The Impact of Physical Channel on Performance of Subspace-Based Channel Estimation in Massive MIMO Systems

Mohammed Teeti, Member, IEEE, Jun Sun, David Gesbert, Fellow, IEEE, and YingZhuang Liu

Abstract

A subspace method for channel estimation has been recently proposed [1] for tackling the pilot contamination effect, which is regarded by some researchers as a bottleneck in massive MIMO systems. It was shown in [1] that if the power ratio between the desired signal and interference is kept above a certain value, the received signal spectrum splits into signal and interference eigenvalues, namely, the “pilot contamination” effect can be completely eliminated. However, [1] assumes an independently distributed (i.d.) channel, which is actually not much the case in practice. Considering this, a more sensible finite-dimensional physical channel model (i.e., a finite scattering environment, where signals impinge on the base station (BS) from a finite number of angles of arrival (AoA)) is employed in this paper. Via asymptotic spectral analysis, it is demonstrated that, compared with the i.d. channel, the physical channel imposes a penalty in the form of an increased power ratio between the useful signal and the interference. Furthermore, we demonstrate an interesting “antenna saturation” effect, i.e., when the number of the BS antennas approaches infinity, the performance under the physical channel with $P$ AoAs is limited by and nearly the same as the performance under the i.d. channel with $P$ receive antennas.

Index Terms

massive MIMO, physical channel, subspace method, random matrix theory, asymptotic eigenvalue distribution

M. Teeti, J. Sun, and Y. Liu are with the Department of Electronics and Information Engineering, Huazhong University of Science & Technology, Wuhan, 430074, China (e-mail: teeti.moh@gmail.com, francissunj@gmail.com and liuyz@mail.hust.edu.cn).

D. Gesbert is with the Mobile Communications Department of EURECOM, EURECOM, 06410 Biot, France (e-mail: david.gesbert@eurecom.fr).
### I. Introduction

Massive MIMO systems that employ a large number of antennas at the base station (BS) have attracted significant interest recently [2], [3], [4], [5]. The main advantage of using massive MIMO lies in the significant improvement of spectral and energy efficiency. However, when we shift our attention to the multicell scenario, the massive MIMO system would, unfortunately, be plagued by the so-called “pilot contamination” effect, which is due to the use of non-fully orthogonal pilot sequences across all the cells. However, it is often hard to achieve full orthogonality of the pilot sequences among the terminals across the cells, due to the limited coherence time of the mobile communication channels. The performance of massive MIMO systems under the full reuse of pilot sequences was studied intensively by Marzetta [2], where it is shown that, by linear processing of the received signal at the BS, the performance is only interference-limited and does not depend on the transmitted power of users, i.e., signal-to-interference ratio (SIR) cannot grow unboundedly.

Several motivating works have been conducted aiming to address the above pilot contamination problem. In [6], authors use a coordinated channel estimation scheme which exploits the information embedded in the second-order statistics (e.g., covariance matrix) of uplink channels. The main idea of the scheme in [6] is to assign pilots to users associated with covariance matrices exhibiting maximum orthogonality of signal subspaces. Although this scheme greatly alleviates the pilot contamination, but it incurs too much coordination since all covariance matrices of all uplink channels must be learned beforehand.

In [7], by leveraging a key observation that the quasi orthogonality among the channel vectors implies that the channel vectors of the desired users are eigenvectors of the covariance matrix of the received signal in the asymptotic limit, Ngo and Larsson [7] proposed a blind channel estimation method which mitigates the need of pilots. However, it is worth noting this scheme heavily hinges on a large number of BS antennas as well as large block length (i.e., sample size), therefore its performance might be unsatisfactory in the “not very large” regime.

Müller et. al. [8] proposed another blind pilot decontamination method which, in essence, aims to distinguish users in the amplitude domain by exploiting the difference between the channel gains of the intended users and the interfering users. Therefore, this work can be regarded as a parallel to [6], which essentially attempts to distinguish users with the same pilot sequence in the angular domain by exploiting the non-isotropy of angles of arrival (AoA) multipaths of...
the channel. By means of random matrix theory (RMT) and free probability theory (FPT), [8] demonstrates that under a mild power ratio the subspace-based estimation scheme is capable of totally removing the interference and hence completely tackling the pilot contamination problem.

We note that the channel considered in [8] consists of independently distributed (i.d.) Rayleigh-faded entries, implicitly implying a rich scattering environment. However, in the real world, it is quite common that the number of scatterers is limited and correspondingly the number of AoAs is finite [9], [6], [10]. Therefore, considering a more sensible physical channel model (or alternatively, called finite-dimensional channel, due to the finiteness of its degrees of freedom) is of significant importance since the finiteness of degrees of freedom of the channel is relevant to the ability of subspace method to identify the eigenvalues of desired and interfering users. In the absence of a priori knowledge of the channel at the BS, it is natural to assume the AoAs in the channel model to be uniformly distributed in the interval \([0, \pi]\). Therefore, the total number of AoAs, which roughly corresponds to the total number of scatterers around the BS, will be the only key parameter (besides the ratio of the channel gains that correspond to the desired users and the interfering users, respectively) that might have crucial impact on the performance of the blind subspace method. Hence a natural question one may ask: What is the impact of finite-dimensional channel on the performance of the subspace-based channel estimation scheme? What is the price paid for this finite dimensionality of the channel? Moreover, does increasing the number of antennas at the BS help to alleviate the performance degradation?

To answer the above questions, the core task is to characterize the spectrum of the observed signal under the physical channel model. Despite its difficulty, we manage that by leveraging tools in RMT and FPT. Intuitively, the case of the i.d. channel can be regarded as a special case of our result, i.e., when the number of AoAs approaches infinity. The contributions of this paper are summarized as follows:

1) It is shown that to guarantee the performance of the subspace-based channel estimation scheme under the physical channel, we should pay a cost of an extra power margin between the intended users and the interfering users with respect to (w.r.t.) a given BS.

2) For multicell multiuser MIMO system where the users of each cell are seen from distinct AoAs w.r.t. a given BS, it is shown that the performance is mainly determined by the cell associated with smaller number of AoAs.

3) It is demonstrated that there exists an antenna saturation effect at the BS, i.e. adding antennas beyond a threshold at the BS is of limited help to enhance the performance under the physical
channel model. To be concrete, the performance under a physical channel with \( P \) AoAs is nearly the same as the performance under the i.d channel model with \( P \) antennas at the BS (see Fig. 1 and Sec. IV for a detailed description of the simulation setup and discussion).

Fig. 1. Antenna saturation effect under physical channel. \( L = 4 \) cells, \( K = 5 \) users per cell, \( P = 200 \) AoAs, coherence time \( N = 400 \) symbol periods and per-user SNR = -5dB. ZF is used for channel estimation and MF for data detection. Array elements are critically-spaced.

Fig. 2. The effect of the number of AoAs on the required power ratio \( P_I/P_s \). \( M = 200 \) receive antennas, \( L = 2 \) cells, \( K = 5 \) users per cell, coherence time \( N = 400 \) symbol periods, per-user SNR = 0dB. ZF is used for channel estimation and MF for data detection. Array elements are critically-spaced.

In the following we provide some intuitive remarks of the above results. When the number of AoAs is decreased, the correlation (or more exactly, the coherence) among the channel vectors will increase correspondingly. As a result, the condition number (or eigenvalue spread) of the channel matrix \( H_s \) (corresponding to inside-cell users) as well as \( H_f \) (corresponding to outside-cell users) will both increase. Therefore, the left and right endpoints of the eigenvalue clusters
corresponding to $H_s$ and $H_I$, respectively, tend to be closer, even overlapping (e.g., see Fig.5). In order to keep the above two eigenvalue clusters apart, we have to pay a cost of boosting the channel gain ratio between the desired users and the interfering users. Fig.2 illustrates the BER performance of the subspace method proposed in [8] when we vary the number of AoAs, denoted $P$. It can be seen from Fig.2 that the smaller $P$ is, the higher channel gain ratio $P_s/P_I$ is required.

As a final remark, it is worthwhile of pointing out the different role that channel correlation plays in the two pilot decontamination methods, namely, the Bayesian channel estimation in [6] and the subspace method in [8] as well as this paper, that is, while the correlation is beneficial in the former method, it is however, disadvantageous in the latter. The key reason for this difference lies in that, in a nutshell, the former is a Bayesian method while the latter is a non-Bayesian one (i.e., no a priori assumed). Specifically, the correlation is advantageous for the linear (MMSE) channel estimation (since it helps to distinguish users in the angular domain) at the cost of acquiring the a priori (namely, the covariance matrices) and coordinating the pilots. On the other hand, as a non-Bayesian method in essence, the subspace method performs the channel estimation in a nonlinear way, which highly relies on the instantaneous property (such as the eigenvalue spread of the channel matrix), rather than the statistical property in the Bayesian method [6]. Since correlation tends to increase the eigenvalue spread, it will impact the ability of subspace method to distinguish the desired users and interfering users via eigenvalues clustering.

The rest of the paper is organized as follows: In Sec. II we introduce the physical channel model and review the subspace-based channel estimation. In Sec. III the asymptotic eigenvalue distribution (AED) of the channel is derived in the Stieltjes domain. This will enable us to derive analytical expressions, from which the support of the distribution can be identified, hence yielding the main results of this paper. Further, Sec. IV leverages the obtained formulae to demonstrate the impact of the physical channel on the spectral spread of signal and interference subspaces, and BER simulation results are presented for performance comparison. Sec. V summarizes the main results and concludes this paper.

II. PROBLEM FORMULATION

A. Channel model

We will consider the uplink in multicell multiuser MIMO communication system with $L$ cells. Each BS is equipped with a uniform linear array with a large number of antennas $M$, serving $K$
single-antenna users. We also assume time-division duplex architecture. The channel is assumed narrowband flat-fading, which remains constant over a coherence time of $N$ symbol periods, and changes independently from one coherence time to another (i.e., block fading channel [11]). Furthermore, we assume time-synchronous users, and full reuse of pilot sequences among all cells to facilitate channel estimation at the BS.

Since the channel is estimated blindly, we assume no a priori information about the channel is available at BSs, including the angular spread, which is in contrast to the Bayesian channel estimation method [6]. Thus, we assume all AoAs are uniformly distributed in $[0, \pi]$, which is a reasonable, yet mathematically tractable assumption. However, the theoretical analysis in our paper gives insights into the performance of the subspace method under a physical channel with small angular spreads as well.

The array response in a given direction is quantified by the so-called steering vector. Hence, the channel from the $k$th user in the $i$th cell to the intended BS, denoted $h_{ki}$, can be modeled as a linear combination of all steering vectors (see e.g., [10], [6])

$$h_{ki} = \frac{1}{\sqrt{P_{ki}}} \sum_{j=1}^{P_{ki}} \alpha_{kij} s(\varphi_{kij})$$

(1)

where $P_{ki}$ is the number of i.i.d. AoA multipaths, $s(\varphi_{kij}) \in \mathbb{C}^{M}$ is the steering vector corresponding to the angle of arrival $\varphi_{kij}$ associated with the $j$th path, $1/\sqrt{P_{ki}}$ serves as a normalization factor and $\alpha_{kij} \sim \mathcal{CN}(0, \beta_{ki})$ denotes the channel gain associated with the $j$th direction, where $\sqrt{\beta_{ki}}$ is the average channel attenuation, due to path loss and shadowing effect. Moreover, it is assumed that all $\varphi_{kij}$ and $\alpha_{kij}$ are independent over user index $k$, cell index $i$ and direction index $j$. The length-$M$ steering vector associated with the angle of arrival $\varphi_{kij}$ is given by $s(\varphi_{kij}) = (e^{-jf_{1}(\varphi_{kij})}, e^{-jf_{2}(\varphi_{kij})}, \ldots, e^{-jf_{M}(\varphi_{kij})})^T$, where we use “$j$” to denote the imaginary unit and $f_{i}(\varphi_{kij})$ is a function of $\varphi_{kij}$.

Let

$$H_{i} \triangleq (S_{1i} \tilde{h}_{1i}, \ldots, S_{Ki} \tilde{h}_{Ki})$$

(2)

be the $M \times K$ effective fast-fading channel from $K$ users in the $i$th cell to the intended BS, where $S_{ki} = (s(\varphi_{k1i}), \ldots, s(\varphi_{kP_{ki}}))/\sqrt{P_{ki}} \in \mathbb{C}^{M \times P_{ki}}$ comprises all steering vectors of the $k$th user in the $i$th cell w.r.t. the intended BS, and $\tilde{h}_{ki} \in \mathbb{C}^{P_{ki} \times 1}$ consists of i.i.d. $\mathcal{CN}(0, 1)$ entries.
The $M \times N$ signal received by the $l$th BS during $N$ consecutive symbol intervals may be written as

$$Y_l = \sum_{i=1}^{L} H_i D_i^{1/2} T_i^{1/2} X_i + W_l$$  \hspace{1cm} (3)$$

where $D_i = \text{diag}(\beta_{i1}, \beta_{i2}, \ldots, \beta_{iK}) \in \mathbb{R}^{K \times K}$ comprises large-scale fading coefficients, $T_i = \text{diag}(p_{i1}, \ldots, p_{Ki}) \in \mathbb{R}^{K \times K}$ consists of the average transmission powers, and $X_i \in \mathbb{C}^{K \times N}$ is the normalized input symbols (preceded by short pilot signals), all for $K$ users in the $i$th cell. The entries of the noise matrix $W_l$ are assumed i.i.d. $\mathcal{CN}(0, 1)$.

For the sake of analytical simplicity, we shall assume that all users belonging to the same (say, $i$th) cell are seen from the same set of directions (whose cardinality is $P_i$, and a corresponding steering matrix $S_i$) w.r.t. the intended BS. Therefore, in this distinct AoAs scenario, (2) takes the form of $H_i = S_i \tilde{H}_i$, where $\tilde{H}_i \in \mathbb{C}^{P_i \times K}$ consists of i.i.d. $\mathcal{CN}(0, 1)$ entries, correspondingly, (3) can now be rewritten as:

$$Y_l = \sum_{i=1}^{L} S_i \tilde{H}_i D_i^{1/2} T_i^{1/2} X_i + W_l.$$  \hspace{1cm} (4)$$

As a first step in addressing the channel model (4), we’ll make a further assumption that all $S_i$ are identical w.r.t. to the intended BS, which means all users in the network are seen from the same set of $P$ AoAs by the intended BS (for the sake of illustration, see Fig.3). Hence, in this identical AoAs scenario, (4) can take a simpler form as:

$$Y_l = S \sum_{i=1}^{L} \tilde{H}_i D_i^{1/2} T_i^{1/2} X_i + W_l.$$  \hspace{1cm} (5)$$

In Sec. III, we will first focus on the channel model (5) and for the general case, i.e., the channel model (4), we will treat it afterwards.
B. Subspace-based channel estimation

In massive MIMO, by using subspace-based estimation approach, each user’s channel can be estimated blindly and accurately up to a scalar ambiguity, specifically, by identifying the signal subspace from interference subspace. This is possible because when the number of receive antennas is very large and the uncorrelated channel is assumed, the signals of different users are projected onto quasi-orthogonal subspaces. Another key condition that makes it possible to split the spectrum into signal and interference eigenvalue clusters, is due to the fact that inside-cell users and outside-cell users exhibit difference in the received power at the intended BS (e.g., due to path loss). These observations have been recently employed in massive MIMO [7], [8] to circumvent the problem of pilot contamination.

Applying singular value decomposition (SVD) to the received signal (3) (see [8] for more details), the first \( K \) eigenvectors, denoted \( U_s \in \mathbb{C}^{M \times K} \), can then be identified, which correspond to the signal subspace. These eigenvectors represent an estimate of the channel up to scaling ambiguity which can be resolved by exploiting a short training sequence (e.g., 1 pilot) sent by each user. Having identified \( U_s \), the received signal is then projected onto \( U_s \), by which most of interference is annihilated and most of the thermal noise is removed. In this process, a new channel model is obtained:

\[
\tilde{Y}_l = U_s^\dagger Y_l = G_l X_l + \tilde{W}_l
\]

where \( G_l \in \mathbb{C}^{K \times K} \) and \( \tilde{W}_l \in \mathbb{C}^{K \times N} \) are the subspace channel and subspace noise matrix, respectively. Note that the original channel \( H_l \) is not needed for the detection of \( X_l \), rather it can be estimated after \( X_l \) being detected for the purpose of downlink precoding, for example.

III. PERFORMANCE ANALYSIS

In this section we will be primarily concerned with deriving the asymptotic eigenvalue distribution (AED) of the channel models (4) and (5) with the aid of RMT and FPT. In fact, the density of the distribution is not explicitly derived. Instead, a fixed-point equation that satisfies the Stieltjes transform\(^1\) [9], [12], [13] of the density is given, which will serve as grounds for the derivation of the formula that helps identify the support of the distribution, in particular, the

\(^1\)Also known as the Cauchy transform, which encodes all the moments of the underlying distribution in a polynomial function. For a distribution function \( f(x) \), its Stieltjes transform is defined by \( G(s) = \int_{-\infty}^{+\infty} \frac{f(x)dx}{x-s}, \Im s \neq 0 \).
gap between the desired signal and interference subspaces. In Sec. III-A we shall start with the channel model (5) and the channel model (4) is treated in Sec. III-B.

A. Physical channel model with identical AoAs

First, we should stress that the steering matrix $S$ in (5) is random, but fixed, which is dependent on the physical environment only. Equation (5) can be written in the compact form as

$$Y_l = S \tilde{H} \tilde{D}^{1/2} X + W_l$$

(7)

where $\tilde{H} = (\tilde{H}_1, \tilde{H}_2, \ldots, \tilde{H}_L) \in \mathbb{C}^{P \times KL}$, $X = (X_1, X_2, \ldots, X_L)^T \in \mathbb{C}^{KL \times N}$, and $\tilde{D} \in \mathbb{R}^{KL \times KL}$ is a diagonal matrix, where the components of the row vectors $(p_{1i}\beta_{1i}, \ldots, p_{Ki}\beta_{Ki}), i = 1, 2, \cdots, L$ are placed on its diagonal.

If the true data covariance matrix $\Sigma$ is assumed available at the BS, then the eigenvalue distribution of the channel can be accurately determined, for instance by $\Sigma = (S \tilde{H})\tilde{D} (S \tilde{H})^\dagger + I_M$, but in practice, one only has access to the sample covariance matrix $\Sigma_N$ computed from $N$ data samples, which serves as an approximation of $\Sigma$. Thus, our objective is to study the eigenvalue distribution of $\Sigma_N$ in the asymptotic sense, where $\Sigma_N$ is given by the following advanced information-plus-noise model:

$$\Sigma_N = (S \tilde{H} \tilde{D}^{1/2} X + W_l) (S \tilde{H} \tilde{D}^{1/2} X + W_l)^\dagger / N$$

(8)

which is difficult to handle by using Stieltjes transform approach [14], [12].

Remark 1. The analysis of the eigenvalue distribution of $\Sigma_N$ under an arbitrary diagonal matrix $\tilde{D}$ does not admit a tractable solution. Note that our primary concern here is to investigate the impact of power difference between the desired signal and the interference which guarantees, along with other system parameters, the separability of the signal and the interference subspaces. Therefore, without loss of generality, we assume the worst case scenario of power imbalance between inside-cell and outside-cell users. Specifically, we set all interference powers to the maximum interference power, denoted $P_I$, among all interfering users. On the other hand, we set the powers of all desired signals to the minimum signal power, denoted $P_s$, among all inside-cell users. Based on the above assumption, the diagonal matrix $\tilde{D}$ is now comprised of two distinct masses, namely $P_s$ and $P_I$, with multiplicity of $K$ and $K(L - 1)$, respectively.

1) Assumption
For the sake of analytical tractability and asymptotic results, we shall make the following assumptions:

- The input matrix $X$ consists of i.i.d. Gaussian entries. This assumption is not strict since the asymptotic result still holds given the entries are independent with finite moments of order greater than 2 \cite[Section 5.1]{12}, which is fulfilled for most signal constellations.

- The diagonal matrix $\tilde{D}$ has bounded eigenvalues, and in the large limit, the eigenvalue distribution converges to a deterministic distribution. Note that this is true since the attenuation due to path loss and shadowing is generally constant over many symbol intervals.

- The entries of matrix $S$ are assumed independent. Note that this assumption is violated in practice, where it is shown in \cite{15} that the moments of Vandermonde matrix are bounded above and below by the moments of Marčenko-Pasture distribution \cite{16} and Poisson distribution, respectively. Nevertheless, when the number of receive antennas grows large, this assumption becomes reasonable \cite{9}, and hence the moments of Marčenko-Pasture distribution will become a good approximation of the moments of $S$.

2) One-sided spectral analysis

While the exact solution of the AED of (8) is difficult to obtain, to simplify the derivation we assume high SNR regime, i.e., $W_l = 0$. Further, we also assume that $K \ll P$ and $P$ is sufficiently large so that orthogonality between users’ channel vectors can still hold. Under these assumptions, the eigenvalue distributions of the signal and the interference can be studied separately.

Proposition 1. Let $S \in \mathbb{C}^{M \times P}$, $\tilde{H}_l \in \mathbb{C}^{P \times K}$, and $X_l \in \mathbb{C}^{K \times N}$ be defined as in (7). Also let $M, K, P, N \rightarrow \infty$ with $K \ll P$, $K/M \rightarrow \alpha$, $P/M \rightarrow \beta$, and $K/N \rightarrow \gamma$. Consider the $N \times N$ Hermitian matrix $F = P_s(S\tilde{H}_lX_l)\dagger(S\tilde{H}_lX_l)/MN$. Then the AED of $F$ converges to a non-random distribution with Stieltjes transform $G_F(s)$ satisfying the following equation:

$$\beta \gamma^2 (sG_F(s) + 1) + P_sG_F(s)(sG_F(s) - \gamma + 1)(\alpha sG_F(s) + \alpha - \beta \gamma) = 0. \quad (9)$$

\footnote{It is shown in \cite{4} that when $M \rightarrow \infty$, spatial correlation yields only a minor penalty on the orthogonality condition compared to the independently distributed channel. Note also for two channels $h_{ki}$ and $h_{lj}$ as defined in (1) with identical steering matrix $S$, $\lim_{P,M \rightarrow \infty} h_{ki}^\dagger h_{lj}/M \rightarrow 0$ almost surely as $K \ll P$.}
Proof: consider the scaled matrix product

\[ D_k = \sqrt{a_k / mpn} A_k B_k C_k \]  \hspace{1cm} (10)

and let

\[ G_1 \overset{\Delta}{=} D_k^\dagger D_k = a_k C_k^\dagger (A_k B_k)^\dagger (A_k B_k) C_k / mpn \]  \hspace{1cm} (11)

where \( a_k \) is a multiplicative constant, and the matrices \( A_k \in \mathbb{C}^{m \times p} \), \( B_k \in \mathbb{C}^{p \times l_k} \), and \( C_k \in \mathbb{C}^{l_k \times n} \) are mutually independent. Moreover, the entries of \( A_k \) are assumed independent with zero mean and unit variance, whereas the entries of \( B_k \) and \( C_k \) are assumed i.i.d. \( \mathcal{CN}(0,1) \). We also assume \( \text{rank} (A_k B_k C_k) = l_k \), i.e., \( G_1 \) has \( l_k \) nonzero eigenvalues.

Following [9], note that \( \text{tr}(C_k^\dagger (A_k B_k)^\dagger (A_k B_k) C_k C_k^\dagger) / mpn \), i.e., both product factors inside the trace operators share the same nonzero eigenvalues and differ only in \( |n - l_k| \) zeros. Thus it would be more adequate to study the eigenvalue distribution of the \( l_k \times l_k \) matrix

\[ G_2 \overset{\Delta}{=} a_k (A_k B_k)^\dagger (A_k B_k) C_k^\dagger / mpn \]  \hspace{1cm} (12)

without the need to obtain the eigenvalue distribution of \( G_1 \) directly. Now let \( G_2 = G_{21} G_{22} \), where \( G_{21} \) and \( G_{22} \) are defined, respectively, as

\[ G_{21} \overset{\Delta}{=} a_k (A_k B_k)^\dagger (A_k B_k) / mp, \]  \hspace{1cm} (13)

\[ G_{22} \overset{\Delta}{=} C_k C_k^\dagger / n. \]  \hspace{1cm} (14)

To find the AED of \( G_2 \), one may first derive the marginal distributions of \( G_{21} \) and \( G_{22} \) by means of RMT, then FPT, namely, the free multiplicative law [17], [18], [13] can be invoked straightforwardly. The usefulness of FPT lies in the possibility of computing the AED of a product and a sum of Hermitian random matrices based solely on the marginal distributions, given the matrices are asymptotically free (see e.g., [13], Sec. 2.4] for details on asymptotic freeness of random matrices). In FPT, the free multiplicative law of a product of random matrices is computed via the so-called \( S \)-transform (see e.g., [13], Def. 2.15). More specifically, when two Hermitian random matrices are asymptotically free, with each matrix has an eigenvalue distribution that converges almost surely in distribution to a compactly supported distribution, then the \( S \)-transform of their product is the product of their corresponding \( S \)-transforms ([12], Thm. 4.7).
Under the assumptions made on $A_k$ and $B_k$, the asymptotic freeness condition is fulfilled (see Appendix for details). From \cite{9}, the Stieltjes transform $G_{G_{21}}(s)$ of the AED of $G_{21}$ satisfies the following equation:

$$a_k \alpha_1^2 s G_{G_{21}}^2(s) + a_k \alpha_1 s (\alpha_2 + 1 - 2\alpha_1) G_{G_{21}}(s) + (\alpha_2 s + a_k (\alpha_1 - 1)(\alpha_1 - \alpha_2)) G_{G_{21}}(s) - \alpha_2 = 0$$

(15)

where $\alpha_1 = l_k/m$, and $\alpha_2 = p/m$ are limiting ratios. Further, from \cite{9} Def. (21), if we substitute $s = -1/\lambda$ in (15), and replace $G_{G_{21}}(-1/\lambda)$ by $-\lambda(\Upsilon(\lambda) + 1)$, we obtain

$$a_k \alpha_1^2 \lambda(\Upsilon(\lambda) + 1)^3 + a_k \alpha_1 \lambda(\Upsilon(\lambda) + 1)^2(\alpha_2 - 2\alpha_1 + 1) - \lambda(\alpha_2/\lambda - a_k (\alpha_1 - \alpha_2)(\alpha_1 - 1))(\Upsilon(\lambda) + 1) + \alpha_2 = 0.$$  

(16)

Replacing $\lambda$ by $\Upsilon^{-1}(z)$, and using \cite{9} Def. (20), it is easy to check that the $S$-transform $S_{G_{21}}(z)$ reads

$$S_{G_{21}}(z) = \frac{\alpha_2}{a_k(\alpha_1 z + \alpha_2)(\alpha_1 z + 1)}.$$  

(17)

Next, since the entries of $C_k/\sqrt{n}$ are assumed i.i.d. each with zero mean and variance $1/n$, then $G_{22}$ is a central Wishart matrix with $n$ degrees of freedom \cite{19}, \cite{12}, \cite{13}. Further, its eigenvalue distribution follows the well-known Marčenko-Pastur law \cite{16}. Thus when $n, l_k \to \infty$ with a fixed ratio $l_k/n \to \alpha_3$, the $S$-transform $S_{G_{22}}(z)$ corresponding to $G_{22}$ reads as \cite{12}, \cite{13}

$$S_{G_{22}}(z) = \frac{1}{(\alpha_3 z + 1)}.$$  

(18)

Returning to (12), it is easy to see that, as a consequence of asymptotic freeness of $G_{21}$ and $G_{22}$ (the validity of this assumption is discussed in details in Appendix), the $S$-transform corresponding to $G_2$ is the product of $S_{G_{21}}(z)$ and $S_{G_{22}}(z)$, that is \cite{12} Thm. 4.7

$$S_{G_2}(z) = \frac{\alpha_2}{a_k(\alpha_1 z + \alpha_2)(\alpha_3 z + 1)(\alpha_1 z + 1)}.$$  

(19)

and from \cite{13} Thm. 2.32, it can be shown by straightforward algebra the $S$-transform corresponding to $G_1$ \cite{11} is

$$S_{G_1}(z) = \frac{\alpha_2}{a_k(\frac{\alpha_2}{\alpha_3} z + \alpha_2)(\frac{\alpha_2}{\alpha_3} z + 1)(z + \alpha_3)}.$$  

(20)

\footnote{Note that different definition of Stieltjes transform is used in \cite{9}.}
Now we are left with the task of determining the Stieltjes transform corresponding to (20). From [9, Def. (20) and (21)], it can be easily verified that the $S$-transform $S_F(z)$ and the Stieltjes transform $G_F(s)$ of a distribution function $F$ satisfies

$$S_F(-sG_F(s) - 1) = \frac{G_F(s)}{sG_F(s) + 1}. \quad (21)$$

Using (20) and with the aid of (21), it follows straightforwardly after some mathematical manipulations that the Stieltjes transform $G_{G_1}(s)$ corresponding to $G_1$ satisfies the following equation:

$$\alpha_2\alpha_3^2(1 + sG_{G_1}) + a_k G_{G_1}(1 - \alpha_3 + sG_{G_1})(\alpha_1 - \alpha_3 + \alpha_1 sG_{G_1}) = 0 \quad (22)$$

Finally, we remark that if we replace $a_k$, $\alpha_1$, $\alpha_2$ and $\alpha_3$ in (22) by $P_s$, $\alpha$, $\beta$, and $\gamma$, respectively, then we obtain (9).

Similarly, following the preceding derivation it can be shown that the eigenvalue distribution of interference satisfies (22) in the Stieltjes transform with $a_k$, $\alpha_1$, $\alpha_2$ and $\alpha_3$ are replaced by $P_I$, $K(L - 1)/M$, $P/M$, and $K(L - 1)/N$, respectively. When the i.i.d. assumption on the channel $\tilde{S}\tilde{H}_l$ holds we have

Corollary 2. Let $M$, $K$, and $N$ are large with $\alpha$, and $\gamma$ fixed yet not very small. Under rich scattering propagation (i.e., $P \to \infty$), the eigenvalue distribution of $F = P_s(\tilde{S}\tilde{H}_lX_l)^\dagger(\tilde{S}\tilde{H}_lX_l)/MN$ converges to a non-random distribution with Stieltjes transform $G_F$ satisfying

$$\gamma(1 + sG_F) - P_s G_F(1 + sG_F - \gamma)(\alpha + \alpha sG_F - \gamma) = 0. \quad (23)$$

Remark 2. Equation (23) exactly reproduces the result in [8, Eq.(75)], where it was derived under uncorrelated channel model and the convention that $P_s = N/K$.

Corollary 3. Assume the channel model (5). Further, assume $K$, $P$ and $N$ are large with $K/P \to \alpha'$ and $\gamma$ fixed but not very small. Then adding more antennas at the BS, the eigenvalue distribution of $F = P_s(\tilde{S}\tilde{H}_lX_l)^\dagger(\tilde{S}\tilde{H}_lX_l)/MN$ converges to a non-random distribution resembling the eigenvalue distribution under uncorrelated channel with $P$ receive antennas. In Stieltjes domain, the AED of $F$ satisfies (23) with $\alpha$ replaced by $\alpha'$.

Remark 3. From Corollaries 2 and 3 we observe that the gap between signal subspace and interference subspace under a physical channel with $P$ AoAs and unlimited number of receive antennas $M$ is equivalent to the gap under uncorrelated channel with $M = P$ antennas at the
BS. Since blind subpace method relies entirely on the distance between the two subspaces, it turns out that its performance depends mainly on $P$ rather than $M$. Nevertheless, the array gain due to the increase of $M$ can still be partly retained, especially when signal subspace is detached from interference subspace. After that, the performance will be limited by the degrees of freedom of the channel, and hence it saturates.

3) The support of the distribution

In this part we will characterize the support of the distribution, which is important for knowing under what values of the system’s parameters the distribution yields two distinct eigenvalues clusters, and hence subspace separability. It is known that the Stieltjes transform is increasing on intervals on the real line outside the support of its distribution function [12]. Further, its inverse function $s(x)$ is also increasing in these intervals only. Therefore, the endpoints of the support can be determined by finding the local extrema of the inverse function $s(x)$. Unfortunately, these local extrema are notoriously difficult to express in a closed-form solution. So this paper is only devoted to deriving approximate formula without explicitly obtaining the extreme points. To that end, we plot $s(x)$ for real $x$ to find the regions where $s(x)$ is increasing, and hence the support of the distribution, denoted $S$ can be defined as the finite union of regions on the real line where $s(x)$ is not increasing, that is

$$S = \mathcal{R} \bigcup_{x_1 < x_2 \in \mathcal{R}} \{ s(x_1), s(x_2) \mid \forall x \in (x_1, x_2), \frac{d}{dx}s(x) > 0 \}.$$  \hspace{1cm} (24)

To find the support of the distribution, it seems more adequate to use the $l_k \times l_k$ matrix $G_2$ (i.e., a matrix of much smaller dimensions) rather than the $n \times n$ matrix $G_1$, since we are only interested in the nonzero eigenvalues. By the aid of (21) one can verify that the Stieltjes transform $G_{G_2}$ corresponding to the $S$-transform [19] satisfies

$$sG_{G_2} + a_kG_{G_2}(\frac{l_k}{m}sG_{G_2} + \frac{l_k}{m} - 1)(\frac{l_k}{n}sG_{G_2} + \frac{l_k}{n} - 1) \times$$

$$\left(\frac{l_k}{p}sG_{G_2} + \frac{l_k}{n} - 1\right) + 1 = 0.$$  \hspace{1cm} (25)

\footnote{The inverse function $s(x)$ is obtained as a solution for variable $s$ in the fixed-point equation satisfying the Stieltjes transform, where $x$ is a real dummy variable.}
By expanding (25) and after simple mathematical manipulations we have

\[ a_k l_k^3 x^4 s^3 - a_k l_k^2 x^3 (-3l_k + M + N + P) s^2 \\
+ x (MNP + a_k l_k x (3l_k^2 + MN + (M + N) P - 2l_k (M + N + P))) s \\
+ MNP - a_k x (-l_k + M) (l_k - N) (l_k - P) = 0 \]

where we have replaced \( G_{G_2} \) by the dummy variable \( x \) and \( m, n, \) and \( p \) by \( M, N, \) and \( P, \) respectively. Note that (26) is a cubic equation in \( s \), and its three roots, denoted \( s(x) = \{ s_1(x), s_2(x), s_3(x) \} \) must be computed, and hence the support is defined by (24). We remark that (26) can be also used to find the support of eigenvalue distribution of interference by noting that \( a_k = P_I \) and \( l_k = K(L - 1) \).

4) Double-sided spectral analysis

So far we have treated the signal subspace and the interference subspace separately. However, as far as systems with finite dimensions are concerned in practice, the two sets of eigenvectors corresponding to the signal and the interference are somehow interconnected. Thus the eigenvalue distribution of the signal and the interference, should be studied jointly. For the sake of mathematical tractability, we again assume high SNR regime, i.e., we set \( W_l = 0 \) in (8). Let

\[ \tilde{G}_1 \triangleq (S\tilde{H})\tilde{D}^{1/2}XX^\dagger\tilde{D}^{1/2}(S\tilde{H})^\dagger/MN. \]

Instead, it is more convenient to study the eigenvalue distribution of

\[ \tilde{G}_2 \triangleq \tilde{D}^{1/2}XX^\dagger\tilde{D}^{1/2}(S\tilde{H})^\dagger(S\tilde{H})/MN \]

without accessing to the information about the eigenvalues of \( \tilde{G}_1 \). Further, let \( \tilde{G}_2 = \tilde{G}_{21}\tilde{G}_{22} \), where \( \tilde{G}_{21} \) and \( \tilde{G}_{22} \) are defined as follows:

\[ \tilde{G}_{21} \triangleq \tilde{D}^{1/2}XX^\dagger\tilde{D}^{1/2}/N, \]

\[ \tilde{G}_{22} \triangleq (S\tilde{H})^\dagger(S\tilde{H})/M \]

Remark 4. Equation (28) is indeed intuitive in the sense that it gives an insight into the behavior of the gap between the signal subspace and the interference subspace as the ratio \( P_s/P_I \) varies. To see this, assume a rich scattering environment, i.e., \( P \to \infty \). From central limit theorem, as \( N, M \to \infty \) with \( K, L \) fixed, \( XX^\dagger/N \), and \( (S\tilde{H})^\dagger(S\tilde{H})/M \) reduce to identity matrices. Thus the final spectrum of (28) is dictated by the spectrum of \( \tilde{D} \), that is, a probability mass function (pmf) with two masses at \( P_s \) and \( P_I \). In this case, subspace separability is possible whenever
$P_s > P_I$. However, when the dimensions of random matrices in (28) grow large, but with fixed ratios the eigenvalue distribution converges to a non-random distribution rather than the pmf. This leads to eigenvalues spread around the center eigenvalue. Therefore, a large difference between those deterministic parameters of the system is important which help to shift the two clusters of eigenvalues away from each other.

From [20], the asymptotic Stieltjes transform of $X^\dagger \tilde{D}X/N$, denoted $G_{X^\dagger \tilde{D}X/N}(s)$, as $KL, N \to \infty$ with $KL/N \to \gamma$ is the unique solution of

$$G_{X^\dagger \tilde{D}X/N}(s) = -\left(s - \gamma \int \frac{\lambda f(\lambda) d\lambda}{1 + \lambda G_{X^\dagger \tilde{D}X/N}(s)}\right)^{-1}$$

(31)

where $f(\lambda)$ is the probability density function (pdf) of eigenvalue of $X^\dagger \tilde{D}X/N$. It would seem difficult to evaluate (31). Since the entries of $X$ are assumed i.i.d. $CN(0, 1)$, this class of matrices is unitarily invariant and thus asymptotically free w.r.t. any Hermitian matrix (see Appendix). Then it follows from [21, Eq.(43)] that

$$S_{G_{21}}(z) = \frac{1}{\gamma z + 1} S_D(z)$$

(32)

which is derived within the framework of FPT. Since the eigenvalue pdf $f_D(x)$ of $\tilde{D}$ converges to a pmf of two distinct eigenvalues, namely, $P_s$ and $P_I$, hence $f_D(x)$ can be written as

$$f_D(x) = \frac{1}{L} \delta(x - P_s) + \frac{L - 1}{L} \delta(x - P_I)$$

(33)

where $\delta(.)$ is the Dirac delta function. By the definition of Stieltjes transform [9], [12], [13], one can verify that $f_D(x)$ admits a Stieltjes transform $G_D(s)$ of the form

$$G_D = \frac{LP_s - Ls + P_I - P_s}{L(P_s - s)(P_I - s)}$$

(34)

and from [12, Def. 3.4], it is easy to check that the corresponding $S$-transform fulfils the following quadratic equation

$$LP_I P_s z S_D^2(z) - (P_s - P_I + LP_I + LP_I z + LP_s z) S_D(z) + (L + Lz) = 0$$

(35)

with the two roots

$$S_D(z) = \frac{b(z) \pm \sqrt{b^2(z) - 4L^2P_I P_s (z + 1)z}}{2LP_I P_s z}$$

(36)

The center eigenvalues are determined by the deterministic quantities in the channel model. For instance, $Y_l Y_l^\dagger / MN$ has two deterministic quantities, namely, $P_s$ and $P_I$ corresponding to the effective signal and interference powers, respectively.
where \( b(z) \triangleq P_s - P_I + LP_I + LP_Iz + LP_s z \). One may prove that the root with minus sign is the true solution to (35).\(^6\) Let \( KL/M \to \alpha \) and \( KL/P \to \eta \) as \( KL, M, \) and \( P \to \infty \). Plugging (36) into (32), combining the result with (17), and assuming the factors (29) and (30) are asymptotically free (see Appendix), then it follows from [13, Thm. 2.32] the \( S\)-transform corresponding to \( \tilde{G}_2 \) can be finally written as

\[
S_{\tilde{G}_2}(z) = \frac{b(z) - \sqrt{b^2(z) - 4L^2P_IP_s(z+1)z}}{2LP_IP_s(z+1)(\alpha z+1)(\eta z+1)}.
\] (37)

By making use of (21) in (37) we have the following result.

**Proposition 4.** Let \( S, \tilde{H}, \tilde{D}, \) and \( X \) be defined as in (7). Let also \( KL, N, P, \) and \( M \to \infty \), with \( KL/M \to \alpha, KL/P \to \eta, \) and \( KL/N \to \gamma \). Further, denote \( G_{\tilde{G}_2} \) the Stieltjes transform corresponding to \( \tilde{G}_2 \) (28). Then \( G_{\tilde{G}_2} \) is the unique solution of the following fixed-point equation:

\[
LP_s(sG_{\tilde{G}_2} + 1) - P_s + P_I(1 + L(sG_{\tilde{G}_2} + 2GP_s(\alpha sG_{\tilde{G}_2}) + \alpha - 1)(\eta sG_{\tilde{G}_2} + \eta - 1)(\gamma sG_{\tilde{G}_2} + \gamma - 1))) + (P_I^2(LsG_{\tilde{G}_2} + 1)^2 + \frac{P_s^2}{L}(LsG_{\tilde{G}_2} + L - 1)^2 - 2P_IP_s(1 - L + L^2(sG_{\tilde{G}_2} + 1)sG_{\tilde{G}_2}))^{1/2} = 0.
\] (38)

Using (38) to find the support of the distribution is difficult, hence, we adopt an approximate solution. Rewriting (38) in terms of the parameters \( K, L, M, N, \) and \( P \) yields

\[
\frac{2K^3LP_IP_sx^3}{MNPN} - 2K^2L^3P_IP_sx\left(\frac{1}{MN} + \frac{1}{MP} + \frac{1}{NP}\right)x^2
+ (2KL^2P_IP_sx\left(\frac{1}{M} + \frac{1}{N} + \frac{1}{P}\right) + L(P_I + P_s))x
+ (P_I^2(L(v - 1) + 1)^2 + \frac{P_s^2}{L}(Lv - 1)^2
+ 2P_IP_s(L - L^2(v - 1)x - 1))^{1/2}
- LP_I + P_I - P_s - 2LP_IP_sx = 0
\] (39)

where we have replaced \( G_{\tilde{G}_2} \) by the dummy variable \( x \) and \( v = sx + 1 \). Note that (39) (from which the all roots \( s(x) \) must be found and all extreme values are identified) is a polynomial of degree 6 in variables \( s \) and \( x \), hence, finding its zeros is indeed intractable. Instead, we adopt an approximate solution such that the above higher order polynomial is reduced to a low-order
polynomial of degree 2 in variable $s$. To that end, notice that the multiplicative coefficients of the terms $\nu^3$, $\nu^2$ and $\nu$ scale like $\alpha \eta \gamma$, $\alpha \gamma + \alpha \eta + \eta \gamma$, and $\alpha + \eta + \gamma$, respectively. It turns out that in order to truncate out the terms $\nu^3$ and $\nu^2$ and hence have a good approximation of the true solution, the following conditions should be satisfied:

$$\frac{\alpha + \eta + \gamma}{\alpha \eta \gamma} \gg 1, \quad \frac{\alpha + \eta + \gamma}{\alpha \gamma + \alpha \eta + \eta \gamma} \gg 1$$

where ignoring the terms $\nu^3$ and $\nu^2$ don’t affect the final answer significantly. Therefore, after truncating the higher order terms, namely, $\nu^3$ and $\nu^2$, we have

$$(2KL^2PI_Ps\left(\frac{1}{M} + \frac{1}{N} + \frac{1}{P}\right) + L \left(P_I + P_s\right))\nu$$

$$+ \left(P_I^2(1 + L(\nu - 1))^2 + P_s^2(L\nu - 1)^2\right)$$

$$+ 2P_IP_s(L - L^2(\nu - 1)(\nu - 1))^{1/2}$$

$$- LP_I + P_I - P_s - 2LP_IP_s = 0$$

and its two roots $s_1(x)$ and $s_2(x)$ can be computed, and thus the support is defined by (24).

B. Physical channel model with distinct AoAs

So far we have considered the case of identical AoAs. However, depending on the complexity of the propagation environment, we may have different signals received from different/shared directions, captured by the channel model (3). Indeed, the analytical analysis of this channel model is extremely intricate. For the sake of simplicity, there is no significant loss of generality incurred by using the channel model (4). Here we recall that the received signals of the $i$th cell impinge upon the array of the intended BS from i.i.d. AoAs of cardinality $P_i$. We also assume that all sets of AoAs over all cells are mutually independent. Without loss of generality, we consider the worst case of power imbalance between the desired signal and the interfering signal as has been stated in Remark 1. Now we may rewrite (4) as

$$Y_l = \sqrt{P_s}S_l\tilde{H}_lX_l + \sqrt{P_I}\sum_{i=1,i\neq l}^{L} S_i\tilde{H}_iX_i + W_l. \quad (42)$$

Studying the eigenvalue distribution of the signal model (42) is intractable even when $W_l = 0$. Therefore, a seemingly natural way is to do one-sided analysis of the distribution of the first and second terms in (42). This, however, would help us get some insights into the behavior of the eigenvalue distribution under the channel model (42) as it will be shown shortly. For the sake of illustration, we let $l = 1$. 
First, note that the support of the eigenvalue distribution of the first term in (42) can be computed by (26). On the other hand, the second term of (42) can be rewritten in a compact way:

\[ Y_I = \sqrt{P_I} S_I H_I X_I \]  

(43)

where \( S_I = (S_2, S_3, \ldots, S_L) \in \mathbb{C}^{M \times n} \) is the composite steering matrix, where \( n = P_2 + P_3 + \cdots + P_L \), \( H_I \in \mathbb{C}^{n \times K(L-1)} \) is a block diagonal matrix whose blocks are \( \tilde{H}_2, \tilde{H}_3, \ldots, \tilde{H}_L \) and \( X_I = (X_2, \ldots, X_L)^T \in \mathbb{C}^{K(L-1) \times N} \) comprises all interference signals. Note that the eigenvalues of \( H_I \) equals the combined eigenvalues of the submatrices on its diagonal. Moreover, the eigenvalue distribution of each Hermitian matrix \( \tilde{H}_i \tilde{H}_i^\dagger \) converges to a non-random distribution as \( K, P_i \to \infty \) with a fixed ratio \( K/P_i \). This implies that the eigenvalue distribution of \( H_I H_I^\dagger \) also converges to a non-random distribution. To proceed with the analysis, the following Lemma enables us to characterize the behavior of eigenvalue distribution asymptotically.

**Lemma 5.** Assume \( K < P_i, i = 2, \ldots, L \) with fixed ratios \( K/P_i \to \beta_i \). Further, let \( \lambda_i = P_i/n \) and \( f_i(x) \) be the eigenvalue distribution of \( \tilde{H}_i \tilde{H}_i^\dagger \). Then, as \( K, P_i \to \infty \), the eigenvalue distribution \( f(x) \) of \( H_I H_I^\dagger \) is related to \( f_i(x), i = 2, 3, \ldots, L \) through the following relationship:

\[ f(x) = \sum_{i=2}^L \lambda_i f_i(x). \]  

(44)

**Proof:** Since the eigenvalues of \( H_I \) are equal to the combined eigenvalues of \( \tilde{H}_2, \ldots, \tilde{H}_L \), we can write

\[ tr(H_I H_I^\dagger) = tr(\tilde{H}_2 \tilde{H}_2^\dagger) + \cdots + tr(\tilde{H}_L \tilde{H}_L^\dagger) \]

and the \( k \)th moment of the eigenvalue distribution of \( H_I H_I^\dagger \) is given by

\[ tr(H_I H_I^\dagger)^k = tr(\tilde{H}_2 \tilde{H}_2^\dagger)^k + \cdots + tr(\tilde{H}_L \tilde{H}_L^\dagger)^k. \]  

(45)

Note that the number of eigenvalues of \( H_I H_I^\dagger \) are \( n \), while each submatrix \( \tilde{H}_i \tilde{H}_i^\dagger \) has \( P_i \) eigenvalues (i.e., the assumption \( K < P_i \) implies \( K \) nonzero eigenvalues and \( P_i - K \) zero eigenvalues). Thus the normalized trace form of (45) reads

\[ tr_n(H_I H_I^\dagger)^k = \frac{1}{n} (P_2 tr_2(\tilde{H}_2 \tilde{H}_2^\dagger)^k + \cdots + P_L tr_L(\tilde{H}_L \tilde{H}_L^\dagger)^k). \]  

(46)

Recall that the Stieltjes transform \( G_E(s) \) of \( n \times n \) Hermitian matrix \( E \) can be expanded in a Laurent series involving the moments of \( E \) as (see e.g., [12, Thm. 3.3])

\[ G_E(s) = -\frac{1}{s} \sum_{k=0}^{\infty} \frac{tr_n(E^k)}{s^k}. \]  

(47)
Combining (47) with (46) yields

\[ G_{H_i H_i^\dagger}(s) = \sum_{i=2}^{L} \lambda_i G_{\hat{H}_i \hat{H}_i^\dagger}(s). \] (48)

Then the eigenvalue distribution is reconstructed from \( G_{H_i H_i^\dagger}(s) \) by applying the inversion formula of Stieltjes transform [13], [12] to obtain (44).

Since each submatrix \( H_i \) consist of i.i.d. complex Gaussian entries, it follows from (44) that the eigenvalue distribution of \( H_i H_i^\dagger \) is a weighted sum of the Marčenko-Pasture laws [16] with different ratios \( \beta_i, i = 2, \ldots, L \). One particular case is when \( P_2 = \cdots = P_L = P \), it follows that the eigenvalue distribution of \( H_i H_i^\dagger \) reduces to a Marčenko-Pasture distribution with a single parameter \( \beta = K/P \).

**Proposition 6.** Consider the channel model (42). Without loss of generality, assume \( P_1 > P_2 > \cdots > P_L \). Then the users in the \( L \)th cell contribute most to the spreading of the eigenvalues of channel. Further, under any system parameters, the power ratio \( P_s/P_1 \) should be adjusted according to the \( L \)th cell, which is required to maintain specified link performance when subspace method is used at BS.

**Remark 5.** We observe that applying the free multiplicative law to the channel model (42) is extremely difficult because of nonidentical dimensions of the matrices \( H_i, i = 1, 2, \ldots, L \). Therefore, we provide here a non-rigorous yet intuitive proof of Proposition 6. Further, the fact that the users experiencing less channel scatters render the eigenvalue distribution getting wider is indeed very intuitive. Notice that the maximum gap between the two subspaces occurs when \( P \to \infty \) (i.e., uncorrelated channel), i.e., \( K/P \to 0 \). In addition, from Lemma 5, the eigenvalue distribution of \( H_i H_i^\dagger \) is a weighted sum of Marčenko-Pasture distributions with different ratios. Assume that all limiting ratios, \( K/P_i \), are infinitely small except one ratio, say the last ratio, which is made large (i.e., the signals of the \( L \)th cell are received at the desired BS from a very small number of AoAs). Thus the support of the resulting distribution is dominated by the \( L \)th distribution. This leads to a wider spread of the eigenvalues of the interference.

To get insight into how the case of distinct AoAs may shape the distribution, and thus change the support of interference eigenvalue distribution, we assume the same number of AoAs per each cell w.r.t. the desired BS. Again by using one-sided spectral analysis and by following similar steps in Sec.III-A with the aid of Lemma 5, one can prove that the eigenvalue distribution of
the $N \times N$ matrix $Y_t^*Y_t/MN$ reads in the Stieltjes domain as

$$
- NP_I(1 + sx)(K - KL + N + Nsx)(N + P - LP \\
+ Nsx)x + M(KPP_Ix(L - 1)^2 - N(L - 1) \times \\
(P + P_Ix(K + P))(1 + sx) + P_Ix(N + Nsx)^2) = 0.
$$

By computing the three roots of (49), $s(x) = \{s_1(x), s_2(x), s_3(x)\}$, then the support is defined by (24).

## IV. Numerical results

In this section, we use finite-size scenarios to show some numerical results that verify the theoretical results obtained in the asymptotic limit. We also provide simulation results for the uncoded bit error rate (BER) and compare the performance of subspace-based channel estimation scheme [8] under the physical channel model and the i.d. channel. The pilot-based channel estimation scheme is also shown. In our simulations, the BS antennas are sparsely spaced at twice the carrier wavelength unless explicitly stated otherwise.

### A. The support of AED

In Figs. 4a and 4b we use (26) and (41), respectively, to show the approximated boundaries of the eigenvalue distribution. The number of receive antennas $M = 400$, $P = 200$ AoAs, $L = 4$ cells, $K = 5$ users per cell, the coherence time $N = 1000$ symbol periods,\(^7\) the desired signal power $P_s = 0.1(-10\text{dB})$, whereas the interference power $P_I = 0.025 (~-16\text{dB})$, and high SNR is assumed (i.e., $W = 0$). Further, the support of distribution of the desired signal and the interference can be read on the right vertical axes, while on the left vertical axes is the exact noise-free support obtained for an i.d. channel with $M = 400$.

As expected, the eigenvalue clusters of the signal and the interference exhibit larger spread from both sides of the support, and hence the gap between the two clusters decreases significantly. In contrast, under the same setting, but i.d. channel, the eigenvalues show less spread and the gap between the two clusters is more pronounced. In practice, especially with noisy samples, it turns out that, realistic physical channel, relative to i.d. channel, incurs higher power ratio

\(^7\)N is taken to be a relatively large in order to highlight the effect of the number of AoAs and power ratio, rather than the effect of the number of measurements.
$P_s/P_I$ such that the spectrum splits into signal and interference eigenvalues in order to maintain a comparable BER. Otherwise, a high frequency reuse factor, for example, would be needed, which leads to poor use of spectrum resources.

The supported boundaries computed by (26) and (41) are compared to the histogram of nonzero eigenvalues in Fig. 5. Obviously, the histogram is composed of two bulks of eigenvalues clustered around two eigenvalues, 25 and 100 (center eigenvalues). The solid and dashed vertical lines correspond to the endpoints approximated by (41) and (26), respectively. Note that the approximated boundaries from (41) are almost in agreement with the boundaries of the two bulks obtained from the histogram of eigenvalues.

Fig. 6 shows the superimposed probability densities of eigenvalue for the physical channel (solid line) and the i.d. channel (dashed line). The parameters $N, K, L, P_s,$ and $P_I$ are the same as those in Fig. 4. In the physical channel, $P = 100$ identical AoAs for all users w.r.t. BS1, and the number of receive antennas $M$ is set to 600. On the other hand, the number of receive antennas in the case of i.d. channel is set to the number of AoAs, that is $M = 100$. Note that the eigenvalue pdf of $Y_l Y_l^\dagger/M$ almost matches the eigenvalue pdf of the corresponding i.d. channel. Thus, this result implies that the gap between the two bulks of eigenvalues (corresponding to the signal subspace and the interference subspace) of the physical channel with $P$ AoAs cannot be further improved beyond what can be achieved when the channel is i.d. with $P$ receive antennas as $M$ increases unboundedly (see Cor. 3). Nevertheless, array gain can be still retained as $M$ increases, but ultimately the performance will saturate as will be shown in the next subsection.
Fig. 5. Normalized histogram of nonzero eigenvalues of $Y_l Y_l^\dagger / M$, $M = 400$, $P = 200$, $N = 1000$, $K = 5$, $L = 4$, $P_s = -10$dB, $P_I = -16$dB. The solid and dashed lines are the approximated boundaries obtained by (41) and (26), respectively.

Fig. 6. Empirical pdfs of nonzero eigenvalues of $Y_l Y_l^\dagger / M$ for physical and i.d. channels, $N = 1000$, $K = 5$, $L = 4$, $P_s = -10$dB, $P_I = -16$dB.

Finally, Figs 7a and 7b show the normalized histogram of eigenvalue distributions for physical channel model with distinct AoAs. The parameters $N, K, L, P_s$, and $P_I$ are the same as those in Fig 4. In Fig 7a we independently generate four sets of equal number of AoAs ($P_1 = P_2 = P_3 = P_4 = 200$) corresponding to cells 1, 2, 3, and 4 w.r.t. BS1. On the other hand, in Fig 7b we fix the number of AoAs of cells 1, 2, and 3 to 200 (i.e., a large number of AoAs), whereas the number of AoAs corresponding to cell 4 is set to 20 (i.e., very small number of AoAs). Compared with the histogram in Fig 7a, it is clear that the fourth cell renders the interference bulk of eigenvalues (left bulk) more wider and hence the gap is less pronounced.
B. Bit error rate

To give an intuitive feel of the effect of the physical channel on the performance of subspace method, in Fig. 1 we show the uncoded BER versus the power ratio $P_I/P_S$. In our simulation we consider the uplink in a four-cell network with 5 users per cell and QPSK modulation scheme. The coherence time of the channel $N = 400$ symbol periods, all signals are received from identical AoAs with cardinality $P = 200$ and per-user SNR = $-5$dB. To show the effect of increasing the number of receive antennas $M$, we use different values: $M = 200, 400, 600$, where the antenna elements are critically-spaced. We consider full reuse of $K$ orthogonal pilot sequences across all cells. In addition, we use linear zero-forcing (ZF) for channel estimation (after applying SVD) and matched filter (MF) for data detection. The performance comparison with the classical pilot-based channel estimator is also shown, in which ZF and MF receivers are used to estimate the channel and detect data, respectively.

From Fig. 1 we notice that the physical channel incurs loss of performance as opposed to the i.d. channel. For instance, when each BS is equipped with 400 antennas and the number of AoAs is 200, to achieve the same BER of $10^{-2}$, the power ratio should be roughly doubled, compared with the case of i.d. channel. Also, subspace-based scheme outperforms the pilot-based scheme in all cases except at the very high interference level, which is unlikely in practical scenarios. Further, the performance of the subspace-based scheme improves gradually with increasing $M$. As expected, its performance under this physical channel becomes closer to its performance under
an i.d. channel with $M = 200$ receive antennas. However, for very small interference levels, it further improves. Actually, this verifies the fact that when $M$ grows large, the eigenvalue distribution of this physical channel with $P$ AoAs becomes identical to that of an i.d. channel with $M = P$ receive antennas. It should be noted that it still benefits partly from array gain due to increasing $M$, and this array gain becomes more useful when the two subspaces start to detach (i.e., when two subspaces overlap, the estimation error of channel will get larger and thus have a dominant impact on the performance). After that point, the performance becomes dominated by the degrees of freedom of the channel only.

In Fig. 8 we show the BER when distinct AoAs are used per each $K$ users in each cell w.r.t. BS1. To highlight the degradation of the performance due to the cell with the smallest number of AoAs, we fix the number of AoAs of cell 1, 2, and 3 to 100, while we vary the number of AoAs of the fourth cell, $P_4 = 10, 20, 50, 100$. It is clear that when the subspace-based scheme is used, the performance improves as $P_4$ increases, whereas it is almost the same for all values of $P_4$ when the pilot-based scheme is used. In subspace method, the gradual improvement of the performance can be interpreted as a consequence of gradual compression of the eigenvalues cluster of interference when $P_4$ increases (see Fig 7a and Fig 7b). However, when $P_4$ becomes sufficiently large, we observe a slight improvement of the performance (i.e., a saturation effect of the performance). This is because the performance becomes limited again by the other interfering cells associated with smaller number of AoAs. Further, it is expected that when $P_4$ decreases, the degradation in performance becomes even worse when the interfering power from the fourth
cell is comparable to the power of the desired users.

Fig. 9. BER when $N \sim KL$ under an i.d. channel, $M = 400$, $K = 15$, per-user SNR = 0dB, $L = 4$, ZF is used for channel estimation and MF for data detection.

Finally, in Fig. we simulate the BER when the number of users of the network is comparable to the coherence time of the channel. In these scenarios the all-orthogonal pilot-based scheme doesn’t work. Instead, we compare its performance with the classical pilot-based scheme. We assume an i.d. channel, $K = 15$, $L = 4$ and the coherence time $N$ varies around $KL = 60$, i.e., $N = 30, 60, 120$. It can be observed from Fig. that in all scenarios, especially the critical cases when $N = 30$ and $N = 60$, the subspace-based scheme outperforms the pilot-based scheme. Its performance also improves with increasing $N$ while the performance of pilot-based scheme is almost the same for all values of $N$. This means that subspace-based scheme exhibits better flexibility than the classical pilot-based scheme and all-orthogonal pilot-based scheme.

V. Conclusion

Pilot contamination is usually considered as a performance bottleneck in massive MIMO systems, if the conventional linear channel estimation method is used. As a departure from the linear estimation framework, the subspace-based scheme proposed in [8] exhibits better flexibility under various scenarios. In addition, it offers the potential of completely eliminating the pilot contamination effect. However, previous works are based on the common assumption of independent channels, which would be violated by real-world channels where the number of scatterers and AoAs might be both limited.

For the more sensible physical channel model, in this paper, we derived approximate analytic expressions in Stieltjes domain for the corresponding eigenvalue distributions in both scenarios
of identical and distinct AoAs among the cells. The results demonstrate that the physical channel will exhibit a larger spectral spread, thus resulting in a smaller gap between the two eigenvalue clusters, which correspond to signal and interference subspaces. Therefore, to obtain the same performance as of the i.d. channels, we must guarantee a larger power ratio between the intended users and the interfering users. In particular, for the scenarios of distinct number of AoAs, the required power difference is determined by the cell with the smallest number of AoAs. Moreover, it is shown that adding more antennas does not significantly affect the above spectral gap, even though, the array gain could be reaped until saturation.

For future research, since the blind subspace-based channel estimation scheme relies primarily on the eigenvalues spread of the channel (non-Bayesian or data-driven method), rather than the statistical properties of the channel (Bayesian method), the exploration of the possibility of combining the two methods is of importance from the perspective of system performance, especially in the scenario of bounded angular spreads.

Finally, the results in this paper demonstrate the significant impact of the physical channel with finite AoAs on the performance of massive MIMO system, especially for the blind subspace-based channel estimation methods.

APPENDIX

DISCUSSION ON THE FREENESS CONDITION

In this appendix, we prove the asymptotic freeness of the products of matrices in (12) and (28). Let \( \{E_i\} \in \mathbb{C}^{v \times v} \) and \( \{\tilde{E}_i\} \in \mathbb{C}^{v \times v} \) be two families of bi-unitarily invariant random matrices, whose AED’s, as \( v \to \infty \), converge almost surely to non-random distributions with compacted supports. Further, let \( \{Z_j\} \in \mathbb{C}^{v \times v} \) and \( \{\tilde{Z}_j\} \in \mathbb{C}^{v \times v} \) be two families of non-random diagonal matrices with almost sure convergence of their AED’s to non-random distributions with bounded eigenvalues as \( v \to \infty \). Then from [19, Thm. 4.3.11], the family \( \{E_i\}, \{\tilde{E}_i\}, \{Z_j\}, \{\tilde{Z}_j\} \), \( \forall i, \hat{i} \in I, j, \hat{j} \in J \) is asymptotically free. It is important to note that it is not necessary for \( E_i \) and \( \tilde{E}_i \) to be square matrices. This can be seen from the fact that \( Z_j \) and \( \tilde{Z}_j \) can have different dimensions, e.g., an arbitrary number of the last entries can be zeros. Equivalently, \( E_i \) and \( \tilde{E}_i \) can be considered as non-square matrices whereas \( Z_j \) and \( \tilde{Z}_j \) are two square matrices, such that the combination of all matrices is defined.

\(^8\)A bi-unitarily invariant matrix is a matrix whose the joint distribution of its entries is invariant when both left- and right-multiplied by unitary matrices.
Now, let us take the three cases: $E_i = \tilde{E}_i$ and $Z_j = \tilde{Z}_j = I$; $E_i = \tilde{E}_i$ and $Z_j = I$; $E_i = \tilde{E}_i$.

Then for each case we define the respective products:

$$E_i E_i^\dagger, \quad i \in I$$  \hspace{1cm} (50)

$$E_i^\dagger (U^\dagger Z_j U) E_i, \quad i \in I, j \in J$$  \hspace{1cm} (51)

$$Z_j E_i E_i^\dagger \tilde{Z}_j^\dagger, \quad i \in I, j, \tilde{j} \in J$$  \hspace{1cm} (52)

where the factor $(UZ_j U^\dagger)$ in (51) follows from the fact that $E_i$ is bi-unitarily invariant and hence can be replaced by $UE_i V^\dagger$, where $U$ and $V$ are unitary matrices. Note that $U^\dagger Z_j U$ is a unitarily invariant Hermitian matrix with AED converges to a non-random distribution (i.e., the eigenvectors are distributed in a maximally random way implying that $V^\dagger Z_j V$ fulfills the condition of [19, Thm. 4.3.11]), and by [19, Thm. 4.3.11] any subset of (50), (51) and (52) forms a family that is asymptotically free.

Now, from (12), since the entries of $B_k$ and $C_k$ are assumed Gaussian with zero mean and unit variance, then they form bi-unitarily invariant random matrices each with the AED converges to the well-known Marčenko-Pasture law [16]. Then the Hermitian matrices $(A_k B_k)^\dagger (A_k B_k) = B_k^\dagger (A_k^\dagger A_k) B_k$ and $C_k C_k^\dagger$ correspond, respectively, to (51) and (50). Further, because the assumption of i.i.d. of $A_k^\dagger A_k$ incurs insignificant penalty compared with the Vandermonde matrix, it follows from (50) and (51) that $B_k^\dagger$, $A_k^\dagger A_k$, $B_k$, and $C_k C_k^\dagger$ form a family that is free asymptotically.

Finally, the product $\tilde{D}^{1/2} X X^\dagger \tilde{D}^{1/2} (S \tilde{H})^\dagger (S \tilde{H})$ in (28) can be treated in the same way. Note that $\tilde{D}^{1/2} X X^\dagger \tilde{D}^{1/2}$ and $(S \tilde{H})^\dagger (S \tilde{H})$ are special instances of (52) and (51), respectively. Therefore, the asymptotic freeness property holds true for the products in (28).

**ACKNOWLEDGMENT**

The authors are indebted to the anonymous reviewers for several helpful comments and suggestions. This work is supported by Natural Science Foundation of China under grant 61231007 as well as the National Science and Technology Major Project under grant 2013ZX03001015-002, and Dr. Gesbert acknowledges the partial support of European project HARP under the FP7 ICYT Objective 1.1.

**REFERENCES**

[1] R. Mueller, M. Vehkapera, and L. Cottatellucci, “Blind pilot decontamination,” in ITG Workshop on Smart Antennas (WSA), Stuttgart, Germany, Mar. 2013.
[2] T. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[3] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, “Achieving massive MIMO spectral efficiency with a not-so-large number of antennas,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, 2012.

[4] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[5] H. Q. Ngo, E. Larsson, and T. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[6] 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.

[7] H. Q. Ngo and E. Larsson, “EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, Kyoto, Japan, Mar. 2012.

[8] R. Mueller, L. Cottatellucci, and M. Vehkapera, “Blind pilot decontamination,” *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, no. 99, pp. 1–1, 2014.

[9] R. R. Muller, “A random matrix model of communication via antenna arrays,” *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2495–2506, Sep. 2006.

[10] H. Ngo, E. Larsson, and T. Marzetta, “The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel,” *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, 2013.

[11] T. L. Marzetta and B. M. Hochwald, “Capacity of a mobile multi-antenna communication link in rayleigh flat fading,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, 1999.

[12] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. NY, USA: Cambridge University Press, 2011.

[13] A. M. Tulino and S. Verdu, “Random matrix theory and wireless communications,” *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.

[14] R. B. Dozier and J. W. Silverstein, “Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices,” *J. Multivar. Anal.*, vol. 98, no. 6, pp. 1099–1122, Jul. 2007.

[15] O. Ryan and M. Debbah, “Random Vandermonde Matrices-Part I: Fundamental results,” *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3115–3148, 2009.

[16] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Maths. of the USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Oct. 1967.

[17] D. Voiculescu, “Limit laws for random matrices and free products,” *Inventiones mathematicae*, vol. 104, no. 1, pp. 201–220, 1991.

[18] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, 2006.

[19] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy (Mathematical Surveys and Monographs No. 77)*. Providence, RI, USA.: American Mathematical Society, 2006.

[20] J. Silverstein and Z. Bai, “On the empirical distribution of eigenvalues of a class of large dimensional random matrices,” *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175 – 192, 1995.

[21] R. Muller, “On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels,” *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 2086–2091, 2002.