IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 25, 2019, accepted June 2, 2019, date of publication June 6, 2019, date of current version July 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921405

Joint Processing of Pilot and Data for Massive MIMO Systems in Ricean Fading Channels

YUANXUE XIN[1], PENGFEI SHI[1], WENRUI TANG[2], DONGMING WANG[2,3], (Member, IEEE), XUEWU ZHANG[1], AND XIAOHU YOU[2,3], (Fellow, IEEE)

[1]College of Internet of Things Engineering, Hohai University, Nanjing 213022, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 211100, China
[3]Purple Mountain Laboratory, Nanjing 211111, China

Corresponding author: Yuanxue Xin (xinyx@hhu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61801168, 61871122, in part by the Fundamental Research Funds for the Central Universities under Grant 2018B04414, in part by the National Natural Science Foundation of China (NSFC) under Grant 61801169, and in part by the Jiangsu Province Natural Science Foundation under Grant BK20170305.

ABSTRACT Besides the pilot symbols, the received data signals also contain some extra channel state information (CSI), which can be exploited to further improve the channel estimation accuracy. Considering a multi-cell multi-user massive multiple input multiple output (MIMO) with Ricean fading, the spectral efficiency for joint processing of the pilot and data strategy is investigated. Through the mutual information analysis of the multi-cell multi-user massive MIMO system, the intractable expression for the ergodic spectral efficiency is lower bounded and decomposed into three terms, where the explicit physical meanings of them are discussed in detail. Facilitated by the decomposing operation, an asymptotic closed-form expression of the spectral efficiency is obtained, and it is proven to be accurate. Comparing with separate processing, numerical results show that the joint processing outperforms the separate processing in spectral efficiency since the additional CSI is captured. The performance advantage for joint processing over separate processing becomes obvious as the coherence block duration of channels (T) increases. For a long enough T, joint processing eventually approaches the case with perfect CSI in terms of spectral efficiency. In addition, the inter-cell interference heavily reduces the system performance of joint processing when T is relatively small.

INDEX TERMS Massive multiple input multiple output (MIMO), Ricean fading, spectral efficiency.

I. INTRODUCTION

By adopting a large antenna array at the base station (BS), massive MIMO exploits extra spatial degrees of freedom to enable extremely narrow beams for many users, leading to a drastic increase in spectral efficiency, communication reliability and energy efficiency with simple signal processing techniques [1], [2]. In addition, for practical implementation, it is very attractive to arrange these antenna elements with cheap, and energy-saving radio hardwares. Benefiting from all these features, massive MIMO becomes one of the key technologies for the next generation wireless communication systems [3].

More antennas incur the hardware physical size problem in massive MIMO, which facilitates the millimeter wave (mmWave) technique with compact dimensions of antenna arrays [4], [5]. The combination of mmWave and massive MIMO has potential to dramatically improve the system throughput, wireless access, energy efficiency, and flexibility [6]. While highly directional nature of propagation for mmWave introduces some new features, strong Line-of-Sight (LOS) propagation plays an important role at mmWave technique [7]. Therefore, the application of massive MIMO should take the strong deterministic LOS links into consideration. Much of the previous works focus on the Rayleigh fading channel in massive MIMO systems [8]–[10]. As for a more general fading model, Ricean fading considering LOS is more beneficial than Rayleigh fading in terms of the average rate by utilizing the random matrix theory [11]. The work of [12] proves that for a set of extreme scenarios under which massive MIMO could fail in Ricean fading channels, the massive MIMO is still effective at averaging out the effects of

The associate editor coordinating the review of this manuscript and approving it for publication was Hamed Ahmadi.

intracell interference, small scale fading and additive noise by any standard scheduling scheme.

Besides channel fading model, the acquisition of channel state information (CSI) plays an important role in signal detection, and the accuracy of the acquired CSI is strongly associated with the system performance. Lund university and the University of Bristol have measured the performance of real-time massive MIMO system operating in LOS, and concluded that the CSI needs to be updated more frequently as the number of antennas increases [13]. Consequently time division duplex (TDD) with channel reciprocity is preferable for massive MIMO systems due to the large number of antennas [14], [15]. By reciprocity, BS performs channel estimation through the predetermined pilot symbols in the uplink and then uses the acquired CSI for both transmit precoding and signal detection, where processing steps are executed separately. In addition, an important result is given in [16], in which the author indicates that when the number of antennas increases without bound, the only factor limiting system performance is the effect of pilot contamination. Therefore the pilot overhead is a significant factor which dominates the accuracy of channel estimation and imposes severe limitations on the whole system performance [17].

Motivated by improving the system performance without increasing the pilot overhead, some researchers moved on to exploit the CSI carried by data symbols rather than just focusing on the design of the pilot or estimator. In [18] the direction of arrival correlated to CSI is estimated by using data symbols as well as pilot signals after projecting the received signal to a lower dimensional subspace, and the estimation accuracy is significantly improved over solely pilot-based channel estimation scheme. Considering the cooperation between BSs, paper [19] concludes that the detected uplink data signals can provide sufficient degree-of-freedom for channel estimation to help mitigating interference in an iterative way. The iterative joint channel estimation and data detection schemes are studied in [20], indicating the data signals can aid the system to suppress the pilot contamination effect and achieve high spectral efficiency. A three-stage iterative data-assisted channel estimation scheme is proposed in [21], which utilizes the previous uplink data symbols to improve the current channel estimation. Obviously, the above approaches develop a concept of joint processing, where the channel estimation is dependent on both pilot and data symbols in contrast to the separate processing in traditional receivers. Furthermore, in comparison with a separate processing by solely pilot-based channel estimation, the performance gain of a joint processing has been evaluated in some prior works [22], [23]. Nevertheless, prior works are of significant importance to verify the value of joint processing in a way of theoretical analysis. However, papers [23] and [22] have focused on the single-input single-output (SISO) and MIMO of a single cell rather than multiple cell scenario. Obviously the multi-cell MIMO network is more realistic and more complex in practical communication systems. In this paper, our work can be seen as an extension of the previous works to the case of a multi-cell multi-user massive MIMO in Ricean fading channels. In particular, some novel and interesting conclusions different from the previous works are obtained. To the best of our knowledge, this is the first time that the joint processing issue is discussed in a multi-cell massive MIMO system with Ricean fading. More specifically, our main contributions are summarized as the follows.

1) Taking Ricean fading channels into consideration, we analyze the mutual information of joint processing and decompose the intractable lower bound expression into three terms with explicit physical meanings, which further indicates the advantage of joint processing: it can compensate for the performance degradation caused by separate processing.
2) Adopting the asymptotic theory of large dimensional matrix in massive MIMO, we derive the asymptotic closed-form expressions for the three terms of the lower bound on the spectral efficiency one by one.
3) Through the simulation results, our proposed asymptotic closed-form expression of spectral efficiency is shown to have a satisfying accuracy. In addition, we verify the spectral efficiency improvement in joint processing compared with the separate processing in Ricean fading channels. We find that massive MIMO system with a multi-cell structure shows some different properties from the single-cell SIMO and MIMO systems.

The reminder of the paper is organized as follows. Section II introduces the system model. In Section III, considering multi-cell multi-user system, we analyze the lower bound of the system capacity in joint processing, and decompose the lower bound into three terms. In the next section, the closed-form expressions of the three terms proposed in Section III are derived, separately. Section V gives the ergodic spectral efficiency with joint processing in multi-cell multi-user massive MIMO system. Numerical results are shown in Section VI. Section VII concludes the paper.

The notation adopted in this paper conforms to the following convention. Vectors are column vectors and are denoted in lower case bold: $\boldsymbol{x}$. Matrices are bold upper case letters, e.g., $\boldsymbol{A}$. $\boldsymbol{I}_K$ denotes the identity matrix with size $K \times K$. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ represent transpose and Hermitian transpose, respectively. $(\cdot)^*$ denotes the conjugate operation. $\det(\boldsymbol{A})$ denotes the determinant of $\boldsymbol{A}$. $diag\,(\boldsymbol{x})$ is a diagonal matrix with $\boldsymbol{x}$ on its diagonal. The operator $\mathcal{E}(\cdot)$ denotes expectation, and the covariance operator is given by $\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) \triangleq \mathcal{E}(\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}) - \mathcal{E}(\boldsymbol{x})\mathcal{E}(\boldsymbol{y}^{\mathrm{H}})$.

II. SYSTEM MODEL

As shown in Fig.1, we consider L cells, and each cell contains one M-antenna BS and serves K single-antenna users simultaneously at the same frequency. The users randomly locate in all cells. For analytical convenience, cell 1 is chosen to be the reference cell, and the BS in it is denoted as BS 1.

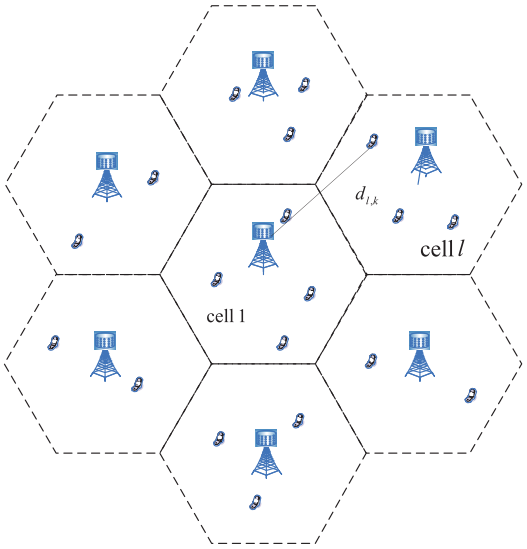

FIGURE 1. Multi-cell multi-user massive MIMO system.

Let $\boldsymbol{G}_l \in \mathbb{C}^{M\times K}$ denote the channel response between all users in the l-th cell and BS 1. The channel matrix $\boldsymbol{G}_l$ is expressed as

$$\boldsymbol{G}_l = \boldsymbol{H}_l \boldsymbol{\Lambda}_l^{1/2}, \tag{1}$$

where $\boldsymbol{\Lambda_l} = \text{diag}\left[\lambda_{l,1} \cdots \lambda_{l,K}\right]$ and the diagonal element $\lambda_{l,k}$ denotes the large-scale fading coefficient. More precisely, we assume $\lambda_{l,k} = d_{l,k}^{-\alpha}$, where $d_{l,k}$ is the distance between the k-the user in the l-th BS and BS 1, and α is the path loss exponent. Here $\boldsymbol{H}_l$ denotes the small-scale fading term.

In this paper, we consider Ricean fading channels, where both an LOS component and a Non-Line-of-Sight (NLOS) component caused by scatters are taken into consideration. For the reference cell, the channel matrix $\boldsymbol{H}_1$ can be decomposed into an LOS component $\bar{\boldsymbol{H}}_1$ and a scattered component $\boldsymbol{H}_{\text{R},1}$. As discussed in [24], $\bar{\boldsymbol{H}}_1$ only exists in the reference cell because of the short range between BS 1 and the users in it. Furthermore, $\bar{\boldsymbol{H}}_1$ is assumed to be a deterministic term [25], and it is given by

$$\left[\bar{\boldsymbol{H}}_1\right]_{m,k} = e^{-j(m-1)\frac{2\pi d}{\lambda}\sin\theta_k}, \tag{2}$$

where d denotes the antenna spacing, λ is the wavelength and θ_k represents the arrival angle of user k which lies in the interval $[-\pi/2, \pi/2]$. The entries in $\boldsymbol{H}_{\text{R},1}$ are modeled as the independently and identically distributed (i.i.d) zero mean circularly symmetric complex Gaussian (ZMCSCG) random variables with unit variance. For the interference cells, since the distances between users in the interference cells and BS 1 are longer than that between users in cell 1 and BS1, there would likely include more scatters and building blocks. As a result, longer distances significantly reduce the possibility of LOS transmission, leading to an assumption that the LOS components do not exist anymore between BS 1 and the users in the interference cells [26]. Therefore, channels between users in the l-th ($l \neq 1$) cell and BS 1, $\boldsymbol{H}_l$ is a collection of NLOS elements which are modeled as i.i.d ZMCSCG random variables with unit variance. For ease of notation, in the following analysis, $\boldsymbol{H}_l$ is rewritten as $\boldsymbol{H}_{\text{R},l}$ when $l \neq 1$.

Let ϑ_k represent the Ricean factor of the k-th user in the reference cell, which denotes the power ratio of the deterministic LOS component to the scattered term. Therefore, the small-scale fading $\boldsymbol{H}_l$ can be re-expressed as [27]

$$\boldsymbol{H}_l = \begin{cases} \bar{\boldsymbol{H}}_1\left[\boldsymbol{\Omega}(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{1/2} + \boldsymbol{H}_{\text{R},1}\left[(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{1/2}, \\ \qquad l = 1 \\ \boldsymbol{H}_{\text{R},l}, \quad l \neq 1 \end{cases} \tag{3}$$

where $\boldsymbol{\Omega} = diag[\vartheta_1 \cdots \vartheta_K]$, and $\boldsymbol{H}_{\text{R},l}$ is the NLOS channels of the users in the l-th cell to BS 1. Furthermore, the channel matrix $\boldsymbol{G}_l$ is re-expressed as

$$\boldsymbol{G}_l = \begin{cases} \bar{\boldsymbol{G}}_1\left[\boldsymbol{\Omega}(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{1/2} + \boldsymbol{G}_{\text{R},1}\left[(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{1/2}, \\ \qquad l = 1 \\ \boldsymbol{G}_{\text{R},l}, \quad l \neq 1 \end{cases} \tag{4}$$

where

$$\bar{\boldsymbol{G}}_1 = \bar{\boldsymbol{H}}_1 \boldsymbol{\Lambda}_1^{1/2}, \tag{5}$$

and

$$\boldsymbol{G}_{\text{R},l} = \boldsymbol{H}_{\text{R},l} \boldsymbol{\Lambda}_l^{1/2}. \tag{6}$$

In this paper, the block fading channel is considered, where the channel remains constant over a coherence interval T and varies independently between one coherence block and the next. In practice, the systems require CSI at the BS for adaptive receive beamforming. Such CSI is typically acquired by a known pilot sequence with a length of τ symbols in T, and it is inserted in the information data symbol sequence [28]. We assume that all users in all cells simultaneously transmit pilot sequence, and the same pilot sequence is reused among different cells. During the uplink training phase, the received pilot signal of BS 1 is

$$\boldsymbol{Y}_{\text{P}} = \boldsymbol{G}_1 \boldsymbol{X}_{\text{P}} + \sum_{l=2}^{L} \boldsymbol{G}_l \boldsymbol{X}_{\text{P}} + \boldsymbol{W}_{\text{P}}, \tag{7}$$

where $\boldsymbol{W}_{\text{P}} \in \mathbb{C}^{M\times\tau}$ denotes additive noise with each element is an i.i.d ZMCSCG random variable with variance σ_{P}^2. Here orthogonal pilot sequences between K users in the same cell are considered, hence

$$\boldsymbol{X}_{\text{P}} \boldsymbol{X}_{\text{P}}^{\text{H}} = \tau \boldsymbol{I}_K. \tag{8}$$

On the other hand, the received data signal at the reference BS is written as

$$\boldsymbol{Y}_{\text{D}} = \boldsymbol{G}_1 \boldsymbol{X}_1 + \sum_{l=2}^{L} \boldsymbol{G}_l \boldsymbol{X}_l + \boldsymbol{W}_{\text{D}}, \tag{9}$$

where $\boldsymbol{X}_l \in \mathbb{C}^{K\times(T-\tau)}$ represents the uplink data transmitted by all users in the l-th cell, and each data symbol is an i.i.d ZMCSCG random variable with unit variance. The addictive

noise matrix $\boldsymbol{W}_{\mathrm{D}}$ is modeled as an i.i.d ZMCSCG random variable with variance σ_{UL}^2.

III. JOINT PROCESSING IN MULTI-CELL MULTI-USER MASSIVE MIMO SYSTEM

Note that the separate processing inevitably introduces channel estimation error, which eventually results in the mismatched decoding. Regarding the joint processing, data transmission also carries some useful channel information so that channel estimation is based on both pilot symbols and data signals. Hence joint processing gives rise to a more accurate channel estimation and brings a capacity improvement in comparison to a separate processing which channel estimation is solely based on pilot sequences. Now, we analyze system capacity of the joint processing of pilot and data sequences in massive MIMO systems. Different from the previous work [22] which concerns the conventional signal-cell MIMO system, we consider a multi-cell multi-user scenario which is a more realistic and complex network structure. Considering the joint processing, the system capacity for the reference cell according to the mutual information is given by

$$\begin{aligned}\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right)&\overset{(a)}{=}\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}},\boldsymbol{G}\right)\\&\quad-\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right)\end{aligned}\tag{10}$$

where (a) follows from the chain rule of mutual information, and

$$\boldsymbol{G}=[\boldsymbol{G}_1\ \cdots\ \boldsymbol{G}_L].\tag{11}$$

The expression of the above mutual information is intractable due to the computational infeasibility as the complexity increases exponentially with the pilot length. Inspired by [23], we move on to further investigate the mutual information. For the first term of the right hand side in (10), we have

$$\begin{aligned}\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}},\boldsymbol{G}\right)&\overset{(a)}{=}\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{G}\right)\\&\overset{(b)}{=}\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_{\mathrm{D}}\,|\boldsymbol{G}\right)+\underbrace{\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{G}\right)}_{=0}\\&=\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_{\mathrm{D}}\,|\boldsymbol{G}\right),\end{aligned}\tag{12}$$

where (a) holds because the knowledge on $\boldsymbol{Y}_{\mathrm{P}}$ and $\boldsymbol{X}_{\mathrm{P}}$ does not impact the mutual information between $\boldsymbol{X}_1$ and $\boldsymbol{Y}_{\mathrm{D}}$ with the perfect CSI $\boldsymbol{G}$. Furthermore, (b) follows from the chain rule of mutual information, and $\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{G}\right)=0$ lies in the fact that $\boldsymbol{G}$ is independent of $\boldsymbol{X}_1$. Conditioned on the channel matrix $\boldsymbol{G}$, the mutual information between the received signal $\boldsymbol{Y}_{\mathrm{D}}$ and the desired transmitted data sequence $\boldsymbol{X}_1$ is the system capacity with perfect CSI. As for the second term in the right hand side of (10), the physical meaning is not so intuitive. By utilizing the information theory, we obtain the following expression

$$\begin{aligned}\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right)&=\mathcal{H}\left(\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right)\\&\quad-\mathcal{H}\left(\boldsymbol{G}|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}},\boldsymbol{X}_1\right).\end{aligned}\tag{13}$$

Following from the principle that conditioning reduces the entropy, we obtain

$$\mathcal{H}\left(\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right)\leq\mathcal{H}\left(\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right),\tag{14}$$

and

$$\mathcal{H}\left(\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\,,\boldsymbol{X}_1\right)\geq\mathcal{H}\left(\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{D}}\right),\tag{15}$$

where $\boldsymbol{X}_{\mathrm{D}}$ denotes data signals from all the users in all L cells

$$\boldsymbol{X}_{\mathrm{D}}\triangleq\begin{bmatrix}\boldsymbol{X}_1\\\vdots\\\boldsymbol{X}_L\end{bmatrix}.\tag{16}$$

In the traditional separate processing, receiver performs the channel estimation only based on the predefined pilot sequence $\boldsymbol{X}_{\mathrm{P}}$ and the observed signals $\boldsymbol{Y}_{\mathrm{P}}$. Thus, for the right hand side of (14), the uncertainty of $\boldsymbol{G}$ conditioned on $\boldsymbol{X}_{\mathrm{P}}$ and $\boldsymbol{Y}_{\mathrm{P}}$ relates to the estimation error while operating the solely pilot-based channel estimation. Hence, we have

$$\begin{aligned}\mathcal{H}\left(\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right)&=\mathcal{H}\left(\boldsymbol{G}\,\middle|\hat{\boldsymbol{G}}_{\mathrm{pil}},\boldsymbol{X}_{\mathrm{P}}\right)\\&=\mathcal{H}\left(\hat{\boldsymbol{G}}_{\mathrm{pil}}+\tilde{\boldsymbol{G}}_{\mathrm{pil}}\,\middle|\hat{\boldsymbol{G}}_{\mathrm{pil}},\boldsymbol{X}_{\mathrm{P}}\right)\\&\overset{(a)}{=}\mathcal{H}\left(\tilde{\boldsymbol{G}}_{\mathrm{pil}}\,|\boldsymbol{X}_{\mathrm{P}}\right),\end{aligned}\tag{17}$$

where $\hat{\boldsymbol{G}}_{\mathrm{pil}}$ and $\tilde{\boldsymbol{G}}_{\mathrm{pil}}$ are defined as the channel estimation matrix and the error matrix for separate processing, separately. When joint multi-user minimum mean square error (MMSE) receiver is employed at BS 1, according to the well-known properties of MMSE estimate, the estimation value is independent of the estimation error, which leads to (a). Equality (17) reveals the fact that the uncertainty of channel is decided by the estimation error.

Similar to the above analysis, we can obtain the following expression

$$\begin{aligned}\mathcal{H}\left(\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{D}}\right)&=\mathcal{H}\left(\boldsymbol{G}\,\middle|\hat{\boldsymbol{G}}_{\mathrm{joint}},\boldsymbol{X}_{\mathrm{P}}\right)\\&=\mathcal{H}\left(\tilde{\boldsymbol{G}}_{\mathrm{joint}}\,|\boldsymbol{X}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{D}}\right),\end{aligned}\tag{18}$$

where $\hat{\boldsymbol{G}}_{\mathrm{joint}}$ and $\tilde{\boldsymbol{G}}_{\mathrm{joint}}$ denote the channel estimation and error matrices for joint processing, respectively.

In order to obtain a more explicit conclusion, we further consider the detailed properties of the above expression. Regarding the fact that the error processes of MMSE channel estimation are zero-mean jointly proper Gaussian [23], we can deduce that the uncertainty of channel estimation error is decided by the covariance matrix of the estimation error. Hence, we have

$$\begin{aligned}&\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{G}\,|\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right)\\&\leq\mathcal{H}\left(\tilde{\boldsymbol{G}}_{\mathrm{pil}}\,|\boldsymbol{X}_{\mathrm{P}}\right)-\mathcal{H}\left(\tilde{\boldsymbol{G}}_{\mathrm{joint}}\,|\boldsymbol{X}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{D}}\right)\\&=\mathcal{E}_{\boldsymbol{X}_{\mathrm{P}}}\left[\log_2\det\left(\pi e\boldsymbol{\Sigma}_{\mathrm{pil}}\right)\right]-\mathcal{E}_{\boldsymbol{X}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{D}}}\left[\log_2\det\left(\pi e\boldsymbol{\Sigma}_{\mathrm{joint}}\right)\right]\\&\overset{(a)}{=}\log_2\det\left(\boldsymbol{\Sigma}_{\mathrm{pil}}\right)-\mathcal{E}_{\boldsymbol{X}_{\mathrm{D}}}\left[\log_2\det\left(\boldsymbol{\Sigma}_{\mathrm{joint}}\right)\right],\end{aligned}\tag{19}$$

where $\boldsymbol{\Sigma}_{\mathrm{pil}}$ and $\boldsymbol{\Sigma}_{\mathrm{joint}}$ denote the covariance matrices of the estimation error for separate processing and joint processing,

respectively. For (19), the expectation operations with respect to $\boldsymbol{X}_{\mathrm{P}}$ are removed because the pilot sequence $\boldsymbol{X}_{\mathrm{P}}$ is determined.

Substituting (19) and (12) into (10), a lower bound on the system capacity for the reference cell with joint processing is given by

$$\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{X}_{\mathrm{P}}\right)\geq\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_{\mathrm{D}}|\boldsymbol{G}\right)-\log_2\det\left(\boldsymbol{\Sigma}_{\mathrm{pil}}\right)+\mathcal{E}_{X_{\mathrm{D}}}\left[\log_2\det\left(\boldsymbol{\Sigma}_{\mathrm{joint}}\right)\right]. \quad (20)$$

Hence, the lower bound of the system capacity is decomposed into three parts: the system capacity with perfect CSI, the channel estimation uncertainty by separate processing and the channel estimation uncertainty by joint processing. Moreover, (20) straightly and clearly reveals the advantage of joint processing scheme: it compensates for the information loss in separate processing, while the compensation part is related to the third term of the right hand side in (20). In the following section, considering the multi-user multi-cell massive MIMO systems, we turn to analyze these three terms individually.

IV. THE CLOSED-FORM EXPRESSIONS OF THE THREE TERMS

A. SYSTEM CAPACITY WITH PERFECT CSI

For analytical tractability, we start with studying the received data signal at a symbol time interval t. The received uplink signal of BS 1 at time t is

$$\boldsymbol{y}_{\mathrm{D}}(t)=\boldsymbol{G}_1\boldsymbol{x}_1(t)+\sum_{l=2}^{L}\boldsymbol{G}_l\boldsymbol{x}_l(t)+\boldsymbol{w}_{\mathrm{D}}(t), \quad (21)$$

where $\boldsymbol{x}_l(t)$ and $\boldsymbol{w}_{\mathrm{D}}(t)$ are the t-th columns of $\boldsymbol{X}_l$ and $\boldsymbol{W}_{\mathrm{D}}$, respectively. According to [29], when the perfect CSI is known, the ergodic capacity for multi-user multi-cell system adopted the MMSE receiver is given as the follows

$$\begin{aligned}C^{perf}&\triangleq\mathcal{E}\left\{I\left(\boldsymbol{x}_1(t);\boldsymbol{y}_{\mathrm{D}}(t)|\boldsymbol{G}_1,\ldots,\boldsymbol{G}_L\right)\right\}\\&\geq\mathcal{E}\left\{\log_2\det\left(\frac{1}{\sigma_{UL}^2}\sum_{l=1}^{L}\boldsymbol{G}_l\boldsymbol{G}_l^H+\boldsymbol{I}_M\right)\right.\\&\quad\left.-\log_2\det\left(\frac{1}{\sigma_{UL}^2}\sum_{l=2}^{L}\boldsymbol{G}_l\boldsymbol{G}_l^H+\boldsymbol{I}_M\right)\right\}. \quad (22)\end{aligned}$$

Although the channel matrices in practical applications are imperfect and outdated due to channel estimation error and feedback delay, here the CSI is assumed to be perfectly known since we mainly focus on analyzing the spectral efficiency performance of joint processing rather than designing the CSI acquisition schemes.

In the following, we consider the massive MIMO systems with a large number of antennas at BS, and derive the closed-form expression for the achievable sum-rate. For notation convenience, we define

$$\boldsymbol{\Phi}=\left[\boldsymbol{\Omega}(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{1/2}, \quad (23)$$

and

$$\boldsymbol{\Gamma}=\left[(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{1/2}. \quad (24)$$

Then $\boldsymbol{G}_1=\bar{\boldsymbol{G}}_1\boldsymbol{\Phi}+\boldsymbol{G}_{\mathrm{R},1}\boldsymbol{\Gamma}$. Considering separating the LOS and NLOS parts, we proceed as the follows

$$\begin{aligned}&\det\left(\frac{1}{\sigma_{\mathrm{UL}}^2}\sum_{l=1}^{L}\boldsymbol{G}_l\boldsymbol{G}_l^H+\boldsymbol{I}_M\right)\\&\overset{(a)}{=}\det\left(\frac{1}{\sigma_{\mathrm{UL}}^2}\begin{bmatrix}\boldsymbol{G}_1^{\mathrm{H}}\boldsymbol{G}_1&\boldsymbol{G}_1^{\mathrm{H}}\boldsymbol{G}_2&\cdots&\boldsymbol{G}_1^{\mathrm{H}}\boldsymbol{G}_L\\\boldsymbol{G}_2^{\mathrm{H}}\boldsymbol{G}_1&\boldsymbol{G}_2^{\mathrm{H}}\boldsymbol{G}_2&\cdots&\boldsymbol{G}_2^{\mathrm{H}}\boldsymbol{G}_L\\\vdots&\vdots&\ddots&\vdots\\\boldsymbol{G}_L^{\mathrm{H}}\boldsymbol{G}_1&\boldsymbol{G}_L^{H}\boldsymbol{G}_2&\cdots&\boldsymbol{G}_L^{\mathrm{H}}\boldsymbol{G}_L\end{bmatrix}+\boldsymbol{I}_{KL}\right)\\&\overset{(b)}{=}\det\left[\frac{1}{\sigma_{\mathrm{UL}}^2}\left(\boldsymbol{U}_{\mathrm{LOS}}+\boldsymbol{U}_{\mathrm{NLOS}}\right)+\boldsymbol{I}_{KL}\right], \quad (25)\end{aligned}$$

where (a) is obtained by applying the determinant identity $\det(\boldsymbol{I}+\boldsymbol{AB})=\det(\boldsymbol{I}+\boldsymbol{BA})$. For (b), $\boldsymbol{U}_{\mathrm{LOS}}$ represents the matrix including the LOS part which is shown in (26), as shown at the bottom of this page. Here $\boldsymbol{U}_{\mathrm{NLOS}}$ refers to the scattered signal part, where

$$\boldsymbol{U}_{\mathrm{NLOS}}=\begin{bmatrix}\boldsymbol{\Gamma}\boldsymbol{G}_{\mathrm{R},1}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},1}\boldsymbol{\Gamma}&\boldsymbol{\Gamma}\boldsymbol{G}_{\mathrm{R},1}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},2}&\cdots&\boldsymbol{\Gamma}\boldsymbol{G}_{\mathrm{R},1}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},L}\\\boldsymbol{G}_{\mathrm{R},2}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},1}\boldsymbol{\Gamma}&\boldsymbol{G}_{\mathrm{R},2}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},2}&\cdots&\boldsymbol{G}_{\mathrm{R},2}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},L}\\\vdots&\vdots&\ddots&\vdots\\\boldsymbol{G}_{\mathrm{R},L}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},1}\boldsymbol{\Gamma}&\boldsymbol{G}_{\mathrm{R},L}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},2}&\cdots&\boldsymbol{G}_{\mathrm{R},L}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},L}\end{bmatrix}. \quad (27)$$

Since assuming the number of antennas at BS grows unlimited is a reasonable approach to study the properties of massive MIMO systems [30], [31], we obtain the following asymptotic expressions from the Theorem 3.4 and Theorem 3.7 in [32]

$$\frac{1}{M}\boldsymbol{G}_{\mathrm{R},l}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},l}-\boldsymbol{\Lambda}_l\underset{M\to\infty}{\longrightarrow}\boldsymbol{0}, \quad (28)$$

and

$$\frac{1}{M}\boldsymbol{G}_{\mathrm{R},l}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},l'}\underset{M\to\infty}{\longrightarrow}\boldsymbol{0},\quad\left(l'\neq l\right). \quad (29)$$

$$\boldsymbol{U}_{\mathrm{LOS}}=\begin{bmatrix}\boldsymbol{\Phi}\bar{\boldsymbol{G}}_1^{\mathrm{H}}\bar{\boldsymbol{G}}_1\boldsymbol{\Phi}+\boldsymbol{\Phi}\bar{\boldsymbol{G}}_1^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},1}\boldsymbol{\Gamma}+\boldsymbol{\Gamma}\boldsymbol{G}_{\mathrm{R},1}^{\mathrm{H}}\bar{\boldsymbol{G}}_1\boldsymbol{\Phi}&\boldsymbol{\Phi}\bar{\boldsymbol{G}}_1^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},2}&\cdots&\boldsymbol{\Phi}\bar{\boldsymbol{G}}_1^{\mathrm{H}}\boldsymbol{G}_{\mathrm{R},L}\\\boldsymbol{G}_{\mathrm{R},2}^{\mathrm{H}}\bar{\boldsymbol{G}}_1\boldsymbol{\Phi}&\boldsymbol{0}_K&\cdots&\boldsymbol{0}_K\\\vdots&\vdots&\ddots&\\\boldsymbol{G}_{\mathrm{R},L}^{\mathrm{H}}\bar{\boldsymbol{G}}_1\boldsymbol{\Phi}&\boldsymbol{0}_K&\cdots&\boldsymbol{0}_K\end{bmatrix} \quad (26)$$

Accordingly the above analysis leads to the asymptotic expression of the NLOS part in (27)

$$\boldsymbol{U}_{\mathrm{NLOS}} - \boldsymbol{U}_{\mathrm{NLOS}}^{inf} \underset{M\to\infty}{\longrightarrow} \mathbf{0}, \tag{30}$$

where

$$\boldsymbol{U}_{\mathrm{NLOS},inf} = M \begin{bmatrix} \boldsymbol{\Gamma}^2 & \\ & \boldsymbol{I}_{K(L-1)} \end{bmatrix} \boldsymbol{\Lambda}. \tag{31}$$

Here we define a matrix $\boldsymbol{\Lambda} \in \mathbb{C}^{KL\times KL}$ as

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & & \\ & \ddots & \\ & & \boldsymbol{\Lambda}_L \end{bmatrix}. \tag{32}$$

For the second term in the right hand side of (22), we can derive the following asymptotic approximation by similar steps above

$$\det\left(\frac{1}{\sigma_{UL}^2}\sum_{l=2}^{L}\boldsymbol{G}_l\boldsymbol{G}_l^H + \boldsymbol{I}_M\right) - \prod_{l=2}^{L}\det\left(\frac{M}{\sigma_{\mathrm{UL}}^2}\boldsymbol{\Lambda}_l + \boldsymbol{I}_K\right) \underset{M\to\infty}{\longrightarrow} \mathbf{0}. \tag{33}$$

By substituting the above two approximated results into (22), the asymptotic expression of the capacity lower bound for massive MIMO systems with perfect CSI is given by

$$C_{\mathrm{LB}}^{perf} = \mathcal{E}\left[\log_2\det\left(\frac{1}{\sigma_{\mathrm{UL}}^2}\left(\boldsymbol{U}_{\mathrm{LOS}} + \boldsymbol{U}_{\mathrm{NLOS},inf}\right) + \boldsymbol{I}_{KL}\right)\right] - \sum_{l=2}^{L}\log_2\det\left(\frac{1}{\sigma_{\mathrm{UL}}^2}\boldsymbol{\Lambda}_l + \boldsymbol{I}_K\right), \tag{34}$$

where the expectation is operated with respect to the scattered channel component.

B. CHANNEL ESTIMATION UNCERTAINTY WITH SEPARATE PROCESSING

This section we focus on the second term of the right hand side for (20), which involves the covariance matrix of the channel estimation error for the separate processing. As the channel estimation is solely based on the pilot sequences, we first investigate the received pilot signal and rewrite (7) as the follows for ease of notation

$$\boldsymbol{Y}_P = \boldsymbol{G}\breve{\boldsymbol{X}}_P + \boldsymbol{W}_P, \tag{35}$$

where $\breve{\boldsymbol{X}}_{\mathrm{P}}$ denotes the pilot sequences of all L cells, and it is defined as

$$\breve{\boldsymbol{X}}_{\mathrm{P}} \triangleq \begin{bmatrix} \boldsymbol{X}_{\mathrm{P}} \\ \vdots \\ \boldsymbol{X}_{\mathrm{P}} \end{bmatrix}. \tag{36}$$

Moreover, we rearrange the channel matrix $\boldsymbol{G}$ in the order of rows to reform a vector, which is

$$\boldsymbol{g} = \begin{bmatrix} \boldsymbol{g}_1 & \cdots & \boldsymbol{g}_M \end{bmatrix}, \tag{37}$$

where $\boldsymbol{g}_m$ corresponds to the m-th row of $\boldsymbol{G}$. More clearly, $\boldsymbol{g}_m$ is given by

$$\boldsymbol{g}_m = \left[\left[\bar{\boldsymbol{G}}_1\right]_m\left[\boldsymbol{\Omega}(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{\frac{1}{2}} + \left[\boldsymbol{G}_{\mathrm{R},1}\right]_m[(\boldsymbol{\Omega}+\boldsymbol{I}_K)]^{-\frac{1}{2}}, \cdots, \left[\boldsymbol{G}_{\mathrm{R},L}\right]_m\right], \tag{38}$$

where $\left[\bar{\boldsymbol{G}}_1\right]_m$ is the m-th row of $\bar{\boldsymbol{G}}_1$ for the reference cell, and $\left[\boldsymbol{G}_{\mathrm{R},l}\right]_m$ denotes the m-th row of $\boldsymbol{G}_{\mathrm{R},l}$. Since the first K elements of $\boldsymbol{g}_m$ indicate the channels between the K users in cell 1 and the m-th antenna at BS 1, $\boldsymbol{g}$ contains both the LOS and NLOS parts. It is obvious that $\boldsymbol{g}$ can be re-expressed as

$$\begin{aligned}\boldsymbol{g}_m &= \left[\left[\bar{\boldsymbol{G}}_1\right]_m\left[\boldsymbol{\Omega}(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{1/2} 0 \cdots 0\right] \\ &\quad + \left[\left[\boldsymbol{G}_{\mathrm{R},1}\right]_m\left[(\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1}\right]^{1/2} \cdots \left[\boldsymbol{G}_{\mathrm{R},L}\right]_m\right] \\ &\triangleq \bar{\boldsymbol{g}}_m + \boldsymbol{g}_{\mathrm{R},m}.\end{aligned} \tag{39}$$

Based on the above re-expression on the channels, next we analyze the received pilot signal of the m-th antenna at BS 1, which is

$$\boldsymbol{y}_{\mathrm{P},m} = \boldsymbol{g}_m\breve{\boldsymbol{X}}_P + \boldsymbol{w}_{\mathrm{P},m}, \tag{40}$$

where $\boldsymbol{w}_{\mathrm{P},m}$ indicates the m-th row of $\boldsymbol{W}_P$. From (40), it is easy to obtain the error covariance matrix when using MMSE estimator, and it is given by

$$\boldsymbol{\Sigma}_{\mathrm{pil},m} = \left(\mathrm{cov}^{-1}\left(\boldsymbol{g}_m^{\mathrm{T}}, \boldsymbol{g}_m^{\mathrm{T}}\right) + \frac{1}{\sigma_{\mathrm{P}}^2}\breve{\boldsymbol{X}}_{\mathrm{P}}^{*}\breve{\boldsymbol{X}}_{\mathrm{P}}^{\mathrm{T}}\right)^{-1}. \tag{41}$$

As the LOS component is conventionally assumed to be known [33], [34], we obtain the follows by substituting (39) into (41)

$$\boldsymbol{\Sigma}_{\mathrm{pil},m} = \left(\mathrm{cov}^{-1}\left(\boldsymbol{g}_{\mathrm{R},m}^{\mathrm{T}}, \boldsymbol{g}_{\mathrm{R},m}^{\mathrm{T}}\right) + \frac{1}{\sigma_{\mathrm{P}}^2}\breve{\boldsymbol{X}}_{\mathrm{P}}^{*}\breve{\boldsymbol{X}}_{\mathrm{P}}^{\mathrm{T}}\right)^{-1}. \tag{42}$$

According to the channel model stated in (4), we have

$$\mathrm{cov}\left(\boldsymbol{g}_{\mathrm{R},m}^{\mathrm{T}}, \boldsymbol{g}_{\mathrm{R},m}^{\mathrm{T}}\right) = \begin{bmatrix} (\boldsymbol{\Omega}+\boldsymbol{I}_K)^{-1} & & \\ & \boldsymbol{I}_K & \\ & & \ddots \\ & & \boldsymbol{I}_K \end{bmatrix}\begin{bmatrix} \boldsymbol{\Lambda}_1 & & & \\ & \boldsymbol{\Lambda}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\Lambda}_L \end{bmatrix}. \tag{43}$$

On the other hand, the orthogonality of pilot sequences leads to the followings

$$\breve{\boldsymbol{X}}_{\mathrm{P}}^{*}\breve{\boldsymbol{X}}_{\mathrm{P}}^{\mathrm{T}} = \tau\begin{bmatrix} \boldsymbol{I}_K & \cdots & \boldsymbol{I}_K \\ \vdots & \ddots & \vdots \\ \boldsymbol{I}_K & \cdots & \boldsymbol{I}_K \end{bmatrix}. \tag{44}$$

From (43) and (44) we conclude that $\boldsymbol{\Sigma}_{\mathrm{pil},m}$ in (42) is irrelevant to m. Therefore, for all different receive antennas, the covariance matrices of MMSE channel estimation error

in (41) have the same value. The reason lies in the fact that the fading processes on individual sub-channels with respect to different receive antennas are independent and identically distributed. Hence, by neglecting the antenna correlation, the covariance matrix for the system can be expressed as

$$\boldsymbol{\Sigma}_{\text{pil}} \triangleq \begin{bmatrix} \boldsymbol{\Sigma}_{\text{pil},1} & & \\ & \ddots & \\ & & \boldsymbol{\Sigma}_{\text{pil},M} \end{bmatrix} = \boldsymbol{I}_M \otimes \boldsymbol{\Sigma}_{\text{pil},m}. \quad (45)$$

Finally, we obtain the channel estimation uncertainty with separate processing which is given by

$$\begin{aligned} \log_2 \det\left(\boldsymbol{\Sigma}_{\text{pil}}\right) &= \log_2 \det\left(\boldsymbol{I}_M \otimes \boldsymbol{\Sigma}_{\text{pil},m}\right) \\ &= \log_2 \left[\left(\det\left(\boldsymbol{I}_M\right)\right)^K \left(\det\left(\boldsymbol{\Sigma}_{\text{pil},m}\right)\right)^M\right] \\ &= M \log_2 \det\left(\boldsymbol{\Sigma}_{\text{pil},m}\right). \end{aligned} \quad (46)$$

Here, the result in (46) obviously indicates that the channel estimation uncertainty with separate processing is a certain term involved the Ricean factors and the large-scale fading coefficients.

C. CHANNEL ESTIMATION UNCERTAINTY WITH JOINT PROCESSING

The advantage of joint processing strategy is supported by making full use of the received data signals to assist the channel estimation. Similar to (35), we rewrite the received pilot signal as

$$\boldsymbol{Y}_{\text{D}} = \boldsymbol{G}\boldsymbol{X}_{\text{D}} + \boldsymbol{W}_{\text{D}}. \quad (47)$$

Here $\boldsymbol{X}_{\text{D}}$ denotes the data transmitted from all users in all L cells, which is

$$\boldsymbol{X}_{\text{D}} \triangleq \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_L \end{bmatrix}. \quad (48)$$

Aiming at exploring the performance gain of the joint processing, we should combine the pilot and data signals in a received signal, yielding the followings

$$\begin{bmatrix} \boldsymbol{Y}_{\text{P}} & \boldsymbol{Y}_{\text{D}} \end{bmatrix} = \boldsymbol{G}\begin{bmatrix} \breve{\boldsymbol{X}}_{\text{P}} & \boldsymbol{X}_{\text{D}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{W}_{\text{P}} & \boldsymbol{W}_{\text{D}} \end{bmatrix}. \quad (49)$$

Thus the joint received signal for the m-th antenna of the reference cell is given by

$$\begin{bmatrix} \boldsymbol{y}_{\text{P},m} & \boldsymbol{y}_{\text{D},m} \end{bmatrix} = \boldsymbol{g}_m \begin{bmatrix} \breve{\boldsymbol{X}}_P & \boldsymbol{X}_{\text{D}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{w}_{\text{P},m} & \boldsymbol{w}_{\text{D},m} \end{bmatrix}, \quad (50)$$

where $\boldsymbol{y}_{\text{D},m}$ and $\boldsymbol{w}_{\text{D},m}$ are the m-th rows of $\boldsymbol{Y}_{\text{D}}$ and $\boldsymbol{W}_{\text{D}}$, respectively. Since the channel estimation uncertainty is determined by the covariance matrix of the channel estimation, it is necessary to take the following analysis

$$\begin{aligned} &\boldsymbol{\Sigma}_{\text{joint},m} \\ &= \left(\text{cov}^{-1}\left(\boldsymbol{g}_m^{\text{T}}, \boldsymbol{g}_m^{\text{T}}\right) + \begin{bmatrix} \breve{\boldsymbol{X}}_{\text{P}}^* & \boldsymbol{X}_{\text{D}}^* \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_{\text{P}}^2}\boldsymbol{I}_{KL} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{\sigma_{\text{UL}}^2}\boldsymbol{I}_{KL} \end{bmatrix} \begin{bmatrix} \breve{\boldsymbol{X}}_{\text{P}}^{\text{T}} \\ \boldsymbol{X}_{\text{D}}^{\text{T}} \end{bmatrix}\right)^{-1} \\ &= \left(\text{cov}^{-1}\left(\boldsymbol{g}_m^T, \boldsymbol{g}_m^T\right) + \frac{1}{\sigma_{\text{P}}^2}\breve{\boldsymbol{X}}_{\text{P}}^*\breve{\boldsymbol{X}}_{\text{P}}^{\text{T}} + \frac{1}{\sigma_{\text{UL}}^2}\boldsymbol{X}_{\text{D}}^*\boldsymbol{X}_{\text{D}}^{\text{T}}\right)^{-1} \\ &= \left(\boldsymbol{\Sigma}_{\text{pil},m}^{-1} + \frac{1}{\sigma_{\text{UL}}^2}\boldsymbol{X}_{\text{D}}^*\boldsymbol{X}_{\text{D}}^{\text{T}}\right)^{-1} \\ &= \boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_{\text{pil},m}, \end{aligned} \quad (51)$$

where

$$\boldsymbol{\Delta} \triangleq \boldsymbol{I}_{KL} + \frac{1}{\sigma_{\text{UL}}^2}\boldsymbol{\Sigma}_{\text{pil},m}\boldsymbol{X}_{\text{D}}^*\boldsymbol{X}_{\text{D}}^{\text{T}}. \quad (52)$$

As discussed in the above Section, we know that $\boldsymbol{\Sigma}_{\text{pil},m}$ is independent of m, while (52) shows that $\boldsymbol{\Delta}$ is also irrelevant to m. Therefore $\boldsymbol{\Sigma}_{\text{joint},m}$ are identical for all antennas deployed at BS 1. According to the independency among fading processes on individual sub-channels with respect to different antennas, the estimation error covariance matrix for $\boldsymbol{g}^{\text{T}}$ is given as the follows with ignoring the antenna correlation

$$\boldsymbol{\Sigma}_{\text{joint}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\text{joint},1} & & \\ & \ddots & \\ & & \Sigma_{\text{joint},M} \end{bmatrix} = \boldsymbol{I}_M \otimes \boldsymbol{\Sigma}_{\text{joint},m}. \quad (53)$$

Substituting (53) into the third term in the right hand side of (20), we have the channel estimation uncertainty with joint processing which is

$$\begin{aligned} &\mathcal{E}_{\boldsymbol{X}_{\text{D}}}\left[\log_2 \det\left(\boldsymbol{\Sigma}_{\text{joint}}\right)\right] \\ &\quad = \mathcal{E}_{\boldsymbol{X}_{\text{D}}}\left[\log_2 \det\left(\boldsymbol{I}_M \otimes \boldsymbol{\Sigma}_{\text{joint},m}\right)\right] \\ &\quad = \mathcal{E}_{\boldsymbol{X}_{\text{D}}}\left\{\log_2\left[\left(\det\left(\boldsymbol{I}_M\right)\right)^K \left(\det\left(\boldsymbol{\Sigma}_{\text{joint},m}\right)\right)^M\right]\right\} \\ &\quad = M\mathcal{E}_{\boldsymbol{X}_{\text{D}}}\left[\log_2 \det\left(\boldsymbol{\Sigma}_{\text{joint},m}\right)\right]. \end{aligned} \quad (54)$$

V. ERGODIC SPECTRAL EFFICIENCY WITH JOINT PROCESSING

Now we investigate the closed-form expression of the spectral efficiency lower bound for joint processing in (20). The analysis of the three terms in (20) are further performed to figure out their closed-form expressions.

For the first term of the right hand side in (20), since we neglect the temporal correlation of the channels, the ergodic capacity of cell 1 with perfect CSI can be computed as

$$\begin{aligned} \mathcal{E}\left\{\mathcal{I}\left(\boldsymbol{X}_1; \boldsymbol{Y}_{\text{D}} | \boldsymbol{G}\right)\right\} &= \sum_{t=\tau}^{T} \mathcal{E}\left\{\mathcal{I}\left(\boldsymbol{x}_1(t); \boldsymbol{y}_{\text{D}}(t) | \boldsymbol{G}_1, \cdots, \boldsymbol{G}_L\right)\right\} \\ &= (T-\tau)\, C^{perf} \\ &\geq (T-\tau)\, C_{\text{LB}}^{perf}. \end{aligned} \quad (55)$$

Considering the other two terms in (20), from (46) and (54), it is found that

$$\begin{aligned}&-M\log_2\det\left(\boldsymbol{\Sigma}_{\text{pil},m}\right)+M\mathcal{E}_{\boldsymbol{X}_\text{D}}\left[\log_2\det\left(\boldsymbol{\Sigma}_{\text{joint},m}\right)\right]\\&=M\mathcal{E}_{\boldsymbol{X}_\text{D}}\left[\log_2\det\left(\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_{\text{pil},m}\right)\right]-M\log_2\det\left(\boldsymbol{\Sigma}_{\text{pil},m}\right)\\&\overset{(a)}{=}-M\mathcal{E}_{\boldsymbol{X}_\text{D}}\left[\log_2\det\left(\boldsymbol{\Delta}\right)\right],\end{aligned}\tag{56}$$

where for (a) we use the fact that $\boldsymbol{\Sigma}_{\text{pil},m}$ is independent with $\boldsymbol{X}_\text{D}$. Obviously it is necessary to derive the closed-form expression of $\det(\boldsymbol{\Delta})$, however it is not easy to address this issue directly. Therefore, we introduce a useful lemma as the follows to help exporting the ultimate closed-form expression.

Lemma 1: Let $\boldsymbol{P}$ be a permutation matrix for rearranging $\boldsymbol{\Sigma}_{\text{pil},m}^{-1}$ into a block diagonalization matrix $\boldsymbol{A}$, we can obtain the follows

$$\mathcal{E}_{\boldsymbol{X}_\text{D}}\left[\log_2\det\left(\boldsymbol{\Delta}\right)\right]=\mathcal{E}_{\breve{\boldsymbol{X}}_\text{D}}\log_2\det\left(\boldsymbol{I}_{KL}+\frac{1}{\sigma_{\text{UL}}^2}\boldsymbol{A}^{-1}\breve{\boldsymbol{X}}_\text{D}^\text{H}\breve{\boldsymbol{X}}_\text{D}\right),\tag{57}$$

where

$$\breve{\boldsymbol{X}}_\text{D}=\boldsymbol{X}_\text{D}^\text{T}\boldsymbol{P}^\text{T}.\tag{58}$$

Proof: See Appendix A. ∎

Utilizing *lemma 1* and substituting (56) and (55) into (20), we obtain the ergodic spectral efficiency as follows

$$\begin{aligned}\frac{1}{T}\mathcal{E}\left[\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_\text{D},\boldsymbol{Y}_\text{P},\boldsymbol{X}_\text{P}\right)\right]\geq\left(1-\frac{\tau}{T}\right)C^{perf}-\frac{M}{T}\mathcal{E}_{\breve{\boldsymbol{X}}_\text{D}}\log_2\\\times\det\left(\boldsymbol{I}_{KL}+\frac{1}{\sigma_{\text{UL}}^2}\boldsymbol{A}^{-1}\breve{\boldsymbol{X}}_\text{D}^\text{H}\breve{\boldsymbol{X}}_\text{D}\right).\end{aligned}\tag{59}$$

By Jensen's inequality, this yields

$$\begin{aligned}&\frac{1}{T}\mathcal{E}\left[\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_\text{D},\boldsymbol{Y}_\text{P},\boldsymbol{X}_\text{P}\right)\right]\\&\geq\left(1-\frac{\tau}{T}\right)C_{\text{LB}}^{perf}-\frac{M}{T}\log_2\det\left[\boldsymbol{I}_{KL}+\frac{1}{\sigma_{\text{UL}}^2}\boldsymbol{A}^{-1}\mathcal{E}_{\breve{\boldsymbol{X}}_\text{D}}\left(\breve{\boldsymbol{X}}_\text{D}^\text{H}\breve{\boldsymbol{X}}_\text{D}\right)\right]\\&\overset{(a)}{=}\left(1-\frac{\tau}{T}\right)C_{\text{LB}}^{perf}-\frac{M}{T}\log_2\det\left(\boldsymbol{I}_{KL}+\frac{T-\tau}{\sigma_{\text{UL}}^2}\boldsymbol{A}^{-1}\right),\end{aligned}\tag{60}$$

where (a) follows from the fact that the elements of $\breve{\boldsymbol{X}}_\text{D}$ are i.i.d ZMCSCG random variables with unit variance. From (58) it can be seen that the element distributional properties of $\breve{\boldsymbol{X}}_\text{D}$ and $\boldsymbol{X}_\text{D}$ are same, this is because the permutation matrix multiplication does not affect the element distributional properties. Motivated by the above analysis, the next theorem provides the closed-form expression of the ergodic spectral efficiency.

Theorem 1: The ergodic spectral efficiency with joint processing of pilot and data symbols in multi-cell multi-user massive MIMO systems is lower bounded by

$$\frac{1}{T}\mathcal{E}\left[\mathcal{I}\left(\boldsymbol{X}_1;\boldsymbol{Y}_\text{D},\boldsymbol{Y}_\text{P},\boldsymbol{X}_\text{P}\right)\right]\geq C^{joint},\tag{61}$$

where

$$\begin{aligned}C^{joint}=\left(1-\frac{\tau}{T}\right)C_{\text{LB}}^{perf}-\frac{M}{T}\sum_{k=1}^{K}\log_2 b_k\\-\frac{M}{T}\sum_{k=1}^{K}\sum_{l=2}^{L}\log_2\left(1+\lambda_{l,k}\frac{T-\tau}{\sigma_{\text{UL}}^2}\right),\end{aligned}\tag{62}$$

and

$$b_k=\frac{\frac{\vartheta_k+1}{\lambda_{1,k}}+\frac{\tau}{\sigma_\text{P}^2}+\frac{T-\tau}{\sigma_{\text{UL}}^2}+\frac{\tau}{\sigma_\text{P}^2}\left(\frac{\vartheta_k+1}{\lambda_{1,k}}+\frac{T-\tau}{\sigma_{\text{UL}}^2}\right)\sum_{l=2}^{L}\left(\frac{1}{\lambda_{l,k}}+\frac{T-\tau}{\sigma_{\text{UL}}^2}\right)^{-1}}{\frac{\vartheta_k+1}{\lambda_{1,k}}+\frac{\tau}{\sigma_\text{P}^2}+\frac{\tau}{\sigma_\text{P}^2}\frac{(\vartheta_k+1)}{\lambda_{1,k}}\sum_{l=2}^{L}\lambda_{l,k}}.\tag{63}$$

Proof: See Appendix B. ∎

VI. NUMERICAL RESULTS

In this section, simulations are conducted to validate the performance of joint processing strategy and verify the accuracy of the proposed closed-form expressions of the spectral efficiency in multi-cell multi-user massive MIMO systems. Furthermore, the advantages of the joint processing strategy are discussed compared with transmission schemes with perfect CSI and separate processing. Note that the assumption of perfect CSI is widely applied for performance analysis in massive MIMO systems [35], and here we consider such an ideal scenario in order to investigate the properties of joint processing strategy. Some interesting and meaningful results can be observed from the numerical results. In this paper, as a comparison, we use the expression of the spectral efficiency by separate processing which is proposed in [36], and it is denoted as C^{sep}. A cellular system with 7 cells is considered, and the cell radius R is 1km. Each BS is located at the center of a cell, and $M=40$. The path loss exponent α is set to be 3.7. To ensure the result repeatability, 4 users are distributed uniformly on a circle of radius 2/3 around their corresponding serving BS [37]. Assuming all users in the same cell have an identical Ricean factor, and the Ricean factor is adopted among three classical values 0, 3, 5 according to [27]. For simplicity, all of the spacings between adjacent antennas at BS are assumed to be 1/2 of the wavelength, and the arrival angles are uniformly distributed, which means $\theta_k=\frac{\pi(k-1)}{K}-\frac{\pi}{2}, k=1,\cdots,K$. In Fig.2-Fig.4 and Fig.6 we choose the channel coherence interval $T=1024$ and the pilot length $\tau=K$.

In Fig.2, the simulated ergodic spectral efficiency lower bound in (59) is compared with the theoretical approximation in (62) to validate the accuracy of the proposed close-form expression. The spectral efficiency saturates along with the growth of SNR due to the pilot contamination and the cell interference. Massive MIMO can operate also at lower SNRs,

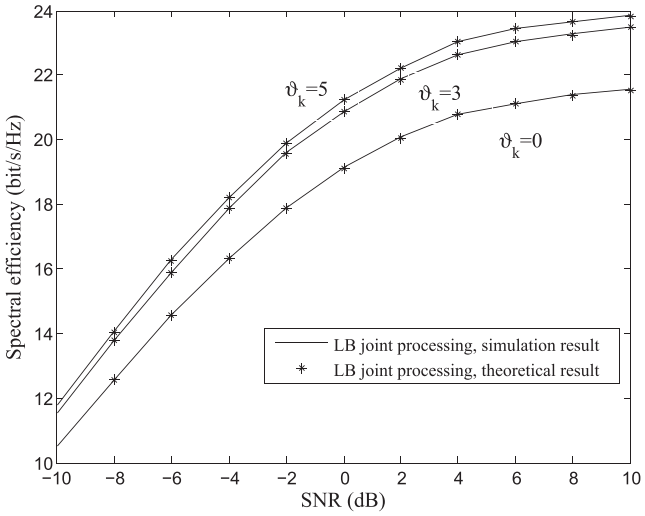

FIGURE 2. Lower bound on the ergodic spectral efficiency versus SNR for joint processing.

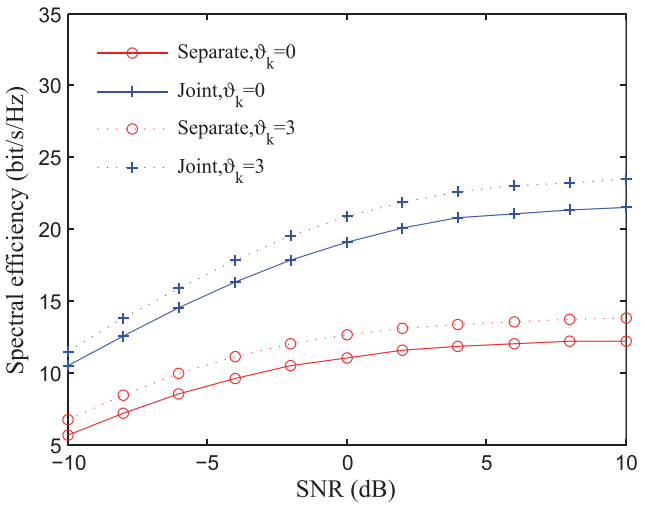

FIGURE 3. Comparison of ergodic spectral efficiency for various signal processing strategies.

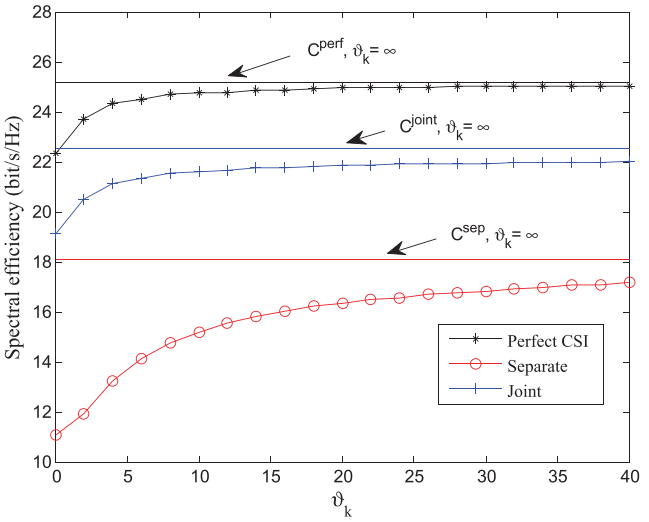

FIGURE 4. Impact of Ricean factor variation on ergodic spectral efficiency for various signal processing strategies.

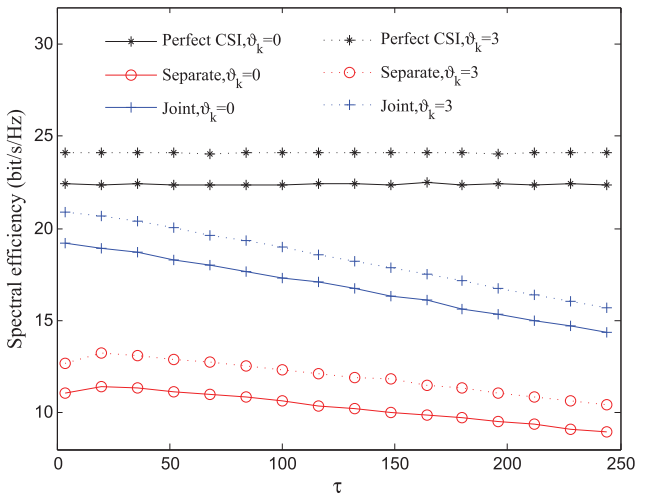

FIGURE 5. Ergodic spectral efficiency versus the pilot sequence length.

but results in a performance loss. When $\vartheta = 0$ the Ricean fading channel turns to a Rayleigh fading channel without LOS component, thus the channel is totally unknown to the receiver. Hence the spectral efficiency has the lowest performance when $\vartheta = 0$. Furthermore, Fig.2 shows the tightness between the simulation and theoretical results, therefore we use the closed-form expression of spectral efficiency for the following numerical work.

As seen from Fig.3, the spectral efficiency gains a better performance as the Ricean factor grows for all the signal processing schemes. Fig.3 indicates that the joint processing provides a better performance in spectral efficiency compared with separate processing. It is reasonable since the received data signal carries a certain amount of CSI, and the joint processing can make use of this part of CSI to improve the channel estimation accuracy.

In order to discuss the advantages of joint processing in spectral efficiency and the effect of Ricean factor on system performance, we depict Fig.4. Fig.4 explicitly shows the liming case $\vartheta = \infty$ achieves the best performance in terms of spectral efficiency as the channel is determined without any scattered signal part. More importantly, from the curves plotted in Fig.4, it can be observed that the spectral efficiency always gains benefits from utilizing joint processing compared with separate processing no matter how ϑ varies. As for the crossover between C^{perf} when $\vartheta = 0$ and C^{joint} when $\vartheta = \infty$, it implies that the joint processing achieves a better performance thanks to the strong LOS signal intensity although all the scattered signals are perfect known for C^{perf}.

Next, Fig.5 investigates how the pilot sequence length affects the spectral efficiency by assuming $K = 4$ and $T = 1024$. When CSI is perfect known, it can be seen

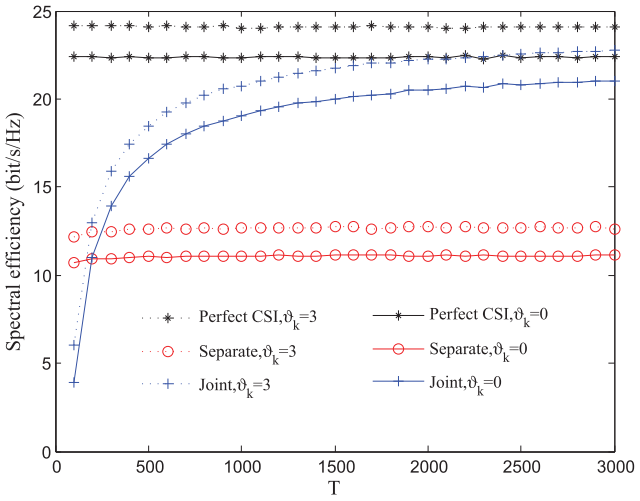

FIGURE 6. Ergodic spectral efficiency versus the coherence interval in multi-cell multi-user massive MIMO system.

that the pilot length does not affect the spectral efficiency. For separate processing, CSI is completely determined by pilot symbols through channel estimation. Hence lengthening the pilot sequence can improve the channel estimation accuracy, and achieve a better spectral efficiency performance. However, the continuous growth of τ even leads to spectral efficiency reduction for separate processing as observed from Fig.5. This is because more pilot symbols will lead to a higher system overhead, the pilot expenditure becomes the main limiting factor of the spectral efficiency performance. Consequently the spectral efficiency gain benefited from adding more pilot symbols cannot compensate the extra system overhead in joint processing strategy. Therefore we can see that the spectral efficiency function firstly increases and then drops with respect to τ in joint processing technique. As for the joint processing, received data signals also carry certain amount of knowledge for CSI, thus channel estimation performance is not completely decided by pilot sequence. Although longer τ may provide more CSI contained in pilot sequence, it will cause the CSI reduction for data signals as well. Moreover, increasing the length of τ mainly gives rise to a higher system overhead. Therefore, for joint processing, the spectral efficiency performance degradation keep going along with increasing τ. Different from the spectral efficiency with perfect CSI, the function of spectral efficiency of joint processing scheme is monotonically decreasing with increasing τ.

In order to provide an assessment of the influence of coherence block length, Fig.6 plots the spectral efficiency as a function for various signal processing strategies. When CSI is perfect known, the spectral efficiency will not be affected by T due to the fact that system capacity is constant at each moment. In separate processing, when T is relatively small, increasing T leads to a pilot overhead reduction, and finally brings the system improvement in spectral efficiency. However, when T reaches a large enough value, the pilot overhead can be neglected comparing with T. Therefore, the spectral efficiency finally tends to a fixed value when using separate processing. For joint processing, fixing the length of pilot sequence, increasing T means lengthening the data signal. The system performance is surely improved since the channel estimation accuracy is promoted.

Interestingly, in Fig.6, there exist some crossovers between joint processing and separate processing strategies which seem to be a different result from previous work [22]. In [22], it discussed the joint processing in SISO and single-cell MIMO scenarios, and the joint processing always shows a better performance in spectral efficiency for any values of T. However, here we investigate the multi-cell MIMO system where it involves much more interference from surrounding cells. The received data signals for cell 1 are interfered by other cells, hence data signals may contain little useful information for channel estimation with insufficient length of T. Then the spectral efficiency performance for BS 1 is mainly limited by the interference, and joint processing may have a worse performance than separate processing due to the signal detection complexity. With the growing of T, the data signal takes more useful information for channel estimation, and leads to spectral efficiency improvement. In addition, Fig.6 indicates that the spectral efficiency for joint processing converges to C^{perf}, and it outperforms separate processing in terms of spectral efficiency.

VII. CONCLUSION

In this paper, we analyze the performance of the joint processing strategy in multi-cell multi-user massive MIMO system with Ricean fading channels. Making full use of the CSI carried by data symbols, the joint processing provides an solution to improve the system spectral efficiency by achieving a higher channel estimation accuracy.

Based on the mutual information analysis, we lower bound the system capacity to obtain a tractable expression and decompose it into three terms with explicit physical meanings. Utilizing the properties of information theory and large dimensional matrix, we derive the closed-form expressions for the three terms. Numerical results verify that the joint processing can compensate for the information loss of channel estimation in separate processing, and the closed-form expression is proven to be accurate. In addition, we study how the lengths of pilot sequence and coherence block interval affect the spectral efficiency in joint processing, respectively. Since channel estimation depends on both pilot and data signals, lengthening pilot sequence leads to performance degradation due to the serious pilot overhead. Moreover, joint processing achieves a better performance by increasing the length of coherence block interval and eventually reaches the system performance with perfect CSI. Our analysis supports the application of joint processing in multi-cell multi-user massive MIMO systems.

$$\boldsymbol{A}_k = \begin{bmatrix} \left[\boldsymbol{\Lambda}_1^{-1}(\boldsymbol{\Omega}+\boldsymbol{I}_K)+\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} & \left[\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} & \cdots & \left[\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} \\ \left[\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} & \left[\boldsymbol{\Lambda}_2^{-1}+\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} & \cdots & \left[\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} \\ \vdots & \vdots & \ddots & \vdots \\ \left[\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} & \left[\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} & \cdots & \left[\boldsymbol{\Lambda}_L^{-1}+\tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K\right]_{k,k} \end{bmatrix}$$
$$= \begin{bmatrix} (\vartheta_k+1)/\lambda_{1,k}+\tau/\sigma_{\mathrm{P}}^2 & \tau/\sigma_{\mathrm{P}}^2 & \cdots & \tau/\sigma_{\mathrm{P}}^2 \\ \tau/\sigma_{\mathrm{P}}^2 & 1/\lambda_{2,k}+\tau/\sigma_{\mathrm{P}}^2 & \cdots & \tau/\sigma_{\mathrm{P}}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau/\sigma_{\mathrm{P}}^2 & \tau/\sigma_{\mathrm{P}}^2 & \cdots & 1/\lambda_{L,k}+\tau/\sigma_{\mathrm{P}}^2 \end{bmatrix} \tag{66}$$

APPENDIX A
PROOFS OF LEMMA 1

As shown in (52) it is obvious that $\boldsymbol{\Delta}$ contains the covariance matrix $\boldsymbol{\Sigma}_{\mathrm{pil},m}$ the first step is to simplify the expression of $\boldsymbol{\Sigma}_{\mathrm{pil},m}$.

Substituting (43) and (44) into (42), straight-forward computation yields to

$$\boldsymbol{\Sigma}_{\mathrm{pil},m} = \begin{bmatrix} \boldsymbol{\Lambda}_1^{-1}(\boldsymbol{\Omega}+\boldsymbol{I}_K)+\frac{\tau}{\sigma_{\mathrm{P}}^2}\boldsymbol{I}_K & \frac{\tau}{\sigma_{\mathrm{P}}^2}\boldsymbol{I}_K & \cdots & \tau/\sigma_{\mathrm{P}}^2\boldsymbol{I}_K \\ \frac{\tau}{\sigma_{\mathrm{P}}^2}\boldsymbol{I}_K & \boldsymbol{\Lambda}_2^{-1}+\frac{\tau}{\sigma_{\mathrm{P}}^2}\boldsymbol{I}_K & \cdots & \frac{\tau}{\sigma_{\mathrm{P}}^2}\boldsymbol{I}_K \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\tau}{\sigma_{\mathrm{P}}^2}\boldsymbol{I}_K & \frac{\tau}{\sigma_{\mathrm{P}}^2}\boldsymbol{I}_K & \cdots & \boldsymbol{\Lambda}_L^{-1}+\frac{\tau}{\sigma_{\mathrm{P}}^2}\boldsymbol{I}_K \end{bmatrix}^{-1}. \tag{64}$$

From the above expression it can be seen that each K-th order sub-matrix of $\boldsymbol{\Sigma}_{\mathrm{pil},m}^{-1}$ is a diagonal matrix. Hence a new block diagonal matrix $\boldsymbol{A}$ can be obtained by introducing a permutation matrix $\boldsymbol{P}$ to generate a columns and rows rearrangement, which is given by

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Sigma}_{\mathrm{P},m}^{-1}\boldsymbol{P}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{A}_1 & & \\ & \ddots & \\ & & \boldsymbol{A}_K \end{bmatrix}, \tag{65}$$

where $\boldsymbol{A}_k$ is shown in (66), as shown at the top of this page. Since a permutation matrix is also an orthogonal matrix, then

$$\boldsymbol{\Sigma}_{\mathrm{pil},m}^{-1} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}^{\mathrm{T}}\begin{bmatrix} \boldsymbol{A}_1 & & \\ & \ddots & \\ & & \boldsymbol{A}_K \end{bmatrix}\boldsymbol{P}. \tag{67}$$

Furthermore, we have

$$\boldsymbol{\Sigma}_{\mathrm{pil},m} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{P} = \boldsymbol{P}^{\mathrm{T}}\begin{bmatrix} \boldsymbol{A}_1^{-1} & & \\ & \ddots & \\ & & \boldsymbol{A}_K^{-1} \end{bmatrix}\boldsymbol{P}. \tag{68}$$

Substituting (68) into (52), it yields to

$$\boldsymbol{\Delta} = \boldsymbol{I}_{KL} + \frac{1}{\sigma_{\mathrm{UL}}^2}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{P}\boldsymbol{X}_{\mathrm{D}}^{*}\boldsymbol{X}_{\mathrm{D}}^{\mathrm{T}}. \tag{69}$$

Considering (56), we have the follows

$$\begin{aligned} &-M\log_2\det\left(\boldsymbol{\Sigma}_{\mathrm{pil},m}\right)+M\mathcal{E}_{\boldsymbol{X}_{\mathrm{D}}}\left[\log_2\det\left(\boldsymbol{\Sigma}_{\mathrm{joint},m}\right)\right] \\ &= -M\mathcal{E}_{\boldsymbol{X}_{\mathrm{D}}}\log_2\det\left(\boldsymbol{I}_{KL}+\frac{1}{\sigma_{\mathrm{UL}}^2}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{P}\boldsymbol{X}_{\mathrm{D}}^{*}\boldsymbol{X}_{\mathrm{D}}^{\mathrm{T}}\right) \\ &= -M\mathcal{E}_{\boldsymbol{X}_{\mathrm{D}}}\log_2\det\left(\boldsymbol{I}_{KL}+\frac{1}{\sigma_{\mathrm{UL}}^2}\boldsymbol{A}^{-1}\boldsymbol{P}\boldsymbol{X}_{\mathrm{D}}^{*}\boldsymbol{X}_{\mathrm{D}}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\right) \\ &= -M\mathcal{E}_{\breve{\boldsymbol{X}}_{\mathrm{D}}}\log_2\det\left(\boldsymbol{I}_{KL}+\frac{1}{\sigma_{\mathrm{UL}}^2}\boldsymbol{A}^{-1}\breve{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\breve{\boldsymbol{X}}_{\mathrm{D}}\right). \end{aligned} \tag{70}$$

This completes the proof.

APPENDIX B
PROOFS OF THEOREM 1

Applying the properties of block diagonal matrix directly to $\boldsymbol{A}^{-1}$, we have

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{A}_1^{-1} & & \\ & \ddots & \\ & & \boldsymbol{A}_K^{-1} \end{bmatrix}. \tag{71}$$

The second term in the right hand side of (60) becomes

$$\begin{aligned} &\frac{M}{T}\log_2\det\prod_{k=1}^{K}\left(\boldsymbol{I}_L+\frac{T-\tau}{\sigma_{\mathrm{UL}}^2}\boldsymbol{A}_k^{-1}\right) \\ &= \frac{M}{T}\sum_{k=1}^{K}\log_2\det\left(\boldsymbol{I}_L+\frac{T-\tau}{\sigma_{\mathrm{UL}}^2}\boldsymbol{A}_k^{-1}\right) \\ &= \frac{M}{T}\left[\sum_{k=1}^{K}\log_2\det\left(\boldsymbol{A}_k+\frac{T-\tau}{\sigma_{\mathrm{UL}}^2}\boldsymbol{I}_L\right)-\sum_{k=1}^{K}\log_2\det\boldsymbol{A}_k\right]. \end{aligned} \tag{72}$$

In the following we focus on getting the closed-form expressions for $\log_2\det\boldsymbol{A}_k$ and $\log_2\det\left(\boldsymbol{A}_k+\frac{T-\tau}{\sigma_{\mathrm{UL}}^2}\boldsymbol{I}_L\right)$.

$$\log_2 \det\left(\boldsymbol{A}_k + \frac{T-\tau}{\sigma_{\mathrm{UL}}^2}\boldsymbol{I}_L\right) = \log_2\left(\frac{\vartheta_k+1}{\lambda_{1,k}} + \frac{\tau}{\sigma_{\mathrm{P}}^2} + \frac{T-\tau}{\sigma_{\mathrm{UL}}^2} + \frac{\tau}{\sigma_{\mathrm{P}}^2}\left(\frac{\vartheta_k+1}{\lambda_{1,k}} + \frac{T-\tau}{\sigma_{\mathrm{UL}}^2}\right)\sum_{l=2}^{L}\frac{1}{\frac{1}{\lambda_{l,k}} + \frac{T-\tau}{\sigma_{\mathrm{UL}}^2}}\right) \tag{75}$$

According to the properties of the determinant, it yields

$$\begin{aligned}
&\det \boldsymbol{A}_k \\
&= \begin{vmatrix} (\vartheta_k+1)/\lambda_{1,k} + \tau/\sigma_{\mathrm{P}}^2 & \tau/\sigma_{\mathrm{P}}^2 & \cdots & \tau/\sigma_{\mathrm{P}}^2 \\ -(\vartheta_k+1)/\lambda_{1,k} & 1/\lambda_{2,k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -(\vartheta_k+1)/\lambda_{1,k} & 0 & \cdots & 1/\lambda_{L,k} \end{vmatrix} \\
&= \begin{vmatrix} \frac{\vartheta_k+1}{\lambda_{1,k}} + \frac{\tau}{\sigma_{\mathrm{P}}^2} + \sum_{l=2}^{L}\frac{\tau}{\sigma_{\mathrm{P}}^2\lambda_{l,k}\lambda_{1,k}} & 0 & \cdots & 0 \\ -(\vartheta_k+1)/\lambda_{1,k} & 1/\lambda_{2,k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -(\vartheta_k+1)/\lambda_{1,k} & 0 & \cdots & 1/\lambda_{L,k} \end{vmatrix}.
\end{aligned} \tag{73}$$

Hence the determinant of the above lower triangular matrix is

$$\log_2 \det \boldsymbol{A}_k = \log_2\left(\frac{\vartheta_k+1}{\lambda_{1,k}} + \frac{\tau}{\sigma_{\mathrm{P}}^2} + \frac{\tau}{\sigma_{\mathrm{P}}^2}\frac{\vartheta_k+1}{\lambda_{1,k}}\sum_{l=2}^{L}\lambda_{l,k}\right) + \sum_{l=2}^{L}\log_2\left(\frac{1}{\lambda_{l,k}}\right). \tag{74}$$

Following the similar steps described above yields (75), as shown at the top of this page. Substituting (74) and (75) into (60), the closed-form of ergodic spectral efficiency in *theorem 1*. This completes the proof.

REFERENCES

[1] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[2] Y. Xin, R. Zhang, D. Wang, J. Li, L. Yang, and X. You, "Antenna clustering for bidirectional dynamic network with large-scale distributed antenna systems," *IEEE Access*, vol. 5, pp. 4037–4047, 2017.

[3] J. Yang, H. Wang, J. Ding, X. Gao, and Z. Ding, "Spectral and energy efficiency analysis for massive MIMO multi-pair two-way relaying networks under generalized power scaling," *Sci. China Inf. Sci.*, vol. 60, no. 10, Mar. 2017, Art. no. 102303.

[4] J. Huang, C.-X. Wang, R. Feng, J. Sun, W. Zhang, and Y. Yang, "Multi-frequency mmWave massive MIMO channel measurements and characterization for 5G wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1591–1605, Jul. 2017.

[5] G. N. Kamga and S. Aïssa, "Wireless power transfer in mmwave massive MIMO systems with/without rain attenuation," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 176–189, Jan. 2019.

[6] L. Li, D. Wang, X. Niu, Y. Chai, L. Chen, L. He, X. Wu, F. Zheng, T. Cui, and X. You, "mmWave communications for 5G: Implementation challenges and advances," *Sci. China Inf. Sci.*, vol. 61, no. 2, Jan. 2018, Art. no. 021301.

[7] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

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

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

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

[11] L. Sanguinetti, A. Kammoun, and M. Debbah, "Asymptotic analysis of multicell massive MIMO over Rician fading channels," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3539–3543.

[12] M. Matthaiou, P. J. Smith, H. Q. Ngo, and H. Tataria, "Does massive MIMO fail in Ricean channels?" *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 61–64, Feb. 2019.

[13] P. Harrism, S. Malkowsky, J.Vieira, E. Bengtsson, F. Tufvesson, W. B. Hasan, L. Liu, M. Beach, S. Armour, and O. Edfors, "Performance characterization of a real-time massive MIMO system with LOS mobile channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1244–1253, Jun. 2017.

[14] H. Q. Ngo, H. Tataria, M. Matthaiou, S. Jin, and E. G. Larsson, "On the performance of cell-free massive MIMO in Ricean fading," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 980–984.

[15] H. Wei, D. Wang, J. Wang, and X. You, "TDD reciprocity calibration for multi-user massive MIMO systems with iterative coordinate descent," *Sci. China Inf. Sci.*, vol. 59, no. 10, Oct. 2016, Art. no. 102306.

[16] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[17] S. Xu, H. Zhang, T. Jie, and P. T. Mathiopoulos, "Pilot reuse and power control of D2D underlaying massive MIMO systems for energy efficiency optimization," *Sci. China Inf. Sci.*, vol. 60, no. 10, Oct. 2017, Art. no. 100303.

[18] S. Alzeer, S. Almatrudi, H. A. Bukhari, Y. Han, Y. Ding, and B. D. Rao, "Millimeter wave channel estimation using data-aided DoA estimation," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1567–1571.

[19] R. Wang, Y. Chen, H. Tan, and Q. Zhang, "Data-assisted massive MIMO uplink transmission with large backhaul cooperation delay: Scheme design and system-level analysis," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–7.

[20] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, Jun. 2014.

[21] Y. Li, R. Wang, Y. Chen, and S. Zhu, "Exploiting temporal channel correlation in data-assisted massive MIMO uplink detection," *IEEE Wireless Commun. Lett.*, vol. 21, no. 2, pp. 430–433, Feb. 2017.

[22] N. Jindal, A. Lozano, and T. L. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels?" in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 2189–2193.

[23] M. Dorpinghaus, A. Ispas, and H. Meyr, "On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2963–2982, May 2012.

[24] Q. Zhang, S. Jin, Y. Huang, and H. Zhu, "Uplink rate analysis of multicell massive MIMO systems in Ricean fading," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 3279–3284.

[25] S.-N. Jin, D.-W. Yue, and H. H. Nguyen, "Equal-Gain transmission in massive MIMO systems under Ricean fading," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9656–9668, Oct. 2018.

[26] I. Boukhedimi, A. Kammoun, and M.-S. Alouini, "LMMSE receivers in uplink massive MIMO systems with correlated Rician fading," *IEEE Trans. Wireless Commun.*, vol. 67, no. 1, pp. 230–243, Jan. 2019.

[27] X. Yu, S.-H. Leung, W. Xu, X. Dang, Z. Luo, and Y. Wang, "Precoding design for distributed antenna systems in spatially correlated Ricean fading channel," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9138–9152, Nov. 2016.

[28] P. Zhu, J. Li, D. Wang, and X. You, "Large system performance and distributed scheme of downlink beamforming in F-RANs with distributed antennas," *IEEE Access*, vol. 7, p. 33441–33453, 2019.

[29] D. Wang, C. Ji, X. Gao, S. Sun, and X. You, "Uplink sum-rate analysis of multi-cell multi-user massive MIMO system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 5404–5408.

[30] J. Li, D. Wang, P. Zhu, and X. You, "Energy efficiency optimization of distributed massive MIMO systems under Ergodic QoS and Per-RAU power constraints," *IEEE Access*, vol. 7, pp. 5001–5013, 2019.

[31] R. Sabbagh, C. Pan, and J. Wang, "Pilot allocation and sum-rate analysis in cell-free massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[32] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, 1st ed. New York, NY, USA: Cambridge Univ. Press, 2011.

[33] M. Matthaiou, P. J. Smith, H. Q. Ngo, and H. Tataria, "Does massive MIMO fail in Ricean channels?" *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 61–64, Feb. 2019.

[34] S. Jin, D. Yue, and H. H. Nguyen, "Multicell massive MIMO: Downlink rate analysis with linear processing under Ricean fading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3777–3791, Apr. 2019.

[35] G. N. Kamga, M. Xia, and S. Aïssa, "Spectral-efficiency analysis of regular- and large-scale (massive) MIMO with a comprehensive channel model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4984–4996, Jun. 2017.

[36] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang, and X. You, "Area spectral efficiency and area energy efficiency of massive MIMO cellular systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3243–3254, May 2016.

[37] J. Cao, D. Wang, J. Li, Q. Sun, and Y. Hu, "Uplink spectral efficiency analysis of multi-cell multi-user massive MIMO over correlated Ricean channel," *Sci. China Inf. Sci.*, vol. 61, no. 8, Aug. 2018, Art. no. 082305.