Capacity Scaling of Massive MIMO in Strong Spatial Correlation Regimes

Junyoung Nam, *Member, IEEE*, Giuseppe Caire, *Fellow, IEEE*, Mérouane Debbah, *Fellow, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

Abstract—This paper investigates the capacity scaling of multicell massive MIMO systems 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. To achieve the asymptotic capacity results, *statistical spatial despreading* 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, a linear growth of the capacity with respect to the number of antennas is achievable under the sparse angular support model. We also note that the achievable rate lower bound based on massive MIMO "channel hardening", widely used in the massive MIMO literature, yields rather loose results in the strong spatial correlation regimes and may significantly underestimate the achievable rate of massive MIMO. This work therefore considers an alternative bounding technique which is better suited to the strong correlation regimes. In fading channels with sparse angular support, it is further shown that spatial despreading (spreading) in uplink (downlink) has a more prominent impact on the performance of massive MIMO than channel hardening.

Index Terms—Large-scale MIMO, asymptotic capacity scaling, multiplexing gain, correlated fading channels.

I. INTRODUCTION

Achieving ever higher spectral efficiency has always been a central problem in wireless networks. To this end, massive multiple-input multiple-output (MIMO) [1], [2], also referred to as large-scale MIMO, 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 holds despite hardware impairments. In order to accurately predict the performance of multicell massive MIMO, it will be important to investigate the asymptotic sum capacity in the limit of a large number of antennas, when not only channel training cost, channel uncertainty, and out-of-cell interference but spatial correlation is also taken into account. This work considers particular regimes where the covariance matrix of each user channel has a rank which scales sublinearly with the number of BS antennas, M.

For the large-scale antenna array regime, in which typically the number of antennas M per cell is larger than the number of all active users in the homogeneous L-cell network with K users per cell and finite L, we can characterize the achievable rate of massive MIMO in several ways. For a simple isotropic channel model with pilot sequence reuse factor 1, where every cell shares the same set of pilot sequences, the spectral efficiency of typical massive MIMO networks can be written by following the line of arguments in [1], [3], [4] as

$$\mathcal{R}_M = \left(1 - \frac{\kappa_1}{T_c}\right)\kappa_1 L \log\left(1 + \frac{1}{\iota(L-1)}\right) + o(1) \quad (1)$$

where $\mathcal{R}_M$ is the asymptotic[1] achievable rate as $M \to \infty$, T_c is the number of channel uses of a time-frequency coherence block, $\kappa_1 = \min\{M, K, \lfloor T_c/2 \rfloor\}$, $\iota \in (0,1]$ is a symmetric intercell interference factor, and $o(1)$ goes to zero as M grows without bound. This result clearly shows performance limits of massive MIMO due to the pilot contamination effect coming from the reuse of pilot sequences among different cells. More specifically, the multiplexing gain (or spatial degrees of freedom) defined by $\lim_{\mathsf{SNR} \to \infty} \frac{\mathcal{C}(\mathsf{SNR})}{\log \mathsf{SNR}}$, where $\mathcal{C}(\mathsf{SNR})$ is the capacity of a channel with SNR being equivalent to the uplink/downlink sum power per cell in this work, vanishes and the power (beamforming) gain is saturated, no matter how large M is. In order to avoid the pilot contamination problem, one may utilize the globally orthogonal pilot sequences across the entire network. As a straightforward consequence of [1], [3], [5], the corresponding massive MIMO network has then

---

[1]In fact, some dependency on M is captured in the $o(1)$ factor, which vanishes in the large M limit.

the capacity scaling law

$$\mathcal{C}_M = \left(1 - \frac{\kappa_2}{T_c}\right)\kappa_2 \log\left(\mathsf{SNR}\,\frac{M}{K}\right) + o(1) \qquad (2)$$

where $\mathcal{C}_M$ is the asymptotic capacity in the limit of M, $\kappa_2 = \min\{M, KL, \lfloor T_c/2 \rfloor\}$, and $o(1)$ follows from the fact that every interference asymptotically vanishes in the absence of pilot contamination. For finite $T_c \leq 2KL$,[2] however, the above multiplexing gain of the L-cell network is approximately limited by $\frac{T_c}{4}$ even if both M and K grow without bound, yielding that every cell only gets multiplexing gain of $\frac{T_c}{4L}$. Therefore, for fixed T_c with large KL, the channel training cost immediately turns out to be a critical limiting factor that undermines the performance gain of MIMO networks [8]. Recently, allocating up to $\lfloor T_c/2 \rfloor$ channel uses to pilots was restated by [9] in the context of massive MIMO along with an analysis of the optimal number of scheduled users in terms of spectral efficiency. In fact, the scaling law of (2) serves as a lower bound on asymptotic achievable rate for the globally orthogonal pilot scheme, regardless of channel statistics, whereas it becomes tight when user channels are isotropic. This implies that for different and more realistic channel models, both (1) and (2) are not necessarily tight. Meanwhile, other capacity scaling results can also be found that investigate non-coherent single-input multiple-output (SIMO) multiple access channel [10] and single-user multiple-input single-output (MISO) channel with feedback of previous channel outputs [11].

Since the use of orthogonal pilot sequences across the entire network is too constraining in terms of the pilot dimensionality overhead (see (2)) and pilot contamination without a further countermeasure yields an interference limited system (see (1)), several techniques to tackle the pilot contamination problem have been proposed in the literature. For instance, [12] proposed multicell cooperative precoding/combining over the entire network, and blind pilot decontamination was given by [13] to separate signal and interference subspaces into disjoint supports. Following [14], [15], 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 (non-overlapping) angular supports. A joint angle and delay domain based pilot decontamination was also developed in [16], [17]. In a different line of work, [18] recently proved that the linear independence of those subspaces is rather surprisingly not a necessary condition for the elimination of contamination with infinitely many M antennas. The more general sufficient

condition therein is an *asymptotic* linear independence of the covariance matrices themselves 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 each cell to estimate all user channels in the multicell network.

Rather than any explicit pilot decontamination technique, this work focuses on the capacity scaling law in the massive MIMO network when the randomness and the sparsity of angular supports of channel covariance matrices are taken into consideration. Different scattering geometries of users located arbitrarily in a network make the angular support of each user channel random. This randomness also captures the fact that due to common scatterers, the angular supports of user channels are often partially overlapped. The sparsity of the angular support means that the number of significant multipaths in angular domain is much smaller than M. This arises in scenarios of limited scattering geometry, which has been observed in several channel measurement campaigns in not only millimeter wave (mm-Wave) but also below 6 GHz bands (see [19], [20] and references therein), where the number of non-negligible angular components of the user channel is fairly smaller than M, although the covariance matrix is mathematically of full rank. In this paper, we consider particularly two strong spatial correlation regimes, in which the rank r of channel covariance matrices grows sublinearly with M. 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 [21]–[24] that the ergodic sum capacity behaves as (3), shown on the top of page 3, where we used (intra-cell) non-orthogonal pilot that consumes only a single channel use per coherence block across the network as an extreme case and also assumed $\lim_{M\to\infty}\frac{K}{M} = 0$ for the former regime and $\limsup_{M\to\infty}\frac{K}{M} < \infty$ for the latter. Note that the above scaling law is asymptotically tight and its prelog factor 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 [3], whose prelog factor is $(1 - T_c^{-1})$.

The main differences of (3) and prior work can be summarized as follows:

1) *Capacity scaling characterization*: To the best of our knowledge, it is not clear in the prior work how the sum rate scales with respect to any of M, K, T_c, and SNR, except for the simple case of orthogonal pilots across the whole network as in (2). In particular, as mentioned earlier, the multiplexing gain of the massive MIMO network is limited by $T_c/4$ when $T_c \leq 2KL$ and globally orthogonal pilots are used, meaning that $K \leq \frac{T_c}{2L}$ users per cell can only be served. Otherwise, pilot contamination eventually suffocates multiplexing

---

[2]In the block-fading channel model, we presume a single channel use per coherence block for channel training. For a desirable channel estimation performance with noisy observations of realistic (non block-fading) channels, it is common to design an "oversampled" allocation of pilot symbols across time and frequency grid by a factor of $O = 2$ or even more, where O is an oversampling factor per time/frequency domain (e.g, [6], [7]). In this case, the pre-log factor becomes $(1 - \kappa_2' O^2/T_c)\kappa_2'$ with $\kappa_2' = \min\{KL, \lfloor \frac{T_c}{2O^2} \rfloor\}$, meaning that the condition $T_c \leq 2KL$ should be replaced with $T_c \leq 2KLO^2$.

$$\mathcal{C}_M = \begin{cases} \left(1 - \frac{1}{T_c}\right) \min\{M, K\} L \log\left(\mathsf{SNR} \frac{M}{K}\right) + o(1) & \text{under strong spatial correlation} \\ \left(1 - \frac{1}{T_c}\right) K L \log\left(\mathsf{SNR} \frac{r}{K}\right) + o(1) & \text{under very strong spatial correlation} \end{cases} \tag{3}$$

gain as in (1). However, the capacity scaling in (3) is not dominated by T_c any more under strong spatial correlation regimes.

2) *Scaling of K with respect to M*: Past work has generally assumed $\lim_{M\to\infty} \frac{K}{M} = 0$, i.e., K is finite or at most grows slower than M. For instance, if $KL > M$, the linear independence of the subspaces of covariance matrices in [14], [15] is never attainable. Likewise, if $\liminf_{M\to\infty} \frac{K}{M} > 0$, then the asymptotic linear independence of covariance matrices in [18] does not hold any longer. Therefore, the results based on the linear independence of the signal subspaces or of the covariance matrices cannot capture the capacity scaling with respect to the ratio $\frac{K}{M} > 0$, as both M and K grow large. This poses a fundamental methodological question since in reality we are in the presence of a finite system with given M and K. Which of the following large system analyses will produce the more meaningful prediction of its behavior: considering finite K and letting $M \to \infty$ [1], [2], [18] or considering both K and $M \to \infty$ with fixed ratio equal to the actual ratio $\frac{K}{M}$ of the practical finite system [4]? We claim that the latter methodology yields a more relevant asymptotics.

3) *Scaling of T_c with respect to M*: The conventional large system analysis including [4] implicitly assumes that the coherence block T_c of the channel scales linearly with K to accommodate the intra-cell orthogonal pilot sequences consuming K channel uses. Moreover, the semi-blind decontamination technique [13] can completely remove pilot contamination, provided that T_c goes to infinity as well as M. However, the channel coherence block depends on Doppler spread due to mobility of users and on frequency selectivity due to delay spread of multipath channels, irrespectively of M, and hence it is finite in practice. Meanwhile, our large system limits hold in finite T_c since the use of non-orthogonal pilot turns out to be feasible in strong spatial correlation regimes.

4) *Unlimited capacity as M and K grow with $\frac{K}{M} > 0$ and finite T_c*: The second capacity characterization in (3) shows that every user can get unlimited spectral efficiency in the very strong spatial correlation regime with finite T_c, as long as K grows no faster than M and the uplink/downlink per-user transmit power $\frac{\mathsf{SNR}}{K}$ is not vanishing. It turns out that spatial despreading in uplink (or spreading in downlink), which acts as statistical spatial filters matched to the angular supports of user channel covariances, makes pilot contamination terms fade away and that an infinite sum of the vanishing contamination terms still vanishes. Moreover, overlapped angular components between any pair of the channel covariance matrices of users only occupy a vanishing

portion of the total signal space dimension if their angular supports become sufficiently sparse as $M \to \infty$.

5) *Beamforming architecture and complexity*: Spatial despreading (spreading) is followed by low-dimensional (i.e., r-dimensional) channel estimation and single-cell combining in uplink (precoding in downlink), which is asymptotically optimal in the strong correlation regimes even if each BS only estimates the low-dimensional effective channels of its own users. We will use the low-dimensional processing to prove the achievability of the large system limits in (3). The low-dimensional minimum mean-squared error (MMSE) channel estimation and combining/precoding show a comparable performance to the conventional M-dimensional MMSE processing at finite M. Furthermore, spatial despreading/spreading is in fact long-term or wideband (i.e., frequency non-selective) processing, while low-dimensional combining/precoding is short-term or narrowband (frequency selective) processing. Therefore, they are amenable to the hybrid beamforming architecture (see [25] and references therein) that is beneficial to realize massive MIMO in practice by saving hardware cost and channel training/feedback overhead.

In order to provide some intuition behind the main results, let us explain why pilot contamination vanishes even though we do not rely on any explicit decontamination technique and only use a finite pilot dimension (e.g., a single symbol) in finite coherence block T_c, such that there is even pilot contamination for the users in the same cell. From a technical perspective, as long as the number of users K and the rank of the channel covariance r grow sublinearly with (i.e., slower than) the number of BS antennas M, overlapped angular support of channel covariances of different users can be managed as only a subset of Lebesgue measure zero in positive reals $\mathbb{R}^+$ under appropriate conditions, as $M \to \infty$, which leads to the almost sure convergence of pilot contamination to zero. It is mainly due to the effect of spatial (de)spreading, which is implicitly conducted by channel propagation itself and suppresses the contamination by a factor of $\frac{M}{r}$, that the user channels come to be "asymptotically orthogonal" to each other. Therefore, the effect of pilot contamination is indeed not a fundamental limiting factor for the user channels with random sparse angular support. This also holds true in the very strong correlation regime, in which K grows linearly with M, as long as r grows rather more slowly than M and the received channel energy also grows slower than M.

Another contribution of this work is an investigation into lower-bounding techniques in terms of massive MIMO with random sparse angular support. We begin with the fact that the well-known channel hardening effect in massive MIMO can be undermined by spatial correlation at finite M. In

particular, the useful signal coefficient does not concentrate on its mean when the fading channels are highly correlated. As a consequence, the widely used non-coherent bound [2] in massive MIMO downlink, which works very well in the case of channel hardening, may significantly underestimate the achievable rate, depending on spatial correlation. Furthermore, for finite M, the coherent bounding technique [2] also widely used in uplink may suffer from channel estimation error due to imperfect CSI at the receiver (CSIR) especially when non-orthogonal pilot is employed. In this work, we further consider a lesser-known alternative non-coherent bounding technique given in [26] to better estimate the performance of massive MIMO in the strong spatial correlation regimes.

The paper is organized as follows: In Sec. II, we describe the system model with two stochastic spatial correlation models and explain spatial (de)spreading. Sec. III addresses our main results on the capacity scaling of massive MIMO along with their implication. In Sec. IV, we present an extension of the deterministic equivalents technique. Sec. V provides the proofs of the main results in Sec. III. Sec. VI contains some numerical results. We conclude this work in Sec. VII.

Notation: We use $\xrightarrow{a.s.}$ for the almost sure convergence such that, for sequences a_n and b_n, $a_n - b_n \xrightarrow{a.s.} 0$ as $n \to \infty$, and $\simeq$ is also used for the sake of compactness. Let $\limsup_{n\to\infty} a_n$ and $\liminf_{n\to\infty} a_n$ denote the limit superior and the limit inferior of a_n, respectively. For a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_2$ and $\mathrm{tr}\boldsymbol{A}$ denote the spectral norm and the trace of $\boldsymbol{A}$, respectively. For a vector $\boldsymbol{a}$, $\|\boldsymbol{a}\|$ denotes the ℓ_2 norm of $\boldsymbol{a}$. $\mathcal{CN}(0,1)$ denotes the zero-mean circularly symmetric complex Gaussian distribution.

II. PRELIMINARIES

A. System Model

We consider an L-cell time-division duplex (TDD) MIMO network where the ℓ-th BS has M antennas and serves K_ℓ single-antenna users. The user channel follows the frequency-flat block-fading model for which it remains constant during the coherence block of T_c but changes independently every interval, where T_c is the number of channel uses or the signal dimension in the time-frequency domain. The fading distribution is known at both the transmitters and the receivers. We do not require any cooperation between multiple BSs such that there is neither channel state information (CSI), data, scheduling information, nor pilot allocation information coordinated across the network through backhaul links. Indexing the kth user in BS ℓ by ℓ_k, $\boldsymbol{h}_{\ell\ell'_k}$ is the uplink channel from user ℓ'_k to BS ℓ, and $\boldsymbol{h}^{\mathsf{H}}_{\ell'\ell_k}$ is the downlink channel from BS ℓ' to user ℓ_k. Using the Karhunen-Loève transform, the channel vector $\boldsymbol{h}_{\ell\ell'_k}$ can be expressed as

$$\boldsymbol{h}_{\ell\ell'_k} = \boldsymbol{U}_{\ell\ell'_k} \boldsymbol{\Lambda}_{\ell\ell'_k} \mathsf{h}_{\ell\ell'_k} \tag{4}$$

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 of

$\boldsymbol{R}_{\ell\ell'_k}$, and $\mathsf{h}_{\ell\ell'_k} \in \mathbb{C}^{r_{\ell\ell'_k}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ is the small-scale channel component. Throughout this work, it is assumed that the rank $r_{\ell\ell'_k}$ of $\boldsymbol{R}_{\ell\ell'_k}$ is assumed to be much smaller than M due to limited scattering environments. Furthermore, BS ℓ has a prior knowledge on the low-rank covariance matrices $\boldsymbol{R}_{\ell\ell_k}$ of its own users and a sum of low-dimensional covariance matrices of the other-cell user channels projected by $\boldsymbol{U}_{\ell\ell'_k}$, depending on the channel training schemes in Subsection II-D.

In the MIMO uplink, the received signal vector at BS ℓ can be given by

$$\boldsymbol{y}_\ell = \sum_k \boldsymbol{h}_{\ell\ell_k} x_{\ell_k} + \sum_{\ell' \neq \ell} \sum_k \boldsymbol{h}_{\ell\ell'_k} x_{\ell'_k} + \boldsymbol{z}_\ell \tag{5}$$

where x_{ℓ_k} is the input signal of user ℓ_k chosen from a Gaussian codebook and satisfies the equal power constraint such that $\mathbb{E}[|x_{\ell_k}|^2] \leq P_{\mathrm{ul}}$, and $\boldsymbol{z}_\ell \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ is the Gaussian noise at the BS antennas. Since the noise power per antenna is normalized to be unity, P_{ul} can be regarded as the transmit SNR per user in uplink. Meanwhile, the received signal vector at user ℓ_k at BS ℓ in the downlink can be given by

$$y_{\ell_k} = \boldsymbol{h}^{\mathsf{H}}_{\ell\ell_k} \boldsymbol{p}_{\ell_k} d_{\ell_k} + \sum_{k' \neq k} \boldsymbol{h}^{\mathsf{H}}_{\ell\ell_k} \boldsymbol{p}_{\ell_{k'}} d_{\ell_{k'}}$$
$$+ \sum_{\ell' \neq \ell} \sum_{k'} \boldsymbol{h}^{\mathsf{H}}_{\ell'\ell_k} \boldsymbol{p}_{\ell'_{k'}} d_{\ell'_{k'}} + z_{\ell_k} \tag{6}$$

where $\boldsymbol{p}_{\ell_k}$ and d_{ℓ_k} are the precoding vector and the input signal of user ℓ_k satisfying $\mathbb{E}[\|\boldsymbol{p}_{\ell_k} d_{\ell_k}\|^2] \leq P_{\mathrm{dl},\ell}$, respectively, and $z_{\ell_k} \sim \mathcal{CN}(0,1)$ is the Gaussian noise. We assumed the equal power allocation within a cell such that $P_{\mathrm{dl},\ell} = \frac{P_{\mathrm{dl}}}{K_\ell}$, where P_{dl} is the sum-power constraint per cell, and hence P_{dl} becomes the normalized transmit SNR in line with the classical MIMO downlink, which is denoted by SNR in this paper.

B. Stochastic Spatial Correlation Models

The channel covariance matrix is determined by the propagation geometry, and in particular by the distribution of the signal power over the angle domain (angular scattering function). It is well known that the propagation geometry changes on a time-scale much larger than the coherence time of the small-scale fading. This means that the channel can be considered as "locally wide-sense stationary (WSS)". In other words, on a relatively large-scale time span, which we call *local stationarity interval* in this paper, it is seen as a snapshot taken from a WSS process with given second-order statistics (e.g., [27], [28]). In this paper, we assume that on each such local stationarity interval, the channel covariance matrices are drawn at random from a distribution with $\boldsymbol{R}_{\ell\ell'_k} = \boldsymbol{U}_{\ell\ell'_k} \boldsymbol{\Lambda}_{\ell\ell'_k} \boldsymbol{U}^{\mathsf{H}}_{\ell\ell'_k}$. While the columns of $\boldsymbol{U}_{\ell\ell'_k}$ span the large-scale angular domain subspace of multipath components, $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$ captures the large-scale fading factors such as path loss and shadow fading, i.e., the channel energy that BS ℓ receives from user ℓ'_k. In the following, we introduce two stochastic models for the eigenvector matrix $\boldsymbol{U}_{\ell\ell'_k}$ to capture the randomness of spatial correlation, which typically

arises in wireless channel propagation due to arbitrary user and scattering geometry.

1) Random partial unitary model: A simple model for $U_{\ell\ell'_k}$ is a random partial unitary matrix. In this model, $U_{\ell\ell'_k}$ is independently and uniformly drawn from a random partial unitary matrix whose column space is in the Grassmann manifold $\mathcal{G}(M, r_{\ell\ell'_k})$, which is the set of all $r_{\ell\ell'_k}$-dimensional subspaces in $\mathbb{C}^M$. Hence, the random partial unitary matrices are mutually independent for all (ℓ, ℓ', k). The ith column of U (denoted by u_i) is a random unit vector uniformly distributed on the M-dimensional complex unit sphere such that $u_i = \frac{x}{\|x\|_2}$ with covariance $\frac{1}{M} I_M$, where $x \sim \mathcal{CN}(\mathbf{0}, I_M)$.

2) Random partial Fourier model: Another interesting spatial correlation model is a random partial (subsampled) Fourier matrix, motivated by the typical uniform linear array (ULA) in multiple antenna systems. Let F_{jk} denote the (j, k)th entry of the discrete Fourier transform (DFT) matrix $F \in \mathbb{C}^{M \times M}$, as shown by

$$F_{jk} = \frac{1}{\sqrt{M}} e^{j2\pi jk/M}, \ j, k = 0, \ldots, M-1.$$

Suppose that $U_{\ell\ell'_k}$ is composed of $r_{\ell\ell'_k}$ column vectors uniformly drawn at random *without replacement* from the Fourier basis functions of F so that different users can have common basis elements, taking into account common scatterers shared by multiple users. The resulting unitary matrix can be represented by

$$U_{\ell\ell'_k} = F G_{\ell\ell'_k} \tag{7}$$

where $G_{\ell\ell'_k} \in \mathbb{C}^{M \times r_{\ell\ell'_k}}$ is the random selection matrix that chooses $r_{\ell\ell'_k}$ columns (angular components) without replacement from M columns of F. This correlation model is a reminiscence of the angular domain representation of MIMO channels [29] or virtual channel representation [30] widely-used in the literature. In particular, the random partial Fourier model can be justified by the known asymptotic behavior of channel covariance in massive MIMO with ULA [15], showing that the eigenvectors of channel covariance matrices are well approximated by the columns of DFT for large M. Notice that the random partial Fourier model has much less degrees of freedom (i.e., highly structured or less randomness) than the random partial unitary model. In particular, the columns of $U_{\ell\ell'_k}$ in (7) are not statistically independent due to sampling without replacement, even if they are linearly independent.

It is important to notice that given the above random realizations of the channel covariance matrices of users, the resulting ergodic achievable rates are conditional to such realizations. Therefore they are random variables unlike most sum-rate analyses in the massive MIMO literature. 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 characterize the ergodic capacity scaling in this work rather than the outage capacity scaling, we make the assumption that the

long-term channel energy captured by the large-scale fading factor $\mathrm{tr}\Lambda_{\ell\ell'_k}$ does not change over user mobility and different scattering geometries for a given value of M such that

$$\mathrm{tr}\Lambda_{\ell\ell'_k} = c_{\ell\ell'_k}(M), \ \forall(\ell, \ell', k) \tag{8}$$

where $c_{\ell\ell'_k}(M)$ is a positive real constant over certain local stationarity intervals[3]. Under this deterministic $\mathrm{tr}\Lambda_{\ell\ell'_k}$, we will show later in Sec. III-D that the conditional ergodic rates converge to a deterministic limit for sufficiently large M, meaning that the limit does not depend any longer on a specific covariance realization but only on the distribution. As a consequence, we will focus on such deterministic limits in our asymptotic sum-rate analysis.

Our spatial correlation models that capture the random (or diverse) nature of angular components in user channels allow different angular supports among users as in [14], [15]. We notice here that in some other works, e.g., [4], [32], antenna correlation was modeled by letting all users to have the same covariance matrix. Such model is physically less justifiable, since it implies that all users share the same multipath components with the same strength, i.e., the users are all co-located.

Given the stochastic models on $U_{\ell\ell'_k}$, we next make several assumptions on the eigenvalue matrices $\Lambda_{\ell\ell'_k}$ with $\mathrm{tr}\Lambda_{\ell\ell'_k}$ deterministic for the large system analysis in the random covariance matrices $R_{\ell\ell'_k}$.

Assumption 1. For all (ℓ, ℓ', k)

$$\limsup_{M \to \infty} \frac{r_{\ell\ell'_k}}{M} \|\Lambda_{\ell\ell'_k}\|_2 < \infty \tag{9a}$$

$$\limsup_{M \to \infty} \frac{\mathrm{tr}\Lambda_{\ell\ell'_k}}{M} < \infty. \tag{9b}$$

Assumption 2 (*Strong Spatial Correlation Regime*)**.** The number $r_{\ell\ell'_k}$ of non-zero eigenvalues of $R_{\ell\ell'_k}$ grows without bound but slower than M such that

$$\frac{r_{\ell\ell'_k}}{M} \triangleq \alpha_{\ell\ell'_k} \xrightarrow[M \to \infty]{} 0 \tag{10a}$$

$$\liminf_{M \to \infty} \frac{\mathrm{tr}\Lambda_{\ell\ell'_k}}{M} > 0, \ \forall(\ell, \ell', k). \tag{10b}$$

First of all, the condition (10a) implies in conjunction with (10b) that the spectral norm[4] $\|\Lambda_{\ell\ell'_k}\|_2$ is *not necessarily* uniformly bounded with respect to M. The uniform boundedness is a necessary condition for the method of deterministic equivalents [21], [23], [33], [34]. Hence, under those conditions the deterministic equivalents technique cannot directly apply any longer. We will extend the technique later on in IV to address this issue.

---

[3] For an illustrative example, 5G new radio (NR) mm-Wave systems may have OFDM symbol duration of 4.5 μsec with 240 KHz subcarrier spacing [31]. In this case, 10^4 symbols only span 45 msec, which corresponds to a local stationarity interval over which channel second order statistics does not change. With a few hundred such intervals, a pedestrian only travels 10 m at a walking speed of 1 m/sec. Therefore it may be reasonable to assume that the long-term channel energy captured by $\mathrm{tr}\Lambda_{\ell\ell'_k}$ remains unchanged over such intervals since the amount of intervals does not substantially change path loss and shadowing.

[4] For positive semidefinite A, it is simply the maximum eigenvalue of A
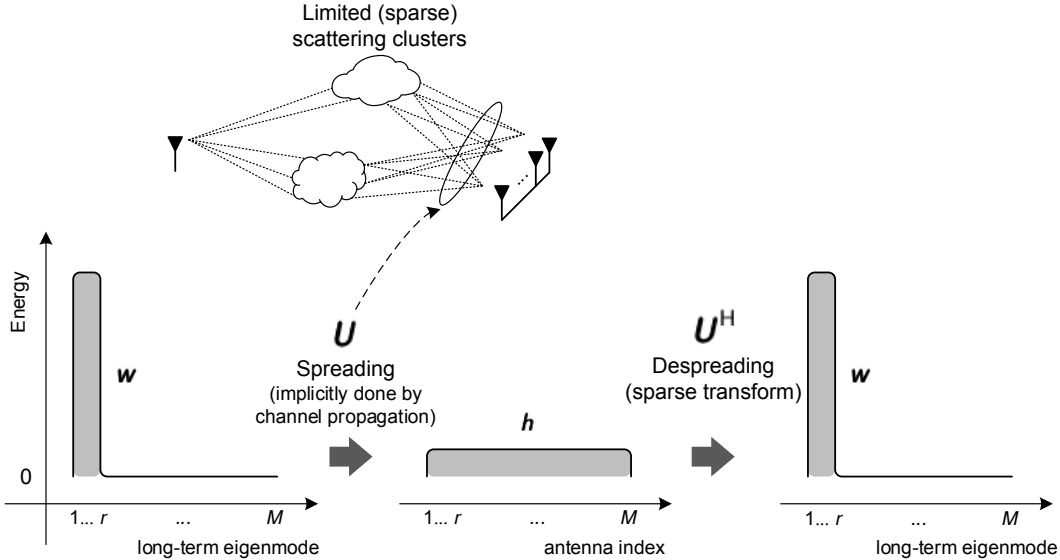
Fig. 1. Change of coordinates (sparse transform of the channel vector $\boldsymbol{h}$) in uplink.

Remark 1. The sublinear sparsity assumption in (10a) hinges on the premise that the number of angular components of each user channel grows without bound, but slower than the number of BS antennas. In the massive MIMO literature, the angular domain model is categorized into two cases: a large but finite number of angular components (e.g., [32]) and infinitely many components with a constant ratio of $\alpha_{\ell\ell'_k} > 0$ (e.g., [4]). The former model is based on the well-known fact [29] that angular resolution of an array is proportional to the array aperture that should be finite in practice. In contrast, the latter can be justified for high carrier frequency systems like mm-Wave, where the angular resolution might continue to increase proportionally with the carrier frequency f_c. In the mm-Wave channels, however, it was observed (e.g., [35]) that the higher f_c, the smaller number of multipaths that can arrive at the receivers due to higher path loss, shadowing, and blockage. This sparsity of surviving multipath components is also verified by mm-Wave propagation measurement campaigns [36] and also by [20] even in below 6 GHz, meaning that non-negligible eigenvalues of $\boldsymbol{R}_{\ell\ell'_k}$ may be very sparse in the large M limit.[5] As a matter of fact, Assumptions 1 and 2 model the above *limited scattering* channel environments in the sense that more energy concentrates upon relatively sparse angular support of $\boldsymbol{R}_{\ell\ell'_k}$ as M increases. Therefore, it is particularly reasonable to assume the sublinear sparsity when the size of the antenna array does not scale with M in practice and rather carrier frequency f_c scales.

An insufficiency of Assumption 2 is that the large-scale fading factor $\text{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$ 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. This means that despite such a sparse angular support in limited scattering clusters, the channel energy does not dissipate so as to concentrate upon the $r_{\ell\ell'_k}$-dimensional subspace spanned by the columns of $\boldsymbol{U}_{\ell\ell'_k}$. Therefore, we also consider another spatial correlation regime, where $\text{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$ 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 of non-zero eigenvalues $r_{\ell\ell'_k}$ grows without bound but much more slowly than M such that

$$\frac{r_{\ell\ell'_k}^4}{M^3} \xrightarrow[M\to\infty]{} 0 \tag{11a}$$

$$\limsup_{M\to\infty} \frac{r_{\ell\ell'_k}^2}{M} < \infty \tag{11b}$$

$$\limsup_{M\to\infty} \frac{\text{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}}{r_{\ell\ell'_k}} < \infty, \ \forall(\ell, \ell', k). \tag{11c}$$

This very strong correlation regime is of importance since it captures significant path loss, shadowing, and penetration loss in high carrier frequency bands like mm-Wave and above. Interestingly, it will be shown later on that the two correlation regimes lead to different sum-rate scaling laws in terms of the ratio of the number of users per cell and the number of BS antennas.

C. Spatial Despreading and Spreading

In this subsection, we introduce statistical spatial despreading in uplink (spreading in downlink) based on channel covariance matrices and explain its implication.

---

[5]Although user terminals in mm-Wave systems have multiple or even large-scale antennas, the analog beamforming architecture turns the MIMO channel into an effective multiple-input single-output channel. Hence, our channel model with sparse angular support can cover the typical mm-Wave beamforming architecture.

- *Sparse transform:* Under our spatial correlation models, $\boldsymbol{h}_{\ell\ell'_k}$ has a sparse representation and $\boldsymbol{U}_{\ell\ell'_k}$ serves as a sparse transformation matrix [6] of $\boldsymbol{h}_{\ell\ell'_k}$ such that

$$\boldsymbol{U}_{\ell\ell'_k}^{\mathsf{H}} \boldsymbol{h}_{\ell\ell'_k} = \boldsymbol{\Lambda}_{\ell\ell'_k}^{\frac{1}{2}} \mathsf{h}_{\ell\ell'_k} \triangleq \boldsymbol{w}_{\ell\ell'_k} \qquad (12)$$

where $\boldsymbol{w}_{\ell\ell'_k} \in \mathbb{C}^{r_{\ell\ell'_k}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Lambda}_{\ell\ell'_k})$ is the projected effective channel vector, whose dimension is much lower than that of the original vector $\boldsymbol{h}_{\ell\ell'_k}$ in $\mathbb{C}^M$.

- *Energy concentration:* In the above transformation, the energy of $\boldsymbol{h}_{\ell\ell'_k}$ is preserved and concentrated on $\boldsymbol{w}_{\ell\ell'_k}$ in (12) since $\mathrm{tr}\boldsymbol{R}_{\ell\ell'_k} = \mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}$, implying $\boldsymbol{U}_{\ell\ell'_k}^{\perp\,\mathsf{H}} \boldsymbol{h}_{\ell\ell'_k} = \boldsymbol{0}$, where $\boldsymbol{U}_{\ell\ell'_k}^{\perp}$ spans the null space of $\boldsymbol{U}_{\ell\ell'_k}$. The energy concentration is simply due to the classical Parseval theorem implying that the change of coordinates (basis) preserves inner products (energy).

The sparse transform naturally gives rise to an interesting interpretation of spatial despreading as follows. One can regard $\{\boldsymbol{U}_{\ell\ell'_k}, \forall \ell, \ell', k\}$ as "statistical spatial matched filters" or "random spreading sequences" in the classical uplink CDMA [39] with asynchronous users since the sparse transform (multiplying $\boldsymbol{h}_{\ell\ell'_k}$ by $\boldsymbol{U}_{\ell\ell'_k}^{\mathsf{H}}$) in uplink and its counterpart (multiplying $\boldsymbol{w}_{\ell\ell'_k}$ by $\boldsymbol{U}_{\ell\ell'_k}$) in downlink are a reminiscence of *despreading* and *spreading*, respectively. Hence the sparse transform and its counterpart will be referred to as spatial despreading and spreading, respectively. Unlike CDMA, the spatial (de)spreading is not controllable at the cost of bandwidth, but it depends on the propagation channel. In particular, under the assumptions made before, the channel with limited scattering offers an unbounded spatial spreading gain (energy concentration) as $M \to \infty$ without incurring any bandwidth cost. Fig. 1 illustrates how spatial spreading is implicitly conducted by channel propagation and how spatial despreading based on the second-order channel statistics $\boldsymbol{U}_{\ell\ell'_k}$ is done by the change of coordinates.

For the homogeneous network where all users have the same $r_{\ell\ell'_k} = r$, the symmetric spatial spreading gain is defined as

$$\zeta = \frac{M}{r}. \qquad (13)$$

Similar to the uplink CDMA with random spreading signatures, for large M, one can then intuitively expect that spatial despreading can suppress interference power and, more importantly, pilot contamination by the factor ζ in the homogeneous network. This effect will be addressed later on in (45) of Sec. V.

Once spatial despreading is performed upon the received signal $\boldsymbol{y}_\ell$ in (5), the transformed vector $\boldsymbol{y}_{\ell_k}$ for user ℓ_k is

---

[6]This is also known as a sparse representation matrix and should not be confused with a sensing (or measurement) matrix in compressed sensing [37], [38].

given by

$$\begin{aligned}
\boldsymbol{y}_{\ell_k} &= \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{y}_\ell \\
&= \boldsymbol{w}_{\ell\ell_k} x_{\ell_k} + \sum_{k' \neq k} \boldsymbol{w}_{\ell_k \ell_{k'}} x_{\ell_{k'}} + \sum_{\ell' \neq \ell} \sum_{k'} \boldsymbol{w}_{\ell_k \ell'_{k'}} x_{\ell'_{k'}} + \boldsymbol{z}_{\ell_k}
\end{aligned}$$
$$(14)$$

where $\boldsymbol{z}_{\ell_k} = \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{z}_\ell$ and

$$\boldsymbol{w}_{\ell_k \ell'_{k'}} = \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell'_{k'}} \boldsymbol{w}_{\ell\ell'_{k'}}, \ \forall(\ell, k, \ell', k'). \qquad (15)$$

The subsequent receiver processing applies to the transformed vector $\boldsymbol{y}_{\ell_k}$ instead of $\boldsymbol{y}_\ell$. Meanwhile, in order to leverage the sparsity in spatially correlated fading channels in downlink, we let the BS perform spatial spreading such that $\boldsymbol{p}_{\ell_k} = \boldsymbol{U}_{\ell\ell_k} \boldsymbol{g}_{\ell_k}, \forall \ell, k$, where $\boldsymbol{g}_{\ell_k} \in \mathbb{C}^{r_{\ell\ell_k}}$. We can then express (6) as

$$y_{\ell_k} = \boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{g}_{\ell_k} d_{\ell_k} + \sum_{(\ell',k') \neq (\ell,k)} \boldsymbol{w}_{\ell'_{k'} \ell_k}^{\mathsf{H}} \boldsymbol{g}_{\ell'_{k'}} d_{\ell'_{k'}} + z_{\ell_k}. \qquad (16)$$

It should be noted that spatial despreading in uplink (or spreading in downlink) by $\boldsymbol{U}_{\ell\ell_k}$ does not incur any loss of optimality from the single-user perspective, but it is suboptimal from the perspective of MU-MIMO because spatial despreading projects the M-dimensional signal space onto the $r_{\ell\ell_k}$-dimensional subspace, in which multiuser combining (or precoding) is performed based on $\{\boldsymbol{w}_{\ell\ell_k}, \boldsymbol{w}_{\ell_k \ell'_{k'}}\}$. As a matter of fact, it is only asymptotically optimal under Assumption 2 (or 3), which will be shown by the main capacity scaling result in Theorem 2. Nevertheless, the purpose of representing the uplink/downlink received signal in the above forms (14) and (15) is three-fold: 1) to explicitly show the role of spatial spreading/despreading, 2) to separate the effect of channel hardening through $\boldsymbol{w}_{\ell\ell_k}$ and that of spatial despreading through $\boldsymbol{U}_{\ell\ell_k}$, and 3) to study the asymptotic performance of the resulting $r_{\ell\ell_k}$-dimensional channel estimation and combining/precoding. In particular, the effect of channel hardening arises when $\frac{\|\boldsymbol{w}_{\ell\ell_k}\|^2}{\mathbb{E}[\|\boldsymbol{w}_{\ell\ell_k}\|^2]} \xrightarrow{a.s.} 1$ for sufficiently large $r_{\ell\ell_k}$ as $M \to \infty$. The almost sure convergence can be given by the well-known trace lemma [21, Lem. 2.7] (see also [23, Thm. 3.4]), although the elements of $\boldsymbol{w}_{\ell\ell_k}$ are non-identically distributed. Therefore, channel hardening depends directly on the dimension $r_{\ell\ell_k}$ of the effective channel $\boldsymbol{w}_{\ell\ell_k}$ in our rank-deficient spatial correlation models.

D. Low-Dimensional Channel Estimation

For channel training through uplink pilot signals, we consider both orthogonal pilot per cell and non-orthogonal pilot over the entire network. In each case, the corresponding MMSE channel estimate is described in the following.

1) (Intra-Cell) Orthogonal Pilot Scheme: Let us consider a pilot-aided multicell MIMO system, where the uplink/downlink system uses a certain amount of channel uses to allow the BSs to estimate their associated user channel vectors. For the conventional orthogonal pilot scheme [1], [2], users in each cell send a set of mutually orthogonal pilot

sequences, which are shared by L cells in the network, so that user k in every cell uses the same pilot sequence, resulting in pilot contamination. The received pilot signal for user k in cell ℓ can then be expressed as

$$\bar{\boldsymbol{s}}_{\ell_k} = \sum_{\ell'=1}^{L} \boldsymbol{h}_{\ell\ell'_k} + \frac{1}{\sqrt{\rho_p}} \boldsymbol{z}_\ell \qquad (17)$$

where $\rho_p = \varrho_p P_{\text{ul}}$ with ϱ_p the power boosting factor (i.e., power gap between the training phase and the communication phase). This requires the overall training cost of $\max_\ell K_\ell$ channel uses in half duplex TDD, where we assume synchronized cells such that in a given time slot all BSs schedule transmissions in the same direction to avoid severe inter-cell interference. In general, the maximum number of users per cell is limited by the uplink pilot dimension in this orthogonal scheme [1]. After spatial despreading for the same purpose as (14), we will make use of the projected pilot signal $\boldsymbol{s}_{\ell_k}$ represented as

$$\boldsymbol{s}_{\ell_k} = \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \bar{\boldsymbol{s}}_{\ell_k} = \boldsymbol{w}_{\ell\ell_k} + \sum_{\ell' \neq \ell} \boldsymbol{w}_{\ell_k \ell_k} + \frac{1}{\sqrt{\rho_p}} \boldsymbol{z}_{\ell_k}. \qquad (18)$$

Given the noisy observation $\boldsymbol{s}_{\ell_k}$ and the prior knowledge on $\boldsymbol{U}_\ell = [\boldsymbol{U}_{\ell\ell_1}, \ldots, \boldsymbol{U}_{\ell\ell_{K_\ell}}]$ and the sum, $\sum_{\ell' \neq \ell} \boldsymbol{U}_{\ell\ell'_k}^{\mathsf{H}} \boldsymbol{R}_{\ell\ell'_k} \boldsymbol{U}_{\ell\ell'_k}$, of low-dimensional covariance matrices of the other-cell user channels after spatial despreading by $\boldsymbol{U}_{\ell\ell'_k}^{\mathsf{H}}$, and conditioned on a realization of $\underline{\boldsymbol{U}} = \{\boldsymbol{U}_{\ell\ell'_k}, \forall(\ell, \ell', k)\}$, the MMSE estimate $\hat{\boldsymbol{w}}_{\ell\ell_k}$ of the effective channel $\boldsymbol{w}_{\ell\ell_k}$ is given by

$$\begin{aligned}
\hat{\boldsymbol{w}}_{\ell\ell_k} &= \mathbb{E}\big[\boldsymbol{w}_{\ell\ell_k} \boldsymbol{s}_{\ell_k}^{\mathsf{H}} | \underline{\boldsymbol{U}}\big] \mathbb{E}\big[\boldsymbol{s}_{\ell_k} \boldsymbol{s}_{\ell_k}^{\mathsf{H}} | \underline{\boldsymbol{U}}\big]^{-1} \boldsymbol{s}_{\ell_k} \\
&= \boldsymbol{\Lambda}_{\ell\ell_k} \boldsymbol{\Xi}_{\ell\ell_k} \boldsymbol{s}_{\ell_k}
\end{aligned} \qquad (19)$$

where $\boldsymbol{\Xi}_{\ell\ell_k} \triangleq \big(\boldsymbol{\Lambda}_{\ell\ell_k} + \sum_{\ell' \neq \ell} \tilde{\boldsymbol{R}}_{\ell_k \ell'_k} + \rho_p^{-1} \boldsymbol{I}_{r_{\ell\ell_k}}\big)^{-1}$ with $\tilde{\boldsymbol{R}}_{\ell_k \ell'_k} \triangleq \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{R}_{\ell\ell'_k} \boldsymbol{U}_{\ell\ell_k} \in \mathbb{C}^{r_{\ell\ell_k} \times r_{\ell\ell_k}}$ being the covariance of $\boldsymbol{w}_{\ell_k \ell'_k}$. The sum of low-dimensional covariance matrices $\boldsymbol{\Xi}_{\ell\ell_k}$ can rather be estimated by an empirical covariance matrix based on $\boldsymbol{s}_{\ell_k}$ instead of the prior knowledge since the fading channel statistics are ergodic and WSS, as long as the scattering geometry of users remains unchanged over the local stationarity interval. The distribution of $\hat{\boldsymbol{w}}_{\ell\ell_k}$ is $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Phi}_{\ell\ell_k})$, where $\boldsymbol{\Phi}_{\ell\ell_k} = \boldsymbol{\Lambda}_{\ell\ell_k} \boldsymbol{\Xi}_{\ell\ell_k} \boldsymbol{\Lambda}_{\ell\ell_k}$.

The effective channel $\boldsymbol{w}_{\ell\ell_k}$ of user ℓ_k can then be written as

$$\boldsymbol{w}_{\ell\ell_k} = \hat{\boldsymbol{w}}_{\ell\ell_k} + \boldsymbol{n}_{\ell\ell_k}$$

where $\boldsymbol{n}_{\ell\ell_k} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{N}_{\ell\ell_k})$ is conditionally independent of $\hat{\boldsymbol{w}}_{\ell\ell_k}$ given $\boldsymbol{U}_\ell$ by the orthogonality property of the MMSE estimate and the joint Gaussianity of $\hat{\boldsymbol{w}}_{\ell\ell_k}$ and $\boldsymbol{n}_{\ell\ell_k}$. From (19), the conditional error covariance matrix is given by

$$\begin{aligned}
\mathbb{E}\big[\boldsymbol{n}_{\ell\ell_k} \boldsymbol{n}_{\ell\ell_k}^{\mathsf{H}} | \underline{\boldsymbol{U}}\big] &= \mathbb{E}\big[\boldsymbol{w}_{\ell\ell_k} \boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}}\big] - \mathbb{E}\big[\hat{\boldsymbol{w}}_{\ell\ell_k} \boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}} | \underline{\boldsymbol{U}}\big] \\
&= \boldsymbol{\Lambda}_{\ell\ell_k} - \boldsymbol{\Lambda}_{\ell\ell_k} \boldsymbol{\Xi}_{\ell\ell_k} \boldsymbol{\Lambda}_{\ell\ell_k}.
\end{aligned}$$

Notice that spatial despreading gives us the $r_{\ell\ell_k}$-dimensional MMSE channel estimation for $\boldsymbol{w}_{\ell\ell_k}$ rather than the M-dimensional one. In fact, spatial despreading (spreading) is implicitly done in the full-dimensional MMSE channel

estimation and combining (precoding) based on $\bar{\boldsymbol{s}}_{\ell_k}$ in (17). The following result shows that the low-dimensional channel estimate $\hat{\boldsymbol{w}}_{\ell\ell_k}$ after spatial despreading is asymptotically equivalent to the full-dimensional channel estimate based on $\bar{\boldsymbol{s}}_{\ell_k}$.

Lemma 1. Under Assumption 2, $\boldsymbol{s}_{\ell_k}$ in (18) is an asymptotically sufficient statistic conditioned on $\boldsymbol{U}_\ell$ to estimate the effective channel $\boldsymbol{w}_{\ell\ell_k}$ from $\bar{\boldsymbol{s}}_{\ell_k}$ as $M \to \infty$.

Proof: Refer to Appendix B. ∎

2) (Intra-Cell) Non-Orthogonal Pilot Scheme: It is known [3], [8] that the conventional orthogonal pilot scheme may significantly limit the sum-rate performance of MU-MIMO systems 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 signal $\boldsymbol{s}'_{\ell_k}$ is non-orthogonal over intra cell as well as inter cell such that it consumes a single channel use[7] per L-cell network per coherence block, given by

$$\boldsymbol{s}'_{\ell_k} = \boldsymbol{w}_{\ell\ell_k} + \sum_{(\ell', k') \neq (\ell, k)} \boldsymbol{w}_{\ell_k \ell'_{k'}} + \frac{1}{\sqrt{\rho_p}} \boldsymbol{z}_{\ell_k}. \qquad (20)$$

Similar to the orthogonal pilot scheme, we have the MMSE channel estimate of $\boldsymbol{w}_{\ell\ell_k}$ given $\boldsymbol{s}'_{\ell_k}$ as follows.

$$\check{\boldsymbol{w}}_{\ell\ell_k} = \boldsymbol{\Lambda}_{\ell\ell_k} \boldsymbol{\Xi}'_{\ell\ell_k} \boldsymbol{s}'_{\ell\ell_k} \qquad (21)$$

where $\boldsymbol{\Xi}'_{\ell\ell_k} \triangleq \big(\boldsymbol{\Lambda}_{\ell\ell_k} + \sum_{(\ell', k') \neq (\ell, k)} \tilde{\boldsymbol{R}}_{\ell_k \ell'_{k'}} + \rho_p^{-1} \boldsymbol{I}_{r_{\ell\ell_k}}\big)^{-1}$. The distribution of $\check{\boldsymbol{w}}_{\ell\ell_k}$ is given by $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Phi}'_{\ell\ell_k})$, where $\boldsymbol{\Phi}'_{\ell\ell_k} = \boldsymbol{\Lambda}_{\ell\ell_k} \boldsymbol{\Xi}'_{\ell\ell_k} \boldsymbol{\Lambda}_{\ell\ell_k}$. As in Lemma 1, one can show that $\boldsymbol{s}'_{\ell\ell_k}$ is an asymptotically sufficient statistic to estimate the effective channel $\boldsymbol{w}_{\ell\ell_k}$ from $\boldsymbol{s}'_\ell$ under Assumption 2.

A natural question arises on how we estimate channel covariance matrices in the non-orthogonal pilot scheme. While non-orthogonal pilot can give a better sum-rate performance by reducing the training overhead for the estimation of effective channels when T_c is relatively small, extra "long-period" orthogonal pilot would be necessary for the channel covariance estimation. As T_c is typically much smaller than the local stationarity interval over which the channel process is WSS [27, Chap. 4], the corresponding training overhead for channel covariances will be rather small relative to "short-period" orthogonal pilot $\boldsymbol{s}_{\ell_k}$. In the practical scenario that one should estimate channel covariance matrices from contaminated pilot sequences, [17] proposed algorithms that exploit sparse angle-delay supports between users to separate the covariance estimates of the different users of interest.

E. Low-Dimensional Multiuser Combining and Precoding

We will only describe single-cell MMSE combining and precoding vectors based on the channel estimates $\hat{\boldsymbol{w}}_{\ell\ell_k}$ in

---

[7]For more general non-orthogonal pilot design, one may use Welch bound equality frames [40], where the cross-correlation coefficient between pilot sequences shared by different cells in the network is non-zero but the same. We can then design non-orthogonal pilot sequences with training cost K' such that $1 \leq K' < \max_\ell K_\ell$ (e.g., see [41] in the context of massive MIMO).

(19) for the orthogonal pilot scheme, not requiring cell ℓ to estimate the effective channels $\boldsymbol{w}_{\ell\ell'_{k'}}$ of users k' in other cells $\ell' \neq \ell$. Replacing $\hat{\boldsymbol{w}}_{\ell\ell_k}$ with $\check{\boldsymbol{w}}_{\ell\ell_k}$ in (21), we can have the MMSE combining and precoding vectors for non-orthogonal pilot. Since we restrict our attention to the noncooperative multiuser combining/precoding and the decoder that take the intercell interference as additional additive noise, the resulting sum rates in this work is achieved by treating interference as noise [42]. A sum-rate gap of such trivial (intercell) interference managements including time-division multiplexing from several capacity upper bounds was studied and turned out to be not significant at practical values of SNR for two-user [43] and K-user [44] Gaussian interference channels.

1) Uplink: Spatial despreading in (14) indicates that we restrict the full-dimensional linear combining vector to $\boldsymbol{U}_{\ell\ell_k}\boldsymbol{v}_{\ell_k} \in \mathbb{C}^M$, where $\boldsymbol{v}_{\ell_k}$ is the $r_{\ell\ell_k}$-dimensional combining vector for user ℓ_k. Given $\hat{\boldsymbol{w}}_{\ell_1}, \ldots, \hat{\boldsymbol{w}}_{\ell_{K_\ell}}$ and letting $\hat{\boldsymbol{w}}_{\ell_k\ell'_{k'}} \triangleq \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell'_{k'}}\hat{\boldsymbol{w}}_{\ell\ell'_{k'}}, \forall(\ell', k')$, the single-cell MMSE combining vector can be written as

$$\boldsymbol{v}_{\ell_k}^{\mathsf{mmse}} = \boldsymbol{\Upsilon}_{\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k} \qquad (22)$$

where

$$\boldsymbol{\Upsilon}_{\ell_k} = \left(\hat{\boldsymbol{w}}_{\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}} + \sum_{k'\neq k}\hat{\boldsymbol{w}}_{\ell_k\ell'_{k'}}\hat{\boldsymbol{w}}_{\ell_k\ell'_{k'}}^{\mathsf{H}} + \boldsymbol{z}_{\ell_k} + P_{\mathsf{ul}}^{-1}\boldsymbol{I}_{r_{\ell\ell_k}}\right)^{-1}$$

where $\boldsymbol{z}_{\ell_k}$ is a system design parameter and typically defined as $\boldsymbol{z}_{\ell_k} = \sum_{\ell'\neq\ell}\sum_k \boldsymbol{\Xi}_{\ell\ell'_k} + \sum_{\ell',k'}\boldsymbol{N}_{\ell_k\ell'_{k'}}$. The above low-dimensional vector $\boldsymbol{v}_{\ell_k}^{\mathsf{mmse}}$ is given based on the received vector $\boldsymbol{y}_{\ell_k}$ in (14), to which spatial despreading of $\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}$ applied.

2) Downlink: For downlink, the low-dimensional MMSE precoding vector can be written as

$$\boldsymbol{g}_{\ell_k}^{\mathsf{mmse}} =$$
$$\left(\hat{\boldsymbol{w}}_{\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}} + \sum_{k'\neq k}\hat{\boldsymbol{w}}_{\ell_{k'}\ell_k}\hat{\boldsymbol{w}}_{\ell_{k'}\ell_k}^{\mathsf{H}} + \boldsymbol{Z}_{\ell\ell_k}' + P_{\mathsf{dl},\ell}^{-1}\boldsymbol{I}_{r_{\ell\ell_k}}\right)^{-1}\hat{\boldsymbol{w}}_{\ell\ell_k}$$
$$(23)$$

where $\boldsymbol{Z}_{\ell\ell_k}' = \sum_{\ell'\neq\ell}\sum_k\left(\boldsymbol{\Lambda}_{\ell\ell'_k} + \sum_{\ell'\neq\ell}\boldsymbol{\Lambda}_{\ell'_{k'}\ell_k} + \rho_p^{-1}\boldsymbol{I}_{r_{\ell\ell_k}}\right)^{-1} + \sum_{\ell',k'}\boldsymbol{N}_{\ell_{k'}\ell_k}$. The final precoding vector is followed by spatial spreading such that $\boldsymbol{p}_{\ell_k} = \boldsymbol{U}_{\ell\ell_k}\boldsymbol{g}_{\ell_k}^{\mathsf{mmse}}$. One can also derive a more general form of the above MMSE precoding, i.e., in a regularized zero-forcing fashion.

By the data processing inequality, the performance of the above $r_{\ell\ell_k}$-dimensional combining (or precoding) vector is upper-bounded by that of the conventional (M-dimensional) MMSE combining (precoding) vector based on $\boldsymbol{y}_\ell$ in (5) (y_{ℓ_k} in (6)). We will see that the performance of $\boldsymbol{v}_{\ell_k}^{\mathsf{mmse}}$ is asymptotically equivalent to the large-dimensional MMSE combining vector as $M \to \infty$, although $\boldsymbol{v}_{\ell_k}^{\mathsf{mmse}}$ shows a performance degradation at finite M. This is also the case with $\boldsymbol{g}_{\ell_k}^{\mathsf{mmse}}$ in downlink.

A 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 [4, Eqn. (11)]). The full-dimensional MMSE processing generally shows the complexity of $O(M^3)$ complex multiplications and $O(M^2)$ complex divisions per cell due to matrix inversion. Meanwhile, the low-dimensional processing involves $O(Kr^3)$ complex multiplications and $O(Kr^2)$ complex divisions per cell in the homogeneous network, where every cell has K users with the same rank of $r_{\ell\ell_k} = r$. This low computational complexity also holds for the low-dimensional MMSE channel estimation in Sec. II-D. Consequently, the proposed MMSE processing may have much lower complexity in strong spatial correlation regimes, where $r \ll M$.

III. MAIN RESULTS

In this section, we provide capacity scaling results in massive MIMO under the system model and assumptions in Sec. II. To this end, we consider three lower bounds on the achievable uplink rate based on an extension of the deterministic equivalents technique [21]–[24], [33], [34]. While the first lower bound is based on coherent demodulation and decoding, the others are non-coherent. The results are then extended to downlink.

A. Uplink Capacity Scaling in the Strong Spatial Correlation Regime

1) Coherent Lower Bound: We first consider the standard lower bound on the achievable rate proposed by Hassibi and Hochwald [45], based on the worst-case uncorrelated additive noise lemma. This lemma has been widely used for coherent detection with imperfect CSI in the vast literature of limited feedback single/multiple-user MIMO systems (e.g., see [46] and reference therein). In this paper, the channel estimate of either (19) or (21) is available at the BS for coherent detection in the uplink.

Using the above bounding technique and the channel estimate in (19), the ergodic achievable rate of user ℓ_k is lower-bounded by (24), shown on the top of page 10, where $\hat{\boldsymbol{w}}_\ell = [\hat{\boldsymbol{w}}_{\ell_1}, \ldots, \hat{\boldsymbol{w}}_{\ell_{K_\ell}}]$. Notice that the above outer expectation is taken over $\{\boldsymbol{w}_{\ell\ell'_k}, \forall(\ell, \ell', k)\}$ conditioned on $\underline{\boldsymbol{U}}$,[8] implying that $\mathcal{R}_{\mathsf{ul},\ell_k}^{(1)}$ is a random variable conditional to a particular covariance realization of $\underline{\boldsymbol{U}}$ since it is a function of $\underline{\boldsymbol{U}}$, which is itself random. As mentioned earlier, it will be shown that such ergodic rates $\mathcal{R}_{\mathsf{ul},\ell_k}^{(1)}$ converge to a deterministic limit under the stochastic spatial correlation models in Sec. II-B. The derivation of (24) follows from the MMSE decomposition of the useful signal channel and from the worst-case uncorrelated additive noise argument. In this subsection, it is assumed that K_ℓ is a finite constant or scales sublinearly

---

[8]One might take the outer expectation also over $\underline{\boldsymbol{U}}$. In order to attain the ergodicity with respect to the channel covariances, we will need codewords that span hundreds of local stationary intervals much longer than small-fading coherence block. However, this is not very realistic in practical systems.

$$\mathcal{R}_{\mathrm{ul},\ell_k}^{(1)} = \mathbb{E}\left[\log\left(1 + \frac{|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\hat{\boldsymbol{w}}_{\ell\ell_k}|^2}{\mathbb{E}\left[|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{n}_{\ell\ell_k}|^2 + \sum_{(\ell',k')\neq(\ell,k)}|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell'_{k'}}|^2 + \frac{1}{P_{\mathrm{ul}}}|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{z}_{\ell_k}|^2\big|\,\hat{\boldsymbol{w}}_\ell,\underline{\boldsymbol{U}}\right]}\right)\bigg|\,\underline{\boldsymbol{U}}\right] \tag{24}$$

$$\mathcal{R}_{\mathrm{ul},\ell_k}^{(2)} = \log\left(1 + \frac{\left|\mathbb{E}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell_k}|\underline{\boldsymbol{U}}]\right|^2}{\frac{1}{P_{\mathrm{ul}}} + \mathrm{var}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell_k}|\underline{\boldsymbol{U}}] + \sum_{(\ell',k')\neq(\ell,k)}\mathbb{E}\left[|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\right]}\right) \tag{27}$$

with M such that $\lim_{M\to\infty}\frac{K_\ell}{M} = 0$ for all ℓ, as usual in the massive MIMO literature. We will address the case where $\limsup_{M\to\infty}\frac{K}{M} < \infty$ in Sec. III-C. Using (24), we obtain the following lower bound on the asymptotic capacity with respect to M.

Theorem 1. For large M with $\lim_{M\to\infty}\frac{K}{M} = 0$ and Assumptions 1 and 2 with the orthogonal pilot scheme, the ergodic sum capacity of MIMO uplink is lower-bounded by

$$\mathcal{C}_M^{\mathrm{ul}} \geq \sum_{\ell=1}^{L}\sum_{k=1}^{\kappa_3}\left(1 - \frac{\kappa_3}{T_c}\right)\log(P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}) + o(1). \tag{25}$$

where $\kappa_3 = \min\{\max_\ell K_\ell, \lfloor\frac{T_c}{2}\rfloor\}$.

Proof: Refer to Sec. V-A. ∎

For the above large system analysis, we extended the standard technique of deterministic equivalents in a few aspects in Appendix A: 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 [21], crucial for deterministic equivalents, is not applicable. The latter is due to the fact that the entries of the columns of $\boldsymbol{U}_{\ell\ell'_k}$ are not i.i.d. any longer and that the column vectors are also not independent of each other.

Suppose the homogeneous network such that

$$K_\ell = K, \ P_{\mathrm{ul}} = P_{\mathrm{dl},\ell} = \frac{\mathsf{SNR}}{K}, \ r_{\ell\ell'_k} = r, \ \forall(\ell,\ell',k). \tag{26}$$

In this homogeneous network, it follows from (25) that the coherent lower bound yields the multiplexing gain per user of $1 - \frac{\kappa_1}{T_c}$. Also multiplexing gain per cell is given by $\left(1 - \frac{\kappa_1}{T_c}\right)\kappa_1$ for the typical (orthogonal) pilot scheme. Accordingly, an intriguing question arises: *Is the sum-rate scaling law with the multiplexing gain of* (25) *optimal in massive MIMO under Assumption 2?* Such multiplexing gain per cell is indeed limited by $\frac{T_c}{4}$ when $K \geq \frac{T_c}{2}$ for relatively small T_c [8]. To answer this question, we need to consider the non-orthogonal pilot $\boldsymbol{s}'_{\ell_k}$ in Subsection II-D, whose training cost is only a single channel use across the L-cell network. In this case, unfortunately, the lower bound $\mathcal{R}_{\mathrm{ul},\ell_k}^{(1)}$ in (24) is not amenable to large system analysis because the trace lemma [21, Lem. 2.7] is not applicable. Furthermore, this coherent bound is expected to suffer from significant channel estimation error in the non-orthogonal pilot scheme. We thus turn our attention to non-coherent bounding techniques.

2) Non-Coherent Lower Bound: Marzetta [47] 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. The bounding technique may be traced back to Médard [48]. Following this technique, [49] developed a lower bound on the achievable rate of the MIMO downlink, where the fading distribution is assumed to be known, but the channel estimates are unavailable at the receivers unless additional dedicated downlink pilots are provided. Although this bound is accordingly very natural to use in downlink, it is applicable in uplink as well (e.g., [50]), where the uplink pilot is used to only construct the coherent combining vector $\boldsymbol{v}_{\ell_k}$ rather than coherent demodulation. Using this bounding technique, we have the ergodic achievable rate in (27), shown on the top of page 10. We can then obtain the following main result of this work.

Theorem 2. For large M with $\lim_{M\to\infty}\frac{K}{M} = 0$ and Assumptions 1 and 2 with the non-orthogonal pilot scheme, the sum capacity of MIMO uplink behaves as

$$\mathcal{C}_M^{\mathrm{ul}} = \left(1 - T_c^{-1}\right)\sum_{\ell=1}^{L}\sum_{k=1}^{K_\ell}\log(P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}) + o(1). \tag{28}$$

Proof: Refer to Sec. V-B. ∎

The factor $P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}$ in the right-hand side (RHS) of (28) represents the post-processing SNR in uplink. For the homogeneous network in (26) with $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k} = M$ and $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k} = \iota M$ for all (ℓ,ℓ',k), the first scaling law in (3) directly follows from (28). Compared to (1) and (2), this scaling function shows multiplexing gain is not limited any longer by $\frac{T_c}{4}$ for $T_c \leq 2LK$ in the strong correlation regime. This capacity scaling is asymptotically tight, as mentioned in the introduction, and shows the maximum multiplexing gain of $\left(1 - T_c^{-1}\right)LK$ from the perspective of either pilot-aided or non-coherent communications.

The lower bound $\mathcal{R}_{\mathrm{ul},\ell_k}^{(2)}$ predicts well the achievable rate of massive MIMO when the useful signal coefficient $\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell_k}$ hardens, i.e., behaves almost deterministically. However, in fading channels with strong spatial correlation, where the dimension $r_{\ell\ell_k}$ of the effective channel $\boldsymbol{w}_{\ell\ell_k}$ could be much smaller than M (i.e., lack of channel hardening), $\mathcal{R}_{\mathrm{ul},\ell_k}^{(2)}$ suffers from the self-interference due to the non-negligible variance term $\mathrm{var}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell_k}|\underline{\boldsymbol{U}}]$ in (27) unless $r_{\ell\ell_k}$ becomes sufficiently large. As a consequence, the non-coherent bound $\mathcal{R}_{\mathrm{ul},\ell_k}^{(2)}$ may substantially underestimate the sum-rate performance of massive MIMO, which will be examined through numerical

$$\mathcal{R}_{\mathrm{dl},\ell_k}^{(1)} = \log\left(1 + \frac{\left|\mathbb{E}[\boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{g}_{\ell_k}|\underline{\boldsymbol{U}}]\right|^2}{\frac{1}{P_{\mathrm{dl},\ell}} + \mathrm{var}[\boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{g}_{\ell_k}|\underline{\boldsymbol{U}}] + \sum_{(\ell',k')\neq(\ell,k)}\mathbb{E}\left[|\boldsymbol{w}_{\ell'_k,\ell_k}^{\mathsf{H}}\boldsymbol{g}_{\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\right]}\right) \tag{31}$$

$$\mathcal{R}_{\mathrm{dl},\ell_k}^{(2)} = \mathbb{E}\left[\log\left(1 + \frac{\left|\boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{g}_{\ell_k}\right|^2}{\frac{1}{P_{\mathrm{dl},\ell}} + \sum_{(\ell',k')\neq(\ell,k)}|\boldsymbol{w}_{\ell'_k,\ell_k}^{\mathsf{H}}\boldsymbol{g}_{\ell'_{k'}}|^2}\right)\Bigg|\,\underline{\boldsymbol{U}}\right] - \frac{1}{T_c}\sum_{(\ell',k')\neq(\ell,k)}\log\left(1 + P_{\mathrm{dl},\ell}\mathrm{var}[\boldsymbol{w}_{\ell'_k,\ell_k}^{\mathsf{H}}\boldsymbol{g}_{\ell'_{k'}}|\underline{\boldsymbol{U}}]\right) \tag{32}$$

---

results in Sec. VI. This is also particularly evident in the large but finite antenna regime.

3) Alternative Non-Coherent Lower Bound: We consider another non-coherent bounding technique very recently derived in [26] for the single-cell downlink scenario. The bound can be traced back to a mutual information decomposition given in the tutorial paper by Biglieri, Proakis, and Shamai [51, Eqn. 3.3.27] to study non-coherent block fading channels. Assuming Gaussian inputs, linear combining/precoding vector, and treating interference as noise, we have the following upper bound on the ergodic achievable rate [26]

$$\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}} = \mathbb{E}\left[\log\left(1 + \frac{\left|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell_k}\right|^2}{\frac{1}{P_{\mathrm{ul}}} + \sum_{(\ell',k')\neq(\ell,k)}|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|^2}\right)\Bigg|\,\underline{\boldsymbol{U}}\right] \tag{29}$$

which is a max-min bound such that the maximum is over the coding/decoding strategy of user ℓ_k and the minimum is over all input distributions of the other users. The third lower bound used in this work is given by [26]

$$\begin{aligned}\mathcal{R}_{\mathrm{ul},\ell_k}^{(3)} &= \mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}} \\ &\quad - \frac{1}{T_c}\sum_{(\ell',k')\neq(\ell,k)}\log\left(1 + P_{\mathrm{ul}}\mathrm{var}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|\underline{\boldsymbol{U}}]\right).\end{aligned} \tag{30}$$

The second term in the RHS of (30) 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_{ul} inside the logarithm. On one hand, this bound comes very close to the max-min upper bound 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, where interference is significantly suppressed by spatial despreading. On the other hand, a negative feature of this bound is that when the first term $\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}}$ is interference limited at finite M, P_{ul} grows to infinity, and/or T_c does not scale well with M, the bound goes to even below zero. Therefore, another lower bound with a complementary behavior can be found in [26, Lemma 4]. One can prove that the non-coherent lower bound $\mathcal{R}_{\mathrm{ul},\ell_k}^{(3)}$ achieves the same capacity scaling as Theorem 2. A sketch of the proof will be given in Sec. V-C.

B. Downlink Capacity Scaling in the Strong Spatial Correlation Regime

Unless we employ downlink dedicated (precoded) pilot signals consuming additional K_ℓ channel uses for the communication phase, the useful signal coefficient necessary for

coherent detection is unknown so that we cannot use $\mathcal{R}_{\mathrm{ul},\ell_k}^{(1)}$ for downlink. On the contrary, we can utilize both non-coherent bounds $\mathcal{R}_{\mathrm{ul},\ell_k}^{(2)}$ and $\mathcal{R}_{\mathrm{ul},\ell_k}^{(3)}$ for downlink as well to derive the capacity scaling result under sparse angular support models. Given the downlink received signal in (16) and conditioned on $\underline{\boldsymbol{U}}$, the lower bounds are given by (31) and (32) on the top of page 11. Using the above bounds under Assumptions 1 and 2 with the orthogonal pilot scheme, we can first derive the achievable ergodic sum rate in MIMO downlink

$$\mathcal{C}_M^{\mathrm{dl}} \geq \sum_{\ell=1}^L\sum_{k=1}^{\kappa_3}\left(1 - \frac{\kappa_3}{T_c}\right)\log(P_{\mathrm{dl},\ell}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}) + o(1) \tag{33}$$

where $P_{\mathrm{dl},\ell}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}$ is the post-processing SNR of user ℓ_k in the downlink. The above sum-rate characterization follows from the same footsteps in the proof of Theorem 1. Similar to Theorem 2, we also have the following result.

Theorem 3. For large M with $\lim_{M\to\infty}\frac{K}{M} = 0$ and Assumptions 1 and 2 with the non-orthogonal pilot scheme, the sum capacity of MIMO downlink behaves as

$$\mathcal{C}_M^{\mathrm{dl}} = \left(1 - T_c^{-1}\right)\sum_{\ell=1}^L\sum_{k=1}^{K_\ell}\log(P_{\mathrm{dl},\ell}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}) + o(1). \tag{34}$$

In order to prove the above result, it suffices to use either $\mathcal{R}_{\mathrm{dl},\ell_k}^{(1)}$ or $\mathcal{R}_{\mathrm{dl},\ell_k}^{(2)}$ with the transmit maximal ratio (or conjugate) precoding, i.e., $\boldsymbol{g}_{\ell_k} = \check{\boldsymbol{w}}_{\ell\ell_k}$. The proof is omitted for the sake of brevity, but follows the footsteps and arguments in the proofs of Theorems 1 and 2. A main difference is that we sometimes use Lemmas 7 and 9 instead of Lemmas 6 and 8, respectively. As expected, under the channel model assumptions of this paper the downlink has the same capacity scaling of the uplink.

C. Very Strong Spatial Correlation Regime

It is important to notice that in order to obtain (28) (or the first scaling law in (3)) in Theorem 2 and (34) in Theorem 3 in the strong correlation regime, we have assumed that L is finite and K is also finite or $\lim_{M\to\infty}\frac{K}{M} = 0$, which is implicit in the massive MIMO literature [1], [2], [4], [18], 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 still be achieved as long as $\limsup_{M\to\infty}\frac{K}{M} < \infty$, i.e., K does not grow faster than M, even though the coherence block size T_c does not scale with K. To show this, we will use $\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}}$ in (29) and the naive channel estimate $\boldsymbol{s}_{\ell_k}'$ in (20) based on non-orthogonal pilot instead of the MMSE estimate

$\breve{\boldsymbol{w}}_{\ell\ell_k}$ for the sake of exposition, since the former channel estimate essentially captures the pilot contamination term and the inter/intra-cell interference term. With $\boldsymbol{v}_{\ell_k} = \boldsymbol{s}'_{\ell_k}$ (ignoring normalization), the overall interference term in (29) for user ℓ_k in uplink is given by

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

$$
\sum_{(\ell',k')\neq(\ell,k)} \left| \left(\boldsymbol{w}_{\ell\ell_k} + \sum_{(j,m)\neq(\ell,k)} \boldsymbol{w}_{\ell_k j_m} + \frac{1}{\sqrt{\rho_p}} \boldsymbol{z}_{\ell_k} \right)^{\mathsf{H}} \boldsymbol{w}_{\ell_k\ell'_{k'}} \right|^2.
\tag{35}
$$

We investigate the asymptotic behaviors of three main terms in the above equation.

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

$$
\frac{\|\boldsymbol{w}_{\ell_k\ell'_{k'}}\|^2}{M} = \frac{\left| \boldsymbol{w}_{\ell\ell'_{k'}}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell'_{k'}}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell_k} \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell'_{k'}} \boldsymbol{w}_{\ell\ell'_{k'}} \right|^2}{M}
$$
$$
\xrightarrow{a.s.} \frac{r}{M^2} \mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}},
\tag{36}
$$

where $\boldsymbol{w}_{\ell_k\ell'_{k'}}$ is given in (15). The convergence follows from Lemma 5 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 $\boldsymbol{U}_{\ell\ell'_{k'}}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell_k} \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell'_{k'}} \xrightarrow{a.s.} \frac{r}{M}\boldsymbol{I}_r$ by Lemma 6 (or 8) for the random partial unitary correlation model (or the random partial Fourier model). Note that the latter convergence shows the role of spatial despreading in uplink, which is more crucial than the former (channel hardening) to obtain the capacity scaling results in this paper. Using the dominated convergence theorem, we have

$$
\frac{1}{M^2} \sum_{(\ell',k')\neq(\ell,k)} |\boldsymbol{w}_{\ell_k\ell'_{k'}}^{\mathsf{H}} \boldsymbol{w}_{\ell_k\ell'_{k'}}|^2 \xrightarrow{a.s.} \frac{LK}{M} \cdot \frac{r^2 \mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}^2}{M^3}
\tag{37}
$$

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 conventional massive MIMO because of finite K and the law of large number in the limit of M, this is not necessarily the case with $\limsup_{M\to\infty} \frac{K}{M} < \infty$. For the residual interference term after spatial despreading, $|\boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{w}_{\ell_k\ell'_{k'}}|^2 = |\boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell'_{k'}} \boldsymbol{w}_{\ell\ell'_{k'}}|^2, \forall(\ell',k') \neq (\ell,k)$, "common" angular components between the channel covariance matrices of any pair of users are relatively very small such that $\lim_{r\to\infty} \frac{\Delta_{\ell_k,\ell'_{k'}}}{r} = 0$, where $\Delta_{\ell_k,\ell'_{k'}}$ is the number of the similar angular components shared by users ℓ_k and $\ell'_{k'}$ at BS ℓ. Specifically, $\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell'_{k'}} \xrightarrow{a.s.} \frac{1}{\sqrt{M}} \boldsymbol{J}_r$ as a straightforward corollary of Lemma 6 (or 8), where $\boldsymbol{J}_r$ is the r-dimensional all-ones matrix, and $\frac{|\mathsf{h}_{\ell\ell_k}^{\mathsf{H}} \mathsf{h}_{\ell\ell'_{k'}}|^2}{r^2} \xrightarrow{a.s.} 0$

by Lemma 4, where $\mathsf{h}_{\ell\ell'_{k'}}$ is defined in (4). Similar to (37), we have

$$
\frac{1}{M^2} \sum_{(\ell',k')\neq(\ell,k)} |\boldsymbol{w}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{w}_{\ell_k\ell'_{k'}}|^2 \xrightarrow{a.s.}
$$
$$
\frac{LK}{M} \cdot \frac{r^2 \mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k} \mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}}{M^2} \cdot \frac{|\mathsf{h}_{\ell\ell_k}^{\mathsf{H}} \mathsf{h}_{\ell\ell'_{k'}}|^2}{r^2}
\tag{38}
$$

which also vanishes under Assumption 3, particularly, (11b) and (11c).

- Noncoherent pilot contamination term ($(j,m) \neq (\ell',k')$): Although the amount of these terms $|\boldsymbol{w}_{\ell_k j_m}^{\mathsf{H}} \boldsymbol{w}_{\ell_k\ell'_{k'}}|^2 = |\boldsymbol{w}_{\ell j_m}^{\mathsf{H}} \boldsymbol{U}_{\ell j_m}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell_k} \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{U}_{\ell\ell'_{k'}} \boldsymbol{w}_{\ell\ell'_{k'}}|^2$ is extremely large due to non-orthogonal pilot, the infinite sum of quickly vanishing terms eventually goes to zero under our channel models and assumptions.

Note that since the deterministic approximation of the normalized signal power $\frac{\|\boldsymbol{w}_{\ell\ell_k}\|^2}{M}$ as well as (37) and (38) depends on realizations of $\{\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\}$, the resulting SINR of user ℓ_k is conditional to random channel covariance matrices of all users in the network. Under the deterministic $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}$ assumption in (8), however, the conditional SINR converges to a deterministic limit by Lemma 5 for sufficiently large M, thus allowing the ergodic rate analysis in this work. Using the same argument as the above derivations, we have the following corollary of Theorems 2 and 3.

Theorem 4. For large M with $\limsup_{M\to\infty} \frac{K}{M} < \infty$ under Assumptions 1 and 3 with the non-orthogonal pilot scheme, the sum capacities of massive MIMO uplink and downlink asymptotically behave as (28) and (34), respectively. For downlink, the sum power P_{dl} is assumed to scale linearly with K.

Using the homogeneous network in (26) with $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k} = r$ and $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}} = \iota r$ for all (ℓ,ℓ',k) in the very strong correlation regime, the second scaling law in (3) immediately follows from the above result. For the consistency (or duality) with uplink, where the uplink per-user transmit power P_{ul} is fixed and the uplink total power automatically scales with K, we assumed that the downlink sum power P_{dl} also scales with K. Otherwise, the post-processing SNR $\frac{P_{\mathrm{dl}}}{K}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}$ will vanish as $K \to \infty$.

D. Interpretation of Main Results

Let us consider the homogeneous network in (26) with $\iota = 1$. The main results in Theorems 2 and 3 show the role of spatial despreading and spreading in strong spatial correlation regimes because the achievability proof is based on the corresponding low-dimensional channel estimation and combining/precoding. We have not invoked any explicit pilot decontamination technique such as covariance-aided coordination with non-overlapping angular supports [14], semi-blind estimation [13], or multicell precoding and combining [9], [12]. As a matter of fact, our assumptions on strong correlation regimes are rather restrictive or favorable in the sense

that each covariance matrix occupies a vanishing fraction of dimensions with respect to the total signal space dimensions M so that a sort of coordination based on covariances is implicitly achieved without information exchange among cells. Accordingly, our asymptotic results in this paper may be regarded as a direct consequence of the channel models and their assumptions. However, since we leverage the asymptotic sum-rate scaling law only as a tool to provide insight into the role of spatial spreading and despreading in terms of M, K, T_c, and r, it is important to see if our asymptotic results are translated into a good prediction in the performance behavior of finite systems with strong spatial correlation. To this end, the linear sum-rate scaling of our results with respect to either SNR or M (equivalently, K for fixed $\frac{K}{M} > 0$) in (25) will be shown valid through numerical examples in Figs. 3 and 4 for *moderate* values of r with respect to finite M, as the large system analysis in general works quite precisely even for rather small system dimensions [4], [5]. It is also pointed out that any advanced precoding technique like block diagonalization [15], [52] over channel covariances is not required to realize the sum-rate performance predicted by the asymptotic capacity scaling under our channel models. We have shown that single-cell linear precoding (combining) based on spatial spreading (despreading) is sufficient to achieve the optimal capacity scaling laws (3) in the limit of M, although multicell precoding can significantly outperform single-cell processing at finite M.

Theorem 4 indicates that *every user can achieve unlimited spectral efficiency* in multicell massive MIMO even when the number of users per cell scales linearly with the number of BS antennas. The result is only achievable under the very strong correlation regime, which stipulates a much larger value of M for r to be sufficiently large relative to the strong correlation regime, e.g., to guarantee the asymptotic orthogonality of $|\mathsf{h}_{\ell\ell_k}^{\mathsf{H}} \mathsf{h}_{\ell\ell_{k'}}|^2 / r^2$ in the RHS of (38). We also used the non-orthogonal pilot scheme whose training overhead is not a function of K and in general much less than T_c, no matter how large K is. This somewhat surprising argument may appear counterintuitive, since it has been presumed in the literature that each data stream requires at least one orthogonal pilot dimension within its own cell, in order to allow the receiver (uplink) or the transmitter (downlink) to separate the streams and provide an interference-free channel. Therefore, the common intuition suggests 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 (see (11a) and (11b)) so that α vanishes in the limit of M and the channel energy captured by $\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_{k'}}$ also grows slower than M (see (11c)), thus making non-orthogonal pilot based channel training feasible.

Finally it deserves to mention the assumption on the downlink sum power P_{dl} in Theorem 4. Due to limited downlink transmit power per cell in practice, the scaling of P_{dl} with

respect to K makes the downlink system power-limited for large K. Consequently, the second scaling law in (3) is more feasible in uplink rather than in downlink. Furthermore, the result does not mean that spatial multiplexing per cell is represented by $(1 - T_c^{-1})K$ since we only have $\log(\alpha\mathsf{SNR})$ with α vanishing in the limit of M.

IV. AN EXTENSION OF DETERMINISTIC EQUIVALENTS

As mentioned earlier, it is essential to extend Theorem 1 in [22], [24] to the case of (9a) and (10a), where the spectral norm of $\boldsymbol{\Lambda}_{\ell\ell_k}$ grows without bound as $M \to \infty$. In this section, we develop the following extension of deterministic equivalents based on Corollary 2 in Appendix A.

Theorem 5. Let

$$\boldsymbol{B}_N = \boldsymbol{X}_N \boldsymbol{X}_N^{\mathsf{H}} + \boldsymbol{A}_N \tag{39}$$

where

1) $N, N_1, \ldots, N_n$, n, and $M_0, \ldots, M_n$ grow with ratios $a_0 = N/n$, $a_i = N/N_i$, and $b_i = N/M_i$, respectively, such that $0 < \liminf_N a_i \le \limsup_N a_i < \infty$ and $0 \le \liminf_N b_i \le \limsup_N b_i \le 1, \forall i$, and $0 < \liminf_N \frac{b_i}{b_j} \le \limsup_N \frac{b_i}{b_j} < \infty, \forall i \ne j$;

2) $\boldsymbol{X}_N \in \mathbb{C}^{N \times n}$ has independent columns $\boldsymbol{x}_i = \boldsymbol{\Psi}_i \boldsymbol{y}_i$, where $\boldsymbol{y}_i \in \mathbb{C}^{N_i}$ has either i.i.d. entries of zero mean, variance $\frac{1}{N}$, and $4 + \epsilon$ moment of order $\mathcal{O}(\frac{1}{N^{2+\epsilon/2}})$, or 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 $\mathcal{O}(\frac{1}{n^4})$;

3) $\boldsymbol{\Psi}_i \in \mathbb{C}^{N \times N_i}$ are deterministic and the spectral norm of $\boldsymbol{\Theta}_i = \boldsymbol{\Psi}_i \boldsymbol{\Psi}_i^{\mathsf{H}}$ is not necessarily uniformly bounded with respect to N, but there exist $M_i \ge N, i = 1, \ldots, n$ satisfying

$$\limsup_{N \to \infty} \|b_i \boldsymbol{\Theta}_i\|_2 < \infty, \forall i;$$

4) $\boldsymbol{A}_N \in \mathbb{C}^{N \times N}$ is non-negative Hermitian;

5) and let $\boldsymbol{Q}_N \in \mathbb{C}^{N \times N}$ be non-negative Hermitian with not necessarily uniformly bounded spectral norm with respect to N, but there exist $M_0 \ge N$ satisfying $\limsup_{N \to \infty} \|b_0 \boldsymbol{Q}_N\|_2 < \infty$.

Define

$$m_N(z) \triangleq \frac{1}{N} \mathrm{tr}\, \boldsymbol{Q}_N \big(\boldsymbol{B}_N - z\boldsymbol{I}_N\big)^{-1}$$

and $\beta_i \triangleq \frac{M_i}{M_{\max}}$ with $M_{\max} = \min_{i=0,1,\ldots,n} M_i$.

For $z \in \mathbb{C} \setminus \mathbb{R}^+$, as $N \to \infty$ with ratios a_i, b_i, and $\beta_i, \forall i$, we have

$$\frac{1}{\beta_0} m_N(z) - \frac{1}{\beta_0 N} \mathrm{tr}\, \boldsymbol{Q}_N \boldsymbol{T}_N(z) \xrightarrow{a.s.} 0 \tag{40}$$

where

$$\boldsymbol{T}_N(z) = \left(\frac{1}{N} \sum_{j=1}^{n} \frac{\boldsymbol{\Theta}_j}{1 + e_{N,j}(z)} + \boldsymbol{A}_N - z\boldsymbol{I}_N \right)^{-1} \tag{41}$$

and $e_{N,1}(z), \ldots, e_{N,n}(z)$ form the unique functional solution of

$$e_{N,i}(z) = \frac{1}{\beta_i N} \text{tr} \; \Theta_i T_N(z) \tag{42}$$

which is the Stieltjes transform of a nonnegative finite measure on $\mathbb{R}^+$ and given by $e_{N,i}(z) = \lim_{t\to\infty} e_{N,i}^{(t)}(z)$ where $e_{N,j}^{(0)}(z) = -\frac{1}{z}$ and for $t \geq 1$

$$e_{N,i}^{(t)}(z) = \frac{1}{\beta_i N} \text{tr} \; \Theta_i \left(\frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j}{1 + e_{N,j}^{(t-1)}(z)} + A_N - z I_N \right)^{-1} \tag{43}$$

Proof: See Appendix C. ∎

It is clear that if all the related matrices have uniformly bounded spectral norm, the above result reduces to Theorem 1 in [22], [24]. The contribution of our result is that we relax the condition of uniformly bounded spectral norms into the more general conditions 1) and 3). Accordingly, the method of deterministic equivalents is still valid and hence we could obtain the capacity scaling law in the previous section.

It is pointed out that, similar to Theorem 5, we can naturally extend other results requiring random or deterministic matrices with uniformly bounded spectral norm in large random matrix theory. For instance, a generalization of [4, Thm. 2] (see also [24, Thm. 2]) is given next and will be used in this work.

Theorem 6. Let $\Omega_N \in \mathbb{C}^{N \times N}$ be non-negative Hermitian with not necessarily uniformly bounded spectral norm with respect to N, but there exist $M_c \geq N$ satisfying $\limsup_{N\to\infty} \| \frac{N}{M_c} \Omega_N \|_2 < \infty$, and also let $\vartheta \triangleq \frac{M_0 M_c}{M_{\max}^2}$ and $\vartheta_{jk} \triangleq \frac{M_j M_k}{M_{\max}^2}$. Under the same conditions as in Theorem 5, we have

$$\frac{1}{\vartheta N} \text{tr} \; Q_N (B_N - z I_N)^{-1} \Omega_N (B_N - z I_N)^{-1}$$
$$- \frac{1}{\vartheta N} \text{tr} \; Q_N T_N'(z) \xrightarrow[N\to\infty]{a.s.} 0 \tag{44}$$

where $T_N'(z)$ is defined as

$$T_N'(z) = T_N(z) \Omega_N T_N(z)$$
$$+ T_N(z) \frac{1}{N} \sum_{j=1}^{n} \frac{\Theta_j e_{N,j}'(z)}{(1 + e_{N,j}(z))^2} T_N(z)$$

with $e'(z) = [e_{N,1}'(z), \ldots, e_{N,n}'(z)]$ is given by

$$e'(z) = (I_n - J(z))^{-1} v(z)$$

where $J(z)$ and $v(z)$ are defined as

$$[J(z)]_{jk} = \frac{1}{\vartheta_{jk} N} \frac{\text{tr} \; \Theta_j T_N(z) \Theta_k T_N(z)}{N (1 + e_{N,j}(z))^2}, \forall j, k$$

$$[v(z)]_j = \frac{1}{\vartheta N} \text{tr} \; \Theta_j T_N(z) \Omega_N T_N(z), \forall j.$$

Finally, the above result directly applies to the random partial Fourier model, in which i.i.d. sequences are replaced with uncorrelated constant modulus sequences defined in Corollary 1.

V. PROOFS OF MAINS RESULTS

We first derive a deterministic equivalent of the SINR $\gamma_{\text{ul},\ell_k}^{\text{mmse}}$ of the low-dimensional MMSE combining in uplink. Our analysis is based on (14) to highlight the role of spatial despreading. Based on Theorems 5 and 6, the deterministic equivalent $\bar{\gamma}_{\text{ul},\ell_k}^{\text{mmse}}$ of the SINR $\gamma_{\text{ul},\ell_k}^{\text{mmse}}$ of the MMSE detector is given by the following result.

Lemma 2. Under Assumption 1 and the orthogonal pilot scheme in Subsection II-D1, we have $\gamma_{\text{ul},\ell_k}^{\text{mmse}} \xrightarrow[M\to\infty]{} \bar{\gamma}_{\text{ul},\ell_k}^{\text{mmse}}$, where $\bar{\gamma}_{\text{ul},\ell_k}^{\text{mmse}}$ is given by

$$\bar{\gamma}_{\text{ul},\ell_k}^{\text{mmse}} = \frac{\delta_{\ell_k}^2}{\frac{1}{P_{\text{ul}}M}\mu_{\ell_k} + \sum_{\ell'\neq\ell} \alpha_{\ell\ell_k}^2 |\nu_{\ell\ell_k'}|^2 + \frac{1}{M}\sum_{\ell'\neq\ell, k'\neq k} \mu_{\ell_k\ell_{k'}}} \tag{45}$$

with $\delta_{\ell_k} = \frac{1}{M}\text{tr} \; \Phi_{\ell\ell_k} T_{\ell_k}$, $\mu_{\ell_k} = \frac{1}{M}\text{tr} \; \Phi_{\ell\ell_k} T_{\ell_k}'$, $\nu_{\ell\ell_k'} = \frac{1}{r_{\ell\ell_k}M}\text{tr} \; \Lambda_{\ell\ell_k'} \Xi_{\ell\ell_k} \Lambda_{\ell\ell_k'} T_{\ell_k}$, $\mu_{\ell_k\ell_{k'}} = \frac{\text{tr}\Lambda_{\ell\ell_{k'}}}{M}\mu_{\ell_k}$, where

1) $T_{\ell_k} = T_N(\frac{1}{P_{\text{ul}}r_{\ell\ell_k}})$ are given by Theorem 5 with $N = r_{\ell\ell_k}$, $M_k = M, b_k = \frac{1}{M}$ (i.e., $\beta_k = 1$), $Q_N = r_{\ell\ell_k}\Phi_{\ell\ell_k}$, $\Theta_k = r_{\ell\ell_k}\Phi_{\ell\ell_k}$, $A_N = \frac{1}{r_{\ell\ell_k}}z_{\ell_k}$, and $e_k = \delta_{\ell_k}, \forall i$;
2) $T_{\ell_k}' = T_N'(\frac{1}{P_{\text{ul}}r_{\ell\ell_k}})$ is given by Theorem 6 with $\Omega_N = I_{r_{\ell\ell_k}}, \forall i$.

Proof: A high-level sketch of the proof is as follows. Given the random realization of $\{U_{\ell\ell_k'}\}$, the SINR $\gamma_{\text{ul},\ell_k}^{\text{mmse}}$ converges to a conditional random variable that does not depend on the particular realization of effective channels $\{w_{\ell\ell_k'}\}$ in the limit of M. Using Lemma 5, we can show that the conditional random variable in fact converges to a *deterministic* limit independent of the specific covariance realization under the distributions of $U_{\ell\ell_k'}$ in Section II-B. For the details, see Appendix D. ∎

Notice that $\bar{\gamma}_{\text{ul},\ell_k}^{\text{mmse}}$ is a limiting value that does not depend on the particular realization of small-scale fading channels $\{w_{\ell\ell_k'}\}$ in the limit of M, but conditional to the random realization of $\{U_{\ell\ell_k'}\}$ through $\Xi_{\ell\ell_k}$ defined in (19) and also through $\Phi_{\ell\ell_k}$.

Using the standard technique in the literature [4], [21]–[24], [33], [34] and some new tools herein, the above result can be naturally extended to other cases such as other lower bounds and receive algorithms like the multicell MMSE detector. We skip the details because our focus is the asymptotic capacity scaling of massive MIMO rather than deterministic approximation of different receive algorithms.

A. Proof of Theorem 1

We begin with the brief derivation of (24) using the standard arguments in [45], [46], [48], [53]. Given $s_\ell = [s_{\ell_1}, \ldots, s_{\ell_{K_\ell}}]$ in the training phase and conditioned on a realization of $\underline{U}$ on the channel second-order statistics of all users in the network, the individual rate $\mathcal{R}_{\text{ul},\ell_k}$ of user ℓ_k can be lower-bounded by

$$\mathcal{R}_{\text{ul},\ell_k} = I(x_{\ell_k}; y_\ell, s_\ell | \underline{U})$$
$$\geq h(x_{\ell_k}) - \mathbb{E}\left[\log\left(\pi e \; \mathbb{E}\left[|x_{\ell_k} - \xi v_{\ell_k}^{\mathsf{H}} y_{\ell_k}|^2 \big| \hat{w}_\ell, \underline{U}\right]\right)\right]$$

14

where we used the data processing inequality, the independence of x_{ℓ_k} and $(\hat{\boldsymbol{w}}_\ell, \underline{\boldsymbol{U}})$, and the fact that Gaussian distribution maximizes the conditional distribution for a given covariance.

The received signal vector $\boldsymbol{y}_{\ell_k}$ in (14) of user ℓ_k can be rewritten as

$$\boldsymbol{y}_{\ell_k} = \hat{\boldsymbol{w}}_{\ell\ell_k} x_{\ell_k} + \boldsymbol{n}_{\ell\ell_k} x_{\ell_k} + \sum_{(\ell',k') \neq (\ell,k)} \boldsymbol{w}_{\ell_k \ell'_{k'}} x_{\ell_k \ell'_{k'}} + \boldsymbol{z}_{\ell_k}.$$

Based on the fact that $\hat{\boldsymbol{w}}_{\ell\ell_k} x_{\ell_k}$ and $\boldsymbol{z}'_{\ell\ell_k} = \boldsymbol{n}_{\ell\ell_k} x_{\ell_k} + \sum_{(\ell',k') \neq (\ell,k)} \boldsymbol{w}_{\ell_k \ell'_{k'}} x_{\ell_k \ell'_{k'}} + \boldsymbol{z}_{\ell_k}$ are uncorrelated, the worst-case uncorrelated noise argument [45], [53] implies that the worst-case $\boldsymbol{z}'_{\ell\ell_k}$ is Gaussian with variance of $\mathbb{E}[\boldsymbol{z}'_{\ell\ell_k} \boldsymbol{z}'^{\mathsf{H}}_{\ell\ell_k}]$. Letting $\xi \boldsymbol{v}^{\mathsf{H}}_{\ell_k} \boldsymbol{y}_{\ell_k}$ equal to the linear MMSE estimate of x_{ℓ_k} given $\boldsymbol{v}^{\mathsf{H}}_{\ell_k} \boldsymbol{y}_{\ell_k}$ and $(\hat{\boldsymbol{w}}_\ell, \underline{\boldsymbol{U}})$ to minimize $\mathbb{E}[|x_{\ell_k} - \xi \boldsymbol{v}^{\mathsf{H}}_{\ell_k} \boldsymbol{y}_{\ell_k}|^2 | \hat{\boldsymbol{w}}_\ell, \underline{\boldsymbol{U}}]$ and using the orthogonality principle, for any $\boldsymbol{v}_{\ell_k}$, the achievable rate of user ℓ_k is lower-bounded by (24).

For the MMSE detector in (22), we now use the deterministic equivalent of the corresponding SINR in Lemma 2. Let $\varphi_{\mathsf{ul}} = P_{\mathsf{ul}} M < \infty$. As $M \to \infty$, we have $N = r_{\ell\ell_k}$, $M_k = M, b_k = \frac{1}{M}$ (i.e., $\beta_k = 1$), $\boldsymbol{Q}_N = r_{\ell\ell_k} \boldsymbol{\Phi}_{\ell\ell_k}$, $\boldsymbol{\Theta}_k = r_{\ell\ell_k} \boldsymbol{\Phi}_{\ell\ell_k}$, $\boldsymbol{A}_N = \frac{1}{r_{\ell\ell_k}} \boldsymbol{z}_{\ell_k}$, and $e_k = \delta_{\ell_k}, \forall i$. Based on Lemma 5, we can further obtain the following asymptotic equivalence

$$\boldsymbol{\Xi}_{\ell\ell_k} \simeq \left(\boldsymbol{\Lambda}_{\ell\ell_k} + \left(\sum_{\ell' \neq \ell} \frac{\mathrm{tr} \boldsymbol{\Lambda}_{\ell\ell_k}}{M} + \rho_p^{-1} \right) \boldsymbol{I}_{r_{\ell\ell_k}} \right)^{-1} \simeq \boldsymbol{\Lambda}_{\ell\ell_k}^{-1} \tag{46}$$

where the first step follows from Lemmas 6 and the second step is due to Assumption 2 (i.e., the eigenvalues of $\boldsymbol{\Lambda}_{\ell\ell_k}$ grow without bound as $M \to \infty$) with finite L and M.

From (41), we have

$$\boldsymbol{T}_{\ell_k}^{-1} = \frac{1}{r_{\ell\ell_k}} \sum_{j=1}^{K_\ell} \frac{r_{\ell\ell_j} \boldsymbol{\Phi}_{\ell\ell_j}}{1 + \delta_{\ell\ell_j}(\varphi_{\mathsf{ul}}^{-1})} + \frac{1}{r_{\ell\ell_k}} \boldsymbol{z}_{\ell_k} - \varphi_{\mathsf{ul}}^{-1} \boldsymbol{I}_{r_{\ell\ell_k}}$$

$$\overset{(a)}{\simeq} \frac{\boldsymbol{\Phi}_{\ell\ell_k}}{1 + \delta_{\ell_k}(\varphi_{\mathsf{ul}}^{-1})} + \frac{1}{r_{\ell\ell_k}} \boldsymbol{z}_{\ell_k} - \varphi_{\mathsf{ul}}^{-1} \boldsymbol{I}_{r_{\ell\ell_k}}$$

$$\overset{(b)}{\simeq} \frac{\boldsymbol{\Lambda}_{\ell\ell_k}}{1 + \delta_{\ell_k}(\varphi_{\mathsf{ul}}^{-1})} + \frac{1}{r_{\ell\ell_k}} \boldsymbol{\Lambda}_{\ell\ell_k}^{-1} - \varphi_{\mathsf{ul}}^{-1} \boldsymbol{I}_{r_{\ell\ell_k}}$$

$$\overset{(c)}{\simeq} \varphi_{\mathsf{ul}}^{-1} \boldsymbol{I}_{r_{\ell\ell_k}} \tag{47}$$

where (a) follows from the fact that $\boldsymbol{\Phi}_{\ell\ell_j} = \boldsymbol{\Lambda}_{\ell_j \ell_k} \boldsymbol{\Xi}_{\ell\ell_j} \boldsymbol{\Lambda}_{\ell_j \ell_k} \simeq 0$ as $\alpha_{\ell\ell_j} \to 0$ for $j \neq k$ by Lemma 6, where $\boldsymbol{\Lambda}_{\ell_j \ell_k}$ is defined in (19), (b) follows from (46), and in (c) we used $\delta_{\ell_k}(\varphi_{\mathsf{ul}}^{-1}) \simeq \frac{P_{\mathsf{ul}}}{r_{\ell\ell_k}} \mathrm{tr}^2 \boldsymbol{\Lambda}_{\ell\ell_k}$ through (43). Similarly, we can have

$$\boldsymbol{T}'_{\ell_k} = \boldsymbol{T}_{\ell_k} \boldsymbol{T}_{\ell_k} + \boldsymbol{T}_{\ell_k} \frac{1}{r_{\ell\ell_k}} \sum_{j=1}^{K_\ell} \frac{r_{\ell\ell_j} \boldsymbol{\Phi}_{\ell\ell_j} \delta'_{\ell\ell_j}(\frac{1}{P_{\mathsf{ul}} r_{\ell\ell_j}})}{(1 + \delta_{\ell\ell_j}(\frac{1}{P_{\mathsf{ul}} r_{\ell\ell_j}}))^2} \boldsymbol{T}_{\ell_k}$$

$$\simeq \varphi_{\mathsf{ul}}^2 \boldsymbol{I}_{r_{\ell\ell_k}}. \tag{48}$$

Substituting (46)–(48) into (45), (49) on the top of Page 16 immediately follows under Assumption 2.

The corresponding ergodic achievable rate $\mathcal{R}_{\mathsf{ul},\ell_k}$ of user ℓ_k is given by

$$\mathcal{R}_{\mathsf{ul},\ell_k} \geq \mathbb{E}\left[\log \left(1 + \gamma_{\mathsf{ul},\ell_k}^{\mathsf{mmse}} \right) \big| \underline{\boldsymbol{U}} \right] \tag{50}$$

$$\simeq \log \left(1 + \bar{\gamma}_{\mathsf{ul},\ell_k}^{\mathsf{mmse}} \right)$$

where the expectation is over $\{\boldsymbol{w}_{\ell\ell'_k}\}$, and we used the dominated convergence theorem. Although the RHS of (50) is conditional to the realization of $\underline{\boldsymbol{U}}$, $\mathcal{R}_{\mathsf{ul},\ell_k}$ eventually converges to the deterministic limit by Lemma 5. The prelog factor in (25) of scheduled users is simply given by finding κ' that maximizes multiplexing gain in the quadratic function $(1 - \frac{\kappa'}{T_c})\kappa'$ for cell ℓ. This completes the achievability.

Remark 2. One can make an important observation from (49) on the role of spatial despreading. The coherent pilot contamination term eventually approximated as a non-zero finite deterministic value $\nu_{\ell\ell'}$ multiplied by the ratio $\alpha_{\ell\ell_k}$ vanishes under the sublinear sparsity in Assumption 2. Due to its "non-coherent" nature (i.e., $\boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{h}_j$ terms with $i \neq j$), the interference term disappears much faster than pilot contamination through spatial despreading. For illustrative purposes, suppose that $r = \sqrt{M}$ in the homogeneous network. When $M = 100$, (49) shows that spatial despreading can suppress pilot contamination and interference in the uplink by a factor of $\zeta = 10$ in (13). Meanwhile, $r = 10$ is not large enough to see the channel hardening effect including asymptotic orthogonality of small-scale or fast fading channel components of different users by the law of large number. Therefore, spatial despreading plays the central role to eliminate pilot contamination and interference. This is even prominent when one cannot take advantage of channel hardening due to small $r_{\ell\ell'_k}$, as will be also shown through numerical results in Sec. VI.

B. Proof of Theorem 2

It suffices to consider the matched filter (MF) receiver to derive the asymptotic capacity scaling in (28), although the MF would perform inferior to the MMSE receiver at finite M. This is because the effect of spatial despreading remains essentially the same regardless of MF and MMSE. We accordingly set $\boldsymbol{v}_{\ell_k} = \check{\boldsymbol{w}}_{\ell\ell_k}$, where $\check{\boldsymbol{w}}_{\ell\ell_k}$ is given by (21), and denote the resulting SINR by $\gamma_{\mathsf{ul},\ell_k}^{\mathsf{MF}}$.

For the MF receiver, it is straightforward to show that

$$\frac{1}{M^2} \left| \mathbb{E}[\check{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{w}_{\ell\ell_k} | \underline{\boldsymbol{U}}] \right|^2 \overset{(a)}{\simeq} \left(\frac{1}{M} \mathrm{tr} \boldsymbol{\Phi}'_{\ell\ell_k} \right)^2$$

$$\frac{1}{M^2} \mathrm{var}[\check{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}} \boldsymbol{w}_{\ell\ell_k} | \underline{\boldsymbol{U}}] \simeq 0$$

where (a) follows from Corollary 2 and Lemma 4. In what follows, we first consider the random partial unitary model for $\boldsymbol{U}_{\ell\ell_k}$ and focus on the dominating components in the in-

$$\bar{\gamma}_{\mathrm{ul},\ell\ell_k}^{\mathrm{mmse}} \simeq \frac{\left(\frac{\varphi_{\mathrm{ul}}}{M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}\right)^2}{\frac{\varphi_{\mathrm{ul}}}{M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}+\sum_{\ell'\neq\ell}\alpha_{\ell\ell_k}^2\left(\frac{\varphi_{\mathrm{ul}}}{M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}\right)^2+\frac{1}{M}\sum_{\ell',k'}\frac{\varphi_{\mathrm{ul}}}{M^2}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}} \simeq \frac{\varphi_{\mathrm{ul}}}{M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k} = P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}. \tag{49}$$

terference power term $\sum_{(\ell',k')\neq(\ell,k)}\frac{1}{M^2}\mathbb{E}\big[|\check{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell_k\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\big]$ that incur the effect of pilot contamination. From

$$\mathbb{E}\big[|\check{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell_k\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\big] = \mathbb{E}\bigg[\Big|\big(\boldsymbol{w}_{\ell\ell_k} + \sum_{(j,m)\neq(\ell,k)}\boldsymbol{w}_{\ell kj_m} + $$
$$\rho_p^{-1/2}\boldsymbol{z}_{\ell_k}\big)^{\mathsf{H}}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{w}_{\ell\ell_k\ell'_{k'}}\Big|^2\Big|\underline{\boldsymbol{U}}\bigg]$$

we extract the pilot contamination component (i.e., $(j,m)=(\ell',k')$), which leads to

$$\mathbb{E}\big[|\boldsymbol{w}_{\ell_k\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{w}_{\ell_k\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\big]$$
$$= \mathbb{E}\big[|\boldsymbol{w}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell_k}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell'_{k'}}\boldsymbol{w}_{\ell\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\big]$$
$$\overset{(a)}{\simeq} \frac{1}{M^2}\mathbb{E}\Big[\big|\mathrm{tr}(\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k})\boldsymbol{w}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell'_{k'}}\boldsymbol{w}_{\ell\ell'_{k'}}\big|^2|\underline{\boldsymbol{U}}\Big]$$
$$\simeq \frac{1}{M^2}\mathbb{E}\Big[\big|\mathrm{tr}(\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k})\boldsymbol{w}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell'_{k'}}\big|^2\Big]$$
$$\overset{(b)}{\simeq} \frac{1}{M^2}\big(\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\big)^2\big(\mathrm{tr}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\big)^2$$
$$= r_{\ell\ell_k}^2\Big(\frac{1}{M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\frac{1}{r_{\ell\ell_k}}\mathrm{tr}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\Big)^2$$
$$= M^2\alpha_{\ell\ell_k}^2\psi_{\ell'_{k'}}^2, \tag{51}$$

where $\psi_{\ell'_{k'}} = \frac{1}{r_{\ell\ell_k}M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\mathrm{tr}\,\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}$. In (a), we used the independence of $\check{\boldsymbol{U}}_{\ell\ell_k}$ and $\boldsymbol{U}_{\ell\ell'_{k'}}$, and the direct consequence of Lemma 6 that $\boldsymbol{U}_{\ell\ell_k}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}} = \frac{\mathrm{tr}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}}{M}\boldsymbol{I}_M$, and (b) follows from the trace lemma. We can see that $0 < \liminf_{M\to\infty}\psi_{\ell'_{k'}}^2 \le \limsup_{M\to\infty}\psi_{\ell'_{k'}}^2 < \infty$ under Assumption 1.

Neglecting the interference components vanishing in the limit of M, under Assumptions 1 and 2, we have

$$\gamma_{\mathrm{ul},\ell_k}^{\mathrm{MF}} \xrightarrow[M\to\infty]{a.s.} \bar{\gamma}_{\mathrm{ul},\ell_k}^{\mathrm{MF}}$$

where

$$\bar{\gamma}_{\mathrm{ul},\ell_k}^{\mathrm{MF}} = \frac{\left(\frac{1}{M}\mathrm{tr}\boldsymbol{\Xi}'_{\ell\ell_k}\right)^2}{\frac{1}{P_{\mathrm{ul}}M}\frac{1}{M}\mathrm{tr}\boldsymbol{\Xi}'_{\ell\ell_k} + \sum_{(\ell',k')\neq(\ell,k)}\alpha_{\ell\ell_k}^2\psi_{\ell'_{k'}}^2}. \tag{52}$$

The above pilot contamination term $\sum_{(\ell',k')\neq(\ell,k)}\alpha_{\ell\ell_k}^2\psi_{\ell'_{k'}}^2$ goes to zero under Assumption 2 because K_ℓ, L, and $\psi_{\ell'_{k'}}$ are all finite. Similar to (46), we can then have a deterministic approximation of $\boldsymbol{\Xi}'_{\ell\ell_k}$ and eventually have

$$\bar{\gamma}_{\mathrm{ul},\ell_k}^{\mathrm{MF}} \simeq P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k}. \tag{53}$$

For the random partial Fourier model, it suffices to use a direct combination of Corollaries 1 and 2 and then Lemma 8 instead of Lemma 6.

For the converse proof, we use a simple cut-set upper bound [54] on the sum rate of the pilot-aided MIMO uplink, where a cut divides the BSs from the users. Inspired by

[3], the work in [8] provided the capacity upper bound of a pilot-aided single-cell frequency-division duplex (FDD) MIMO downlink system, in which the BS perfectly obtains CSIT through a delay-free and error-free feedback channel. Likewise, assuming that CSIR is perfectly acquired by the BS through a contamination-free and noise-free uplink training channel, the per-user (or per-link) achievable rate is upper-bounded by the full CSI capacity of SIMO channel with a single channel training cost over the coherence block T_c such that

$$\mathcal{R}_{\mathrm{ul},\ell_k} \le \Big(1 - \frac{1}{T_c}\Big)\mathbb{E}\Big[\log\big(1 + P_{\mathrm{ul}}\boldsymbol{h}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{h}_{\ell\ell_k}\big)\Big]$$
$$= \Big(1 - \frac{1}{T_c}\Big)\log P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k} + o(1). \tag{54}$$

Following the above footsteps and using the non-coherent communication argument in [3], one can easily show the same capacity scaling as (28).

C. Sketch of Achievability Proof of $\mathcal{R}_{ul,\ell_k}^{(3)}$

Following along the lines and arguments of Thoerems 1 and 2, and noticing (36), it is straightforward to show that $\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}}$ in (29) behaves as $(1 - T_c^{-1})\log P_{\mathrm{ul}}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell_k} + o(1)$. Since the first term $\mathcal{R}_{\mathrm{ul},\ell_k}^{\mathrm{ub}}$ in (30) is the upper bound of all three lower bounds in this paper as in (36), we only need to see if

$$\mathrm{var}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|\underline{\boldsymbol{U}}] = \mathbb{E}\big[|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\big] - \big|\mathbb{E}[\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|\underline{\boldsymbol{U}}]\big|^2$$
$$\xrightarrow[M\to\infty]{} 0. \tag{55}$$

With $\boldsymbol{v}_{\ell_k} = \check{\boldsymbol{w}}_{\ell\ell_k}$, the deterministic approximation of $\mathbb{E}\big[|\check{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell_k\ell'_{k'}}|^2|\underline{\boldsymbol{U}}\big]$ is given in (51). Following the same steps as (51) and using that

$$\boldsymbol{w}_{\ell_k\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{w}_{\ell_k\ell'_{k'}}$$
$$= \boldsymbol{w}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell_k}\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell'_{k'}}\boldsymbol{w}_{\ell\ell'_{k'}}$$
$$\simeq \frac{1}{M}\mathrm{tr}\big(\boldsymbol{\Xi}'_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\big)\boldsymbol{w}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell'_{k'}}$$

where we used Lemma 6 (or 8) for the random partial unitary model (or partial Fourier model), we have $\check{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{w}_{\ell\ell_k\ell'_{k'}} \xrightarrow{a.s.} M\alpha_{\ell\ell_k}\psi_{\ell'_{k'}}$. Applying the continuous mapping theorem and the dominated convergence theorem gives rise to (55). This concludes that $\mathcal{R}_{\mathrm{ul},\ell_k}^{(3)}$ achieves the scaling law in (28).

VI. 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
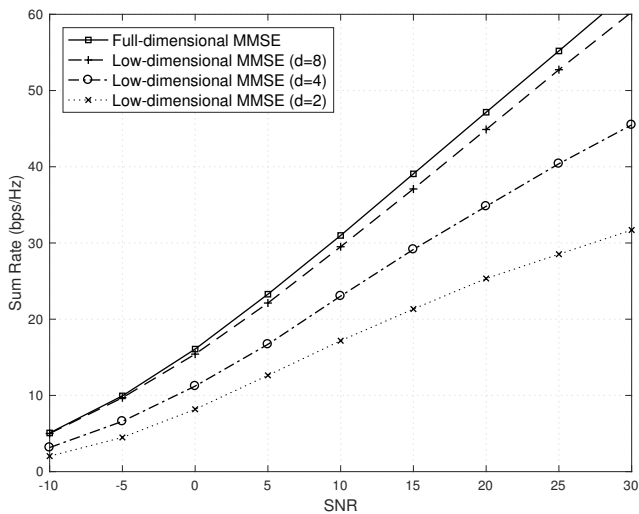
Fig. 2. Sum-rate curves of different precoding schemes versus SNR (dB) in MIMO downlink, where $M = 100, L = 4, K = 5, r = 8, T_c = 500$. The random partial Fourier model in (7) was used for channel covariance matrices. For low-dimensional MMSE precoding, we used (23).

geometry of users is assumed such that we normalize channel covariance matrices to satisfy $\mathrm{tr}\boldsymbol{R}_{\ell\ell'_k} = M$ for all (ℓ, ℓ', k) and $r_{\ell\ell_k} = r$ for all (ℓ, k). We assume that the number $r_{\ell\ell'_k}$ 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 depends on the distance between sender and receiver. The larger the distance, the smaller multipath components survive after several specular reflections and diffusion. We used pilot power gap (boosting) of $\varrho_p = 2$ (i.e., 3 dB). The random partial Fourier correlation model in Sec. II-B is only considered. Under the random partial unitary correlation model, the sum-rate performance shows degradation compared to the Fourier model, since the eigenmodes of channel covariance matrices of different users are not orthogonal but linearly independent of each other. In what follows, we evaluate all uplink and downlink lower bounds in this paper for finite M and SNR to verify the asymptotic sum-rate scaling results.

In Fig. 2, we investigate the lower bound on the per-cell achievable rate in (32) of different MMSE precoding schemes in downlink, where $r = 8$. While the 'full-dimensional MMSE' represents the conventional M-dimensional MMSE channel estimation and precoding scheme based on $\bar{\boldsymbol{s}}_{\ell_k}$ in (17) without spatial (de)spreading, the 'low-dimensional MMSE' is given by (19) and (23), where d indicates the number of Fourier coefficients used out of r ones under the Fourier correlation model for spatial (de)spreading. The 8-dimensional ($d = 8$) channel estimation and precoding are shown to cause only graceful performance degradation[9], compared to the 100-

[9] In a dense user scenario, e.g., $K = 20$, the performance degradation will be more significant especially at high SNR.

dimensional vector operation. However, significant performance loss is observed when we only utilize $d(< r)$ Fourier coefficients uniform randomly chosen from the r elements of angular support that have equal path gain. This implies that one should make use of as many dominant elements of angular support as possible for spatial (de)spreading in practical systems to realize the capacity scaling.

Fig. 3 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. $\mathcal{R}_{\mathrm{ul},\ell_k}^{(1)}$ ('lower bound 1') in (24) and $\mathcal{R}_{\mathrm{ul},\ell_k}^{(3)}$ ('lower bound 3') in (30) turn out to yield a linear growth of the ergodic sum rate with SNR (dB) for both orthogonal and non-orthogonal pilot schemes. In contrast, $\mathcal{R}_{\mathrm{ul},\ell_k}^{(2)}$ ('lower bound 2') in (27) does not show linear growth due to lack of hardening of the effective channels $\boldsymbol{w}_{\ell\ell_k}$, whose dimension is r. The coherent lower bound suffers from channel estimation error represented by the parallel shift of capacity versus SNR curves, also known as power offset. It can be further seen that almost the same multiplexing gain as (25) 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 $\boldsymbol{w}_{\ell\ell_k}$ than spatial despreading of $\boldsymbol{U}_{\ell\ell_k}$. Hence, the larger r turns out rather beneficial in the low SNR regime.

Fig. 4 verifies the scalability of the sum-rate scaling of MIMO downlink in (33) 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 both K and r as well as M, although r is not much smaller than M. Furthermore, this linear growth is observed even for $M < KL$.

In Fig. 5, we see that the non-orthogonal pilot scheme in Subsection II-D2 that consumes only a single channel use per network per coherence block can be beneficial in terms of the per-cell sum rate in uplink, even though the network is dense such that $L = 7$ and $K = 20$ for $M = 100$. For small coherence block size of $T_c = 50$, the non-orthogonal pilot helps only in fading channels with strong spatial correlation. As mentioned earlier, $\mathcal{R}_{\mathrm{ul},\ell_k}^{(3)}$ has a shortcoming of sum-rate underestimation for this small T_c unless r is sufficiently small. Fig. 6 shows the impact of the coherence block size T_c in the strong spatial correlation cases of $r = 4, 8$. For $r = 4$, the non-orthogonal pilot scheme turns out useful in a small to moderate range of T_c. In addition, the coherent lower bound $\mathcal{R}_{\mathrm{ul},\ell_k}^{(1)}$ widely used in uplink suffers from considerable channel estimation error due to imperfect CSIR in case of non-orthogonal pilot. A general non-orthogonal pilot scheme based on Welch bound equality frames is expected to outperform the naive scheme in (20).
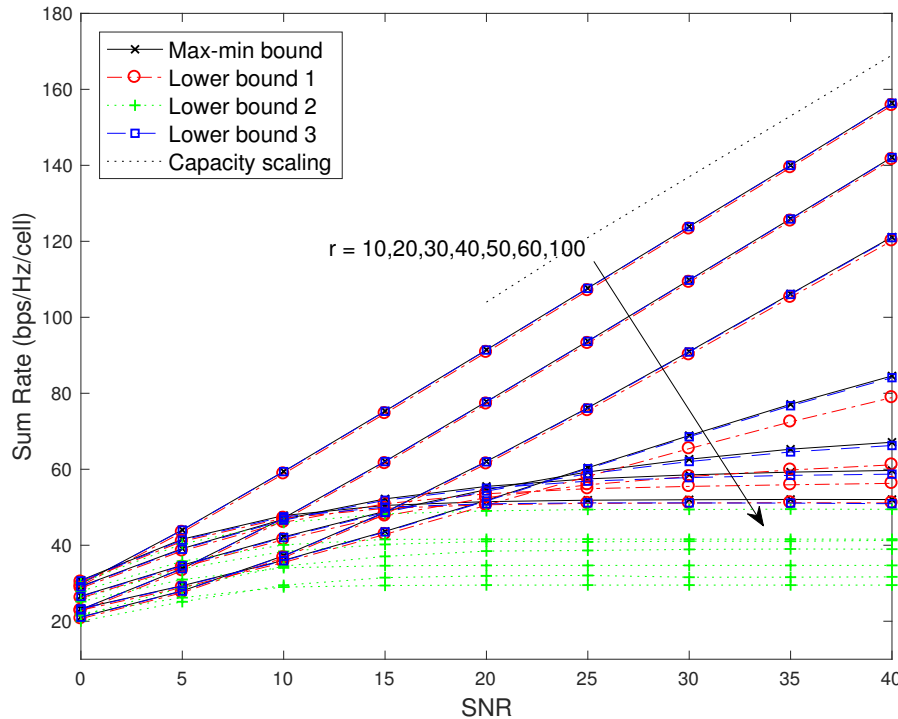
Fig. 3. The impact of spatial despreading on sum-rate scaling of MIMO uplink with respect to the sparsity of angular support (r), where $M = 200$, $L = 4, K = 10, T_c = 500$. The max-min bound is given by (29). While 'Lower bound 1' is the coherent bound in (24), 'Lower bound 2' and 'Lower bound 3' indicate the noncoherent ones in (27) and (30), respectively. While $r = 100, 60, 10, 20, 50, 30, 40$ from top to bottom for Lower bound 2, $r = 10, 20, 30, 40, 50, 60, 100$ for the rest. The asymptotic capacity scaling is given by (25) with $o(1) = 0$.
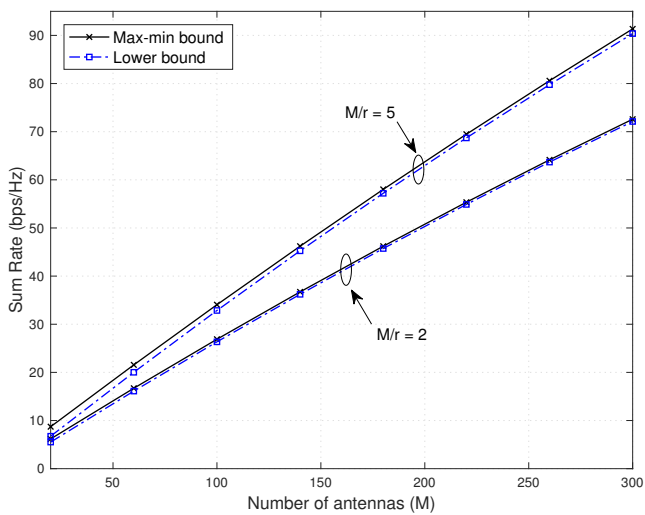
Fig. 4. The impact of spatial spreading on sum-rate scaling with respect to M in MIMO downlink with the ratios $\frac{M}{K} = 5$ and $\zeta = \frac{M}{r}$ fixed, where the 'Lower bound' is given by (32).
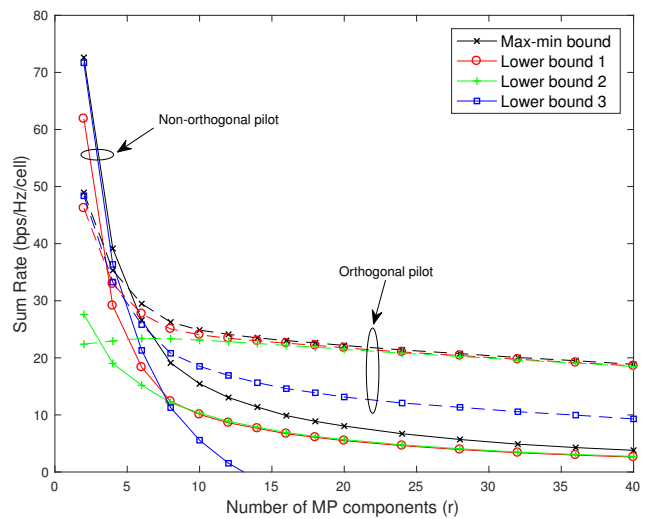
Fig. 5. The performance of the non-orthogonal pilot scheme with respect to r for small coherence block size $T_c = 50$, where $M = 100, L = 7$, $K = 20$, and SNR= 20 dB.

VII. CONCLUSION

Channel hardening has been traditionally considered as an essential source of massive MIMO gain. This is not necessar-ily 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 laws in this work. Although the
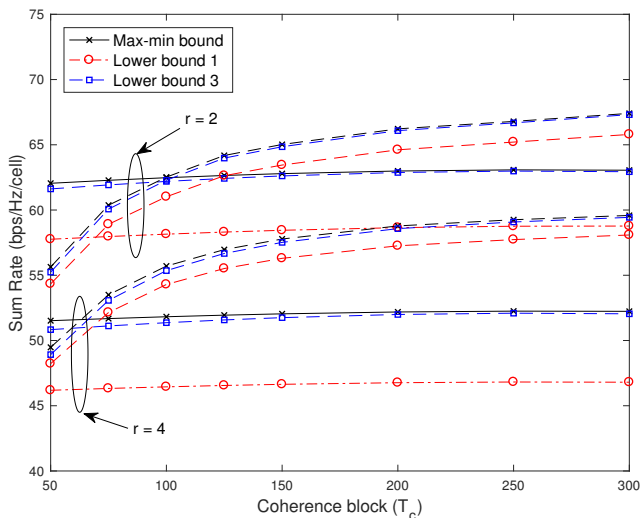
Fig. 6. The impact of the coherence block size T_c on both orthogonal and non-orthogonal pilot schemes, where $M = 100$, $L = 7$, $K = 10$, and SNR$= 20$ dB. While the solid lines indicate the non-orthogonal pilot scheme, the dashed lines represent the orthogonal pilot sequences.

exact capacity scaling of massive MIMO 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. Some important implications of this work can be summarized as follows.

- Once the multiple antenna channels in the network satisfy a certain sparsity level of angular support, one can incorporate spatial (de)spreading into system designs such as non-orthogonal pilot and low-dimensional channel estimation and precoding (or combining) instead of orthogonal pilot and full-dimensional MMSE processing to realize the very promising sum-rate performance of massive MIMO in this paper.
- The potential of serving as many users as the number of large-scale BS antennas through interference-free links in strong spatial correlation regimes would be of importance particularly in uplink, where the per-cell sum power scales linearly with the number of users. Our results have shown that one may *simultaneously* provide "super-massive connectivity" and very high per-user data rate for the next-generation wireless network with very high carrier frequencies, where meeting the uplink data requirement is sometimes more challenging than downlink.
- The three lower bounds on the achievable rate of multi-cell MU-MIMO that we considered show mutually complementary behaviors. In order to better understand and predict the performance of massive MIMO, therefore, one should carefully select a proper bound, depending on the channel and system parameters such as sparsity of angular support, coherence block size, and orthogonal/non-orthogonal pilot sequences.

- Finally, our large system analysis holds true in finite coherence block and even when the number of users per cell scales linearly with the number of BS antennas. Meanwhile, large system analysis in the literature implicitly has assumed that coherence block grows unbounded and the number of users is finite or grows slower than the number of antennas.

For further study, we are investigating the performance behavior of low-dimensional precoding/combining based on spatial spreading/despreading in more realistic spatial correlation models based on arbitrary angle of arrivals of users, whose subspaces are not mutually orthogonal but linearly independent of each other with high probability. The effect of spatial (de)spreading has an important implication in practical massive MIMO systems such as 5G NR, which is based on downlink training and feedback like FDD settings even if the NR system is mainly on TDD spectrum. For both below 6GHz and mm-Wave bands, 5G NR massive MIMO presumes hybrid beamforming due to power consumption and implementation cost. Since spatial (d)spreading corresponds to wideband/analog beamforming matched to space-frequency channel covariances of users in practical systems, our results strongly suggest that given r dominant eigen-directions or DFT columns of users, a massive MIMO system should be designed to use as many wideband beamforming dimensions (i.e., $d \leq r$ RF chains) as possible to realize the very promising spectral efficiency in this work, as shown in Fig. 2. In other words, even if an mm-Wave channel has asymptotically unlimited capacity, no one can leverage such promising benefit with insufficient wideband beamforming dimensions. Although having d RF chains entails the cost of high power consumption and implementation cost at the BS, it may be justified from a viewpoint of system performance enhancement. We plan to address performance benefit of having spatial spreading matched to space-frequency channel covariances of users in a follow-up paper under the 5G NR channel model [55].

APPENDIX A
USEFUL LEMMAS

We collect here some known or new lemmas to be used throughout this work. Silverstein and Bai derived the following well-known result in large random matrix theory, which is pivotal in the method of deterministic equivalents.

Lemma 3. (Trace Lemma [21, Lem. 2.7], [23, Thm. 3.4]) 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 $\mathcal{O}(\frac{1}{n^4})$. Let $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$, with $\boldsymbol{A}_n \in \mathbb{C}^{n \times n}$, be a series of matrices with uniformly bounded spectral norm with respect to n, independent of $\boldsymbol{x}_n$. Then

$$\boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{A}_n \boldsymbol{x}_n - \frac{1}{n} \operatorname{tr} \boldsymbol{A}_n \xrightarrow{a.s.} 0 \qquad (56)$$

as $n \to \infty$.

In this paper, we provide a direct generalization of the above result for the two cases: one is the case where $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$, are infinite sequences whose entries are *uncorrelated* with each other and not necessarily identically distributed; the other is that the spectral norm of $\boldsymbol{A}_n$ is *not necessarily uniformly bounded*.

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 $i = 1, \ldots, n$. With $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$, in Lemma 3, we have

$$\boldsymbol{y}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \operatorname{tr} \boldsymbol{A}_n \xrightarrow{a.s.} 0 \qquad (57)$$

as $n \to \infty$.

Proof: It follows from the Markov inequality that for any $\epsilon > 0$

$$\Pr\left(\left| \boldsymbol{y}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \operatorname{tr} \boldsymbol{A}_n \right| > \epsilon \right) \leq \frac{\mathbb{E}\left[\left| \boldsymbol{y}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \operatorname{tr} \boldsymbol{A}_n \right|^4 \right]}{\epsilon^4}. \qquad (58)$$

In order to show that $\mathbb{E}\left[\left| \boldsymbol{y}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \operatorname{tr} \boldsymbol{A}_n \right|^4 \right] \leq \frac{C_1}{n^2}$, where $C_1 > 0$ is a constant independent of n and $\boldsymbol{A}_n$, following the footsteps in [33, Lem. 3.1] and [23, Thm. 3.4], we begin with

$$\mathbb{E}\left[\left| \boldsymbol{y}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \operatorname{tr} \boldsymbol{A}_n \right|^4 \right]$$

$$\leq 8 \left(\mathbb{E}\left[\sum_i A_{n,ii}(|y_{n,i}|^2 - n^{-1}) \right]^4 + \mathbb{E}\left[\sum_{i \neq j} A_{n,ij} y_{n,i}^* y_{n,j} \right]^4 \right)$$

$$= 8 \left(\mathbb{E}\left[\sum_{i \neq j} A_{n,ij} y_{n,i}^* y_{n,j} \right]^4 \right)$$

$$= \mathcal{O}\left(\frac{1}{n^2} \right) \qquad (59)$$

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 the eighth order moment of $\frac{1}{n^4}$ 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 $\mathbb{E}[y_i^* y_j] = 0, \forall i \neq j$. Plugging (59) into (58), we have

$$\sum_{n=1}^{\infty} \Pr\left(\left| \boldsymbol{y}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{y}_n - \frac{1}{n} \operatorname{tr} \boldsymbol{A}_n \right| > \epsilon \right) < \infty.$$

The almost sure convergence in (57) immediately follows from the first Borel-Cantelli lemma [56]. ∎

Although the above constant modulus requirement of $|y_{n,i}|^2$ is a strong condition, some popular structured random matrices like random partial Fourier and Hadamard matrices [57], [58] satisfy the condition.

Corollary 2. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ be as in Lemma 3. Let $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$, with $\boldsymbol{A}_n \in \mathbb{C}^{n \times n}$, be a series of matrices independent of $\boldsymbol{x}_n$. If there exists $m_n \geq n$ such that

$$\limsup_{n \to \infty} \frac{n}{m_n} \| \boldsymbol{A}_n \|_2 < \infty \qquad (60)$$

then

$$\frac{n}{m_n} \boldsymbol{x}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{x}_n - \frac{1}{m_n} \operatorname{tr} \boldsymbol{A}_n \xrightarrow{a.s.} 0 \qquad (61)$$

as $n \to \infty$.

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 (60), 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 Lemma 3, for any $\epsilon > 0$, there exists n_0 such that for all $n \geq n_0$, $\left| \boldsymbol{x}_n^\mathsf{H} \boldsymbol{B}_n \boldsymbol{x}_n - \frac{1}{n} \operatorname{tr} \boldsymbol{B}_n \right| < \epsilon$. ∎

We can also combine the above two corollaries so as to make the trace lemma still valid in the more general case where $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$, are uncorrelated random vectors with constant modulus and/or the spectral norms of $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$, are not necessarily uniformly bounded. The following lemma is a generalization of [23, Thm. 3.7].

Lemma 4. Let $\boldsymbol{x}_n \in \mathbb{C}^m$ and $\boldsymbol{y}_n \in \mathbb{C}^n$, be two independent random vectors whose entries satisfy the conditions in either Lemma 3 or Corollary 1, with a constant ratio $\frac{m}{n}$ as $p \to \infty$, where $p = \min(m, n)$. Also let $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, independent of $\boldsymbol{x}_n$ and $\boldsymbol{y}_n$. As p goes to infinity, we have

$$\boldsymbol{x}^\mathsf{H} \boldsymbol{A} \, \boldsymbol{y} \xrightarrow{a.s.} 0. \qquad (62)$$

Proof: Following the line of arguments in [33, Lem. 3.1] and [23, Thm. 3.7], it is not difficult to show that $\mathbb{E}\left[|\boldsymbol{x}^\mathsf{H} \boldsymbol{A} \, \boldsymbol{y}|^8 \right] < \frac{c}{p^4}$ for some $c > 0$ independent of p. Then, (62) follows from the Markov inequality, Borel-Cantelli lemma, and the Tonelli theorem [56], [59]. ∎

The above results require that a random vector either have i.i.d. entries or satisfy the conditions in Lemma 3. Let us consider the third type of random vector $\boldsymbol{u}_n$, which is a column of an $n \times n$ Haar distributed unitary matrix. Even though the random unit vector $\boldsymbol{u}_n$ does not have i.i.d. entries, it can be rewritten as $\boldsymbol{u}_n = \frac{\boldsymbol{x}_n}{\|\boldsymbol{x}_n\|}$, where $\boldsymbol{x}_n$ has i.i.d. zero-mean Gaussian distributed entries with unit variance and $\|\boldsymbol{x}_n\|$ is χ^2 distributed with $2n$ degrees of freedom. It can then be shown [60, Sec. VI] that $\boldsymbol{u}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{u}_n$ and $\frac{1}{\|\boldsymbol{x}_n\|^2} \boldsymbol{x}_n^\mathsf{H} \boldsymbol{A}_n \boldsymbol{x}_n$ have the same asymptotic behavior, where $\boldsymbol{A}_n$ has uniformly bounded spectral norm with respect to n, independent of $\boldsymbol{u}_n$. As a consequence, the above trace lemma and its variants directly apply to the random partial unitary correlation model in II-B.

The following lemma is about the convergence of a mixture of two independent random measures with the distributions considered in this paper.

Lemma 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 either Lemma 3 or Corollary 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 $n \to \infty$

$$\boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{\Lambda}_n \boldsymbol{x}_n - \frac{1}{n} \operatorname{tr}\mathring{\boldsymbol{\Lambda}}_n \xrightarrow{a.s.} 0. \tag{63}$$

Proof: It suffices to show that

$$
\begin{aligned}
&\Pr\left(\lim_n \boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{\Lambda}_n \boldsymbol{x}_n = \frac{1}{n} \operatorname{tr}\mathring{\boldsymbol{\Lambda}}_n\right) \\
&\geq \Pr\left(\lim_n \boldsymbol{x}_n^{\mathsf{H}} \boldsymbol{\Lambda}_n \boldsymbol{x}_n = \frac{1}{n} \operatorname{tr}\mathring{\boldsymbol{\Lambda}}_n, \lim_n \boldsymbol{x}_n^{\mathsf{H}} \mathring{\boldsymbol{\Lambda}}_n \boldsymbol{x}_n = \frac{1}{n} \operatorname{tr}\mathring{\boldsymbol{\Lambda}}_n\right) \\
&= \Pr\left(\lim_n \boldsymbol{x}_n^{\mathsf{H}}(\boldsymbol{\Lambda}_n - \mathring{\boldsymbol{\Lambda}}_n)\boldsymbol{x}_n = 0, \lim_n \boldsymbol{x}_n^{\mathsf{H}} \mathring{\boldsymbol{\Lambda}}_n \boldsymbol{x}_n = \frac{1}{n} \operatorname{tr}\mathring{\boldsymbol{\Lambda}}_n\right) \\
&\overset{(a)}{\geq} \Pr\left(\lim_n \boldsymbol{\Lambda}_n - \mathring{\boldsymbol{\Lambda}}_n = 0, \lim_n \boldsymbol{x}_n^{\mathsf{H}} \mathring{\boldsymbol{\Lambda}}_n \boldsymbol{x}_n = \frac{1}{n} \operatorname{tr}\mathring{\boldsymbol{\Lambda}}_n\right) = 1
\end{aligned}
\tag{64}
$$

where (a) follows from the continuous mapping theorem since $\boldsymbol{x}_n^{\mathsf{H}}(\boldsymbol{\Lambda}_n - \mathring{\boldsymbol{\Lambda}}_n)\boldsymbol{x}_n$ is a linear function of $\operatorname{vec}(\boldsymbol{\Lambda}_n - \mathring{\boldsymbol{\Lambda}}_n)$ and also $\boldsymbol{x}_n$ and $\boldsymbol{\Lambda}_n - \mathring{\boldsymbol{\Lambda}}_n$ are independent, and the second event in $\Pr(\cdot)$ in the RHS of (64) is almost sure by Lemma 3 and Corollary 1. Then the almost sure convergence in (63) immediately follows due to the fact that the intersection of two almost sure events is also almost sure. ∎

Notice that the continuous mapping theorem does not directly apply to the above lemma since $\boldsymbol{x}_n$ does not converge to a random vector. In what follows, we derive some matrix analogs of the trace lemma for random matrices $\boldsymbol{X}$ rather than random vectors $\boldsymbol{x}$. For the first results, we use the random partial unitary model in II-B.

Lemma 6. Let $\boldsymbol{U} \in \mathbb{C}^{p \times m}$ and $\boldsymbol{V} \in \mathbb{C}^{p \times n}$ be m and n columns of two independent $p \times p$ Haar distributed unitary matrices, respectively, where $p > \max(m,n)$. Also let $\boldsymbol{D} \in \mathbb{C}^{n \times n}$ be an arbitrary matrix with $\limsup_{p \to \infty} \|\boldsymbol{D}\|_2 < \infty$, independent of $\boldsymbol{U}$ and $\boldsymbol{V}$. Then, as p grows without bound, we have

$$\boldsymbol{U}^{\mathsf{H}} \boldsymbol{V} \boldsymbol{D} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{U} - \frac{\operatorname{tr}\boldsymbol{D}}{p} \boldsymbol{I}_m \xrightarrow{a.s.} 0. \tag{65}$$

Proof: We denote the ith columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ as $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$, respectively. Since $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are taken from two independent Haar distributed unitary matrices, respectively, we can apply Lemmas 3 and 4, as mentioned earlier. The matrix multiplication $\boldsymbol{U}^{\mathsf{H}} \boldsymbol{V} \boldsymbol{D} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{U}$ unfolds as

$$[\boldsymbol{U}^{\mathsf{H}} \boldsymbol{V} \boldsymbol{D} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{U}]_{k,\ell} = \sum_{i=1}^n \sum_{j=1}^n d_{ij} \boldsymbol{u}_k^{\mathsf{H}} \boldsymbol{v}_i \boldsymbol{v}_j^{\mathsf{H}} \boldsymbol{u}_\ell, \ \forall (k,\ell).$$

The diagonal and off-diagonal entries can be written respectively as

$$
\begin{aligned}
&\sum_{i,j} d_{ij} \boldsymbol{u}_k^{\mathsf{H}} \boldsymbol{v}_i \boldsymbol{v}_j^{\mathsf{H}} \boldsymbol{u}_k = \sum_i d_{ii} |\boldsymbol{v}_i^{\mathsf{H}} \boldsymbol{u}_k|^2 + \boldsymbol{u}_k^{\mathsf{H}}\left(\sum_{i \neq j} d_{ij} \boldsymbol{v}_i \boldsymbol{v}_j^{\mathsf{H}}\right)\boldsymbol{u}_k \\
&\sum_{i,j} d_{ij} \boldsymbol{u}_k^{\mathsf{H}} \boldsymbol{v}_i \boldsymbol{v}_j^{\mathsf{H}} \boldsymbol{u}_\ell = \\
&\quad \boldsymbol{u}_k^{\mathsf{H}}\left(\sum_i d_{ii} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{H}}\right)\boldsymbol{u}_\ell + \boldsymbol{u}_k^{\mathsf{H}}\left(\sum_{i \neq j} d_{ij} \boldsymbol{v}_i \boldsymbol{v}_j^{\mathsf{H}}\right)\boldsymbol{u}_\ell.
\end{aligned}
\tag{66}
$$

For $k = \ell$, we have

$$
\begin{aligned}
&\sum_i d_{ii} |\boldsymbol{v}_i^{\mathsf{H}} \boldsymbol{u}_k|^2 + \boldsymbol{u}_k^{\mathsf{H}}\left(\sum_{i \neq j} d_{ij} \boldsymbol{v}_i \boldsymbol{v}_j^{\mathsf{H}}\right)\boldsymbol{u}_k \\
&\overset{(a)}{\simeq} \frac{\operatorname{tr}\sum_i d_{ii} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{H}}}{p} + \frac{\operatorname{tr}\sum_{i \neq j} d_{ij} \boldsymbol{v}_i \boldsymbol{v}_j^{\mathsf{H}}}{p} \\
&= \frac{\sum_i d_{ii} \boldsymbol{v}_i^{\mathsf{H}} \boldsymbol{v}_i}{p} + \frac{\sum_{i \neq j} d_{ij} \boldsymbol{v}_j^{\mathsf{H}} \boldsymbol{v}_i}{p} \\
&= \frac{\operatorname{tr}\boldsymbol{D}}{p}
\end{aligned}
$$

where (a) follows from Lemma 3. For $k \neq \ell$, we obtain by Lemma 4

$$\boldsymbol{u}_k^{\mathsf{H}}\left(\sum_i d_{ii} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{H}}\right)\boldsymbol{u}_\ell + \boldsymbol{u}_k^{\mathsf{H}}\left(\sum_{i \neq j} d_{ij} \boldsymbol{v}_i \boldsymbol{v}_j^{\mathsf{H}}\right)\boldsymbol{u}_\ell \simeq 0$$

which yields (71). ∎

For finite dimensional matrices $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{D}$, we can have the following similar result.

Lemma 7. Let $\boldsymbol{U} \in \mathbb{C}^{p \times m}$ and $\boldsymbol{V} \in \mathbb{C}^{p \times n}$ be m and n columns of two independent $p \times p$ Haar distributed unitary matrices, respectively, where $p > \max(m,n)$. Also let $\boldsymbol{D} \in \mathbb{C}^{n \times n}$ be a random matrix, independent of $\boldsymbol{U}$ and $\boldsymbol{V}$. Then we have

$$\mathbb{E}\left[\boldsymbol{U}^{\mathsf{H}} \boldsymbol{V} \boldsymbol{D} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{U}\right] = \frac{\operatorname{tr}\mathbb{E}[\boldsymbol{D}]}{p} \boldsymbol{I}_m. \tag{67}$$

Proof: Notice that $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are two independent random unit vectors uniformly distributed on the $2p$-dimensional complex unit sphere such that $\boldsymbol{u}_i = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}$ with covariance $\frac{1}{p}\boldsymbol{I}_p$, where $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_p)$, and so $\boldsymbol{v}_i$ is, and that $|\boldsymbol{v}_i^{\mathsf{H}} \boldsymbol{u}_k|^2$ is a Beta-distributed random variable with parameter $(1, p-1)$. By the independence assumption on $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{D}$ and the law of total expectation, the proof of (67) immediately follows similar to Lemma 6. ∎

For the random partial Fourier model for $\boldsymbol{U}$ in II-B, we have the following two results analogous to the above lemmas.

Lemma 8. 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 $\limsup_{p \to \infty} \|\boldsymbol{D}\|_2 < \infty$, independent of $\boldsymbol{U}$ and $\boldsymbol{V}$. Then we have

$$\boldsymbol{U}^{\mathsf{H}} \boldsymbol{V} \boldsymbol{D} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{U} - \frac{\operatorname{tr}\boldsymbol{D}}{p} \boldsymbol{I}_m \xrightarrow{a.s.} 0. \tag{68}$$

Proof: The proof immediately follows from the same footsteps in the proof of Lemma 6 using Corollary 1 and Lemma 4. ∎

Note that sampling *without* replacement in the random partial Fourier model guarantees the orthonormality between columns so that the above result holds for the channel covariance model.

Lemma 9. Given the same $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{D}$ as in Lemma 8, we have

$$\mathbb{E}\left[\boldsymbol{U}^{\mathsf{H}} \boldsymbol{V} \boldsymbol{D} \boldsymbol{V}^{\mathsf{H}} \boldsymbol{U}\right] \longrightarrow \frac{\operatorname{tr}\mathbb{E}[\boldsymbol{D}]}{p} \boldsymbol{I}_m. \tag{69}$$

Proof: Following the same reasoning and procedure as the proofs of Lemmas 6 and 8, it suffices to be aware of the distribution of the binary random variable

$$|\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{u}_k|^2 = \begin{cases} 1 & \text{if } \boldsymbol{v}_i = \boldsymbol{u}_k \\ 0 & \text{otherwise} \end{cases}.$$

We can see that $|\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{u}_k|^2$'s are i.i.d. Bernoulli with $1/p$, thus yielding $\mathbb{E}\left[|\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{u}_k|^2\right] = 1/p$. Using the independence assumption and the law of total expectation again, we have

$$\mathbb{E}\left[\sum_i d_{ii}|\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{u}_k|^2\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_i d_{ii}|\boldsymbol{v}_i^{\mathsf{H}}\boldsymbol{u}_k|^2\Big|\boldsymbol{D}\right]\right] = \frac{\mathbb{E}[\mathrm{tr}\boldsymbol{D}]}{p}.$$

Noticing that the expectation of the remaining terms in (66) equals to zero, we have (69). ∎

APPENDIX B
PROOF OF LEMMA 1

Let $\tilde{\boldsymbol{w}}_{\ell\ell_k}$ denote the MMSE estimate of $\boldsymbol{w}_{\ell\ell_k}$ from $\bar{\boldsymbol{s}}_{\ell_k}$ instead of $\boldsymbol{s}_{\ell_k}$ given the knowledge of $\boldsymbol{U}_\ell$. We only prove the random partial unitary correlation model and omit the random partial Fourier model because one can follow the same footsteps. The estimate $\tilde{\boldsymbol{w}}_{\ell\ell_k}$ is given by

$$\tilde{\boldsymbol{w}}_{\ell\ell_k} = \boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\left(\boldsymbol{R}_{\ell\ell_k} + \sum_{\ell'\neq\ell}\boldsymbol{R}_{\ell\ell_k'} + \rho_p^{-1}\boldsymbol{I}_M\right)^{-1}\bar{\boldsymbol{s}}_{\ell_k}. \quad (70)$$

Letting $\boldsymbol{R}_{-\ell} \triangleq \sum_{\ell'\neq\ell}\boldsymbol{R}_{\ell\ell_k'}$, we have

$$\left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{R}_{\ell\ell_k} + \boldsymbol{R}_{-\ell}\right)^{-1}$$
$$= \left(\boldsymbol{I}_M + \left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{U}_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\right)^{-1}\boldsymbol{R}_{-\ell}\right)^{-1} \times$$
$$\left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{R}_{\ell\ell_k}\right)^{-1}$$
$$\overset{(a)}{=} \left(\boldsymbol{I}_M + \left(\rho_p\boldsymbol{I}_M - \rho_p\boldsymbol{U}_{\ell\ell_k}(\boldsymbol{\Lambda}_{\ell\ell_k}^{-1} + \rho_p\boldsymbol{I}_{r_{\ell\ell_k}})^{-1}\rho_p\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\right)\times$$
$$\boldsymbol{R}_{-\ell}\right)^{-1}\left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{R}_{\ell\ell_k}\right)^{-1}$$
$$\overset{(b)}{\simeq} \left(\boldsymbol{I}_M + \rho_p\boldsymbol{R}_{-\ell}\right)^{-1}\left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{R}_{\ell\ell_k}\right)^{-1}$$
$$\overset{(c)}{\simeq} \prod_{\ell'\neq\ell}\left(\boldsymbol{I}_M + \rho_p\boldsymbol{R}_{\ell\ell_k'}\right)^{-1}\left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{R}_{\ell\ell_k}\right)^{-1} \quad (71)$$

where (a) follows from the matrix inversion lemma, (b) follows from the fact that $\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{R}_{-\ell} = \boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\sum_{\ell'\neq\ell}\boldsymbol{U}_{\ell\ell_k'}\boldsymbol{\Lambda}_{\ell\ell_k'}\boldsymbol{U}_{\ell\ell_k'}^{\mathsf{H}} \simeq \boldsymbol{0}$ by Lemma 4 as $M \to \infty$, and we also used $\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell_k} = \boldsymbol{I}_{r_{\ell\ell_k}}$. In (c), we repeated the same decomposition upon $\left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{R}_{\ell j_k} + \boldsymbol{R}_{-\{\ell,j\}}\right)^{-1}$ for all $j \neq \ell$, where $\boldsymbol{R}_{-\{\ell,j\}} \triangleq \sum_{\ell'\neq\{\ell,j\}}\boldsymbol{R}_{\ell\ell_k'}$.

Substituting (71) into (70), we have

$$\tilde{\boldsymbol{w}}_{\ell\ell_k} \simeq \boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\prod_{\ell'\neq\ell}\left(\boldsymbol{I}_M + \rho_p\boldsymbol{R}_{\ell\ell_k'}\right)^{-1}\left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{R}_{\ell\ell_k}\right)^{-1}\bar{\boldsymbol{s}}_{\ell_k}$$
$$\overset{(a)}{\simeq} \boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\left(\rho_p^{-1}\boldsymbol{I}_M + \boldsymbol{R}_{\ell\ell_k}\right)^{-1}\bar{\boldsymbol{s}}_{\ell_k}$$
$$\overset{(b)}{\simeq} \boldsymbol{\Lambda}_{\ell\ell_k}\left(\rho_p^{-1}\boldsymbol{I}_{r_{\ell\ell_k}} + \boldsymbol{\Lambda}_{\ell\ell_k}\right)^{-1}\boldsymbol{s}_{\ell_k}$$
$$\overset{(c)}{\simeq} \hat{\boldsymbol{w}}_{\ell\ell_k}$$

where (a) follows from the matrix inversion lemma and Lemma 4, (b) follows from the matrix inversion lemma again. In (c), we applied Lemma 6 to $\tilde{\boldsymbol{R}}_{\ell_k\ell_{k'}}$ in (19) to have

$$\tilde{\boldsymbol{R}}_{\ell_k\ell_{k'}} - \frac{\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}}{M}\boldsymbol{I}_{r_{\ell\ell_k}} \overset{a.s.}{\longrightarrow} 0$$

and then $\tilde{\boldsymbol{R}}_{\ell_k\ell_{k'}}$ vanishes as in (46). The desired result then follows from the well-known fact that the MMSE estimate $\tilde{\boldsymbol{w}}_{\ell\ell_k}$ is a sufficient statistic for the estimation of $\boldsymbol{w}_{\ell\ell_k}$ (equivalently, $\boldsymbol{h}_{\ell\ell_k}$) from $\bar{\boldsymbol{s}}_{\ell_k}$ (e.g., see [61]).

APPENDIX C
PROOF OF THEOREM 5

The complete proof is lengthy but somehow straightforward after realizing that all the assumptions in [22, Thm. 1] and [24, Thm. 1] except the uniformly bounded spectral norm assumption apply to our case. For compactness, therefore, we only describe some essential parts without all the details.

Let $\tilde{\boldsymbol{\Psi}}_i = \sqrt{b_i}\boldsymbol{\Psi}_i$, and $\tilde{\boldsymbol{\Theta}}_i = \tilde{\boldsymbol{\Psi}}_i\tilde{\boldsymbol{\Psi}}_i^{\mathsf{H}} = b_i\boldsymbol{\Theta}_i$. To make sure that the random variable $m_N(z)$ converges to its deterministic equivalent, we need the generalized trace lemma in Corollary 2 to hold. Following the same line of arguments in [22, Thm. 1] and [24, Thm. 1], we arrive at obtaining an asymptotic approximation of $\boldsymbol{y}_i^{\mathsf{H}}\boldsymbol{\Psi}_i^{\mathsf{H}}\left(\boldsymbol{B}_{[i]} - z\boldsymbol{I}_N\right)^{-1}\boldsymbol{\Psi}_i\boldsymbol{y}_i$, where $\boldsymbol{B}_{[i]} = \left(\boldsymbol{B}_N - \boldsymbol{\Psi}_i\boldsymbol{y}_i\boldsymbol{y}_i^{\mathsf{H}}\boldsymbol{\Psi}_i^{\mathsf{H}}\right)^{-1}$. Although the $\boldsymbol{\Psi}_i^{\mathsf{H}}\left(\boldsymbol{B}_{[i]} - z\boldsymbol{I}_N\right)^{-1}\boldsymbol{\Psi}_i$ may not have a uniformly bounded spectral norm, we can see that $\limsup_{N\to\infty}\|\tilde{\boldsymbol{\Psi}}_i^{\mathsf{H}}\left(\boldsymbol{B}_{[i]} - z\boldsymbol{I}_N\right)^{-1}\tilde{\boldsymbol{\Psi}}_i\| < \infty$ due to assumption 3) and the fact that $\boldsymbol{B}_{[i]}$ is Hermitian and $z \in \mathbb{C}\setminus\mathbb{R}^+$. Then, it immediately follows from Corollary 2 that

$$\boldsymbol{y}_i^{\mathsf{H}}\tilde{\boldsymbol{\Psi}}_i^{\mathsf{H}}\left(\boldsymbol{B}_{[i]} - z\boldsymbol{I}_N\right)^{-1}\tilde{\boldsymbol{\Psi}}_i\boldsymbol{y}_i - \frac{1}{N}\mathrm{tr}\,\tilde{\boldsymbol{\Theta}}_i\left(\boldsymbol{B}_{[i]} - z\boldsymbol{I}_N\right)^{-1} \overset{a.s.}{\longrightarrow} 0. \quad (72)$$

For $z \in \mathbb{C}\setminus\mathbb{R}^+$, we further obtain by the rank-1 perturbation lemma [34, Lem. 2.1]

$$\frac{1}{N}\mathrm{tr}\,\tilde{\boldsymbol{\Theta}}_i\left(\boldsymbol{B}_{[i]} - z\boldsymbol{I}_N\right)^{-1} - \frac{1}{N}\mathrm{tr}\,\tilde{\boldsymbol{\Theta}}_i\left(\boldsymbol{B}_N - z\boldsymbol{I}_N\right)^{-1} \overset{a.s.}{\longrightarrow} 0.$$

Notice that the rank-1 perturbation lemma does not depend on the asymptotic properties of the perturbation so that $\boldsymbol{\Psi}_i\boldsymbol{y}_i\boldsymbol{y}_i^{\mathsf{H}}\boldsymbol{\Psi}_i^{\mathsf{H}}$ may even have a large Euclidean norm.

It is important to guarantee that the fixed-point algorithm in (43) converges. We already showed that $\frac{1}{N}\mathrm{tr}\,\tilde{\boldsymbol{\Theta}}_i\left(\boldsymbol{B}_N - z\boldsymbol{I}_N\right)^{-1} < \infty, \forall i,$. For some finite constant $b^* > 0$, therefore, $\frac{1}{N}\mathrm{tr}\,\tilde{\boldsymbol{\Theta}}_i\left(b^*\boldsymbol{B}_N - b^*z\boldsymbol{I}_N\right)^{-1}$ is further required not to trivially

go to zero as well. Letting $\boldsymbol{C} = \tilde{\boldsymbol{\Theta}}_i$ and $\boldsymbol{D} = \left(b^*\boldsymbol{B}_N - b^*z\boldsymbol{I}_N\right)^{-1}$, we have

$$
\begin{aligned}
\frac{1}{N}\mathrm{tr}\,\tilde{\boldsymbol{\Theta}}_i\left(b^*\boldsymbol{B}_N - b^*z\boldsymbol{I}_N\right)^{-1} &= \frac{1}{N}\mathrm{tr}\,\boldsymbol{C}\boldsymbol{D} \\
&\overset{(a)}{\geq} \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\boldsymbol{C})\lambda_{N-i+1}(\boldsymbol{D}) \\
&\geq \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\boldsymbol{C})\lambda_N(\boldsymbol{D}) \\
&= \frac{\mathrm{tr}\boldsymbol{C}}{N}\lambda_N(\boldsymbol{D})
\end{aligned}
$$

where (a) follows from [62, H.1.h.]. To make sure that $\frac{1}{N}\mathrm{tr}\,\tilde{\boldsymbol{\Theta}}_i\left(b^*\boldsymbol{B}_N - b^*z\boldsymbol{I}_N\right)^{-1} > 0$, we just need to let $b^* = b_{\min}$ (equivalently, $\beta_i = \frac{b_{\min}}{b_i}$), where $b_{\min} = \min_{i\in\mathcal{N}} b_i$ with $\mathcal{N} = [1:n]$, by noticing that

$$
\begin{aligned}
\lambda_1(\boldsymbol{D}^{-1}) &= b_{\min}\lambda_1\left(\sum_{j\in\mathcal{N}\backslash i}\boldsymbol{x}_j\boldsymbol{x}_j^{\mathsf{H}} + \boldsymbol{A}_N - z\boldsymbol{I}\right) \\
&\leq \sum_{j\in\mathcal{N}\backslash i}\lambda_1(b_{\min}\boldsymbol{x}_j\boldsymbol{x}_j^{\mathsf{H}}) + \lambda_1(b_{\min}\boldsymbol{A}_N) - b_{\min}z < \infty
\end{aligned}
$$

where the first inequality follows from [63], and the last step comes from the fact that $b_{\min}\max_{j\in\mathcal{N}}\lambda_1(\boldsymbol{x}_j\boldsymbol{x}_j^{\mathsf{H}}) < \infty$ due to Corollary 2 and assumption 3). Then, one can show by the same arguments in [22, Thm. 1] and [24, Thm. 1] that (43) converges to a fixed point.

It should be further noticed that $\frac{1}{N}\mathrm{tr}\,\boldsymbol{Q}_N\left(\boldsymbol{B}_N - z\boldsymbol{I}_N\right)^{-1}$ may trivially go to zero or infinity. This can be similarly avoided by appropriate scaling for $\boldsymbol{Q}_N$ under assumption 5). The remainder of the proof then follows from the steps in [22], [24], and we eventually have (40).

APPENDIX D
PROOF OF LEMMA 2

In what follows, we will derive the deterministic equivalents of each term in (24) for the MMSE detector with $\boldsymbol{v}_{\ell_k} = \boldsymbol{v}_{\ell_k}^{\mathrm{mmse}}$ in (22) conditioned on $\underline{\boldsymbol{U}}$. In the sequel, the superscript $^{\mathrm{mmse}}$ is omitted for notational convenience. For channel covariance, we consider only the random partial unitary model in Sec. II-B; A proof for the random partial Fourier model is straightforwardly given by using Corollary 1 and Lemma 4 and Lemma 8 instead of Lemmas 3, 4, and 6.

A. Signal Power

For the asymptotic approximation of $|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\hat{\boldsymbol{w}}_{\ell\ell_k}|^2$, we begin with

$$
\begin{aligned}
\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\hat{\boldsymbol{w}}_{\ell\ell_k} &= \hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k} \\
&\overset{(a)}{=} \frac{\hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}}{1 + \hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}} \\
&\overset{(b)}{\simeq} \frac{\frac{1}{r_{\ell\ell_k}}\mathrm{tr}\,\tilde{\boldsymbol{\Phi}}_{\ell\ell_k}\tilde{\boldsymbol{\Upsilon}}_{\ell_k}}{1 + \frac{1}{r_{\ell\ell_k}}\mathrm{tr}\,\tilde{\boldsymbol{\Phi}}_{\ell\ell_k}\tilde{\boldsymbol{\Upsilon}}_{\ell_k}} \\
&\overset{(c)}{\simeq} \frac{\delta_{\ell_k}}{1 + \delta_{\ell_k}}
\end{aligned}
\tag{73}
$$

where $\boldsymbol{\Upsilon}_{-\ell\ell_k} = \left(\boldsymbol{\Upsilon}_{\ell_k}^{-1} - \hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\hat{\boldsymbol{w}}_{\ell\ell_k}\right)^{-1}$, $\tilde{\boldsymbol{\Phi}}_{\ell\ell_k} = \frac{r_{\ell\ell_k}}{M}\boldsymbol{\Phi}_{\ell\ell_k}$, $\tilde{\boldsymbol{\Upsilon}}_{\ell_k} = M\boldsymbol{\Upsilon}_{\ell_k}$. (a) follows from the matrix inversion lemma [33, Eq. (2.2)], (b) is from Corollary 2 and the rank-1 perturbation lemma [34, Lem. 2.1], and (c) is from Theorem 5. In (b), we used the fact that the spectral norm of $\tilde{\boldsymbol{\Phi}}_{\ell\ell_k}$ is always uniformly bounded, whereas $\boldsymbol{\Phi}_{\ell\ell_k}$ is not necessarily the case, and that $\liminf_{M\to\infty}\frac{1}{r_{\ell\ell_k}}\boldsymbol{\Upsilon}_{\ell_k} > 0$ and $\limsup_{M\to\infty}\frac{1}{r_{\ell\ell_k}}\boldsymbol{\Upsilon}_{\ell_k} < \infty$. Notice that the step (b) is invalid for the case of the non-orthogonal pilot $\boldsymbol{s}'_{\ell_k}$ in (20) over the network since $\check{\boldsymbol{w}}_{\ell\ell_k}$ in (21) and $\boldsymbol{\Upsilon}_{-\ell\ell_k}$, where $\hat{\boldsymbol{w}}_{\ell\ell_k}$ is replaced with $\check{\boldsymbol{w}}_{\ell\ell_k}$, are not independent any longer. This is also the case with the remaining terms including $\boldsymbol{\Upsilon}_{-\ell\ell_k}$.

It follows from (73) and the continuous mapping theorem [64, Thm. 2.3] that, for any $\epsilon > 0$, there exists $r_0 > 0$ such that, for all $r_{\ell\ell_k} \geq r_0$, $\left|\left|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\hat{\boldsymbol{w}}_{\ell\ell_k}\right|^2 - \frac{\delta_{\ell_k}^2}{(1+\delta_{\ell_k})^2}\right| < \epsilon$. Then, using the dominated convergence theorem [56, Thm. 16.4], we obtain

$$
|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\hat{\boldsymbol{w}}_{\ell\ell_k}|^2 - \frac{\delta_{\ell_k}^2}{(1+\delta_{\ell_k})^2} \xrightarrow[M\to\infty]{} 0.
\tag{74}
$$

B. Noise Term

Noticing that $\mathbb{E}\left[|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{z}_{\ell_k}|^2\,\big|\,\hat{\boldsymbol{w}}_\ell\right] = \|\boldsymbol{v}_{\ell_k}\|_2^2$ and applying [33, Eq. (2.2)] twice and Theorem 6, we have

$$
\begin{aligned}
\|\boldsymbol{v}_{\ell_k}\|_2^2 &= \frac{\hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}^2\hat{\boldsymbol{w}}_{\ell\ell_k}}{\left(1 + \hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}\right)^2} \\
&\simeq \frac{\frac{1}{r_{\ell\ell_k}M}\mathrm{tr}\,\tilde{\boldsymbol{\Phi}}_{\ell\ell_k}\tilde{\boldsymbol{\Upsilon}}_{\ell_k}^2}{\left(1 + \frac{1}{r_{\ell\ell_k}}\mathrm{tr}\,\tilde{\boldsymbol{\Phi}}_{\ell\ell_k}\tilde{\boldsymbol{\Upsilon}}_{\ell_k}\right)^2} \\
&\simeq \frac{\mu_{\ell_k}}{M(1+\delta_{\ell_k})^2}.
\end{aligned}
\tag{75}
$$

For the channel estimation error, we have

$$
\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{n}_{\ell\ell_k} = \frac{\hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\boldsymbol{n}_{\ell\ell_k}}{1 + \hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}} \simeq 0
$$

which follows from [33, Eq. (2.2)] and Lemma 4. Applying the continuous mapping theorem and the dominated convergence theorem again leads to

$$
\mathbb{E}\left[|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{n}_{\ell\ell_k}|^2\,\big|\,\hat{\boldsymbol{w}}_\ell\right] \longrightarrow 0.
\tag{76}
$$

C. Pilot Contamination and Interference Terms

It follows from [33, Eq. (2.2)] that

$$\left|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{k\ell\ell_{k'}}\right|^2 = \frac{\hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\boldsymbol{w}_{k\ell\ell_{k'}}\boldsymbol{w}_{k\ell\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}}{1 + \hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}} \quad (77)$$

where $\boldsymbol{w}_{\ell_k\ell\ell_{k'}}$ is given by (15). Since only the effective channels having the same user (pilot) index k are correlated with each other in our setting, we address the following two cases for the numerator in the RHS of (77).

Let us first consider the case of $\ell' \neq \ell$ and $k' = k$, which takes the major portion of pilot contamination and causes a *nonvanishing*[10] interference power in the limit of large M. We can have

$$\hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\boldsymbol{w}_{k\ell\ell'_k}$$
$$= \left(\boldsymbol{w}_{\ell\ell_k} + \sum_{j \neq \ell}\boldsymbol{w}_{\ell_k j_k} + \frac{1}{\sqrt{\rho_p}}\boldsymbol{z}_{\ell_k}\right)^{\mathsf{H}}\boldsymbol{\Xi}_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell\ell_k}\boldsymbol{w}_{k\ell\ell'_k}$$
$$\stackrel{(a)}{\simeq} \boldsymbol{w}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Xi}_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell\ell_k}\boldsymbol{w}_{k\ell\ell'_k}$$
$$= \boldsymbol{w}_{\ell\ell'_{k'}}^{\mathsf{H}}\underbrace{\boldsymbol{U}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell_k}\boldsymbol{\Xi}_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell\ell_k}\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell'_{k'}}}_{\stackrel{(b)}{\simeq}\frac{1}{M}\mathrm{tr}\left(\boldsymbol{\Xi}_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell\ell_k}\right)\boldsymbol{I}_{r_{\ell\ell'_{k'}}}}\boldsymbol{w}_{\ell\ell'_{k'}} \quad (78)$$
$$\stackrel{(c)}{\simeq} \frac{1}{M^2}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}\mathrm{tr}\,\boldsymbol{\Xi}_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\tilde{\boldsymbol{\Upsilon}}_{\ell_k}$$
$$\stackrel{(d)}{\simeq} \frac{r_{\ell\ell_k}}{M}\nu_{\ell\ell'_k}$$
$$= \alpha_{\ell\ell_k}\nu_{\ell\ell'_k} \quad (79)$$

where (a) follows from Lemma 4 for $j = \ell'$, (b) from Lemma 6, and (c) from Corollary 2. As a matter of fact, we used the almost sure convergence in Lemma 5 to obtain (c) since (78) is a mixture of random vector $\boldsymbol{w}_{\ell\ell'_{k'}}$ and random matrix product $\boldsymbol{U}_{\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell_k}$ independent of each other. For the rest of this work, we use Lemma 5 in the same kind of mixtures of random variables. Noticing that $\limsup_{M \to \infty}\max\left\{\frac{1}{M}\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_k}, \|\tilde{\boldsymbol{\Upsilon}}_{\ell_k}\|_2, \frac{1}{r_{\ell\ell_k}}\mathrm{tr}\boldsymbol{\Xi}_{\ell\ell_k}\boldsymbol{\Lambda}_{\ell\ell_k}\right\} < \infty$ by Assumption 1 and using the trace inequality [65] that $\mathrm{tr}\boldsymbol{AB} \leq \|\boldsymbol{A}\|_2\mathrm{tr}\boldsymbol{B}$ for positive semidefinite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and also the fact that the spectral norm is a submultiplicative norm, we have $\limsup_{M \to \infty}\nu_{\ell\ell'_k} < \infty$. Similarly, $\liminf_{M \to \infty}\nu_{\ell\ell'_k} > 0$. In (d) we used Theorem 5.

For the residual interference term, where $k' \neq k$, we can rewrite by Corollary 2 the numerator in the RHS of (77) as

$$\hat{\boldsymbol{w}}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\boldsymbol{w}_{k\ell\ell'_{k'}}\boldsymbol{w}_{k\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell\ell_k}\hat{\boldsymbol{w}}_{\ell\ell_k}$$
$$\simeq \frac{1}{r_{\ell\ell_k}M}\mathrm{tr}\,\tilde{\boldsymbol{\Phi}}_{\ell\ell_k}\tilde{\boldsymbol{\Upsilon}}_{\ell_k}\boldsymbol{w}_{k\ell\ell'_{k'}}\boldsymbol{w}_{k\ell\ell'_{k'}}^{\mathsf{H}}\tilde{\boldsymbol{\Upsilon}}_{\ell_k}$$
$$= \boldsymbol{w}_{k\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{\ell_k}\boldsymbol{\Phi}_{\ell\ell_k}\boldsymbol{\Upsilon}_{\ell_k}\boldsymbol{w}_{k\ell\ell'_{k'}}. \quad (80)$$

[10]More precisely, the pilot contamination component is nonvanishing unless $\alpha_{\ell\ell_k} \to 0$.

Let $\boldsymbol{\Upsilon}_{-\ell_{kk'}} = \left(\boldsymbol{\Upsilon}_{-\ell\ell_k}^{-1} - \hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\right)^{-1}$, then by [4, Lem. 2] we have

$$\boldsymbol{\Upsilon}_{-\ell\ell_k} = \boldsymbol{\Upsilon}_{-\ell_{kk'}} - \frac{\boldsymbol{\Upsilon}_{-\ell_{kk'}}\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}}{1 + \hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}}.$$

We can further rewrite (80) as

$$\boldsymbol{w}_{k\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{\ell_k}\boldsymbol{\Phi}_{\ell\ell_k}\boldsymbol{\Upsilon}_{\ell_k}\boldsymbol{w}_{k\ell\ell'_{k'}} = \boldsymbol{w}_{k\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{\Phi}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{w}_{k\ell\ell'_{k'}}$$
$$+ \frac{|\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{w}_{k\ell\ell'_{k'}}|^2\,\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{\Phi}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}}{(1 + \hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}})^2}$$
$$- 2\mathrm{Re}\left\{\frac{\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{w}_{k\ell\ell'_{k'}}\boldsymbol{w}_{k\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{\Phi}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}}{1 + \hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}}\right\}. \quad (81)$$

We can further have the asymptotic approximations

$$\boldsymbol{w}_{k\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{\Phi}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{w}_{k\ell\ell'_{k'}} \stackrel{(a)}{\simeq} \frac{\mathrm{tr}\boldsymbol{\Lambda}_{\ell\ell'_{k'}}}{M^2}\frac{1}{r_{\ell\ell_k}}\mathrm{tr}\,\tilde{\boldsymbol{\Phi}}_{\ell\ell_k}\tilde{\boldsymbol{\Upsilon}}_{\ell_k}^2$$
$$= \frac{1}{M}\mu_{\ell_k\ell_{k'}}$$

where (a) follows from Lemma 6, and

$$\hat{\boldsymbol{w}}_{\ell_k\ell_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{w}_{k\ell\ell'_{k'}}$$
$$\simeq \boldsymbol{w}_{\ell_k'\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Xi}_{\ell_k'}\boldsymbol{\Lambda}_{\ell_k'}\boldsymbol{U}_{\ell_k'}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell_k}\boldsymbol{\Upsilon}_{-\ell_{kk'}}\boldsymbol{w}_{k\ell\ell'_{k'}}$$
$$\simeq \frac{1}{M}\mathrm{tr}\,\underbrace{\boldsymbol{U}_{\ell\ell_k}^{\mathsf{H}}\boldsymbol{U}_{\ell_k'}\boldsymbol{\Lambda}_{\ell\ell_{k'}}\boldsymbol{U}_{\ell\ell_k'}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell_k}}_{\xrightarrow[a.s.]{(a)}0}\boldsymbol{\Xi}_{\ell_k'}\boldsymbol{\Lambda}_{\ell_k'}\underbrace{\boldsymbol{U}_{\ell_k'}^{\mathsf{H}}\boldsymbol{U}_{\ell\ell_k}}_{\xrightarrow[a.s.]{(b)}0}\boldsymbol{\Upsilon}_{-\ell_{kk'}}$$
$$\simeq 0 \quad (82)$$

where (a) and (b) follow from a direct consequence of Lemma 6 due to the fact that $\boldsymbol{U}_{\ell\ell_k}$ and $\boldsymbol{U}_{\ell_k'}$ are mutually independent. It is important to note that this beneficial effect of interference elimination is enabled by spatial despreading. Combining the above asymptotic approximations into (81) yields

$$\boldsymbol{w}_{k\ell\ell'_{k'}}^{\mathsf{H}}\boldsymbol{\Upsilon}_{\ell_k}\boldsymbol{\Phi}_{\ell\ell_k}\boldsymbol{\Upsilon}_{\ell_k}\boldsymbol{w}_{k\ell\ell'_{k'}} \simeq \frac{1}{M}\mu_{\ell_k\ell_{k'}}. \quad (83)$$

Plugging (79) and (83) into (77) and noticing that (77) is bounded away from infinity, we have by the dominated convergence theorem

$$\sum_{(\ell',k') \neq (\ell,k)}\left|\boldsymbol{v}_{\ell_k}^{\mathsf{H}}\boldsymbol{w}_{k\ell\ell'_{k'}}\right|^2 \simeq \sum_{\ell' \neq \ell}\left(\frac{r_{\ell\ell_k}}{M}\right)^2\frac{|\nu_{\ell\ell'_k}|^2}{(1 + \delta_{\ell_k})^2}$$
$$+ \sum_{\ell' \neq \ell, k' \neq k}\frac{1}{M}\frac{\mu_{\ell_k\ell_{k'}}}{(1 + \delta_{\ell_k})^2}. \quad (84)$$

Combining all the results in (74) and (76) into (84), we arrive at (45).

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] 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.

[4] 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.

[5] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation and reduced-complexity scheduling," *IEEE Trans. on Inform. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[6] D. Kalofonos, M. Stojanovic, and J. Proakis, "Performance of adaptive MC-CDMA detectors in rapidly fading rayleigh channels," *IEEE Trans. on Wireless Commun.*, vol. 2, no. 2, pp. 229–239, Mar. 2003.

[7] K. Fazel and S. Kaiser, *Multi-Carrier Spread Spectrum Systems*. 1st ed. John Wiley & Sons Ltd, 2003.

[8] 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.

[9] E. Björnson, E. G. Larsson, and M. Debbah, "Massive mimo for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. on Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308.

[10] M. Chowdhury, A. Manolakos, and A. Goldsmith, "Scaling laws for noncoherent energy-based communications in the SIMO MAC," *IEEE Trans. on Inform. Theory*, vol. 62, no. 4, pp. 1980–1992, Apr. 2016.

[11] J. Chen, "On the MISO channel with feedback: Can infinitely massive antennas achieve infinite capacity?" *IEEE Trans. on Inform. Theory*, vol. 64, no. 10, pp. 6894–6916, Oct. 2018.

[12] 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.

[13] R. Müller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[14] 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. 264273, Feb. 2013.

[15] 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.

[16] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Muller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. on Sig. Proc.*, vol. 64, no. 11, pp. 2990–3003, Jun. 2016.

[17] S. Haghighatshoar and G. Caire, "Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation," vol. 16, no. 12, pp. 8316 – 8332, Dec. 2017.

[18] 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.

[19] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.

[20] 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.

[21] 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.

[22] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the capacity analysis of correlated MIMO multiple access channels," *IEEE Trans. on Inform. Theory*, vol. 57, no. 6, pp. 3493–3514, 2011.

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

[24] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels with limited feedback," *IEEE Trans. on Inform. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[25] A. F. Molisch, V. V. Ratman, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO – a survey," Sep. 2016. [Online]. Available: arXiv:1609.05078

[26] 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.

[27] T. S. Rappaport, *Wireless communications: Principles and Practice*. Prentice Hall PTR, 2002.

[28] M. Chowdhury, J. Nam, and A. J. Goldsmith, "Coherence time of wireless channels with large antenna arrays," 2017, pp. 1–6.

[29] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[30] A. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[31] 3GPP, "5G NR; Physical channels and modulation," *TS 38.211*, Apr. 2017.

[32] 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.

[33] J. W. Silverstein and Z. D. 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.

[34] Z. D. Bai and J. W. Silverstein, "On the signal-to-interference ratio of CDMA systems in wireless communications," *Ann. Appl. Probabil.*, vol. 17, no. 1, pp. 81–101, Feb. 2007.

[35] H. Zhang, S. Venkateswaran, and U. Madhow, "Channel modeling and MIMO capacity for outdoor millimeter wave links," *Proc. IEEE Wireless Comm. and Net. Conference (WCNC)*, pp. 1–6, 2010.

[36] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, 2013.

[37] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[38] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Inform. Theory*, vol. 52, no. 4, p. 12891306, 2006.

[39] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 622–640, 1999.

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

[41] J.-C. Shen, J. Zhang, and K. Letaief, "Downlink user capacity of massive MIMO under pilot contamination," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 6, pp. 3183–3193, 2015.

[42] C. Geng, N. Naderializadeh, A. S. Avestimehr, and S. A. Jafar, "On the optimality of treating interference as noise," *IEEE Trans. on Inform. Theory*, vol. 61, no. 4, pp. 1753–1767, 2015.

[43] J. Nam, "On the high-SNR capacity of the Gaussian interference channel and new capacity bounds," *IEEE Trans. on Inform. Theory*, vol. 63, no. 8, pp. 5266–5285, Aug. 2017.

[44] ——, "Capacity bounds for the K-user Gaussian interference channel," *IEEE Trans. on Inform. Theory*, vol. 63, no. 10, pp. 6416–6439, Oct. 2017.

[45] 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.

[46] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. on Inform. Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.

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

[48] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. on Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[49] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell tdd systems," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[50] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3464–3467, May 2011.

[51] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[52] J. Nam, J.-Y. Ahn, A. Adhikary, and G. Caire, "Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information," *Proc. Conf. on Inform. Sciences and Systems (CISS)*, pp. 1–6, Mar. 2012.

[53] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect" side information be?" *IEEE Trans. on Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[54] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

[55] 3GPP, "5G; Study on channel model for frequencies from 0.5 to 100 GHz," *TR 38.901*, Apr. 2017.

[56] P. Billingsley, *Probability and Measure*. John Wiley & Sons, 1995.

[57] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[58] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Process.*, vol. 86, no. 3, pp. 549–571, Mar. 2006.

[59] L. Tonelli, "Sull'integrazione per part," *Atti della Accademia Nazionale dei Lincei*, vol. 18, no. 2, pp. 246–253, 1909.

[60] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. on Inform. Theory*, vol. 49, no. 5, pp. 1293–1311, May 2003.

[61] G. D. Forney Jr., "Shannon meets Wiener II: On MMSE estimation in successive decoding schemes," 2004. [Online]. Available: http://arxiv.org/abs/cs/0409011

[62] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and its Applications*. Springer, 2011.

[63] K. Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," *in Proc. Nat. Academy of Sciences of the United States of America*, vol. 37.

[64] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge University Press, New York, 2000.

[65] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

Junyoung Nam (M'17) received the B.Sc. in Statistics from Inha University, Incheon, Korea, in 1997 and the M.Sc. and the Ph.D. degrees in Electrical Engineering (Information and Communication) from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2008 and 2015, respectively. He was with the Communications R&D Center, Samsung Electronics, Seoul, Korea, from 1997 to 2001, and also with the Communications Lab., Seodu InChip, Seoul, from 2001 to 2006. In Summer 2006, he joined Electronics and Telecommunications Research Institute (ETRI), Deajeon, Korea, where he was a principal member of research staff. From January to June 2017, he was with the department of Wireless Communications and Networks, Fraunhofer Heinrich Hertz Institute (HHI), Berlin, Germany. In Summer 2017, he joined Intel Labs, Santa Clara, CA, where he was a senior research scientist. Since March 2019, he has been with Qualcomm, San Jose, CA, where he is currently a senior staff system engineer. Dr. Nam was a 3GPP RAN1 delegate with more than 30 patents and 10 standards contributions. His inventions led to a core part of MIMO in 3GPP Release 13/14 and 5G NR. His research interests are in the areas of wireless communications, information theory, advanced cellular system design, and compressed sensing.

Giuseppe Caire (S'92–M'94–SM'03–F'05) was born in Torino in 1965. He received the B.Sc. in Electrical Engineering from Politecnico di Torino in 1990, the M.Sc. in Electrical Engineering from Princeton University in 1992, and the Ph.D. from Politecnico di Torino in 1994. He has been a post-doctoral research fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) in 1994-1995, Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, a Professor of Electrical Engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, and he is currently an Alexander von Humboldt Professor with the Faculty of Electrical Engineering and Computer Science at the Technical University of Berlin, Germany.

He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, the IEEE Communications Society & Information Theory Society Joint Paper Award in 2004 and in 2011, the Leonard G. Abraham Prize for best IEEE JSAC paper in 2019, the Okawa Research Award in 2006, the Alexander von Humboldt Professorship in 2014, the Vodafone Innovation Prize in 2015, and an ERC Advanced Grant in 2018. Giuseppe Caire is a Fellow of IEEE since 2005. He has served in the Board of Governors of the IEEE Information Theory Society from 2004 to 2007, and as officer from 2008 to 2013. He was President of the IEEE Information Theory Society in 2011. His main research interests are in the field of communications theory, information theory, channel and source coding with particular focus on wireless communications.

Mérouane Debbah (S01M04SM08F15) received the M.Sc. and Ph.D. degrees from the Ecole Normale SuprieureParis-Saclay, France. He was with Motorola Labs, Saclay, France, from 1999 to2002, and also with the Vienna Research Center for Telecommunications, Vienna,Austria, until 2003. From 2003 to 2007, he was an Assistant Professor with theMobile Communications Department, Institut Eurecom, Sophia Antipolis, France.In 2007, he was appointed Full Professor at CentraleSupelec, Gif-sur-Yvette,France. From 2007 to 2014, he was the Director of the Alcatel-Lucent Chair onFlexible Radio. Since 2014, he has been a Vice-President of the Huawei FranceResearch Center and the Director of the Mathematical and Algorithmic SciencesLab. He has managed 8 EU projects and more than 24 national and internationalprojects. His research interests lie in fundamental mathematics, algorithms,statistics, information, and communication sciences research. He is an IEEEFellow, a WWRF Fellow, and a Membre mrite SEE.

He was a recipient of the ERCGrant MORE (Advanced Mathematical Tools for Complex Network Engineering) from2012 to 2017. He was a recipient of the Mario Boella Award in 2005, the IEEEGlavieux Prize Award in 2011, the Qualcomm Innovation Prize Award in 2012 andthe 2019 IEEE Radio Communications Committee Technical Recognition Award. Hereceived 20 best paper awards, among which the 2007 IEEE GLOBECOM Best PaperAward, the Wi-Opt 2009 Best Paper Award, the 2010 Newcom++ Best Paper Award,the WUN CogCom Best Paper 2012 and 2013 Award, the 2014 WCNC Best Paper Award,the 2015 ICC Best Paper Award, the 2015 IEEE Communications Society Leonard G.Abraham Prize, the 2015 IEEE Communications Society Fred W. Ellersick Prize,the 2016 IEEE Communications Society Best Tutorial Paper Award, the 2016European Wireless Best Paper Award, the 2017 Eurasip Best Paper Award, the 2018IEEE Marconi Prize Paper Award, the 2019 IEEE Communications Society YoungAuthor Best Paper Award and the Valuetools 2007, Valuetools 2008, CrownCom2009, Valuetools 2012, SAM 2014, and 2017 IEEE Sweden VT-COM-IT Joint Chapterbest student paper awards. He is an Associate Editor-in-Chief of the journalRandom Matrix: Theory and Applications. He was an Associate Area Editor andSenior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2011 to2013 and from 2013 to 2014, respectively.

H.Vincent Poor (S72–M77–SM82–F87) received the Ph.D. degree in electrical engineering and computerscience from Princeton University in 1977.From 1977 until 1990, he was on the faculty of the University ofIllinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton,where he is the Michael Henry Strater University Professor of ElectricalEngineering. During 2006 to 2016, he served as Dean of Princetons School ofEngineering and Applied Science. He has also held visiting appointments atseveral other institutions, most recently at Berkeley and Cambridge. Hisresearch interests are in the areas of information theory and signal processing,and their applications in wireless networks, energy systems and related fields.Among his publications in these areas is the recent book Information Theoretic Security and Privacy of Information Systems (CambridgeUniversity Press, 2017).

Dr. Poor is a member of the National Academy ofEngineering and the National Academy of Sciences, and is a foreign member ofthe Chinese Academy of Sciences, the Royal Society, and other national andinternational academies. Recent recognition of his work includes the 2017 IEEE Alexander Graham BellMedal, the2019 ASEE Benjamin Garver Lamme Award, a D.Sc. honoris causa from Syracuse University, awarded in 2017, and a D.Eng.honoris causa from the University of Waterloo, awarded in 2019.