# Impact of Strong Spatial Correlation on the Capacity Scaling of Massive MIMO

Junyoung Nam, Giuseppe Caire, Merouane Debbah, H Vincent Poor

# Impact of Strong 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, the capacity scaling of multicell massive MIMO systems is investigated in the presence of spatially correlated fading. In particular, we focus on the strong spatial correlation regimes where the covariance matrix of each user channel vector has a rank that scales sublinearly with the number of base station antennas, as the latter grows to infinity. We also consider the case where the covariance eigenvectors corresponding to the non-zero eigenvalues span randomly selected subspaces. For this channel model, referred to as the "random sparse angular support" model, we characterize the asymptotic capacity scaling law in the limit of large number of antennas. In order to achieve the capacity results, spatial (de)spreading based on the second-order channel statistics plays a pivotal role in terms of pilot decontamination and interference suppression. A remarkable result is that even when the number of users scales linearly with base station antennas, unlimited capacity is achievable under the sparse angular support model as long as the effective signal-to-noise ratio is away from zero.

## I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) [1], [2] 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 despite hardware impairments. To accurately predict the performance of multicell massive MIMO, it will be intriguing to investigate the sum capacity in the limit of a large number of antennas with spatial correlation taken into account. This work considers particular regimes where the covariance matrix of each user channel has a rank that scales sublinearly with the number of BS antennas, $M$. Such strong correlation may be justified by channel measurements even in below 6 GHz bands (e.g., [3] and references therein), where the number of dominant angular components is fairly smaller than $M$, although the covariance matrix is mathematically of full rank.

Since pilot contamination has been a fundamental bottleneck in massive MIMO, several techniques have been proposed to tackle the problem. For example, [4] proposed multicell cooperative precoding/combining over the entire network. Following [5], [6], many pilot decontamination techniques have 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, [7] 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.

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 arises through channel propagation in the typical limited scattering geometry [3]. 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 [8], [9] that the ergodic sum capacity behaves as

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

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 [10], 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 SNR, and how much multiplexing gain one can achieve. 2) Past work has generally assumed $\lim_M \frac{K}{M} = 0$, e.g., if $\lim\inf_M \frac{K}{M} > 0$, then the asymptotic linear independence in [7] does not hold so that pilot contamination cannot be completely eliminated. In contrast, $\lim\sup_M \frac{K}{M} < \infty$ is admissible for every user to achieve unlimited spectral efficiency at finite SNR and $T_c$ under the random sparse angular support model, as long as the per-user effective SNR does not vanish. 3) Neither explicit pilot decontamination techniques nor multicell precoding is not required to obtain (1). Rather, the effect of spatial (de)spreading based on channel covariances turns out crucial.

## II. PRELIMINARIES

### A. System Model

We consider an $L$-cell network with $M$ antennas at each BS serving $K_\ell$ users. Indexing the $k$th user in BS $\ell$ by $\ell_k$, $\boldsymbol{h}_{\ell\ell'_k}$ represents the channel from user $\ell'_k$ to BS $\ell$ 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 $\ell$. Likewise the subscript $\ell\ell_k$ will be hereafter replaced with $\ell_k$ for all other notations. Using the Karhunen-Loève transform, the channel vector $\boldsymbol{h}_{\ell\ell'_k}$ can be expressed as $\boldsymbol{h}_{\ell\ell'_k} = \boldsymbol{U}_{\ell\ell'_k} \boldsymbol{\Lambda}_{\ell\ell'_k} \mathsf{h}_{\ell\ell'_k}$, 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 the channel covariance matrix $\boldsymbol{R}_{\ell\ell'_k}$, $\boldsymbol{U}_{\ell\ell'_k} \in \mathbb{C}^{M \times r_{\ell\ell'_k}}$ is the eigenvector matrix, and $\mathsf{h}_{\ell\ell'_k} \in \mathbb{C}^{r_{\ell\ell'_k}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$. The received signal vector at BS $\ell$ 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 $x_{\ell_k}$ is the input signal of user $\ell_k$ chosen from a Gaussian codebook and satisfies the power constraint such that $\mathbb{E}[x_{\ell_k}^\mathsf{H} x_{\ell_k}] \leq P_{\mathrm{ul}}$, and $\boldsymbol{z}_\ell \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ is the Gaussian noise, where $P_{\mathrm{ul}}$ is the uplink transmit power. Since the noise power per antenna is normalized, $P_{\mathrm{ul}}$ can be regarded as the transmit SNR per user, and SNR denotes the normalized transmit SNR per cell. Although we only describe the uplink MIMO, main results in this work are extended to the downlink [11].

### B. Random Partial Fourier Correlation Model

For spatial correlation, we introduce a random partial (sub-sampled) 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 $F_{jk} = \frac{1}{\sqrt{M}} e^{2\pi jjk/M}$, $j, k = 0, \dots, M-1$. Suppose that $\boldsymbol{U}_{\ell_k}$ is composed of $r_{\ell_k}$ 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 $r_{\ell_k}$ 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., [12], [13]), the above model allows different angular supports as in [5], [6].

It is important to notice that given the above random realizations of the users channel covariances, the resulting ergodic achievable rates are conditional to such realizations. Therefore one might then be interested in their distribution, in particular, the outage probability defined by the distribution function of such ergodic rates conditioned on the covariances. In order to seek after the ergodic capacity scaling in this work, we assume that the channel energy captured by the large-scale fading factor $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$ does not change over user mobility and different scattering geometries for large $M$ such that $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k} = c_{\ell\ell'_k}$, $\forall(\ell, \ell', k)$, where $c_{\ell\ell'_k}$ is a positive real constant. Under this deterministic $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$, we will show later in Sec. III-D that the conditional ergodic rates converge to a deterministic limit for large $M$.

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

**Assumption 1.** For all $(\ell, \ell', k)$

$$\limsup_{M \to \infty} \frac{r_{\ell\ell'_k}}{M} \|\boldsymbol{\Lambda}_{\ell\ell'_k}\|_2 < \infty, \; \limsup_{M \to \infty} \frac{\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}}{M} < \infty.$$

**Assumption 2** (*Strong Spatial Correlation Regime*)**.** 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 $\forall(\ell, \ell', k)$

$$\frac{r_{\ell\ell'_k}}{M} \triangleq \alpha_{\ell\ell'_k} \xrightarrow[M \to \infty]{} 0, \; \liminf_{M \to \infty} \frac{\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}}{M} > 0.$$

In conjunction with Assumption 2, the first condition of Assumption 1 implies that $\|\boldsymbol{\Lambda}_{\ell\ell'_k}\|_2$ is *not necessarily* uniformly bounded with respect to $M$. The uniform boundedness is a necessary condition for the method of deterministic equivalents [8], [9]. Hence, under these conditions the deterministic equivalents cannot be directly applied any longer. An insufficiency of Assumption 2 is that the large-scale fading factor $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$ grows at the same speed as $M \to \infty$, even though the rank $r_{\ell\ell'_k}$ of $\boldsymbol{\Lambda}_{\ell\ell'_k}$ grows only sublinearly with $M$. Therefore, we also consider another regime, where $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$ grows no faster than $r_{\ell\ell'_k}$ as $M \to \infty$ and spatial correlation is rather stronger.

**Assumption 3** (*Very Strong Spatial Correlation Regime*)**.** The number $r_{\ell\ell'_k}$ of non-zero eigenvalues of $\boldsymbol{R}_{\ell\ell'_k}$ grows without bound but much slower than $M$ such that $\forall(\ell, \ell', k)$

$$\frac{r_{\ell\ell'_k}^4}{M^3} \xrightarrow[M \to \infty]{} 0, \; \limsup_{M \to \infty} \frac{r_{\ell\ell'_k}^2}{M} < \infty, \; \limsup_{M \to \infty} \frac{\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}}{r_{\ell\ell'_k}} < \infty.$$

### C. Spatial Despreading and Spreading

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 $\boldsymbol{U}_{\ell_k}^\mathsf{H} \boldsymbol{h}_{\ell_k} = \boldsymbol{\Lambda}_{\ell_k}^{\frac{1}{2}} \mathsf{h}_{\ell_k} \triangleq \boldsymbol{w}_{\ell_k} \in \mathbb{C}^{r_{\ell_k}}$, where $\boldsymbol{w}_{\ell_k} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Lambda}_{\ell_k})$ is the projected effective channel vector, whose dimension is fairly smaller than the original vector $\boldsymbol{h}_{\ell_k}$ in $\mathbb{C}^M$. Based on the sparse transform, we can exploit the sparsity intrinsic in channel vectors due to limited scattering and utilize 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 $\boldsymbol{U}_{\ell_k}^\mathsf{H}$) 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 (de)spreading* is not controllable, but depends on limited scattering channel propagation without a cost of bandwidth.

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

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

where $\boldsymbol{z}_{\ell_k} = \boldsymbol{U}_{\ell_k}^\mathsf{H} \boldsymbol{z}_\ell$ and $\boldsymbol{w}_{\ell_k \ell'_{k'}} \triangleq \boldsymbol{U}_{\ell_k}^\mathsf{H} \boldsymbol{U}_{\ell\ell'_{k'}} \boldsymbol{w}_{\ell\ell'_{k'}}$, $\forall(\ell', k')$. Spatial despreading does not incur any loss of optimality from the single-user perspective, but it is suboptimal from the perspective of MU-MIMO because multiuser combining is performed only based on $\{\boldsymbol{w}_{\ell_k}, \boldsymbol{w}_{\ell_k \ell'_{k'}}\}$. In fact, it is only asymptotically optimal under Assumption 2 or 3, which will be shown by Theorem 2. However, the purpose of the transformed

received signal in (2) is to explicitly show the role of spatial despreading. In downlink, spatial spreading is performed.

### D. Low-Dimensional Channel Estimation

*1) (Intra-cell) Orthogonal Pilot Scheme:* For this typical scheme, user $k$ in each cell shares the same pilot sequence so that the receive pilot signal for user $\ell_k$ is given by $\bar{s}_{\ell_k} = \sum_{\ell'} h_{\ell\ell'_k} + \frac{1}{\sqrt{\rho_p}} z_\ell$, where $\rho_p = \varrho_p P_{\mathrm{ul}}$ with $\varrho_p$ the power boosting factor. For the minimum mean square error (MMSE) channel estimation of user $\ell_k$, we make use of $s_{\ell_k} = U_{\ell_k}^{\mathsf{H}} \bar{s}_{\ell_k} = w_{\ell_k} + \sum_{\ell' \neq \ell} w_{\ell_k \ell'_k} + \frac{1}{\sqrt{\rho_p}} z_{\ell_k}$. Given the observation $s_{\ell_k}$ and the prior knowledge of $U_\ell$, the estimate $\hat{w}_{\ell_k}$ of $w_{\ell_k} \in \mathbb{C}^{r_{\ell_k}}$ is

$$\hat{w}_{\ell_k} = \Lambda_{\ell_k} \Xi_{\ell_k} s_{\ell_k} \tag{3}$$

where $\Xi_{\ell_k} \triangleq \left(\Lambda_{\ell_k} + \sum_{\ell' \neq \ell} \Lambda_{\ell_k \ell'_k} + \rho_p^{-1} I_{r_{\ell_k}}\right)^{-1}$ with $\Lambda_{\ell_k \ell'_{k'}} \triangleq U_{\ell_k}^{\mathsf{H}} U_{\ell'_k} \Lambda_{\ell\ell'_k} U_{\ell'_{k'}}^{\mathsf{H}} U_{\ell_k}$ being the covariance matrix of $w_{\ell_k \ell'_{k'}}$, and $\Xi_{\ell_k}$ can be estimated using a sample mean by the ergodicity of fading channels. The distribution of $\hat{w}_{\ell_k}$ is $\mathcal{CN}(0, \Phi_{\ell_k})$, where $\Phi_{\ell_k} = \Lambda_{\ell_k} \Xi_{\ell_k} \Lambda_{\ell_k}$. The effective channel $w_{\ell_k}$ can be written as $w_{\ell_k} = \hat{w}_{\ell_k} + n_{\ell_k}$, where $n_{\ell_k} \sim \mathcal{CN}(0, N_{\ell_k})$ is conditionally independent of $\hat{w}_{\ell_k}$ given $U_\ell$.

*2) (Intra-cell) Non-Orthogonal Pilot Scheme:* It is known [10], [14] that the orthogonal pilot scheme may significantly limit the sum-rate performance unless $\max_\ell K_\ell$ is smaller than $T_c/2$. To overcome this limiting factor, we also consider the non-orthogonal pilot scheme. This pilot (denoted by $s'_{\ell_k}$) non-orthogonal over intra cell as well as inter cell is given by $s'_{\ell_k} = w_{\ell_k} + \sum_{(\ell',k') \neq (\ell,k)} w_{\ell_k \ell'_{k'}} + \frac{1}{\sqrt{\rho_p}} z_{\ell_k}$. Similar to the orthogonal pilot, we have the MMSE channel estimate of $w_{\ell_k}$

$$\check{w}_{\ell_k} = \Lambda_{\ell_k} \Xi'_{\ell_k} s'_{\ell_k} \tag{4}$$

where $\Xi'_{\ell_k} \triangleq \left(\Lambda_{\ell_k} + \sum_{(\ell',k') \neq (\ell,k)} \Lambda_{\ell_k \ell'_{k'}} + \rho_p^{-1} I_{r_{\ell_k}}\right)^{-1}$. The distribution of $\check{w}_{\ell_k}$ is $\mathcal{CN}(0, \Phi'_{\ell_k})$, where $\Phi'_{\ell_k} = \Lambda_{\ell_k} \Xi'_{\ell_k} \Lambda_{\ell_k}$. For a more general non-orthogonal pilot design, one may use Welch bound equality frames [15].

*3) Low-dimensional Combining/Precoding:* One practical issue of the conventional full-dimensional MMSE combining/precoding is prohibitive computational complexity in the multicell massive MIMO network, where the $M \times M$ covariance matrix of intercell interference should be taken into account to mitigate the intercell interference (e.g., see [12, Eqn. (11)]). Based on the above low-dimensional channel estimation, the following $r_{\ell_k}$-dimensional MMSE processing may have much lower complexity in strong spatial correlation regimes, where $r_{\ell_k} \ll M$. Letting $\hat{w}_{\ell_k \ell'_{k'}} \triangleq U_{\ell_k}^{\mathsf{H}} U_{\ell\ell'_{k'}} \cdot \hat{w}_{\ell\ell'_{k'}}, \forall (\ell', k')$, the single-cell MMSE combining vector is written as $v_{\ell_k}^{\mathrm{mmse}} = \Upsilon_{\ell_k} \hat{w}_{\ell_k}$, where $\Upsilon_{\ell_k} = \left(\hat{w}_{\ell_k} \hat{w}_{\ell_k}^{\mathsf{H}} + \sum_{k' \neq k} \hat{w}_{\ell_k \ell'_{k'}} \hat{w}_{\ell_k \ell'_{k'}}^{\mathsf{H}} + Z_{\ell_k} + P_{\mathrm{ul}}^{-1} I_{r_{\ell_k}}\right)^{-1}$ with $Z_{\ell_k} = \sum_{\ell' \neq \ell} \sum_k \Xi'_{\ell_k} + \sum_{\ell',k'} N_{\ell_k \ell'_{k'}}$.

### 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 [8], [9] under the system model and assumptions in Sec. II-A. While the first lower bound relies on coherent detection, the others are based on non-coherent detection. We also compare them in the same uplink scenario.

### A. Coherent Lower Bound

Using the well-known bounding technique [16], the ergodic achievable rate can be lower-bounded by (5), shown on the top of page 4, where $v_{\ell_k}$ is a linear combining vector for user $\ell_k$, $\hat{w}_\ell = [\hat{w}_{\ell_1}, \ldots, \hat{w}_{\ell_K}]$. 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 sum capacity of MIMO uplink is lower-bounded by

$$\mathcal{C}_M^{\mathrm{ul}} \geq \sum_{\ell=1}^{L} \sum_{k=1}^{\kappa} \left(1 - \frac{\kappa}{T_c}\right) \log(P_{\mathrm{ul}} \mathrm{tr} \Lambda_{\ell_k}) + o(1). \tag{8}$$

where $\kappa = \min\{K_\ell, \lfloor \frac{T_c}{2} \rfloor\}$.

*Proof:* (Sketch of Proof) Using (5), the deterministic equivalent $\bar{\gamma}_{\mathrm{ul},\ell_k}^{\mathrm{mmse}}$ of the SINR $\gamma_{\mathrm{ul},\ell_k}^{\mathrm{mmse}}$ 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

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

with $\delta_{\ell_k} = \frac{1}{M} \mathrm{tr}\ \Phi_{\ell_k} T_{\ell_k}$, $\mu_{\ell_k} = \frac{1}{M} \mathrm{tr}\ \Phi_{\ell_k} T'_{\ell_k}$, $\nu_{\ell\ell'_k} = \frac{1}{r_{\ell_k} M} \mathrm{tr}\ \Lambda_{\ell\ell'_k} \Xi_{\ell_k} \Lambda_{\ell_k} T_{\ell_k}$, $\mu_{\ell_k \ell'_{k'}} = \frac{\mathrm{tr} \Lambda_{\ell\ell'_{k'}}}{M} \mu_{\ell_k}$, where the definitions of $T_{\ell_k}$ and $T'_{\ell_k}$ 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. For this, we extended the standard technique of deterministic equivalents in a few aspects in [11]: 1) the unbounded spectral norm of channel covariance matrices with respect to $M$ and 2) the partial random Fourier correlation model for which the trace lemma in [8, Lem. 2.7] (see also [9, 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 $\varphi_{\mathrm{ul},\ell_k} \triangleq P_{\mathrm{ul}} M < \infty$. As $M \to \infty$ with $\alpha_{\ell_k} \to 0$ under Assumption 2, we have $\Xi_{\ell_k} \simeq \Lambda_{\ell_k}^{-1}, T_{\ell_k} \simeq \varphi_{\mathrm{ul},\ell_k} I_{r_{\ell_k}}, T'_{\ell_k} \simeq \varphi_{\mathrm{ul},\ell_k} I_{r_{\ell_k}}$, where $\simeq$ refers to equivalence in the limit. Plugging those into (9), we have

$$\bar{\gamma}_{\mathrm{ul},\ell_k}^{\mathrm{mmse}} \simeq \frac{\left(\frac{\varphi_{\mathrm{ul},\ell_k}}{M} \mathrm{tr} \Lambda_{\ell_k}\right)^2}{\frac{\varphi_{\mathrm{ul},\ell_k}}{M} \mathrm{tr} \Lambda_{\ell_k} + \sum_{\ell' \neq \ell} \alpha_{\ell_k}^2 \left(\frac{\varphi_{\mathrm{ul},\ell_k}}{M} \mathrm{tr} \Lambda_{\ell\ell'_k}\right)^2 + \frac{1}{M} \sum_{\ell',k'} \frac{\varphi_{\mathrm{ul},\ell_k}}{M^2} \mathrm{tr} \Lambda_{\ell\ell'_k} \mathrm{tr} \Lambda_{\ell_k}}$$

$$\simeq \frac{\varphi_{\mathrm{ul},\ell_k}}{M} \mathrm{tr} \Lambda_{\ell_k} = P_{\mathrm{ul}} \mathrm{tr} \Lambda_{\ell_k}. \tag{10}$$

Th rest of the proof is somewhat standard [10], [12], [14]. ∎

**Remark 1.** We can make an important observation 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 $\alpha_{\ell_k}$ vanishes under Assumption 2. Likewise, the interference term completely disappears through spatial despreading.

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

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

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

---

It follows from (8) that the coherent lower bound yields the multiplexing gain per cell is given by $\left(1 - \frac{\kappa}{T_c}\right)\kappa$ for the typical (orthogonal) pilot scheme. Accordingly, 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 $K \geq \frac{T_c}{2}$ [14]. 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 [8, Lem. 2.7]. We thus turn our attention to non-coherent bounding techniques.

### B. Non-Coherent Lower Bound

Marzetta [17] 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 of 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 this technique, we have the ergodic achievable rate in (6), shown on the top of page 4. We can then obtain the following 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 sum capacity of MIMO uplink behaves as

$$\mathcal{C}_M^{\mathrm{ul}} = \left(1 - T_c^{-1}\right)\sum_{\ell=1}^{L}\sum_{k=1}^{K_\ell}\log(P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell_k}) + o(1). \tag{12}$$

*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 $\gamma_{\mathrm{ul},\ell_k}^{\mathrm{MF}}$. Similar to Lemma 1, we have $\gamma_{\mathrm{ul},\ell_k}^{\mathrm{MF}} \xrightarrow[M\to\infty]{a.s.} \bar{\gamma}_{\mathrm{ul},\ell_k}^{\mathrm{MF}} = \frac{(\frac{1}{M}\mathrm{tr}\boldsymbol{\Xi}'_{\ell_k})^2}{\frac{1}{P_{\mathrm{ul}}M}\frac{1}{M}\mathrm{tr}\boldsymbol{\Xi}'_{\ell_k} + \sum_{(\ell',k')\neq(\ell,k)}\alpha_{\ell_k}^2\psi_{\ell'_{k'}}^2}$, where $\psi_{\ell'_{k'}} = \frac{1}{r_{\ell_k}M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_{k'}}\mathrm{tr}\,\boldsymbol{\Xi}'_{\ell_k}\boldsymbol{\Lambda}_{\ell_k}$. Similar to (10), $\boldsymbol{\Xi}'_{\ell_k} \simeq \boldsymbol{\Lambda}_{\ell_k}^{-1}$ and we have $\bar{\gamma}_{\mathrm{ul},\ell_k}^{\mathrm{MF}} \simeq P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell_k}$, where we used a direct combination of [11, Cor. 1 and 2] and then [11, Lem. 8]. 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). In fading channels with strong spatial correlation, where the dimension $r_{\ell_k}$ of the effective channel $\boldsymbol{w}_{\ell_k}$ could be much small than $M$ (i.e., lack of channel hardening),

$\mathcal{R}_{\mathrm{ul},\ell_k}^{(2)}$ suffers from the self-interference due to a non-negligible variance term $\mathrm{var}[\boldsymbol{g}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k}|\underline{\boldsymbol{U}}]$ unless $r_{\ell_k}$ becomes sufficiently large. As a consequence, $\mathcal{R}_{\mathrm{ul},\ell_k}^{(2)}$ 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 [18]. The third lower bound is given by (7), shown on the top of page 4, . The first term in the right-hand side (RHS) of (7) represents the max-min upper bound denoted by $\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}}$, where the max is over the coding/decoding strategy of user $\ell_k$ and the min is over all input distributions of the other users. The second term consists of the prelog factor $\frac{\sum_\ell K_\ell - 1}{T_c}$ and the variances of coherent interference $\mathrm{var}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell_{k'}}|\underline{\boldsymbol{U}}]$ multiplied by the transmit power $P_{\mathrm{ul}}$ inside the logarithm. Basically this bound comes very close to $\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}}$ when coherence block is sufficiently large or coherent interference is limited. The latter is the case with our main scenario under the strong spatial correlation regimes. One can prove that $\mathcal{R}_{\mathrm{ul},\ell_k}^{(3)}$ achieves the same capacity scaling as Theorem 2 and the same scaling law is achievable in downlink as well.

### D. Interpretation of Main Results

Let us consider the homogeneous network with $\iota = 1$. It is important to note that in order to obtain (1) (or (12)) in the strong correlation regime, we have assumed that both $K$ and $L$ are finite, which is implicit in the massive MIMO literature [1], [2], [7], [12] together with $T_c$ growing linearly with $K$. Rather, we will see that under the very strong correlation regime in Assumption 3, the linear sum capacity scaling with respect to $M$ at finite SNR can be achieved as long as $\limsup_{M\to\infty}\frac{K}{M} < \infty$, even if the coherence block $T_c$ is finite. This argument may appear counterintuitive, since $K$ users in each cell share at least $K$ orthogonal pilot dimension so that the number of data streams is upper-bounded by $\min\{M, K, T_c/2\}$. While this is indeed the case for isotropic fading or, more in general, for channel covariances with rank $r = \alpha M$, with fixed $\alpha > 0$, it is not the case under the strong correlation regimes, where $r$ grows without bound but slower than $M$ so that $\alpha$ vanishes in the limit of $M$, thus making non-orthogonal pilot based channel training feasible. To show this, we will use $\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}}$ and $\boldsymbol{v}_{\ell_k} = \boldsymbol{s}'_{\ell_k}$ (ignoring normalization). The overall interference term in $\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}}$ for user $\ell_k$ is given by

$$\frac{1}{M^2}\sum_{(\ell',k')\neq(\ell,k)}|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell_{k'}}|^2 = \frac{1}{M^2}\times$$

$$\sum_{(\ell',k')\neq(\ell,k)}\left|\left(\boldsymbol{w}_{\ell_k}+\sum_{(j,m)\neq(\ell,k)}\boldsymbol{w}_{\ell_k j_m}+\frac{1}{\sqrt{\rho_p}}\boldsymbol{z}_{\ell_k}\right)^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}\right|^2.$$

We investigate the asymptotic behaviors of two main terms from the above equation.

- Coherent pilot contamination term ($(j,m)=(\ell',k')$): In the limit of $M$, we can derive the following almost sure convergence

$$\frac{\|\boldsymbol{w}_{\ell_k\ell'_{k'}}\|^2}{M}\xrightarrow{a.s.}\frac{r}{M^2}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\qquad(14)$$

where $\boldsymbol{w}_{\ell_k\ell'_{k'}}$ is given in (2), and the convergence follows from Lemma 2 since $\frac{\|\boldsymbol{w}_{\ell\ell'_{k'}}\|^2}{M}\xrightarrow{a.s.}\frac{1}{M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}$ by the trace lemma and $\boldsymbol{U}^{\mathsf{H}}_{\ell\ell'_{k'}}\boldsymbol{U}_{\ell_k}\boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{U}_{\ell\ell'_{k'}}\xrightarrow{a.s.}\frac{r}{M}\boldsymbol{I}_r$ by [11, Lem. 8]. Note that the latter convergence shows the role of spatial despreading in uplink, which is much more crucial than the former (channel hardening) to obtain the capacity scaling results in this paper. Using the dominated convergence theorem, we have

$$\frac{1}{M^2}\sum_{(\ell',k')\neq(\ell,k)}\left|\boldsymbol{w}^{\mathsf{H}}_{\ell_k\ell'_{k'}}\boldsymbol{w}_{\ell_k\ell'_{k'}}\right|^2\xrightarrow{a.s.}\frac{LK}{M}\cdot\frac{r^2\mathrm{tr}\boldsymbol{\Lambda}^2_{\ell\ell'_{k'}}}{M^3}\qquad(15)$$

which vanishes under Assumption 3 as long as $\limsup_{M\to\infty}\frac{K}{M}<\infty$ and $L$ is finite.

- Inter/intra-cell interference term: Although interference does not matter in the typical massive MIMO because of finite $K$ and the law of large number in the limit of $M$, this is not the case with $\limsup_{M\to\infty}\frac{K}{M}<\infty$. For the residual interference term after spatial despreading, $|\boldsymbol{w}^{\mathsf{H}}_{\ell_k}\boldsymbol{w}_{\ell_k\ell'_{k'}}|^2=|\boldsymbol{w}^{\mathsf{H}}_{\ell_k}\boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{U}_{\ell\ell'_{k'}}\boldsymbol{w}_{\ell\ell'_{k'}}|^2,\forall(\ell',k')\neq(\ell,k)$, "common" angular components between the channel covariance matrices of any pair of users are relatively very small such that $\lim_{r\to\infty}\frac{\Delta_{\ell_k,\ell\ell'_{k'}}}{r}=0$, where $\Delta_{\ell_k,\ell\ell'_{k'}}$ is the number of the similar angular components shared by users $\ell_k$ and $\ell\ell'_{k'}$. Specifically, $\boldsymbol{U}^{\mathsf{H}}_{\ell_k}\boldsymbol{U}_{\ell\ell'_{k'}}\xrightarrow{a.s.}\frac{1}{\sqrt{M}}\boldsymbol{J}_r$ as a straightforward corollary of [11, Lem. 8], where $\boldsymbol{J}_r$ is the $r$-dimensional all-ones matrix, and $\frac{|\mathsf{h}^{\mathsf{H}}_{\ell_k}\mathsf{h}_{\ell\ell'_{k'}}|^2}{r^2}\xrightarrow{a.s.}0$. Similar to (15), we have

$$\frac{1}{M^2}\sum_{(\ell',k')\neq(\ell,k)}\left|\boldsymbol{w}^{\mathsf{H}}_{\ell_k}\boldsymbol{w}_{\ell_k\ell'_{k'}}\right|^2\xrightarrow{a.s.}$$
$$\frac{LK}{M}\cdot\frac{r^2\mathrm{tr}\boldsymbol{\Lambda}^2_{\ell\ell'_{k'}}}{M^2}\cdot\frac{|\mathsf{h}^{\mathsf{H}}_{\ell_k}\mathsf{h}_{\ell\ell'_{k'}}|^2}{r^2}\qquad(16)$$

which also vanishes under Assumption 3.

We can observe that $r$ should grow without bound for the interference term in the RHS of (16) to fade away as $M$ tends to infinity. This in turn demands a very large value of $M$ for $r$ to be sufficiently large in the very strong spatial correlation regime. However, $r$ is not required to go to infinity for the system to be free from pilot contamination and interference as in Theorem 2 as long as $\lim_{M\to\infty}\frac{K}{M}=0$. Also, note that since a deterministic approximation of the normalized signal power $\frac{\|\boldsymbol{w}_{\ell_k}\|^2}{M}$ as well as (15) and (16) depends on realizations of $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$, the resulting SINR of user $\ell_k$ is conditional to
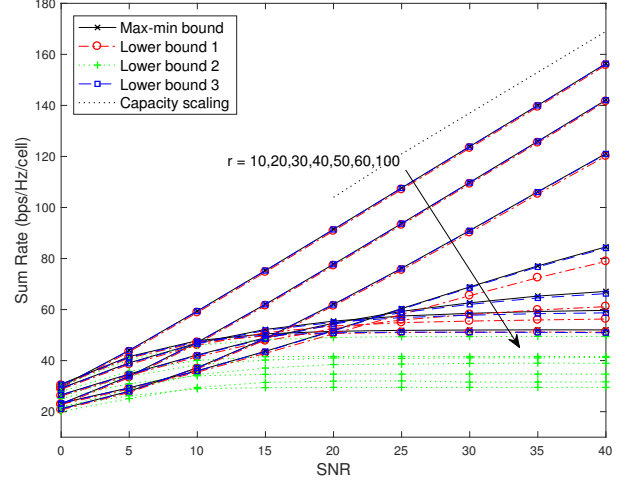


Fig. 1. The impact of spatial despreading on sum-rate scaling with respect to the sparsity of angular supports ($r$), where $M=200$, $L=4$, $K=10$, $T_c=500$. 'Lower bound 1', 'Lower bound 2', and 'Lower bound 3' are given by (5), (6), and (7), respectively. The asymptotic capacity scaling is (8) with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

random channel covariance matrices of all users in the network. However, under the deterministic $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$ assumption, it follows from Lemma 2 that the conditional SINR converges to a deterministic limit for sufficiently large $M$, thus allowing the ergodic rate analysis in this work. Using a similar argument, we have the following corollary of Theorem 2.

**Corollary 1.** For large $M$ with $\limsup_{M\to\infty}\frac{K}{M}<\infty$ under Assumptions 1 and 3 with non-orthogonal pilot, the capacity of massive MIMO uplink asymptotically behaves as (12).

Unlike uplink, it is required in downlink that the sum power $P_{\mathrm{dl}}$ grow linearly with $\frac{K}{\mathrm{tr}\boldsymbol{\Lambda}_{\ell_k}}$ under Assumption 3 with $\limsup_{M\to\infty}\frac{K}{M}<\infty$, making the downlink system power-limited as $K$ increases. Accordingly, this result does not mean that spatial multiplexing per BS is represented by $(1-T_c^{-1})\max\{M,K\}$.

## IV. NUMERICAL RESULTS

For numerical examples in this section, we only consider the homogeneous scenario that has $L$ cells serving $K$ users each with inter-cell interference factor $\iota=0.2$ and use the single-cell MMSE combining or precoding. The symmetric geometry of users is assumed such that we normalize channel covariance matrices to satisfy $\mathrm{tr}\boldsymbol{R}_{\ell\ell'_k}=M$ for all $(\ell,\ell',k)$ and $r_{\ell_k}=r$ for all $(\ell,k)$. We used pilot power gap of $\varrho_p=2$ (i.e., 3 dB).

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 (or multipath components in angular domain) of channel covariance matrices in uplink. We used the orthogonal pilot sequences. Unlike $\mathcal{R}^{(1)}_{\mathrm{ul},\ell_k}$ and $\mathcal{R}^{(3)}_{\mathrm{ul},\ell_k}$, $\mathcal{R}^{(2)}_{\mathrm{ul},\ell_k}$ ('lower bound 2') in (6) does not show linear growth due to lack of hardening of the effective channels $\boldsymbol{w}_{\ell_k}$, whose dimension is $r$. The coherent lower bound suffers from channel estimation error represented by the parallel shift of capacity versus SNR curves. At low SNR, both interference suppression and pilot decontamination effects
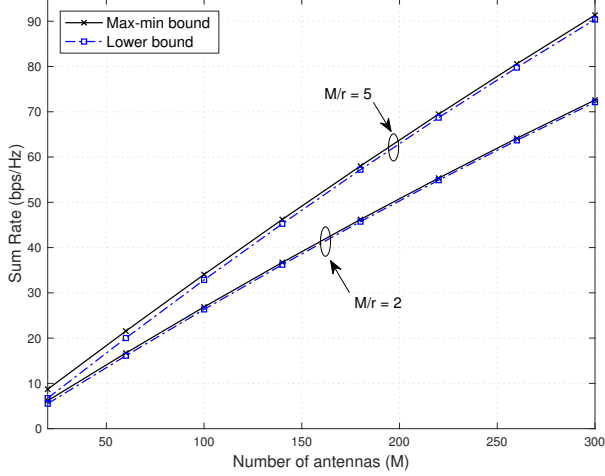
Fig. 2. Sum-rate growth with respect to the number of BS antennas at SNR = 10 dB in MIMO downlink with the ratios $\frac{M}{K} = 5$ and $\zeta = \frac{M}{r}$ fixed, where the 'Lower bound' is given by the downlink version of (5).

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 $\boldsymbol{U}_{\ell_k}$.

Fig. 2 verifies the scalability of the sum-rate scaling of MIMO downlink with respect to $M$ with the ratios $\frac{M}{K}$ and $\zeta = \frac{M}{r}$ fixed, where $M$ is in the linear scale, and $L = 7$. At SNR = 10 dB, we observe that given the fixed ratios of $M, K, r$, the sum rate scales almost *linearly* with $M$. This implies that the effect of spatial spreading scales well with respect to $M$, although $r$ is not much smaller than $M$. Furthermore, this linear growth is observed even for $M < KL$.

## V. CONCLUSION

Channel hardening has been traditionally considered as an essential source of massive MIMO gain. This is not necessarily the case with strong spatial correlation under the random sparse angular support models. Rather, one can observe that the effect of spatial (de)spreading is indeed central to achieve the ultimate capacity scaling law. Although the capacity scaling is achieved under the sublinear sparsity assumption in this work, the effect of spatial (de)spreading is shown to be still valid at finite $M$ with not-so-sparse angular support.

## APPENDIX

**Lemma 2.** ( [11, Lem. 5]) Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots,$ with $\boldsymbol{x}_n \in \mathbb{C}^n$, be random vectors whose entries satisfy the conditions in [11, Cor. 1]. Let $\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \ldots,$ with $\boldsymbol{\Lambda}_n \in \mathbb{C}^{n \times n}$, be a series of random matrices with uniformly bounded spectral norm with respect to $n$, independent of $\boldsymbol{x}_n$ and convergent such that $\boldsymbol{\Lambda}_n - \mathring{\boldsymbol{\Lambda}}_n \xrightarrow{a.s.} 0$. Then as $n \to \infty$, $\boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{\Lambda}_n \boldsymbol{x}_n - \frac{1}{n} \operatorname{tr} \mathring{\boldsymbol{\Lambda}}_n \xrightarrow{a.s.} 0$.

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

$$\hat{\boldsymbol{w}}_{\ell_k}^{\mathsf{H}} \boldsymbol{\Upsilon}_{-\ell_k} \boldsymbol{w}_{\ell_k \ell_k}$$
$$= \left( \boldsymbol{w}_{\ell_k} + \sum_{j \neq \ell} \boldsymbol{w}_{\ell_k j_k} + \sqrt{\rho_p}^{-1} \boldsymbol{z}_{\ell_k} \right)^{\mathsf{H}} \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{\Upsilon}_{-\ell_k} \boldsymbol{w}_{\ell_k \ell_k'}$$

$$\stackrel{(a)}{\simeq} \boldsymbol{w}_{\ell_k \ell_k'}^{\mathsf{H}} \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{\Upsilon}_{-\ell_k} \boldsymbol{w}_{\ell_k \ell_k'}$$
$$= \boldsymbol{w}_{\ell\ell_k'}^{\mathsf{H}} \underbrace{\boldsymbol{U}_{\ell\ell_k'}^{\mathsf{H}} \boldsymbol{U}_{\ell_k} \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{\Upsilon}_{-\ell_k} \boldsymbol{U}_{\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell_k'}}_{\stackrel{(b)}{\simeq} \frac{1}{M} \operatorname{tr} (\boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \boldsymbol{\Upsilon}_{-\ell_k}) \boldsymbol{I}_{r_{\ell_k'}'}} \boldsymbol{w}_{\ell\ell_k'} \qquad (17)$$

$$\stackrel{(c)}{\simeq} \frac{1}{M^2} \operatorname{tr} \boldsymbol{\Lambda}_{\ell\ell_k'} \operatorname{tr} \boldsymbol{\Xi}_{\ell_k} \boldsymbol{\Lambda}_{\ell_k} \tilde{\boldsymbol{\Upsilon}}_{\ell_k} \simeq \frac{r_{\ell_k}}{M} \nu_{\ell\ell_k'} = \alpha_{\ell_k} \nu_{\ell\ell_k'} \qquad (18)$$

where $(a)$ follows from asymptotic orthogonality for $j \neq \ell'$. In $(b)$, we have used Lemma 2 since (17) is a mixture of random vector $\boldsymbol{w}_{\ell\ell_k'}$ and random matrix product $\boldsymbol{U}_{\ell\ell_k'}^{\mathsf{H}} \boldsymbol{U}_{\ell_k}$ independent of each other, and by [11, Lem. 8] we have $\boldsymbol{\Lambda}_{\ell_k \ell_k'} - \frac{\operatorname{tr} \boldsymbol{\Lambda}_{\ell\ell_k'}}{M} \boldsymbol{I}_{r_{\ell_k'}'} \xrightarrow{a.s.} 0$. 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] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[3] Z. Jianhua, W. Chao, W. Zhongyuan, and Z. Weite, "A survey of massive MIMO channel measurements and models," *ZTE Communications*, vol. 15, no. 1, Feb. 2017.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

[11] J. Nam, G. Caire, M. Debbah, and H. V. Poor, "Capacity scaling of massive MIMO in strong spatial correlation regimes," *submitted to IEEE Trans. on Inform. Theory*, Dec. 2018. [Online]. Available: https://arxiv.org/abs/1812.08898

[12] 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.

[13] 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.

[14] 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.

[15] D. V. Sarwate, "Meeting the welch bound with equality," *in Sequences and their Applications. London, U.K.: Springer-Verlag*, pp. 79–102, 1998.

[16] 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.

[17] 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.

[18] G. Caire, "On the ergodic rate lower bounds with applications to massive MIMO," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 5, pp. 3258 – 3268, May 2018.