Impact of Random Spatial Correlation on the Capacity Scaling of Massive MIMO

Junyoung Nam*, Giuseppe Caire†, Mérouane Debbah‡, and H. Vincent Poor§
*Wireless Communications Research (WCR), Intel Labs, Santa Clara, CA, junyoung.nam@intel.com
†Communication Systems Department, Technical University of Berlin, Germany, caire@tu-berlin.de
‡Large Networks and Systems Group (LANEAS), CentraleSupélec, France, merouane.debbah@centralesupelec.fr
§Department of Electrical Engineering, Princeton University, Princeton, NJ, poor@princeton.edu

***Abstract*—In this paper, how random spatial correlation affects the capacity scaling of massive MIMO is investigated. Assuming that spatial correlation is sufficiently high such that the number of non-zero eigenvalues of the channel covariance matrices grows slower than the number of antennas, we establish the asymptotic capacity of massive MIMO. To achieve this capacity scaling, *statistical spatial despreading* based on the second-order channel statistics, implicitly done by coherent combining/precoding and channel propagation in limited scattering environments, is shown to play a pivotal role in terms of pilot decontamination and interference suppression. It is further noted that the channel hardening effect on which a well-known lower bound widely used in the massive MIMO literature hinges is weakened by finitely many antennas with spatial correlation. In particular, the bound may significantly underestimate the achievable rate of massive MIMO. This work therefore considers an alternative bounding technique to better understand the performance of massive MIMO.**

I. Introduction

Achieving ever higher spectral efficiency has always been a central problem in wireless networks. To this end, massive MIMO [1] is a viable technology that avoids centralized processing of multiple base station (BS) sites and yet provides unprecedented spectral efficiency, provided that every BS has a sufficiently large-scale antenna array and that uplink/downlink channel reciprocity is sustainable. In order to accurately predict the performance gain of massive MIMO, therefore, the asymptotic sum capacity has been of utmost interest in the limit of large number of antennas.

Since pilot contamination has been a fundamental bottleneck in massive MIMO, several techniques have been proposed to tackle the problem. For example, [2] proposed multicell cooperative (joint) precoding/combining over the entire network. Following [3], [4], many pilot decontamination techniques has exploited the linear independence between the subspaces spanned by the eigenvectors of the rank-deficient channel covariance matrices of users so that one can find some useful structure of subspaces with orthogoanl supports. In a different line of work, [5] recently proved that the linear independence of those subspaces is rather surprisingly not a necessary condition for the elimination of contamination with infinity many M antennas. The more general sufficient condition therein is an *asymptotic* linear independence of the covariance matrices other than that of their subspaces. This leads to the linear independence of all user channels that is then utilized to eliminate pilot contamination through multicell non-cooperative precoding/combining, requiring every cell to estimate all user channels in the network.

Rather than any explicit pilot decontamination technique, this work focuses on the capacity scaling law in the massive MIMO network with the randomness and the sparsity of angular supports of channel covariance matrices taken into consideration. Different scattering geometries of users located randomly in the network give rise to the randomness of channel covariance matrices of such users. The sparsity of angular supports means that the number of significant multipaths in angular domain may be much smaller than M, which arises through channel propagation in the typical limited scattering geometry. For such correlated fading channels in the homogeneous L-cell network, where every BS has sufficiently large M antennas and serves K users with common signal-to-noise ratio (SNR) and with the same coherence block size T_c, we show by some extensions of the method of deterministic equivalents [6], [7] that the ergodic sum capacity behaves as

$$\mathcal{C}_M = \left(1 - {T_c}^{-1}\right) KL \log(\mathsf{SNR}M) + o(1) \qquad (1)$$

where we used non-orthogonal pilot and $o(1) \to 0$ in the limit of M. Note that this scaling law is asymptotically tight and its multiplexing gain (defined by $\lim_{\mathsf{SNR}\to\infty} \frac{\mathcal{C}(\mathsf{SNR})}{\log \mathsf{SNR}}$) is indeed the best one can ever expect through a cut-set upper bound from the perspective of either pilot-aided or non-coherent communication with a single antenna in block fading [8], whose prelog factor is $(1 - {T_c}^{-1})$. The beamforming gain of M is also optimal.

The main differences of (1) and prior work can be summarized as follows: 1) It is not clear in prior work how the sum rate scales with M or K and how much multiplexing gain one can achieve. 2) Past work has generally assumed $\lim_M \frac{K}{M} = 0$, e.g., if $\liminf_M \frac{K}{M} > 0$, then the asymptotic linear independence in [5] does not hold so that pilot contamination cannot be completely eliminated. In contrast, $\limsup_M \frac{K}{M} < \infty$ is admissible to get (1). No matter how many users the network serves, "common" angular components between any pair of the channel covariance matrices are finite and the infinite sum of vanishing contamination terms also vanishes, as long as angular supports are sufficiently sparse as $M \to \infty$. This in fact breaks the long-standing spectral efficiency barrier of the prelog factor $\min(M, K)$ in multiuser MIMO with *independent fading*.

II. Preliminaries

A. System Model

We consider an L-cell network with M antennas at each BS. Indexing the kth user in BS ℓ by ℓ_k, $\boldsymbol{h}_{\ell\ell'_k}$ is the channel from user ℓ'_k to BS ℓ in the uplink MIMO, and, for notational brevity, let $\boldsymbol{h}_{\ell_k} \triangleq \boldsymbol{h}_{\ell\ell_k}\ \forall(\ell,k)$ for channels from users to their serving cell ℓ. Likewise the subscript $\ell\ell_k$ will be hereafter replaced with ℓ_k for all other notations. The covariance matrix of the channel vector $\boldsymbol{h}_{\ell\ell'_k}$ is denoted by $\boldsymbol{R}_{\ell\ell'_k} = \boldsymbol{U}_{\ell\ell'_k}\boldsymbol{\Lambda}_{\ell\ell'_k}\boldsymbol{U}^{\mathsf{H}}_{\ell\ell'_k}$, where $\boldsymbol{\Lambda}_{\ell\ell'_k} \in \mathbb{C}^{r_{\ell\ell'_k}\times r_{\ell\ell'_k}}$ is the diagonal matrix whose elements are the non-zero eigenvalues of $\boldsymbol{R}_{\ell\ell'_k}$, and $\boldsymbol{U}_{\ell\ell'_k} \in \mathbb{C}^{M\times r_{\ell\ell'_k}}$ is the eigenvector matrix. Let $\mathsf{SNR}^{\mathrm{ul}}_{\ell_k} = \frac{\mathrm{tr}\boldsymbol{\Lambda}_{\ell_k}}{M}P_{\mathrm{ul}}$ be the average uplink SNR per antenna of user ℓ_k, where P_{ul} is the uplink transmit power. The received signal vector at BS ℓ can then be given by $\boldsymbol{y}_\ell = \sum_k \boldsymbol{h}_{\ell_k}x_{\ell_k} + \sum_{\ell'\neq\ell}\sum_k \boldsymbol{h}_{\ell\ell'_k}x_{\ell'_k} + \boldsymbol{z}_\ell$, where x_{ℓ_k} is the input signal of user ℓ_k chosen from a Gaussian codebook and satisfies the power constraint such that $\mathbb{E}[x^{\mathsf{H}}_{\ell_k}x_{\ell_k}] \leq P_{\mathrm{ul}}$, and $\boldsymbol{z}_\ell \sim \mathcal{CN}(\boldsymbol{0},\boldsymbol{I})$ is the Gaussian noise.

B. Random Partial Fourier Correlation Model

For spatial correlation, we introduce a random partial (subsampled) Fourier model, motivated by the typical uniform linear array in multiple antenna systems. Let F_{jk} denote the (j,k)th entry of the discrete Fourier transform (DFT) matrix $\boldsymbol{F} \in \mathbb{C}^{M\times M}$ as $F_{jk} = \frac{1}{\sqrt{M}}e^{2\pi \jmath jk/M},\ j,k = 0,\ldots,M-1$. Suppose that $\boldsymbol{U}_{\ell_k}$ is composed of r_{ℓ_k} column vectors uniformly drawn at random *without replacement* from the Fourier basis functions of $\boldsymbol{F}$ so that different users can have common bases (angular supports), taking into account common scatterers shared by multiple users. The resulting partial unitary matrix can be represented by $\boldsymbol{U}_{\ell_k} = \boldsymbol{F}\boldsymbol{G}_{\ell_k}$, where $\boldsymbol{G}_{\ell_k} \in \mathbb{C}^{M\times r_{\ell_k}}$ is the random selection matrix that chooses r_{ℓ_k} columns without replacement from M columns of $\boldsymbol{F}$. While the channel covariance matrices of all users share the same angular components in some previous results (e.g., [9], [10]), the above model allows different angular supports as in [3], [4].

In this work, BS ℓ only knows the covariance matrices $\boldsymbol{R}_{\ell_k}$ of its own served users with the following assumptions.

Assumption 1. For all (ℓ,ℓ',k), $\limsup_{M\to\infty}\frac{r_{\ell\ell'_k}}{M}\|\boldsymbol{\Lambda}_{\ell\ell'_k}\|_\infty < \infty$, $\liminf_{M\to\infty}\frac{\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}}{M} > 0$, $\limsup_{M\to\infty}P_{\mathrm{ul}}M < \infty$, $\lim_{M\to\infty}\frac{\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}}{r_{\ell_k}\lambda_{\ell\ell'_k,i}} = 1$, where $\lambda_{\ell\ell'_k,i}$ be the ith non-zero eigenvalue of $\boldsymbol{R}_{\ell\ell'_k}$.

Assumption 2 (*Sublinear sparsity*)**.** The number $r_{\ell\ell'_k}$ of non-zero eigenvalues of $\boldsymbol{R}_{\ell\ell'_k}$ grows without bound but slower than M such that $\frac{r_{\ell\ell'_k}}{M} \triangleq \alpha_{\ell\ell'_k} \xrightarrow[M\to\infty]{} 0,\ \forall(\ell,\ell',k)$.

In conjunction with Assumption 2, the first condition of Assumption 1 implies that $\|\boldsymbol{\Lambda}_{\ell\ell'_k}\|_\infty$ is *not necessarily* uniformly bounded with respect to M. The uniform boundedness is a necessary condition for the method of deterministic equivalents [6], [7]. Hence, under these conditions the deterministic equivalents cannot be directly applied any longer. Also, this implies that $\lambda_{\ell\ell'_k,i}$ grows without bound for all i.

C. Sparse Transform and Spatial Despreading

Under the random partial Fourier model and Assumptions 1 and 2, in fact $\boldsymbol{U}_{\ell_k}$ serves as a sparse transformation matrix of $\boldsymbol{h}_{\ell_k}$ such that $\boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{h}_{\ell_k} = \boldsymbol{\Lambda}^{\frac{1}{2}}_{\ell_k}\boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{h}_{\mathrm{w}_{\ell_k}} \triangleq \boldsymbol{w}_{\ell_k} \in \mathbb{C}^{r_{\ell_k}\times 1}$, where $\boldsymbol{h}_{\mathrm{w}_{\ell_k}} \sim \mathcal{CN}(\boldsymbol{0},\boldsymbol{I})$, and $\boldsymbol{w}_{\ell_k} \sim \mathcal{CN}(\boldsymbol{0},\boldsymbol{\Lambda}_{\ell_k})$ is the projected effective channel vector, which has a smaller dimension than the original vector $\boldsymbol{h}_{\ell_k}$ in $\mathbb{C}^{M\times 1}$. Based on the sparse transform, we can exploit the sparsity intrinsic in channel vectors due to spatial correlation and use the randomness of $\boldsymbol{U}_{\ell_k}$ inherent in wireless multi-user communications due to arbitrary scattering geometry and mobility. In particular, one can interpret $\{\boldsymbol{U}_{\ell_k},\forall\ell,k\}$ as "random spreading sequences" in the classical uplink CDMA with asynchronous users since the sparse transform (multiplying $\boldsymbol{h}_{\ell_k}$ by $\boldsymbol{U}^{\mathsf{H}}_{\ell_k}$) and its counterpart (multiplying $\boldsymbol{w}_{\ell_k}$ by $\boldsymbol{U}_{\ell_k}$) are a reminiscence of *despreading* and *spreading*, respectively. Unlike CDMA, the spatial spreading is not controllable at the cost of bandwidth, but depends on channel propagation due to limited scattering without a cost.

We conduct the spatial despreading upon the received signal $\boldsymbol{y}_\ell$ such that the transformed vector $\boldsymbol{y}_{\ell_k}$ is given by

$$\boldsymbol{y}_{\ell_k} = \boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{y}_\ell = \boldsymbol{w}_{\ell_k}x_{\ell_k} + \sum_{(\ell',k')\neq(\ell,k)}\boldsymbol{w}_{\ell_k\ell'_{k'}}x_{\ell'_{k'}} + \boldsymbol{z}_{\ell_k} \qquad (2)$$

where $\boldsymbol{z}_{\ell_k} = \boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{z}_\ell$ and $\boldsymbol{w}_{\ell_k\ell'_{k'}} \triangleq \boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{U}_{\ell\ell'_{k'}}\boldsymbol{w}_{\ell\ell'_{k'}},\ \forall(\ell',k')$. The transformed received signal (2) will be used hereafter to explicitly show the role of spatial despreading.

D. Channel Estimation

1) (Intra-cell) Orthogonal Pilot Scheme: The typical pilot signal for uplink training is given by $\boldsymbol{s}_\ell = \sum_{\ell'}\boldsymbol{h}_{\ell\ell'_k} + \frac{1}{\sqrt{\rho_p P_{\mathrm{ul}}}}\boldsymbol{z}_\ell$, where ρ_p is the power boosting factor, with training cost of K channel uses. For the minimum mean square error (MMSE) channel estimation of user ℓ_k, we make use of the training signal $\boldsymbol{s}_{\ell_k}$ given by $\boldsymbol{s}_{\ell_k} = \boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{s}_\ell = \boldsymbol{w}_{\ell_k} + \sum_{\ell'\neq\ell}\boldsymbol{w}_{\ell_k\ell'_k} + \frac{1}{\sqrt{\rho_p P_{\mathrm{ul}}}}\boldsymbol{z}_{\ell_k}$. It is shown in [11] that $\boldsymbol{s}_{\ell_k}$ is a sufficient statistic for estimating $\boldsymbol{w}_{\ell_k}$ conditioned on $\boldsymbol{U}_\ell = [\boldsymbol{U}_{\ell_1},\ldots,\boldsymbol{U}_{\ell_K}]$. Given the observation $\boldsymbol{s}_{\ell_k}$ and the knowledge of $\boldsymbol{U}_\ell$, the estimate $\hat{\boldsymbol{w}}_{\ell_k}$ of the *effective channel* $\boldsymbol{w}_{\ell_k}$ is

$$\hat{\boldsymbol{w}}_{\ell_k} = \boldsymbol{\Lambda}_{\ell_k}\boldsymbol{\Xi}_{\ell_k}\boldsymbol{s}_{\ell_k} \qquad (3)$$

where $\boldsymbol{\Xi}_{\ell_k} \triangleq \left(\boldsymbol{\Lambda}_{\ell_k} + \sum_{\ell'\neq\ell}\boldsymbol{\Lambda}_{\ell_k\ell'_{k'}} + \rho_p^{-1}\boldsymbol{I}_{r_{\ell_k}}\right)^{-1}$ with $\boldsymbol{\Lambda}_{\ell_k\ell'_{k'}} \triangleq \boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{U}_{\ell'_{k'}}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\boldsymbol{U}^{\mathsf{H}}_{\ell'_{k'}}\boldsymbol{U}_{\ell_k}$ being the covariance matrix of $\boldsymbol{w}_{\ell_k\ell'_{k'}}$, and $\boldsymbol{\Xi}_{\ell_k}$ can be estimated using a sample mean by the ergodicity of fading channels. The distribution of $\hat{\boldsymbol{w}}_{\ell_k}$ is $\mathcal{CN}(\boldsymbol{0},\boldsymbol{\Phi}_{\ell_k})$, where $\boldsymbol{\Phi}_{\ell_k} = \boldsymbol{\Lambda}_{\ell_k}\boldsymbol{\Xi}_{\ell_k}\boldsymbol{\Lambda}_{\ell_k}$. The effective channel $\boldsymbol{w}_{\ell_k}$ can be written as $\boldsymbol{w}_{\ell_k} = \hat{\boldsymbol{w}}_{\ell_k} + \boldsymbol{n}_{\ell_k}$, where $\boldsymbol{n}_{\ell_k} \sim \mathcal{CN}(\boldsymbol{0},\boldsymbol{N}_{\ell_k})$ is conditionally independent of $\hat{\boldsymbol{w}}_{\ell_k}$ given $\boldsymbol{U}_\ell$.

2) (Intra-cell) Non-Orthogonal Pilot Scheme: It is known [8], [12] that the orthogonal pilot scheme may significantly limit the sum-rate performance unless K is smaller than $T_c/2$. To overcome this limiting factor, we also consider the non-orthogonal pilot scheme. The pilot (denoted by $\boldsymbol{s}'_{\ell_k}$) is non-orthogonal over intra cell as well as inter cell is given by $\boldsymbol{s}'_{\ell_k} = \boldsymbol{w}_{\ell_k} + \sum_{(\ell',k')\neq(\ell,k)}\boldsymbol{w}_{\ell_k\ell'_{k'}} + \frac{1}{\sqrt{\rho_p P_{\mathrm{ul}}}}\boldsymbol{z}_{\ell_k}$. Similar to the orthogonal pilot, we have the MMSE channel estimate of $\boldsymbol{w}_{\ell_k}$

$$\check{\boldsymbol{w}}_{\ell_k} = \boldsymbol{\Lambda}_{\ell_k}\boldsymbol{\Xi}'_{\ell_k}\boldsymbol{s}'_{\ell_k} \qquad (4)$$

where $\boldsymbol{\Xi}'_{\ell_k} \triangleq \left(\boldsymbol{\Lambda}_{\ell_k} + \sum_{(\ell',k')\neq(\ell,k)}\boldsymbol{\Lambda}_{\ell_k\ell'_{k'}} + \rho_p^{-1}\boldsymbol{I}_{r_{\ell_k}}\right)^{-1}$. The distribution of $\check{\boldsymbol{w}}_{\ell_k}$ is $\mathcal{CN}(\boldsymbol{0},\boldsymbol{\Phi}'_{\ell_k})$, where $\boldsymbol{\Phi}'_{\ell_k} = \boldsymbol{\Lambda}_{\ell_k}\boldsymbol{\Xi}'_{\ell_k}\boldsymbol{\Lambda}_{\ell_k}$.

III. Sum-Rate Bounds and the Capacity Scaling

In this paper, we consider three lower bounds on the achievable uplink rate based on an extension of the deterministic equivalents technique [6], [7] under the system model and assumptions in Sec. II-A. While the first lower bound is based on coherent detection, the others are rooted in non-coherent detection. We also compare them in the same uplink scenario.

A. Coherent Lower Bound

We first consider the standard lower bound on the achievable rate proposed by Hassibi and Hochwald, [13] based on the worst-case uncorrelated additive noise lemma. Using this well-known bounding technique, the ergodic achievable rate is lower-bounded by (5), shown on the top of page 4, where $\boldsymbol{v}_{\ell_k}$ is a linear combining vector for user ℓ_k, $\hat{\boldsymbol{w}}_\ell = [\hat{\boldsymbol{w}}_{\ell_1}, \ldots, \hat{\boldsymbol{w}}_{\ell_K}]$, and the outer expectation is taken over $\underline{\boldsymbol{U}} = [\boldsymbol{U}_1, \ldots, \boldsymbol{U}_L]$ as well to address the randomness of channel covariance. Throughout this paper, an ergodic achievable rate is averaged over realizations of channel covariance. The derivation of (5) basically follows from the MMSE decomposition of the useful signal channel and from the worst-case uncorrelated additive noise argument. Based on (5), we derive the following asymptotic capacity result.

Theorem 1. For large M and Assumptions 1 and 2 with the orthogonal pilot scheme and the spatial correlation model in Sec. II, the capacity of MIMO uplink is lower-bounded by

$$\mathcal{C}_M^{\text{ul}} \geq \sum_{\ell=1}^{L} \sum_{k=1}^{K_\ell} \Big(1 - \frac{\min\{K_\ell, \lfloor \frac{T_c}{2} \rfloor\}}{T_c}\Big) \log(\mathsf{SNR}_{\ell_k}^{\text{ul}} \text{tr} \boldsymbol{\Lambda}_{\ell_k}) + o(1). \tag{8}$$

Proof: (Sketch of Proof) Letting $\hat{\boldsymbol{w}}_{\ell_k \ell'_{k'}} \triangleq \boldsymbol{U}_{\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell \ell'_{k'}} \cdot \hat{\boldsymbol{w}}_{\ell \ell'_{k'}}, \forall (\ell', k')$, the single-cell MMSE combining vector is written as $\boldsymbol{v}_{\ell_k}^{\mathsf{mmse}} = \boldsymbol{\Upsilon}_{\ell_k} \hat{\boldsymbol{w}}_{\ell_k}$, where $\boldsymbol{\Upsilon}_{\ell_k} = \big(\hat{\boldsymbol{w}}_{\ell_k} \hat{\boldsymbol{w}}_{\ell_k}^{\mathsf{H}} + \sum_{k' \neq k} \hat{\boldsymbol{w}}_{\ell_k \ell_{k'}} \hat{\boldsymbol{w}}_{\ell_k \ell_{k'}}^{\mathsf{H}} + \boldsymbol{Z}_{\ell_k} + {P_{\text{ul}}}^{-1} \boldsymbol{I}_{r_{\ell_k}}\big)^{-1}$ with $\boldsymbol{Z}_{\ell_k} = \sum_{\ell' \neq \ell} \sum_k \boldsymbol{\Xi}_{\ell'_k} + \sum_{\ell', k'} \boldsymbol{N}_{\ell_k \ell'_{k'}}$. Using (5), the deterministic equivalent $\bar{\gamma}_{\text{ul},\ell_k}^{\mathsf{mmse}}$ of the SINR $\gamma_{\text{ul},\ell_k}^{\mathsf{mmse}}$ of the MMSE detector is given by the following result.

Lemma 1. Under Assumption 1 and the orthogonal pilot scheme, as $M \to \infty$, we almost surely have

$$\bar{\gamma}_{\text{ul},\ell_k}^{\mathsf{mmse}} = \frac{\delta_{\ell_k}^2}{\frac{1}{P_{\text{ul}} M} \mu_{\ell_k} + \sum_{\ell' \neq \ell} \alpha_{\ell_k}^2 |\nu_{\ell \ell'_k}|^2 + \frac{1}{M} \sum_{\ell' \neq \ell, k' \neq k} \mu_{\ell_k \ell'_{k'}}} \tag{9}$$

with $\delta_{\ell_k} = \frac{1}{M} \text{tr } \boldsymbol{\Phi}_{\ell_k} \boldsymbol{T}_{\ell_k}$, $\mu_{\ell_k} = \frac{1}{M} \text{tr } \boldsymbol{\Phi}_{\ell_k} \boldsymbol{T}'_{\ell_k}$, $\nu_{\ell \ell'_k} = \frac{1}{r_{\ell_k} M} \text{tr } \boldsymbol{\Lambda}_{\ell \ell'_k} \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{T}_{\ell_k}$, $\mu_{\ell_k \ell'_{k'}} = \frac{\text{tr} \boldsymbol{\Lambda}_{\ell \ell'_{k'}}}{M} \mu_{\ell_k}$, where the definitions of $\boldsymbol{T}_{\ell_k}$ and $\boldsymbol{T}'_{\ell_k}$ can be found in [11].

A proof of the second term caused by pilot contamination in the denominator of (9) is particularly given in Appendix B. For this, we extended the standard technique of deterministic equivalents in a few aspects in Appendix A: 1) non-uniform boundedness of the spectral norm of channel covariance matrices with respect to M and 2) the partial random Fourier correlation model for which the trace lemma in [6, Lem. 2.7] (see also [7, Thm. 3.4]) crucial for deterministic equivalents is not applicable. The latter is due to the fact that its resulting random sequences (column vectors and also their entries) are not i.i.d. any longer.

Let $\varphi_{\text{ul},\ell_k} \triangleq \frac{1}{P_{\text{ul}} \text{tr} \boldsymbol{\Lambda}_{\ell_k}} > 0$ by Assumption 1. As $M \to \infty$ with $\alpha_{\ell_k} \to 0$ under Assumption 2, we get $\boldsymbol{\Xi}_{\ell_k} \simeq \boldsymbol{\Lambda}_{\ell_k}^{-1}, \boldsymbol{T}_{\ell_k} \simeq \varphi_{\text{ul},\ell_k}^{-1} \boldsymbol{I}_{r_{\ell_k}}, \boldsymbol{T}'_{\ell_k} \simeq \varphi_{\text{ul},\ell_k}^{-2} \boldsymbol{I}_{r_{\ell_k}}$, where $\simeq$ refers to equivalence in the limit. Plugging these asymptotic approximations into (9), we have

$$\bar{\gamma}_{\text{ul},\ell_k}^{\mathsf{mmse}} \simeq \frac{\big(\frac{\varphi_{\text{ul},\ell_k}^{-1}}{M} \text{tr} \boldsymbol{\Lambda}_{\ell_k}\big)^2}{\frac{\varphi_{\text{ul},\ell_k}^{-1}}{M} \text{tr} \boldsymbol{\Lambda}_{\ell_k} + \sum\limits_{\ell' \neq \ell} \alpha_{\ell_k}^2 \big(\frac{\varphi_{\text{ul},\ell_k}^{-1}}{M} \text{tr} \boldsymbol{\Lambda}_{\ell \ell'_k}\big)^2 + \frac{1}{M} \sum\limits_{\ell', k'} \frac{\varphi_{\text{ul},\ell_k}^{-1}}{M^2} \text{tr} \boldsymbol{\Lambda}_{\ell \ell'_k} \text{tr} \boldsymbol{\Lambda}_{\ell_k}}$$

$$\simeq \frac{\varphi_{\text{ul},\ell_k}^{-1}}{M} \text{tr} \boldsymbol{\Lambda}_{\ell_k} = \mathsf{SNR}_{\ell_k}^{\text{ul}} \text{tr} \boldsymbol{\Lambda}_{\ell_k}. \tag{10}$$

Th rest of the proof is somewhat standard [8], [9], [12]. ∎

It follows from (8) that the coherent lower bound yields the multiplexing gain of $1 - \frac{\min\{K, \lfloor T_c/2 \rfloor\}}{T_c}$ and the maximum beamforming gain of M for the typical (orthogonal) pilot scheme with the homogeneous network, where $K_\ell = K$, $\mathsf{SNR}_{\ell_k}^{\text{ul}} = \mathsf{SNR}$, $\text{tr} \boldsymbol{\Lambda}_{\ell_k} = M$, and $r_{\ell_k} = r$, $\forall \ell, k$.

Remark 1. We can make important observations from (10) on the role of spatial despreading. The coherent pilot contamination term approximated as a non-zero finite deterministic value $\nu_{\ell \ell'_k}$ multiplied by the ratio α_{ℓ_k} vanishes under the sublinear sparsity in Assumption 2. Likewise, the interference term completely disappears through spatial despreading [11].

From Theorem 1 an intriguing question arises: *Is the sum-rate scaling law with the multiplexing gain of* (8) *optimal in massive MIMO under Assumption 2?* Such multiplexing gain is indeed limited by $\frac{T_c L}{4}$ when $K \geq \frac{T_c}{2}$ [12]. To answer this question, we need to consider the non-orthogonal pilot $\boldsymbol{s}'_{\ell_k}$ in Sec. II, whose training cost is only a single channel use across the L-cell network. In this case, unfortunately, the coherent lower bound in (5) does not lend itself to the trace lemma [6, Lem. 2.7]. We thus turn our attention to non-coherent bounding techniques.

B. Non-Coherent Lower Bound

Marzetta [14] proposed a non-coherent bound based on separating the useful signal coefficient into a deterministic part and a random fluctuation part, not requiring the coherent detection. Although this bound is very natural to use in downlink since in general the receivers are assumed to only know the fading distribution, it is applicable in uplink as well, where the uplink pilot is used to construct the coherent combining vector $\boldsymbol{v}_{\ell_k}$, but not for the coherent detection. By the technique, we have the ergodic achievable rate in (6), shown on the top of page 4. We can then obtain the following main result.

Theorem 2. For large M and Assumptions 1 and 2 with the non-orthogonal pilot scheme and the spatial correlation model in Sec. II, the capacity of MIMO uplink behaves as

$$\mathcal{C}_M^{\text{ul}} = \big(1 - T_c^{-1}\big) \sum_{\ell=1}^{L} \sum_{k=1}^{K_\ell} \log(\mathsf{SNR}_{\ell_k}^{\text{ul}} \text{tr} \boldsymbol{\Lambda}_{\ell_k}) + o(1). \tag{12}$$

$$\mathcal{R}^{(1)}_{\text{ul},\ell_k} = \mathbb{E}\left[\log\left(1+\frac{|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\hat{\boldsymbol{w}}_{\ell_k}|^2}{\mathbb{E}\left[|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{n}_{\ell_k}|^2+\sum_{(\ell',k')\neq(\ell,k)}|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|^2+\frac{1}{P_{\text{ul}}}|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{z}_{\ell_k}|^2 \,\middle|\, \hat{\boldsymbol{w}}_{\ell},\boldsymbol{U}_{\ell}\right]}\right)\right]. \tag{5}$$

$$\mathcal{R}^{(2)}_{\text{ul},\ell_k} = \mathbb{E}\left[\log\left(1+\frac{\left|\mathbb{E}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k}|\underline{\boldsymbol{U}}]\right|^2}{\frac{1}{P_{\text{ul}}}\|\boldsymbol{v}_{\ell_k}\|^2+\text{var}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k}|\underline{\boldsymbol{U}}]+\sum_{(\ell',k')\neq(\ell,k)}\mathbb{E}\left[|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\right]}\right)\right]. \tag{6}$$

$$\mathcal{R}^{(3)}_{\text{ul},\ell_k} = \mathbb{E}\left[\log\left(1+\frac{\left|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k}\right|^2}{\frac{1}{P_{\text{ul}}}+\sum_{(\ell',k')\neq(\ell,k)}|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|^2}\right)\right]-\frac{1}{T_c}\sum_{(\ell',k')\neq(\ell,k)}\log\left(1+P_{\text{ul}}\text{var}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|\underline{\boldsymbol{U}}]\right). \tag{7}$$

Proof: (Sketch of Proof) It suffices to consider the simple matched filter receiver for the desired capacity scaling law. We set $\boldsymbol{v}_{\ell_k} = \check{\boldsymbol{w}}_{\ell_k}$, where $\check{\boldsymbol{w}}_{\ell_k}$ is given by (4), and denote the resulting SINR by $\gamma^{\text{MF}}_{\text{ul},\ell_k}$. Similar to Lemma 1, we have $\gamma^{\text{MF}}_{\text{ul},\ell_k} \xrightarrow[M\to\infty]{a.s.} \bar{\gamma}^{\text{MF}}_{\text{ul},\ell_k} = \frac{(\frac{1}{M}\text{tr}\boldsymbol{\Xi}'_{\ell_k})^2}{\frac{1}{P_{\text{ul}}M}\frac{1}{M}\text{tr}\boldsymbol{\Xi}'_{\ell_k}+\sum_{(\ell',k')\neq(\ell,k)}\alpha^2_{\ell_k}\psi^2_{\ell'_{k'}}}$, where $\psi_{\ell'_{k'}} = \frac{1}{r_{\ell_k}M}\text{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\text{tr}\,\boldsymbol{\Xi}'_{\ell_k}\boldsymbol{\Lambda}_{\ell_k}$. Similar to (10), $\boldsymbol{\Xi}'_{\ell_k} \simeq \boldsymbol{\Lambda}^{-1}_{\ell_k}$ and we get $\bar{\gamma}^{\text{MF}}_{\text{ul},\ell_k} \simeq \text{SNR}^{\text{ul}}_{\ell_k}\text{tr}\boldsymbol{\Lambda}_{\ell_k}$, where we used a direct combination of Corollaries 1 and 2 and then Lemma 2.

For the converse proof, we can use a simple cut-set upper bound argument on the sum rate of massive MIMO uplink, where a cut divides the BSs from the users. ■

The scaling law in (1) for the homogeneous network directly follows from (12). It should be pointed out that any effect of a finite number of overlapped Fourier basis functions between any two users that incurs pilot contamination vanishes under the sublinear sparsity assumption, similarly to (16) and (10). Therefore, even the infinite sum of the vanishing contamination and residual interference terms due to finite overlap between angular supports selected by different users sharing the non-orthogonal pilot still does not affect the asymptotic scaling in Theorem 2, although K may grow faster than M. In fading channels with strong spatial correlation, where the dimension r_{ℓ_k} of the effective channel $\boldsymbol{w}_{\ell_k}$ could be much small than M (i.e., lack of channel hardening), $\mathcal{R}^{(2)}_{\text{ul},\ell_k}$ suffers from the self-interference due to a non-negligible variance term $\text{var}[\boldsymbol{g}^{\mathsf{H}}_{\ell_k}\boldsymbol{w}_{\ell_k}|\underline{\boldsymbol{U}}]$ unless r_{ℓ_k} becomes sufficiently large. As a consequence, $\mathcal{R}^{(2)}_{\text{ul},\ell_k}$ may substantially underestimate an achievable rate of massive MIMO.

C. Alternative Non-Coherent Lower Bound

We consider another non-coherent bounding technique very recently derived by Caire [15]. The third lower bound is given by (7), shown on the top of page 4, where the expectation is again taken over $\underline{\boldsymbol{U}}$ as well as $\{\boldsymbol{w}_{\ell_k}\}$. The second term in (7) consists of the prelog factor $\frac{\sum_\ell K_\ell - 1}{T_c}$ and the variances of coherent interference $\text{var}[\boldsymbol{v}^{\mathsf{H}}_{\ell_k}\boldsymbol{w}_{\ell_k\ell'_{k'}}|\underline{\boldsymbol{U}}]$ multiplied by the transmit power P_{ul} inside the logarithm. Basically this bound comes very close to the so-called max-min bound[1] [15] when coherence block is sufficiently large or coherent interference is limited. The latter is the case with our main scenario under the sublinear sparsity assumption. One can prove that $\mathcal{R}^{(3)}_{\text{ul},\ell_k}$ achieves the same capacity scaling as Theorem 2 and the same scaling law is achievable in downlink as well.

[1] The max-min upper bound on achievable rate is given by the first term in (7), where the max is over the coding/decoding strategy of user ℓ_k and the min is over all input distributions of the other users.

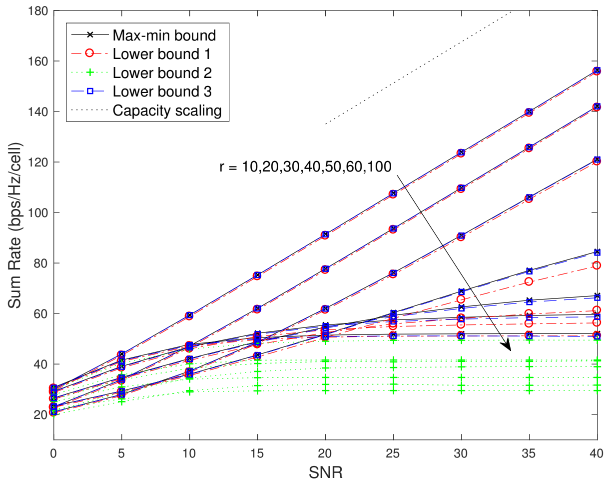

Fig. 1. The impact of spatial despreading on sum-rate scaling with respect to the sparsity of angular supports (r), where $M = 200$, $L = 4$, $K = 10$, $T_c = 500$. 'Lower bound 1', 'Lower bound 2', and 'Lower bound 3'are given by (5), (6), and (7), respectively. The asymptotic capacity scaling is (8) with $o(1) = 0$.

IV. Numerical Results

In this section, the (single-cell) MMSE receiver in the homogeneous uplink scenario is only considered with inter-cell interference factor of $\sqrt{0.2}$. We assume that the number of non-zero eigenvalues of the covariance matrices of channels from other cells $\ell' \neq \ell$ is equal to $r/2$ to take into account the fact that spatial correlation also depends on the distance between sender and receiver. We used pilot power boosting of $\rho_p = 2$. In what follows, we evaluate the three lower bounds for finite M and SNR to verify the capacity scaling result.

Fig. 1 shows how strong spatial correlation we need to achieve linear sum-rate scaling with respect to SNR (dB) in terms of the number of non-zero eigenvalues r of channel covariance matrices. It can be seen that almost the same multiplexing gain as (8) (or (1)) is achievable up to $r = 30$. At low SNR, both interference suppression and pilot decontamination effects of spatial despreading are diluted by noise and the sum-rate performance depends more on channel hardening of $\boldsymbol{w}_{\ell_k}$ than spatial despreading of $\boldsymbol{U}_{\ell_k}$. Hence, the larger r turns out rather beneficial at low SNR.

As a concluding remark, channel hardening has been traditionally considered as an essential source of massive MIMO gain. This is not the case with sparse angular supports (small r). Rather, one can observe that the effect of spatial despreading is indeed central to achieve the ultimate scaling law in (1).

Appendix A

The well-known trace lemma [6], [7] is generalized for the following two cases.

Corollary 1. Let $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$, with $\boldsymbol{y}_n \in \mathbb{C}^n$, be random vectors with uncorrelated entries of zero mean, variance $1/n$, $|y_{n,i}|^2 = 1/n$, where $y_{n,i}$ is the ith entry of $\boldsymbol{y}_n$ and $i = 1, \ldots, n$, and eighth order moment of order $\mathcal{O}(\frac{1}{n^4})$. $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$, with $\boldsymbol{A}_n \in \mathbb{C}^{n \times n}$, denotes a series of matrices with uniformly bounded spectral norm with respect to n, independent of $\boldsymbol{y}_n$. We have

$$\boldsymbol{y}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \, \mathrm{tr} \boldsymbol{A}_n \xrightarrow[n \to \infty]{a.s.} 0. \tag{14}$$

Proof: It follows from the Markov inequality that for every $\epsilon > 0$, $\Pr\left(|\boldsymbol{y}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \, \mathrm{tr} \boldsymbol{A}_n| > \epsilon\right) \leq \frac{\mathbb{E}[|\boldsymbol{y}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \, \mathrm{tr} \boldsymbol{A}_n|^4]}{\epsilon^4}$. In order to show that $\mathbb{E}\left[|\boldsymbol{y}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \, \mathrm{tr} \boldsymbol{A}_n|^4\right] \leq \frac{C_1}{n^2}$, where $C_1 > 0$ is a constant independent of n and $\boldsymbol{A}_n$, following the footsteps in [6, Lem. 3.1] and [7, Thm. 3.4], we begin with $\mathbb{E}[|\boldsymbol{y}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \, \mathrm{tr} \boldsymbol{A}_n|^4] \leq 8(\mathbb{E}[\sum_i A_{n,ii}(|y_{n,i}|^2 - n^{-1})\Big]^4 + \mathbb{E}[\sum_{i \neq j} A_{n,ij} y_{n,i}^* y_{n,j}]^4) = 8(\mathbb{E}[\sum_{i \neq j} A_{n,ij} y_{n,i}^* y_{n,j}]^4) = \mathcal{O}(\frac{1}{n^2})$, where we used $|y_{n,i}|^2 = 1/n, \forall i$. The last equality follows from the facts: 1) each term in the sum is finite since $y_{n,i}$ has eighth order moment of order $\mathcal{O}(\frac{1}{n^4})$ and $\boldsymbol{A}_n$ has uniformly bounded norm, and 2) such terms amount to $\mathcal{O}(n^2)$ since the nonzero contribution to the sum arises if $y_{n,i}$ appears an even number of times due to $\mathbb{E}[y_i^* y_j] = 0, \forall i \neq j$. We can then get $\sum_{n=1}^{\infty} \Pr\left(|\boldsymbol{y}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \, \mathrm{tr} \boldsymbol{A}_n| > \epsilon\right) < \infty$. The almost sure convergence in (14) immediately follows from the first Borel-Cantelli lemma. ■

Some structured random matrices like partial Fourier and Hadamard matrices satisfy the above requirement of $|y_{n,i}|^2$.

Corollary 2. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$, with $\boldsymbol{x}_n \in \mathbb{C}^n$, be random vectors with i.i.d. entries of zero mean, variance $1/n$, and eighth order moment of order $\mathcal{O}(\frac{1}{n^4})$. Also let $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$ be a series of matrices independent of $\boldsymbol{x}_n$. If there exists $m_n \geq n$ such that $\limsup_{n \to \infty} \frac{n}{m_n} \, \|\boldsymbol{A}_n\|_\infty < \infty$, then

$$\frac{n}{m_n} \boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{x}_n - \frac{1}{m_n} \, \mathrm{tr} \boldsymbol{A}_n \xrightarrow[n \to \infty]{a.s.} 0. \tag{15}$$

Proof: Even though the spectral norm of $\boldsymbol{A}_n$ is not uniformly bounded across n, one may find a sequence $m_n \geq n$ to satisfy the above condition on m_n, which implies that the spectral norm of $\frac{n}{m_n} \boldsymbol{A}_n$ is uniformly bounded with respect to n. In this case, m_n and n increase at the ratio $\frac{n}{m_n}$, where m_n may grow faster than n such that $\frac{n}{m_n} \to 0$. By noticing $\frac{n}{m_n} \boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{x}_n = \boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{B}_n \boldsymbol{x}_n$, where $\boldsymbol{B}_n = \frac{n}{m_n} \boldsymbol{A}_n$, and by the trace lemma, for any $\epsilon > 0$, there exists n_0 such that for all $n \geq n_0$, $\left|\boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{B}_n \boldsymbol{x}_n - \frac{1}{n} \, \mathrm{tr} \boldsymbol{B}_n\right| < \epsilon$. ■

The following lemma is a matrix analog of the trace lemma for random matrices $\boldsymbol{X}$. The proofs are given in [11].

Lemma 2. Let $\boldsymbol{U} \in \mathbb{C}^{p \times m}$ and $\boldsymbol{V} \in \mathbb{C}^{p \times n}$ be independent random orthonormal matrices whose columns have the same entries as $\boldsymbol{y}_n$ in Corollary 1, where $p \geq \max(m, n)$. Also let $\boldsymbol{D} \in \mathbb{C}^{n \times n}$ be an arbitrary matrix with $\limsup_{p \to \infty} \|\boldsymbol{D}\|_\infty < \infty$, independent of $\boldsymbol{U}$ and $\boldsymbol{V}$. Then we have $\boldsymbol{U}^{\mathsf{H}} \boldsymbol{V} \boldsymbol{D} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{U} - \frac{\mathrm{tr} \boldsymbol{D}}{p} \boldsymbol{I}_m \xrightarrow[p \to \infty]{a.s.} 0$.

Appendix B

The coherent pilot contamination term in (9) can be approximated in the limit of M as follows.

$$\begin{aligned}
&\hat{\boldsymbol{w}}_{\ell_k}^{\mathsf{H}} \boldsymbol{\Upsilon}_{-\ell_k} \boldsymbol{w}_{\ell_k \ell'_k} \\
&\quad = \Big(\boldsymbol{w}_{\ell_k} + \sum_{j \neq \ell} \boldsymbol{w}_{\ell_k j_k} + \sqrt{\rho_p}^{-1} \boldsymbol{z}_{\ell_k}\Big)^{\mathsf{H}} \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{\Upsilon}_{-\ell_k} \boldsymbol{w}_{\ell_k \ell'_k} \\
&\quad \overset{(a)}{\simeq} \boldsymbol{w}_{\ell_k \ell'_k}^{\mathsf{H}} \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{\Upsilon}_{-\ell_k} \boldsymbol{w}_{\ell_k \ell'_k} \\
&\quad = \boldsymbol{w}_{\ell \ell'_{k'}}^{\mathsf{H}} \underbrace{\boldsymbol{U}_{\ell \ell'_{k'}}^{\mathsf{H}} \boldsymbol{U}_{\ell_k} \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{\Upsilon}_{-\ell_k} \boldsymbol{U}_{\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell \ell'_{k'}}}_{\overset{(b)}{\simeq} \frac{1}{M} \mathrm{tr}\left(\boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{\Upsilon}_{-\ell_k}\right) \boldsymbol{I}_{r_{\ell'_{k'}}}} \boldsymbol{w}_{\ell \ell'_{k'}} \\
&\simeq \frac{1}{M^2} \mathrm{tr} \boldsymbol{\Lambda}_{\ell \ell'_k} \mathrm{tr} \, \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \tilde{\boldsymbol{\Upsilon}}_{\ell_k} \overset{(c)}{\simeq} \frac{r_{\ell_k}}{M} \nu_{\ell \ell'_k} = \alpha_{\ell_k} \nu_{\ell \ell'_k}
\end{aligned} \tag{16}$$

where (a) follows from [11, Lem. 4] for $j = \ell'$. In (b), we have used the fact that the asymptotic covariance of $\boldsymbol{w}_{\ell_k \ell'_{k'}}$ is almost surely given by $\frac{1}{M} \mathrm{tr} \boldsymbol{\Lambda}_{\ell \ell'_k}$ since $\boldsymbol{w}_{\ell_k \ell'_{k'}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Lambda}_{\ell_k \ell'_{k'}})$, and by Lemma 2 we get $\boldsymbol{\Lambda}_{\ell_k \ell'_{k'}} - \frac{\mathrm{tr} \boldsymbol{\Lambda}_{\ell \ell'_k}}{M} \boldsymbol{I}_{r_{\ell'_{k'}}} \xrightarrow{a.s.} 0$. In (c) we used [11, Thm. 4].

References

[1] 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.

[2] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, pp. 1137–1141, Jul. 2012.

[3] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, p. 264–273, Feb. 2013.

[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] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[6] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *The Annals of Probability*, vol. 26, no. 1, pp. 316–345, Jan. 1998.

[7] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. New York, NY, USA: Cambridge University Press, 2011.

[8] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Inform. Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[9] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

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

[11] J. Nam, G. Caire, M. Debbah, and H. V. Poor, "Capacity scaling of massive MIMO with random sparse angular supports: breaking the spectral efficiency barrier," 2018. [Online]. Available: http://arxiv.org

[12] J. Nam, G. Caire, and J. Ha, "On the role of transmit correlation diversity in multiuser MIMO systems," *IEEE Trans. on Inform. Theory*, vol. 63, no. 1, pp. 336–354, Jan. 2017.

[13] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. on Inform. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[14] T. L. Marzetta, "How much training is required for multiuser mimo?" *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers (ACSSC)*, pp. 359–363, 2006.

[15] G. Caire, "On the ergodic rate lower bounds with applications to massive MIMO," 2017. [Online]. Available: https://arxiv.org/abs/1705.03577