \\ \mathbf{u}_i & \text{if } m = m_i \in \mathcal{B} \end{cases}$$

where  $\mathbf{u}_i$  denotes a  $M' \times 1$  vector with all zero components but a single “1” in the  $i$ -th position. Using the above property and (10), the effective DL channel vectors take on the form

$$\tilde{\mathbf{h}}_{\text{eff}}^{(k)} = \mathbf{B} \sum_{m \in \mathcal{S}_k} g_m^{(k)} \sqrt{[\boldsymbol{\lambda}^{(k)}]_m} \mathbf{f}_m = \sum_{i: m_i \in \mathcal{B} \cap \mathcal{S}_k} \sqrt{[\boldsymbol{\lambda}^{(k)}]_{m_i}} g_{m_i}^{(k)} \mathbf{u}_i. \quad (16)$$

In words, the effective channel of user  $k$  is a vector with non-identically zero elements only at the positions corresponding to the intersection of the beam directions in  $\mathcal{S}_k$ , along which the physical

channel of user  $k$  carries positive energy, and in  $\mathcal{B}$ , selected by the sparsifying precoder. The non-identically zero elements are independent Gaussian coefficients  $\sim \mathcal{CN}(0, [\boldsymbol{\lambda}^{(k)}]_{m_i})$ . Notice also that, by construction, the number of non-identically zero coefficients are  $|\mathcal{B} \cap \mathcal{S}_k| \leq T_{\text{dl}}$  and their positions (encoded in the vectors  $\mathbf{u}_i$  in (16)), plus an estimate of their variances  $[\boldsymbol{\lambda}^{(k)}]_{m_i}$  are known to the BS. Hence, the effective channel vectors can be estimated from the  $T_{\text{dl}}$ -dimensional DL pilot observation (11) with an estimation MSE that vanishes as  $1/\text{SNR}$ . The pilot observation in the form (11) is obtained at the user  $k$  receiver. In this work, we assume that each user sends its pilot observations using  $T_{\text{dl}}$  channel uses in the UL, using analog unquantized feedback, as analyzed for example in [12, 13]. At the BS receiver, after estimating the UL channel from the UL pilots, the BS can apply linear MMSE estimation and recovers the channel state feedback which takes on the same form of (11) with some additional noise due to the noisy UL transmission.<sup>9</sup>

With the above precoding, we have  $\mathbf{B}\mathbf{B}^H = \mathbf{I}_{M'}$ . Also, we can choose the DL pilot matrix  $\boldsymbol{\Psi}$  to be proportional to a random unitary matrix of dimension  $T_{\text{dl}} \times M'$ , such that  $\boldsymbol{\Psi}\boldsymbol{\Psi}^H = P_{\text{dl}}\mathbf{I}_{T_{\text{dl}}}$ . In this way, the DL pilot phase power constraint (12) is automatically satisfied. The estimation of  $\tilde{\mathbf{h}}_{\text{eff}}^{(k)}$  from the DL pilot observation (11) (with suitably increased AWGN variance due to the noisy UL feedback) is completely straightforward and shall not be treated here in details.

For the sake of completeness, we conclude this section with the DL precoded data phase and the corresponding sum rate performance metric that we shall use in Section V for numerical analysis and comparison with other schemes. 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  $\mathbf{V}$  given by the column-normalized version of the Moore-Penrose pseudoinverse of the estimated channel matrix, i.e.,  $\mathbf{V} = (\hat{\mathbf{H}}_{\text{eff}})^\dagger \mathbf{J}^{1/2}$ , where  $(\hat{\mathbf{H}}_{\text{eff}})^\dagger = \hat{\mathbf{H}}_{\text{eff}} (\hat{\mathbf{H}}_{\text{eff}}^H \hat{\mathbf{H}}_{\text{eff}})^{-1}$  and  $\mathbf{J}$  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

$$y^{(k)} = (\mathbf{h}^{(k)})^H \mathbf{B}^H \mathbf{V} \mathbf{P}^{1/2} \mathbf{d} + n^{(k)}, \quad (17)$$

where  $\mathbf{d} \in \mathbb{C}^{K' \times 1}$  is a vector of unit-energy user data symbols and  $\mathbf{P}$  is a diagonal matrix

<sup>9</sup>As an alternative, one can consider quantized feedback using  $T_{\text{dl}}$  channel uses in the UL (see [12, 13] and references therein). Digital quantized feedback yields generally a better end-to-end estimation MSE in the absence of feedback errors. However, the effect of decoding errors on the channel state feedback is difficult to characterize in a simple manner since it depends on the specific joint source-channel coding scheme employed. Hence, in this work we restrict to the simple analog feedback.

defining the power allocation to the DL data streams. The transmit power constraint is given by

$$\text{tr}(\mathbf{B}^H \mathbf{V} \mathbf{P} \mathbf{V}^H \mathbf{B}) = \text{tr}(\mathbf{V}^H \mathbf{V} \mathbf{P}) = \text{tr}(\mathbf{P}) = P_{\text{dl}},$$

where we used  $\mathbf{B} \mathbf{B}^H = \mathbf{I}_{M'}$  and the fact that  $\mathbf{V}^H \mathbf{V}$  has unit diagonal elements by construction. In particular, in the results of Section V we use the simple uniform power allocation  $P_k = P_{\text{dl}}/K'$  to each  $k$ -th user data stream. In the case of perfect ZF beamforming, i.e., for  $\widehat{\mathbf{H}}_{\text{eff}} = \mathbf{H}_{\text{eff}}$ , we have that (17) reduces to  $y^{(k)} = \sqrt{J_k P_k} d_k + n^{(k)}$ , where  $J_k$  is the  $k$ -th diagonal element of the norm normalizing matrix  $\mathbf{J}$ ,  $P_k$  is the  $k$ -th diagonal element of the power allocation matrix  $\mathbf{P}$ , and  $d_k$  is the  $k$ -th user data symbol. Since in general  $\widehat{\mathbf{H}}_{\text{eff}} \neq \mathbf{H}_{\text{eff}}$ , due to non-zero estimation error, the received symbol at user  $k$  receiver is given by  $y^{(k)} = b_{k,k} d_k + \sum_{k' \neq k} b_{k,k'} d_{k'} + n^{(k)}$ , where the coefficients  $(b_{k,1}, \dots, b_{k,K'})$  are given by the elements of the  $1 \times K'$  row vector  $(\mathbf{h}^{(k)})^H \mathbf{B}^H \mathbf{V} \mathbf{P}^{1/2}$  in (17). Of course, in the presence of an accurate channel estimation we expect that  $b_{k,k} \approx \sqrt{J_k P_k}$  and  $b_{k,k'} \approx 0$  for  $k' \neq k$ . For simplicity, in this paper we compare the performance of the proposed scheme with that of the state-of-the-art CS-based scheme in terms of ergodic sum rate, assuming that all coefficients  $(b_{k,1}, \dots, b_{k,K'})$  are known to the corresponding receiver  $k$ . Including the DL training overhead, this yields the rate expression (see [56])

$$R_{\text{sum}} = \left(1 - \frac{T_{\text{dl}}}{T}\right) \sum_{k \in \mathcal{K}} \mathbb{E} \left[ \log \left( 1 + \frac{|b_{k,k}|^2}{1 + \sum_{k' \neq k} |b_{k,k'}|^2} \right) \right]. \quad (18)$$

## V. SIMULATION RESULTS

In this section we compare the performance of the proposed approach for FDD massive MIMO to two of the most recent CS-based methods proposed in [19] and [25] in terms of channel estimation error and sum-rate. In [19], the authors proposed a method based on common probing of the DL channel with random Gaussian pilots. The DL pilot measurements  $\mathbf{y}^{(k)}$  at users  $k = 1, \dots, K$  are fed back and collected by the BS, which recovers the channel vectors using a joint orthogonal matching pursuit (J-OMP) technique able to exploit the possible common sparsity between the user channels (see channel model in Section II).

In [25], a method based on dictionary learning for sparse channel estimation was proposed. In this scheme, the BS jointly *learns* sparsifying dictionaries for the UL and DL channels by collecting channel measurements at different cell locations (e.g., via an off-line learning phase).

The actual user channel estimation is posed as a norm-minimization convex program using the trained dictionaries and with the constraint that UL and DL channels share the same support over their corresponding dictionaries. Following the terminology used in [25], we refer to this method as JDLCM.

For this comparison, we considered  $M = 128$  antennas at the BS,  $K = 13$  users, and resource blocks of size  $T = 128$  symbols. For our proposed method, the BS computes the users' sample UL covariance matrices by taking  $N_{\text{ul}} = 1000$  UL pilot observations and then applies the scheme explained in Section III. Given the obtained DL channel covariance matrix estimates, we first perform the circulant approximation and extract the vector of approximate eigenvalues as in (9). Then, we compute the sparsifying precoder  $\mathbf{B}$  via the MILP solution as given in Section IV-B. In the results presented here, we set the parameter  $P_0$  in the MILP to a small value in order to favor a high rank of the resulting effective channel matrix over the beamforming gain.<sup>10</sup> After probing the effective channel of the selected users along these active beam directions via a random unitary pilot matrix  $\Psi$ , we calculate their MMSE estimate using the estimated DL covariance matrices. Eventually, for all the three methods, we compute the ZF beamforming matrix based on the obtained channel estimates. In addition, instead of considering all selected users, in both cases we apply the Greedy ZF user selection approach of [57], that yields a significant benefit when the number of users is close to the rank of the effective channel matrix. As said before, the DL SNR is given by  $\text{SNR} = P_{\text{dl}}/N_0$  and during the simulations we consider ideal noiseless feedback for simplicity, i.e., we assume that the BS receives the measurements in (11) without extra feedback noise to the system.<sup>11</sup> The sparsity order of each channel vector is given as an input to the J-OMP method, but not to the other two methods. This represents a genie-aided advantage for J-OMP, that we introduce here for simplicity.

As the simulation geometry, we consider three MPC clusters with random locations within the angular range (parametrized by  $\xi$  rather than  $\theta$ )  $[-1, 1)$ . We denote by  $\Xi$  the  $i$ -th interval and set each interval size to be  $|\Xi_i| = 0.2$ ,  $i = 1, 2, 3$ . The ASF for each user is obtained by selecting at random two out of three such clusters, such that the overlap of the angular

<sup>10</sup>This approach is appropriate in the medium to high-SNR regime. For low SNR, it is often convenient to increase  $P_0$  in order to serve less users with a larger beamforming energy transfer per user.

<sup>11</sup>Notice that by introducing noisy feedback the relative gain w.r.t. J-OMP is even larger, since CS schemes are known to be more noise-sensitive than plain MMSE estimation using estimated DL covariance matrices.

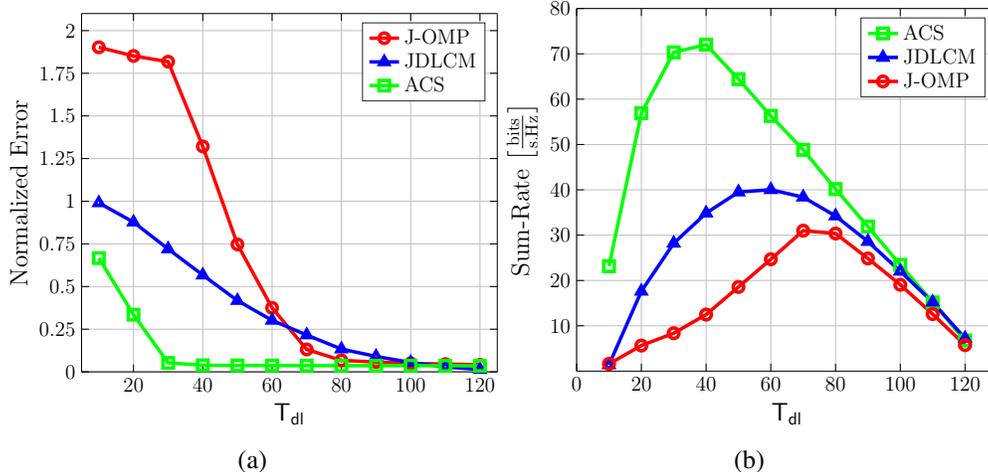


Fig. 3: (a) Normalized channel estimation error, and (b) achievable sum-rate as a function of DL pilot dimension with SNR = 20 dB,  $M = 128$  and  $K = 13$ .

components among users is large. The ASF is non-zero over the angular intervals corresponding to the chosen MPCs and zero elsewhere, i.e.,  $\gamma_k(d\xi) = \beta \mathbf{1}_{\Xi_{i_1} \cup \Xi_{i_2}}$ , where  $\beta = 1 / \int_{-1}^1 \gamma_k(d\xi)$  and  $i_1, i_2 \in \{1, 2, 3\}$ . The described arrangement results in each generated channel vector being roughly  $s_k = 0.2 \times M \approx 26$ -sparse. To measure channel estimation error we use the normalized Euclidean distance as follows. Let  $\mathbf{H} \in \mathbb{C}^{M \times K'}$  define the matrix whose columns correspond to the channel vectors of the  $K'$  served users and let  $\hat{\mathbf{H}}$  denote the estimation of  $\mathbf{H}$ . Then the normalized error is defined as

$$e = \mathbb{E} \left[ \frac{\|\mathbf{H} - \hat{\mathbf{H}}\|^2}{\|\mathbf{H}\|^2} \right].$$

### A. Comparisons

Fig. 3a shows the normalized channel estimation error for the J-OMP, JDLCM and our proposed Active Channel Sparsification (ACS) method as a function of the DL pilot dimension  $T_{\text{dl}}$  with SNR = 20 dB. Our ACS method outperforms the other two by a large margin, especially for low DL pilot dimensions. When the pilot dimension is below channel sparsity order, CS-based methods perform very poorly, since the number of channel measurements is less than the inherent channel dimension. Fig. 3b compares the achievable sum-rate for the three methods. Again our ACS method shows a much better performance compared to J-OMP and JDLCM. This figure also shows that there is an optimal DL pilot dimension that maximizes the sum-rate. This optimal value is  $T_{\text{dl}} \approx 40$  for our proposed method,  $T_{\text{dl}} \approx 60$  for JDLCM and  $T_{\text{dl}} \approx 70$

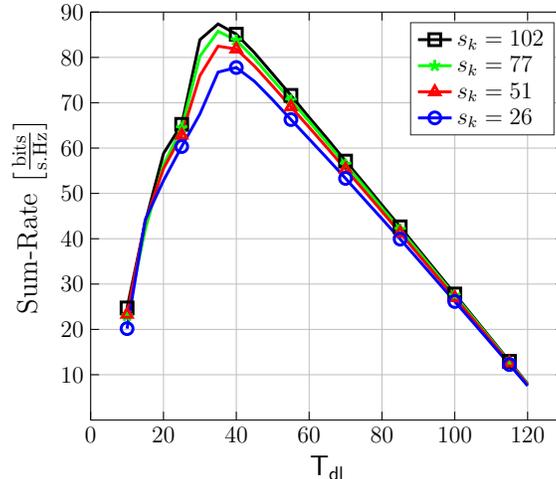


Fig. 4: Sum-rate vs  $T_{dl}$  for various channel sparsity orders. Here SNR = 20 dB,  $M = 128$  and  $K = 13$ .

for the J-OMP method.

### B. The effect of channel sparsity order

Depending on the geometry and user location, channels may show different levels of sparsity in the angular domain. In contrast to CS-based methods, our proposed method is highly flexible with regards to various channel sparsity orders, thanks to the active sparsification method. In this section, we investigate how sparsity order effects channel estimation error as well as sum-rate within the framework of our proposed method. We use the same setup as in section V-A, i.e. user ASFs consist of two clusters chosen at random among the three. But now we vary the size of the angular interval each of the clusters occupies ( $|\Xi_i| = 0.2, 0.4, 0.6, 0.8$ ) and see how it effects the error and sum-rate metrics. The sparsification, channel probing and transmission are performed as described before. Since each ASF consists of two clusters and  $M = 128$  channel sparsity order (roughly) takes on the values  $s_k = 26, 51, 77, 102$  for all users  $k \in [K']$ . For each value of the pilot dimension we perform a Monte Carlo simulation to empirically calculate the sum-rate. Fig. 4 illustrates the results. Notice that in these results we fix the channel coefficient power along each scattering component, such as richer (less sparse) channels convey more signal energy. This corresponds to the physical fact that the more scattered signal energy is collected at the receiving antennas the higher the received signal energy is. As we can see in Fig. 4, for a fixed  $T_{dl}$ , when the number of non-zero channel coefficients increases (i.e., the channel is less sparse), we generally have a larger sum-rate. The main reason is that, with less sparse channels, the beamforming gain is larger due to the fact that more scattering components contribute to the

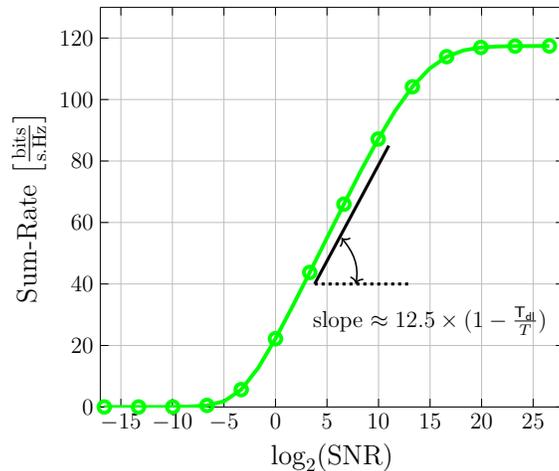


Fig. 5: Sum-rate as a function of  $\log_2(\text{SNR})$  with  $M = 128$  and  $K = 13$ .

channel. Therefore, we can generally say that with our method, for a fixed pilot dimension, less sparse channels are *better*. Of course, this is not the case for CS-based techniques, or techniques based on the “sparsity assumption” of a small number of discrete angular components, which tend to collapse and yield very bad results when such sparsity assumptions are not satisfied.

### C. Relevance of the pre-log factor

An interesting final observation is to examine the system sum-rate vs. SNR with our proposed method, and in particular show that there is indeed a regime of intermediate SNR for which the slope of the sum-rate curve yields quite faithfully the number of spatially multiplexed data streams. We performed a simulation with  $M = 128$  antennas and  $K = 13$  users and a pilot dimension of  $T_{\text{dl}} = 60$ . The pre-log factor determines the slope of the sum-rate vs  $\log_2(\text{SNR})$  curve, in an intermediate regime where the sum-rate is not saturated, and yet the spectral efficiency is large.<sup>12</sup> As illustrated in Fig. 5, this slope is equal to  $12.5 \times (1 - \frac{T_{\text{dl}}}{T})$ . Notice that the Greedy ZF scheme decides to serve a number of users that may be less than  $K$  in an opportunistic fashion, such that the expected number of served users (DL data streams) in this SNR regime is indeed slightly less than the maximum possible  $K = 13$ . Hence, the agreement between the sum-rate slope in this regime and the number of served DL data streams is exactly what can be expected, thus showing the relevance of maximizing the rank of the effective matrix in the proposed optimization of the sparsifying precoder.

<sup>12</sup>This saturation is due to the non-vanishing covariance estimation error and happens at around  $\text{SNR} = 60$  dB.

## VI. CONCLUSION

We presented a novel approach for FDD massive MIMO systems. Our approach exploits the reciprocity of the angular scattering function to estimate the covariance matrix of the users' DL channels from the UL pilots sent by the users to the BS. The estimated DL covariance matrices of all users can be approximately expressed in terms of a common system of covariance eigenvectors (beam-space representation). For the ULA setting, such eigenvectors are the columns of a DFT matrix, and this representation incurs a vanishing error for large number of BS antennas  $M$ . This beam-space information allows the BS to smartly select a set of beams and users such that communication over the resulting effective channels is efficient even with a limited DL pilot dimension. This beam-user selection procedure is referred to here as *active channel sparsification* and is achieved via a newly formulated mixed integer linear program (MILP). Our simulation results show that the proposed method performs well even in cases where the available DL pilot dimension is far less than the inherent dimension of the channel vectors. This represents a fundamental improvement with respect to the state-of-the-art CS-based method (in particular, exploiting common sparsity or learned sparsifying dictionaries), for which the DL pilot dimension should always be larger than the inherent channel sparsity in the angle domain. We conclude by mentioning that in this paper we focused on purpose on a simple single-cell scenario. When multiple cells are considered, inter-cell interference should be taken into account. However, unlike TDD systems where UL and DL across different cells are synchronous, and the limited pilot dimension yields pilot contamination (see [6–8]), in FDD systems there is no need for tight inter-cell synchronization and the inter-cell incoherent interference simply results in a higher level of the background noise, but can be taken into account in a completely straightforward manner (as always traditionally done in the analysis of cellular systems) since no coherently beamformed interference due to pilot contamination appears in FDD systems.

Future work along the lines presented in this paper may consist of generalizing the active channel sparsification method to a broader category of array geometries. While such generalization is straightforward for UPAs, leveraging the block-Toeplitz covariance structure, for other geometries one must find efficient methods for UL-DL covariance transformation and efficient “beam-space representation” for the design of the sparsifying precoder.

## VII. APPENDICES

## A. Proof of Lemma 1

The proof follows by using the representation  $\mathbf{h}^{(k)} = \sum_{m \in \mathcal{S}_k} g_m^{(k)} \sqrt{[\boldsymbol{\lambda}^{(k)}]_m} \mathbf{f}_m$  (see (10)), which holds exactly by assumption. Estimating  $\mathbf{h}^{(k)}$  is equivalent to estimating the vector of KL Gaussian i.i.d. coefficients  $\mathbf{g}^{(k)} = (g_m^{(k)} : m \in \mathcal{S}_k) \in \mathbb{C}^{s_k \times 1}$ . Define the  $M \times s_k$  DFT submatrix  $\mathbf{F}_{\mathcal{S}_k} = (\mathbf{f}_m : m \in \mathcal{S}_k)$ , and the corresponding diagonal  $s_k \times s_k$  matrix of the non-zero eigenvalues  $\boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)}$ . After some simple standard algebra, the MMSE estimation error covariance of  $\mathbf{g}^{(k)}$  from  $\mathbf{y}^{(k)}$  in (11) with  $\mathbf{B} = \mathbf{I}_M$  can be written in the form

$$\tilde{\mathbf{R}}_e = \mathbf{I}_{s_k} - \left( \boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)} \right)^{1/2} \mathbf{F}_{\mathcal{S}_k}^H \boldsymbol{\Psi}^H \left( \boldsymbol{\Psi} \mathbf{F}_{\mathcal{S}_k} \boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)} \mathbf{F}_{\mathcal{S}_k}^H \boldsymbol{\Psi}^H + N_0 \mathbf{I}_{T_{\text{dl}}} \right)^{-1} \boldsymbol{\Psi} \mathbf{F}_{\mathcal{S}_k} \left( \boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)} \right)^{1/2}. \quad (19)$$

Using the fact that  $\mathbf{R}_e = \mathbf{F}_{\mathcal{S}_k} (\boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)})^{1/2} \tilde{\mathbf{R}}_e (\boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)})^{1/2} \mathbf{F}_{\mathcal{S}_k}^H$ , such that  $\text{tr}(\mathbf{R}_e) = \text{tr}(\boldsymbol{\Lambda}_{\mathcal{S}_k} \tilde{\mathbf{R}}_e)$ , we have that  $\text{tr}(\mathbf{R}_e)$  and  $\text{tr}(\tilde{\mathbf{R}}_e)$  have the same vanishing order with respect to  $N_0$ . In particular, it is sufficient to consider the behavior of  $\text{tr}(\tilde{\mathbf{R}}_e)$  as a function of  $N_0$ . Now, using the Sherman-Morrison-Woodbury matrix inversion lemma [58], after some algebra omitted for the sake of brevity we arrive at

$$\text{tr}(\tilde{\mathbf{R}}_e) = s_k - \sum_{i=1}^{s_k} \frac{\mu_i}{N_0 + \mu_i}, \quad (20)$$

where  $\mu_i$  is the  $i$ -th eigenvalue of the  $s_k \times s_k$  matrix  $\mathbf{A} = (\boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)})^{1/2} \mathbf{F}_{\mathcal{S}_k}^H \boldsymbol{\Psi}^H \boldsymbol{\Psi} \mathbf{F}_{\mathcal{S}_k} (\boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)})^{1/2}$ . Next, notice that

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{F}_{\mathcal{S}_k}^H \boldsymbol{\Psi}^H \boldsymbol{\Psi} \mathbf{F}_{\mathcal{S}_k}) = \text{rank}(\mathbf{F}_{\mathcal{S}_k} \mathbf{F}_{\mathcal{S}_k}^H \boldsymbol{\Psi}^H) \leq \min\{s_k, T_{\text{dl}}\}. \quad (21)$$

In fact,  $\boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)}$  is diagonal with strictly positive diagonal elements, such that left and right multiplication by  $(\boldsymbol{\Lambda}_{\mathcal{S}_k}^{(k)})^{1/2}$  yields rank-preserving row and column scalings, the matrix  $\mathbf{F}_{\mathcal{S}_k} \mathbf{F}_{\mathcal{S}_k}^H$  is the orthogonal projector onto the  $s_k$ -dimensional column-space of  $\mathbf{F}_{\mathcal{S}_k}$  and has rank  $s_k$ , while the matrix  $\boldsymbol{\Psi}^H \in \mathbb{C}^{M \times T_{\text{dl}}}$  has the same rank of  $\boldsymbol{\Psi}^H \boldsymbol{\Psi}$ , that is at most  $T_{\text{dl}}$ .

For  $T_{\text{dl}} \geq s_k$  the existence of matrices  $\boldsymbol{\Psi}$  such that the rank upper bound (21) holds with equality (i.e., for which  $\text{rank}(A) = s_k$  for any support set  $\mathcal{S}_k$  of size  $s_k$ ) is shown as follows. Generate a random  $\boldsymbol{\Psi}$  with i.i.d. elements  $\sim \mathcal{CN}(0, 1)$ . Then, the columns of  $\mathbf{F}_{\mathcal{S}_k}^H \boldsymbol{\Psi}^H$  form a collection of  $T_{\text{dl}} \geq s_k$  mutually independent  $s_k$ -dimensional Gaussian vectors with i.i.d.  $\sim \mathcal{CN}(0, 1)$  components. The event that these vectors span a space of dimension less than  $s_k$  is a

null event (zero probability). Hence, such randomly generated matrix satisfies the rank equality in (21) with probability 1. As a consequence, for  $T_{\text{dl}} \geq s_k$  we have that  $\mu_i > 0$  for all  $i \in [s_k]$  and (20) vanishes as  $O(N_0)$  as  $N_0 \downarrow 0$ . In contrast, if  $T_{\text{dl}} < s_k$ , by (21) for any matrix  $\Psi$  at most  $T_{\text{dl}}$  eigenvalues  $\mu_i$  in (20) are non-zero and  $\lim_{N_0 \downarrow 0} s_k - \sum_{i=1}^{s_k} \frac{\mu_i}{N_0 + \mu_i} \geq s_k - T_{\text{dl}} > 0$ . ■

### B. Proof of Lemma 3

The determinant of  $\mathbf{Q}$  is given by the expansion  $\det(\mathbf{Q}) = \sum_{\iota \in \pi_r} \text{sgn}(\iota) \prod_i [\mathbf{Q}]_{i, \iota(i)}$ , where  $\iota$  is a permutation of the set  $\{1, 2, \dots, r\}$ , where  $\pi_r$  is the set of all such permutations and where  $\text{sgn}(\iota)$  is either 1 or -1. The product  $\prod_i [\mathbf{Q}]_{i, \iota(i)}$  is non-zero only for the perfect matchings in the bipartite graph. Hence, if the bipartite graph contains a perfect matching, then  $\det(\mathbf{Q}) \neq 0$  with probability 1 (and  $\text{rank}(\mathbf{Q}) = r$ ), since the non-identically zero entries of  $\mathbf{W}$  are drawn from a continuous distribution. If it does not contain a perfect matching, then  $\det(\mathbf{Q}) = 0$  and therefore  $\text{rank}(\mathbf{Q}) < r$ . ■

### C. Proof of Theorem 1

First, without loss of generality let assume that  $\mathcal{L}$  contains no isolated nodes (since these would be discarded anyway). As before the  $|\mathcal{A}| \times |\mathcal{K}|$  weighted adjacency matrix is denoted by  $\mathbf{W}$  where  $[\mathbf{W}]_{m,k} = w_{m,k}$ . An example of the bipartite graph  $\mathcal{L}$  and its corresponding weighted adjacency matrix  $\mathbf{W}$  is illustrated in Figs. 2a and 2b. Given the bipartite graph  $\mathcal{L}(\mathcal{A}, \mathcal{K}, \mathcal{E})$ , we select the subgraph  $\mathcal{L}'(\mathcal{A}', \mathcal{K}', \mathcal{E}')$ , so that the constraint (13b) is satisfied. We introduce the binary variables  $\{x_m, m \in \mathcal{A}\}$  and  $\{y_k, k \in \mathcal{K}\}$  to indicate if beam  $m$  and user  $k$  are selected, respectively. As such, the constraint (13b) is equivalent to the set of constraints:

$$x_m \leq \sum_{k \in \mathcal{K}} [\mathbf{A}]_{m,k} y_k \quad \forall m \in \mathcal{A} \quad (22a) \quad y_k \leq \sum_{m \in \mathcal{A}} [\mathbf{A}]_{m,k} x_m \quad \forall k \in \mathcal{K} \quad (22b)$$

$$\sum_{m \in \mathcal{A}} [\mathbf{A}]_{m,k} x_m \leq T_{\text{dl}} y_k + M(1 - y_k) \quad \forall k \in \mathcal{K} \quad (22c)$$

In particular, (22a) ensures that if the beam  $m$  is selected (i.e.,  $x_m = 1$ ), there must be some  $k \in \mathcal{K}$  such that  $(m, k) \in \mathcal{E}$  is selected as well, whereas if beam  $m$  is not selected, then this constraint is redundant. Similarly, in (22b) if user  $k$  is selected (i.e.,  $y_k = 1$ ), there must be some  $m \in \mathcal{A}$  such that  $(m, k) \in \mathcal{E}$  is selected as well. Furthermore, (22c) guarantees that if user  $k$  is chosen (i.e.,  $y_k = 1$ ), the number of chosen beams with  $x_m = 1$  is no more than  $T_{\text{dl}}$ , and

otherwise this constraint is redundant. Meanwhile, the constraint (13c) is written as:

$$P_0 y_k \leq \sum_{m \in \mathcal{A}} [\mathbf{W}]_{m,k} x_m \quad \forall k \in \mathcal{K} \quad (23)$$

which ensures that if user  $k$  is chosen (i.e.,  $y_k = 1$ ) then the sum weights of the selected beams (i.e.,  $m \in \mathcal{N}_{\mathcal{L}'}(k)$  if  $x_m = 1$ ) is no less than  $P_0$ , while if user  $k$  is not chosen (i.e.,  $y_k = 0$ ) then this constraint is not required and redundant. A closer look reveals that the constraint (23) renders the one (22b) redundant, because when  $y_k = 1$  in (23) there must exist at least one  $m \in \mathcal{A}$  with  $x_m = 1$ . Second, given the selected subgraph  $\mathcal{L}'(\mathcal{A}', \mathcal{K}', \mathcal{E}')$ , we find a matching  $\mathcal{M}(\mathcal{A}', \mathcal{K}')$  with maximum cardinality. To this end, we introduce another set of binary variables  $\{z_{mk}, m \in \mathcal{A}, k \in \mathcal{K}\}$  to indicate if an edge  $(a, k) \in \mathcal{E}$  is chosen to form the maximum matching in  $\mathcal{L}'(\mathcal{A}', \mathcal{K}', \mathcal{E}')$ . Following the canonical linear program formulation of the maximum cardinality matching for bipartite graphs, we translate the objective in (13) into the following optimization:

$$\text{maximize}_{z_{m,k} \in \{0,1\}} \sum_{m \in \mathcal{A}'} \sum_{k \in \mathcal{K}'} [\mathbf{A}]_{m,k} z_{m,k} \quad (24a)$$

$$\text{subject to} \quad \sum_{k \in \mathcal{K}'} [\mathbf{A}]_{m,k} z_{m,k} \leq 1 \quad \forall m \in \mathcal{A}', \quad (24b)$$

$$\sum_{m \in \mathcal{A}'} [\mathbf{A}]_{m,k} z_{m,k} \leq 1 \quad \forall k \in \mathcal{K}', \quad (24c)$$

Now, to transport the optimization problem on  $\mathcal{L}'$  to the original setting on  $\mathcal{L}$ , we need to guarantee that  $\mathcal{M}(\mathcal{A}', \mathcal{K}') \subseteq \mathcal{E}'$ , i.e.,  $z_{mk} = 1$  only if  $m \in \mathcal{A}'$  ( $x_m = 1$ ), and  $k \in \mathcal{K}'$  ( $y_k = 1$ ). This is obtained for a given configuration of the variables  $\{x_m\}$  and  $\{y_k\}$  which define  $\mathcal{L}'$ , by adding constraints to (24) and yields

$$\text{maximize}_{z_{m,k} \in \{0,1\}} \sum_{m \in \mathcal{A}} \sum_{k \in \mathcal{K}} [\mathbf{A}]_{m,k} z_{m,k} \quad (25a)$$

$$\text{subject to} \quad \sum_{k \in \mathcal{K}} [\mathbf{A}]_{m,k} z_{m,k} \leq 1 \quad \forall m \in \mathcal{A}, \quad (25b)$$

$$\sum_{m \in \mathcal{A}} [\mathbf{A}]_{m,k} z_{m,k} \leq 1 \quad \forall k \in \mathcal{K}, \quad (25c)$$

$$[\mathbf{A}]_{m,k} z_{m,k} \leq x_m \quad \forall k \in \mathcal{K}, m \in \mathcal{A}, \quad (25d)$$

$$[\mathbf{A}]_{m,k} z_{m,k} \leq y_k \quad \forall k \in \mathcal{K}, m \in \mathcal{A}, \quad (25e)$$

where (25d)-(25e) impose that the edge set  $\{(m, k) : z_{m,k} = 1\}$  should be a subset of  $\mathcal{E}'$ . A further inspection on these constraints yields the following equivalent simplified form:

$$\underset{z_{m,k} \in \{0,1\}}{\text{maximize}} \quad \sum_{m \in \mathcal{A}} \sum_{k \in \mathcal{K}} z_{m,k} \quad (26a)$$

$$\text{subject to} \quad z_{m,k} \leq [\mathbf{A}]_{m,k}, \quad \forall m \in \mathcal{A}, k \in \mathcal{K}, \quad (26b)$$

$$\sum_{k \in \mathcal{K}} z_{m,k} \leq x_m, \quad \forall m \in \mathcal{A}, \quad (26c)$$

$$\sum_{m \in \mathcal{A}} z_{m,k} \leq y_k, \quad \forall k \in \mathcal{K}, \quad (26d)$$

where the additional constraint (26b) turns all the terms of the type  $[\mathbf{A}]_{m,k} z_{m,k}$  in (25) to  $z_{m,k}$  in (26), the constraint (26c) results from the combination of the constraints (25b) and (25d), and (26d) results from the combination of (25c) with (25e). The formulation in (26) can be seen as a modified maximum cardinality bipartite matching with selective vertices, in which the vertices with  $x_m = 1$  and  $y_k = 1$  are selected to participate in the maximum cardinality matching. The eventual mixed integer linear program is given as in (14). Notice that we have relaxed the binary constraint on  $\{z_{m,k}, m \in \mathcal{A}, k \in \mathcal{K}\}$  to the linear constraint (14i) based on the following lemma.

**Lemma 4:** The problem  $\mathcal{P}_{\text{MILP}}$  as stated in (14) always has binary-valued solutions for  $\{z_{m,k}, m \in \mathcal{A}, k \in \mathcal{K}\}$ .  $\square$

*Proof:* It suffices to show that  $z_{m,k}$  are binary, given that  $x_m$  and  $y_k$  are binary. First, if either  $x_m, m \in \mathcal{A}$  or  $y_k, k \in \mathcal{K}$  are 0, then  $z_{m,k} = 0$ . So, we only need to focus on the case where  $x_m = y_k = 1, m \in \mathcal{A}, k \in \mathcal{K}$ . In that case, the constraints of  $\mathcal{P}_{\text{MILP}}$  with respect to  $z_{m,k}, m \in \mathcal{A}, k \in \mathcal{K}$  form a convex polytope. This polytope is called the bipartite matching polytope, which is integral, i.e. all of its extreme points have integer (and in this case binary) values (see [59, Corollary 18.1b. and Theorem 18.2.]). Therefore, given  $x_m, y_k \in \{0, 1\}, \forall m \in \mathcal{A}, k \in \mathcal{K}$ ,  $\mathcal{P}_{\text{MILP}}$  reduces to a linear program with respect to the variables  $z_{m,k}$  and the optimal solutions are the integral extreme points of the corresponding polyhedra and the proof is complete.  $\blacksquare$

## REFERENCES

- [1] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [2] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.

- [3] 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.
- [4] 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.
- [5] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5213–5226, 2013.
- [6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [7] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [8] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO.* Cambridge University Press, 2016.
- [9] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.
- [10] S. Sesia, M. Baker, and I. Toufik, *LTE-the UMTS long term evolution: from theory to practice.* John Wiley & Sons, 2011.
- [11] S. Malkowsky, J. Vieira, L. Liu, P. Harris, K. Nieman, N. Kundargi, I. C. Wong, F. Tufvesson, V. Öwall, and O. Edfors, "The World's First Real-Time Testbed for Massive MIMO: Design, Implementation, and Validation," *IEEE Access*, vol. 5, pp. 9073–9088, 2017.
- [12] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.
- [13] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2228–2240, 2011.
- [14] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, 2013.
- [15] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE transactions on information theory*, vol. 49, no. 10, pp. 2735–2747, 2003.
- [16] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on information theory*, vol. 52, no. 11, pp. 5045–5060, 2006.
- [17] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2868–2882, 2015.
- [18] P. W. Chan, E. S. Lo, R. R. Wang, E. K. Au, V. K. Lau, R. S. Cheng, W. H. Mow, R. D. Murch, and K. B. Letaief, "The evolution path of 4G networks: FDD or TDD?" *IEEE Communications Magazine*, vol. 44, no. 12, pp. 42–50, 2006.
- [19] 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.
- [20] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.
- [21] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.
- [22] 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.
- [23] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath, “Compressed channel feedback for correlated massive MIMO systems,” *Journal of Communications and Networks*, vol. 18, no. 1, pp. 95–104, 2016.
- [24] 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.
- [25] Y. Ding and B. D. Rao, “Dictionary learning based sparse channel representation and estimation for FDD massive MIMO systems,” *arXiv preprint arXiv:1612.06553*, 2016.
- [26] J. Fang, X. Li, H. Li, and F. Gao, “Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1935–1947, 2017.
- [27] J. Dai, A. Liu, and V. K. Lau, “FDD massive MIMO channel estimation with arbitrary 2D-array geometry,” *arXiv preprint arXiv:1711.06548*, 2017.
- [28] H. Xie, F. Gao, S. Zhang, and S. Jin, “A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3170–3184, 2017.
- [29] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [30] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [31] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.
- [32] Y. C. Eldar and H. Rauhut, “Average case analysis of multichannel sparse recovery using convex relaxation,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 505–519, 2010.
- [33] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, “Correlation analysis based on MIMO channel measurements in an indoor environment,” *IEEE Journal on Selected areas in communications*, vol. 21, no. 5, pp. 713–720, 2003.
- [34] F. Kaltenberger, D. Gesbert, R. Knopp, and M. Kountouris, “Correlation and capacity of measured multi-user MIMO channels,” in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*. IEEE, 2008, pp. 1–5.
- [35] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, “Channel measurements for large antenna arrays,” in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*. IEEE, 2012, pp. 811–815.
- [36] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, “Linear pre-coding performance in measured very-large MIMO channels,” in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*. IEEE, 2011, pp. 1–5.
- [37] K. Hugl, K. Kalliola, and J. Laurila, “Spatial reciprocity of uplink and downlink radio channels in FDD systems,” *Proc. COST 273 Technical Document TD (02)*, vol. 66, p. 7, 2002.
- [38] A. Ali, N. González-Prelcic, and R. W. Heath Jr, “Millimeter wave beam-selection using out-of-band spatial information,” *arXiv preprint arXiv:1702.08574*, 2017.
- [39] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, “Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing,” *arXiv preprint arXiv:1710.00704*, 2017.
- [40] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, “Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling,” *IEEE J. of Sel. Topics in Sig. Proc. (JSTSP)*, vol. 8, no. 5, pp. 876–890, 2014.
- [41] B. K. Chalise, L. Haering, and A. Czylik, “Robust uplink to downlink spatial covariance matrix transformation for

- downlink beamforming,” in *Communications, 2004 IEEE International Conference on*, vol. 5. IEEE, 2004, pp. 3010–3014.
- [42] 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.
- [43] T. Asté, P. Forster, L. Fety, and S. Mayrargue, “Downlink beamforming avoiding doa estimation for cellular mobile communications,” in *IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS SPEECH AND SIGNAL PROCESSING*, vol. 6. INSTITUTE OF ELECTRICAL ENGINEERS INC (IEE), 1998, pp. VI–3313.
- [44] D. Vasisht, S. Kumar, H. Rahul, and D. Katabi, “Eliminating channel feedback in next-generation cellular networks,” in *Proceedings of the 2016 conference on ACM SIGCOMM 2016 Conference*. ACM, 2016, pp. 398–411.
- [45] L. Miretti, R. L. Cavalcante, and S. Stanczak, “FDD massive MIMO channel spatial covariance conversion using projection methods,” *arXiv preprint arXiv:1804.04850*, 2018.
- [46] S. Haghighatshoar, M. B. Khalilsarai, and G. Caire, “Multi-band covariance interpolation with applications in massive MIMO,” *arXiv preprint arXiv:1801.03714*, 2018.
- [47] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. De Doncker, “The COST 2100 MIMO channel model,” *IEEE Wireless Communications*, vol. 19, no. 6, pp. 92–99, 2012.
- [48] “ETSI TS 136 101 V14.3.0 (2017-04) - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (3GPP TS 36.101 version 14.5.0 Release 14).”
- [49] D. P. Bertsekas and A. Scientific, *Convex optimization algorithms*. Athena Scientific Belmont, 2015.
- [50] S. Haghighatshoar and G. Caire, “Channel vector subspace estimation from low-dimensional projections,” *arXiv preprint arXiv:1509.07469*, 2015.
- [51] 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.
- [52] A. G. Davoodi and S. A. Jafar, “Aligned image sets under channel uncertainty: Settling conjectures on the collapse of degrees of freedom under finite precision CSIT,” *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5603–5618, 2016.
- [53] R. Diestel, *Graph theory (Graduate texts in mathematics)*. Springer Heidelberg, 2005, vol. 173.
- [54] S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, “A theory of pseudoskeleton approximations,” *Linear algebra and its applications*, vol. 261, no. 1-3, pp. 1–21, 1997.
- [55] W. T. Tutte, “The factorization of linear graphs,” *Journal of the London Mathematical Society*, vol. 1, no. 2, pp. 107–111, 1947.
- [56] G. Caire, “On the ergodic rate lower bounds with applications to massive MIMO,” *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2018.
- [57] G. Dimic and N. D. Sidiropoulos, “On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm,” *IEEE Transactions on Signal processing*, vol. 53, no. 10, pp. 3857–3868, 2005.
- [58] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.
- [59] A. Schrijver, *Combinatorial optimization: polyhedra and efficiency*. Springer Science & Business Media, 2003, vol. 24.