The Value of Channel Prediction in CoMP Systems with Large Backhaul Latency

Liyan Su, Chenyang Yang, and Shengqian Han

***Abstract*—The potential of coordinated multi-point (CoMP) transmission in providing high spectral efficiency for cellular systems largely depends on the channel quality at the cooperated base stations. In this paper, we investigate whether channel prediction is useful for downlink CoMP systems with backhaul latency in time-varying channels, where both the centralized and decentralized CoMP joint processing (CoMP-JP) as well as the CoMP coordinated beamforming (CoMP-CB) are considered. Toward this goal, we resort to large system analysis with large number of transmit antennas to derive closed form expressions of the average per-user rate of the CoMP systems, when predicted or estimated channels are employed for downlink precoding. By comparing with Non-CoMP systems, we find that channel prediction provides much higher performance gain over channel estimation for CoMP systems, depending on the strategy of the cooperation and the user location. As a result, CoMP systems can perform fairly well for mobile users even under the large backhaul latency, if channel prediction will be used instead of channel estimation. Simulation results are provided to validate our analysis.**

***Index Terms*—Coordinated multi-point transmission, channel prediction, backhaul latency.**

I. Introduction

INTER-CELL interference (ICI) is a major bottleneck to improve spectral efficiency of cellular networks, especially when multi-input multi-output (MIMO) techniques are applied. Recently, coordinated multi-point (CoMP) transmission in the context of 3GPP long term evolution (LTE) [1], also known as network MIMO in literature [2], has attracted much attention [3].

Depending on the type of information shared among the coordinated base stations (BSs), CoMP systems can be roughly divided into CoMP joint processing (CoMP-JP) and CoMP coordinated beamforming (CoMP-CB). When both data and channel state information (CSI) are shared among the BSs, CoMP-JP[1] can exploit the abundant spatial resources provided by multiple BSs with joint multi-user MIMO (MU-MIMO) precoding [2], where the ICI is converted into useful signals. When only the CSI of mobile stations (MSs) are shared, CoMP-CB avoids ICI with individual MU-MIMO precoding at each BS [4]. The potential of both cooperative strategies largely relies on the channel quality obtained at the BSs.

A typical centralized CoMP system consists of a central unit (CU) that is connected with multiple BSs via backhaul links. Under such a framework, the CSI is first obtained by each BS either by uplink training in time division duplex (TDD) systems [5] or by feedback in frequency division duplex (FDD) systems [6]. Then the CU collects the CSI from all coordinated BSs through backhaul links. With all the CSI, the centralized cooperative systems enable either CoMP-JP or CoMP-CB transmission depending on whether or not the data intended to all active MSs are available at the CU. When the backhaul links are with low latency and high capacity, CoMP-JP outperforms CoMP-CB, and both are superior to Non-CoMP systems. CoMP system can also operate in a decentralized manner, where each BS serves as a CU and the BSs are connected via backhaul. Again, the CSI is first obtained at each BS and then shared among the BSs.

In currently deployed cellular systems and emerging wireless standards, the backhaul links are with limited capacity. Moreover, they are with large latency, called X2 interface latency in 3GPP-LTE, which may reach 10 ms or more [7, 8]. Such a large latency will cause severely outdated CSI, which leads to performance deterioration even though the MSs served by CoMP are expected to move in a low speed [1, 8]. Despite that upgrading the backhaul links in existing networks has no technical challenges, this however will increase the infrastructure costs considerably.

In time-varying wireless channels, channel prediction is a popular approach to provide up-to-date channel information, which has been well explored for traditional single cell systems [9–11]. Considering that the small scale channels between multiple BSs and each MS are uncorrelated, existing channel prediction methods can be extended to CoMP systems in a straightforward manner. However, this does not mean that the performance gain of a CoMP system from channel prediction will be the same as that of a Non-CoMP system. On one hand, the performance of channel prediction and its impact on the performance of a system depend on the channel statistics [5]. Because the BSs are not co-located, the channel of CoMP-JP is different from single cell systems. Specifically, the average channel gains from different BSs to each MS differ, and the statistics of the CoMP channel are different from single cell systems, which depends on the location of

Manuscript received January 9, 2013; revised June 15 and August 11, 2013. The editor coordinating the review of this paper and approving it for publication was H. Dai.

The authors are with the School of Electronics and Information Engineering, Beihang University, Beijing China (e-mail: liyansu@ee.buaa.edu.cn, cyyang@buaa.edu.cn, sqhan@ee.buaa.edu.cn).

Part of this work was published in IEEE WCNC, 2012. This work was supported in part by National Natural Science Foundation of China (NSFC) under Grant 61120106002, and by Key project of Next Generation Wideband Wireless Communication Network, Ministry of Industry and Information (MII) 2011ZX03003- 001.

Digital Object Identifier 10.1109/TCOMM.2013.091813.130027

[1]In Release 11 of 3GPP-LTE [1], CoMP-JP is further divided into CoMP joint transmission (CoMP-JT) and dynamic point selection/dynamic point blanking (DPS/DPB). Only one BS transmits data to each user with DPS/DPB. Using CoMP-JT, multiple BSs transmit to a user with coherent cooperation, which is exactly the CoMP-JP referred in this paper or network MIMO in [2].

the MS. On the other hand, there are two factors that cause the outdated CSI in CoMP systems: the delay between uplink training and downlink transmission and the backhaul latency. The first factor is the same as in single cell systems, whose impact only depends on the relationship of the delay and the coherence time of the channel, which is well-understood [9, 12]. The second factor is unique in CoMP systems, whose impact remains unclear. Considering both factors, the overall delay in CSI for CoMP precoders may exceed the coherence time. At the first sight, one may concern from intuition that channel prediction will be no longer useful for CoMP under time-varying channels.

In this paper, we will show that with channel prediction downlink CoMP systems can perform surprisingly well for mobile users even under the large backhaul latency. To this end, we first derive closed form expressions of the average per-user data rate of CoMP-JP (both centralized and decentralized), CoMP-CB and Non-CoMP systems with large system analysis, where the number of transmit antennas approaches infinity. Large system analysis has been widely used to analyze the performance of single and multi-cell systems, see [13–15] and references therein. In [13, 14], the performance of CoMP system was analyzed where channel estimation errors are taken into account for large numbers of transmit antennas and users. The major difference of this work with [13, 14] comes from the goal of the analysis. Since we aim at demonstrating the gain from channel prediction for CoMP with large backhaul latency, we derive the explicit expressions of average rate of typical CoMP strategies with outdated CSI for large number of antennas. Next, we analyze the rate gain of channel prediction over channel estimation of the CoMP and Non-CoMP systems. From the analytical results we can find that CoMP systems can benefit more from channel prediction than Non-CoMP systems. This is just because the outdated CSI is more detrimental to the CoMP transmission.

The rest of the paper is organized as follows. In Section II, we present the system model. In Section III, we analyze the performance gain of using channel prediction by deriving the average per-user rate of downlink CoMP-JP, CoMP-CB and Non-CoMP systems. Simulation and numerical results are provided in Section IV, followed by conclusions in Section V.

II. System Model

Consider downlink CoMP systems, where B BSs each equipped with N_t antennas cooperatively serve M single-antenna MSs. For centralized CoMP, a CU is connected to the coordinated BSs via backhaul links. For decentralized CoMP, there is no CU and the BSs are connected by the backhaul. Denote the backhaul latency and the delay between the uplink channel estimate at each BS and downlink transmission as T and τ, respectively.

We consider both centralized and decentralized CoMP-JP systems, where the BSs serve the MSs with joint precoding. We consider centralized CoMP-CB, where the CU selects multiple MSs for each BS and each BS serves the MSs with individual precoding.

We investigate the impact of the outdated CSI caused by both T and τ on CoMP precoding. We take TDD systems as an example, and the conclusion is also valid for FDD systems. To facilitate downlink transmission, each MS in the cooperating cluster first sends training signals to all BSs in the uplink frame. Then each BS estimates the downlink channels from itself to all MSs exploiting channel reciprocity.

A. Uplink Training

Denote $\mathbf{g}_{mb}(t) = \sqrt{\alpha_{mb}}\mathbf{h}_{mb}(t)$ as the channel vector between BS_b and MS_m, where α_{mb} is the large scale fading gain including path loss and shadowing, $\mathbf{h}_{mb}(t) \in \mathbb{C}^{N_t \times 1}$ is the time-varying small scale fading channel vector whose entries are assumed independent and identically distributed (i.i.d.) with unit variance. When the received signal strength at MS_m from BS_b is stronger than those from other BSs, $\mathbf{g}_{mb}(t)$ is called local channel of MS_m. Otherwise, $\mathbf{g}_{mb}(t)$ is called cross channel of MS_m, whose master BS is not BS_b.

To show the potential of using channel prediction to deal with the outdated CSI in CoMP systems, we consider optimal linear channel predictor, Wiener predictor [16]. Therefore, we assume that a *priori* information of the channel statistics including temporal correlation function and uplink signal-to-noise-ratio (SNR) is known, as is commonly assumed for channel prediction, e.g., as in [9–11]. In practice, the downlink channels can be predicted at each BS in TDD systems and at each MS (or each BS) in FDD systems, after the channel statistics are obtained from multiple received signals.[2] An example of the channel prediction method can be found in [11].

In order to show the gain from channel prediction, we compare the results with those using optimal linear channel estimator under the same criterion. To assist optimal channel prediction or channel estimation, the uplink training signals from multiple MSs are assumed orthogonal. Since the channels between different BSs and MSs are uncorrelated, we only need to consider the channel acquisition for a single link, $\mathbf{g}_{mb}(t)$.

When each BS employs the received training symbols (i.e., the delayed measurements) to estimate the uplink channel (where only the measurement noises can be reduced) and then simply regards it as the downlink channel, we refer to the method of CSI acquisition as *channel estimation*.

When each BS employs the same received training symbols to extrapolate the downlink channel by exploiting the temporal correlation of the channel [9,11], both the measurement noises and the delay for the CSI can be reduced. We refer to this method of CSI acquisition as *channel prediction*.

With minimum mean square error (MMSE) channel estimator, the estimation error vector follows complex Gaussian distribution with covariance $\varepsilon_{mb}^{CE} \cdot \mathbf{I}$, where ε_{mb}^{CE} is the mean square error (MSE) of the estimate. Intuitively, the MSEs of the estimated cross channel vectors should exceed that of the estimated local channel vector. However, this is true only for small scale fading channel vectors $\mathbf{h}_{mb}(t)$ [5]. To be consistent with the understanding for channel estimation obtained from

[2]The channel is predictable with known channel statistics when the channel is a stationary random process (at least in the duration of channel prediction) and the prediction horizon is less than the coherence time of the channel. This happens when the users move in a low varying scattering environment with lower speed with respect to the duration of a uplink training frame and downlink transmission frame.

single cell systems, we use normalized mean square error (NMSE) to measure the estimation performance for $\mathbf{g}_{mb}(t)$, which is $\bar{\varepsilon}_{mb}^{CE} = \varepsilon_{mb}^{CE}/\alpha_{mb}$. Similarly, the performance of the MMSE channel predictor for $\mathbf{g}_{mb}(t)$ is measured by

$$\bar{\varepsilon}_{mb}^{CP} = \varepsilon_{mb}^{CP}/\alpha_{mb}, \tag{1}$$

where ε_{mb}^{CP} is the MSE of the predicted channel vector.[3]

Although channel estimation and channel prediction have the same expression of MSE, their performance differs, depending on the channel variation. If the channels are static, channel prediction and channel estimation will provide same quality for CSI [16]. For time-varying channels, however, channel prediction will provide the CSI with less errors when the temporal correlation of the channel can be exploited [12], because channel estimation does not employ this channel statistic.

B. Downlink transmission

1) CoMP-JP: We consider zero forcing (ZF) precoding at the CU, which is of low complexity. ZF precoding consists of ZF beamforming (ZFBF) and power allocation. Under sum power constraints, it has been proved that a channel pseudo-inverse based ZFBF together with an optimal power allocation that maximizes the sum rate of multiple users is the optimal ZF precoding [19]. Under per-BS power constraint (PBPC) in CoMP-JP systems [2], such a ZFBF in conjunction with the optimal power allocation that maximizes the sum rate has minor performance loss from the optimal ZF precoding when there are many users in each cell [20]. Therefore, we consider the pseudo-inverse based ZFBF.

○ Precoder for Centralized CoMP-JP

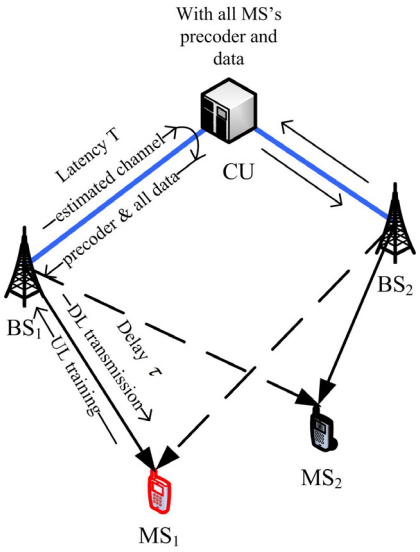
For *centralized CoMP-JP*, all coordinated BSs forward the estimated/predicted channels to the CU via backhaul links. The CU constructs the estimated/predicted global channel vector of MS_m as $\hat{\mathbf{g}}_m = [\hat{\mathbf{g}}_{m1}^H \cdots \hat{\mathbf{g}}_{mB}^H]^H$.[4] The estimated/predicted global channel vectors of all the MSs can be expressed as $\hat{\mathbf{G}}_{JP} = [\hat{\mathbf{g}}_1 \cdots \hat{\mathbf{g}}_M]^H$. Then, the CU selects multiple MSs and computes the precoders for the MSs, and forwards the precoding vectors to all BSs for downlink transmission. It is not hard to see that a delay of $2T+\tau$ will exist in the precoders at the BSs with respect to the realistic channels in downlink transmission if channel estimation is used, as shown in Fig. 1 (a). The joint ZF precoder for MS_m in centralized CoMP-JP systems is

$$\mathbf{w}_{c,m}^{JP} = \frac{\sqrt{p_m^{JP}}\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m}{\|\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m\|}, \tag{2}$$

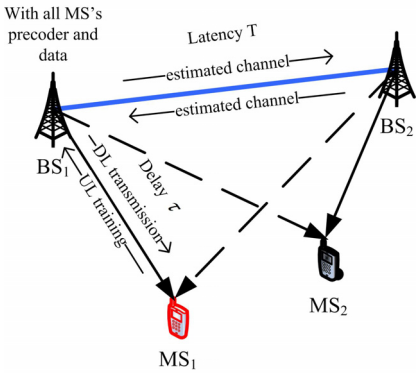
where $\hat{\mathbf{Q}}_{c,m} = \mathbf{I} - \bar{\mathbf{G}}_{c,m}^H(\bar{\mathbf{G}}_{c,m}\bar{\mathbf{G}}_{c,m}^H)^{-1}\bar{\mathbf{G}}_{c,m}$, $\bar{\mathbf{G}}_{c,m} = [\hat{\mathbf{g}}_1 \cdots \hat{\mathbf{g}}_{m-1}\hat{\mathbf{g}}_{m+1} \cdots \hat{\mathbf{g}}_M]^H$, i.e., $\hat{\mathbf{Q}}_{c,m}$ is the null space

[3]Except for the additive errors we considered, there also exist multiplicative errors in the acquired CSI, which for example come from the randomness of analog circuitry, imperfect channel reciprocity in TDD [17] or phase ambiguity in FDD CoMP-JP systems [18]. These multiplicative errors are more detrimental to the performance of CoMP-JP than the additive errors, and are not predictable. Fortunately, the adverse impact of these multiplicative errors can be significantly reduced by a stable analog circuitry design, antenna calibration or phase ambiguity feedback [17, 18], which are important for channel prediction to be useful.

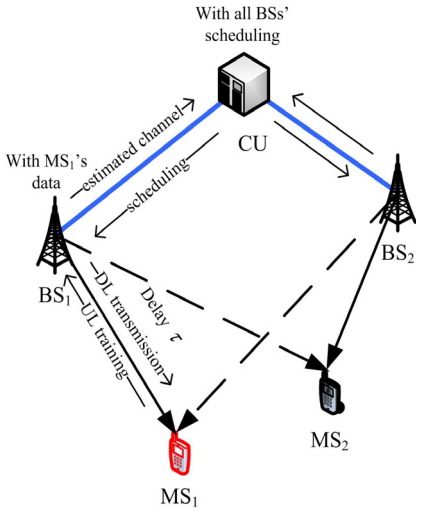
[4]For notational simplicity, we omit the time variable of the time-varying channels in the sequel.



(a) Centralized CoMP-JP



(b) Decentralized CoMP-JP



(c) CoMP-CB

Fig. 1: An example of CoMP systems, where the channel estimation method is applied. The backhaul links are with latency T, and the delay between the estimated channel at each BS and the realistic channel during downlink transmission is τ.

spanned by the global channel vectors of all the MSs except MS_m, and p_m^{JP} is the power allocated to MS_m.

○ Precoder for Decentralized CoMP-JP

For *decentralized CoMP-JP*, each BS forwards its estimated/predicted channels to other BSs via backhaul links. With the estimated/predicted channels and the forwarded channels from other BS, each BS constructs the global channel matrix $\hat{\mathbf{G}}_{JP}$, selects multiple users, and computes the precoding for downlink transmission. A delay of τ exists between the estimated channels and the realistic channels during downlink transmission, and a latency of $T+\tau$ exists between the forwarded channels and the realistic channels, as shown in Fig. 1 (b).

For BS_b, its constructed global channel vector for MS_m is $\hat{\mathbf{g}}_m^{(b)} = [\hat{\mathbf{g}}_{m1}^{(b)H} \cdots \hat{\mathbf{g}}_{mB}^{(b)H}]^H$, and its constructed global channel matrix for all the MSs can be expressed as $\hat{\mathbf{G}}_{JP}^{(b)} = [\hat{\mathbf{g}}_1^{(b)} \cdots \hat{\mathbf{g}}_M^{(b)}]^H$. Note that a part of the global channel vectors $\hat{\mathbf{g}}_{1b}^{(b)} \cdots \hat{\mathbf{g}}_{Mb}^{(b)}$ is obtained at BS_b by channel estimation or prediction without backhaul latency, and the rest of channel vectors $\hat{\mathbf{g}}_{1j}^{(b)} \cdots \hat{\mathbf{g}}_{Mj}^{(b)}, j \neq b$ are forwarded from other BSs with latency T. As a result, multiple BSs will use different channel matrices $\hat{\mathbf{G}}_{JP}^{(b)}, b = 1, \cdots, B$ to compute their joint precoders. With these global channel matrices, each BS, say BS_b, can simply compute the overall ZF precoding matrix at the B BSs for the M MSs as the pseudo-inverse of $\hat{\mathbf{G}}_{JP}^{(b)}$. Denote $\mathbf{w}_{1j}^{(b)}, \cdots, \mathbf{w}_{Mj}^{(b)}$ as the sub-vectors computed at BS_b that will be used at BS_j for the transmission to the M MSs, $j = 1, \cdots, B, j \neq b$. Owing to the backhaul latency, $\mathbf{w}_{mj}^{(b)} \neq \mathbf{w}_{mj}^{(a)}$ when $b \neq a$. That is to say, the joint precoders from different BSs to each MS are not the same, which comes from the different "view" on the channels at different BSs.

In fact, each BS only needs to compute its own sub-vectors for the M MSs, since other sub-vectors in the joint precoding vectors will be used at other BSs for transmission. Then, the joint precoding vector for MS_m is an aggregation of multiple sub-vectors computed at the B coordinated BSs, which is

$$\mathbf{w}_{d,m}^{JP} = [\mathbf{w}_{m1}^{(1)H} \cdots \mathbf{w}_{mB}^{(B)H}]^H. \tag{3}$$

Such a precoder will lead to multi-user interference (MUI), and the performance will be degraded even when the backhaul latency is not large.

To deal with this problem, we consider that each BS, say BS_b, only employs the channels from the M MSs to itself to compute the ZF precoder, instead of using $\hat{\mathbf{G}}_{JP}^{(b)}$. This is equivalent to the scenario where the backhaul latency approaches infinity. Then, the sub-vector of BS_b in the joint precoder for MS_m can be derived as follows

$$\mathbf{w}_{mb}^{(b)} = \frac{\sqrt{p_m^{(b)}}\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}}{\|\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}\|}, \tag{4}$$

where $\hat{\mathbf{Q}}_{mb} = \mathbf{I} - \bar{\mathbf{G}}_{mb}^H(\bar{\mathbf{G}}_{mb}\bar{\mathbf{G}}_{mb}^H)^{-1}\bar{\mathbf{G}}_{mb}$, $\bar{\mathbf{G}}_{mb} = [\hat{\mathbf{g}}_{1b} \cdots \hat{\mathbf{g}}_{m-1,b}\hat{\mathbf{g}}_{m+1,b} \cdots \hat{\mathbf{g}}_{Mb}]^H$, and $p_m^{(b)}$ is the transmit power allocated from BS_b to MS_m.

In this way, the decentralized joint ZF precoder only uses the locally estimated/predicted channel vectors, where the estimated channels have a delay of τ with respect to the realistic channels in the downlink transmission.

○ Received Signal

With the joint ZF precoder, the received signal of MS_m is

$$y_m = \mathbf{g}_m^H\mathbf{w}_m^{JP}x_m + \sum_{j\neq m}\mathbf{g}_m^H\mathbf{w}_j^{JP}x_j + z_m, \tag{5}$$

where the second term is the MUI from the co-scheduled users for MS_m caused by the imperfect CSI, and z_m is additive white Gaussian noise (AWGN) with zero mean and variance σ^2. For centralized CoMP-JP, $\mathbf{w}_m^{JP} = \mathbf{w}_{c,m}^{JP}$. For decentralized CoMP-JP, $\mathbf{w}_m^{JP} = \mathbf{w}_{d,m}^{JP}$ with $\mathbf{w}_{mb}^{(b)}$ defined in (4).

2) CoMP-CB: In CoMP-CB systems, the CU selects multiple MSs based on the collected channels from all the BSs. Then, each BS computes the ZF precoder with the channels from itself to all the active MSs to avoid the ICI. The individual ZF precoder for MS_m at its serving BS, BS_{b_m}, is given by

$$\mathbf{w}_m^{CB} = \frac{\sqrt{p_m^{CB}}\hat{\mathbf{Q}}_{mb_m}\hat{\mathbf{g}}_{mb_m}}{\|\hat{\mathbf{Q}}_{mb_m}\hat{\mathbf{g}}_{mb_m}\|}. \tag{6}$$

The received signal at MS_m is

$$y_m = \mathbf{g}_{mb_m}^H\mathbf{w}_m^{CB}x_m + \sum_{j\neq m}\mathbf{g}_{mb_j}^H\mathbf{w}_j^{CB}x_j + z_m, \tag{7}$$

where BS_{b_j} is the serving BS of MS_j, and the second term is the MUI for MS_m caused by the imperfect CSI.

Comparing (6) with (4), we can see that $\mathbf{w}_m^{CB}$ is almost the same as $\mathbf{w}_{mb}^{(b)}$ except for the allocated power, which however differentiates CoMP-CB from the decentralized CoMP-JP. In CoMP-CB, the data of the MSs are not shared among BSs, each BS only serves its local MSs, and generates nulls (i.e., transmits with zero powers) to the MSs in other cells. In decentralized CoMP-JP, by contrast, the data of all the MSs are shared among BSs but the CSI is actually unnecessary for sharing, and each BS serves all the MSs with the joint precoder shown in (3) with sub-vectors in (4).

3) Non-CoMP: In Non-CoMP systems, each BS serves the MSs in its own cell based on the local channels of these MSs. To avoid the MUI for the MSs in the same cell, each BS employs ZF precoding independently. The ZF precoder for MS_m at its serving BS, BS_{b_m}, is

$$\mathbf{w}_m^{NC} = \frac{\sqrt{p_m^{NC}}\hat{\mathbf{Q}}_m^{NC}\hat{\mathbf{g}}_{mb_m}}{\|\hat{\mathbf{Q}}_m^{NC}\hat{\mathbf{g}}_{mb_m}\|}, \tag{8}$$

where $\hat{\mathbf{Q}}_m^{NC} = \mathbf{I} - \bar{\mathbf{G}}_{NC}^H(\bar{\mathbf{G}}_{NC}\bar{\mathbf{G}}_{NC}^H)^{-1}\bar{\mathbf{G}}_{NC}$, $\bar{\mathbf{G}}_{NC} = [\hat{\mathbf{g}}_{S_1b} \cdots \hat{\mathbf{g}}_{S_sb}]^H$, $S_i \in \mathcal{S}_{b_m}, i = 1 \cdots s$, i.e., $\hat{\mathbf{Q}}_m^{NC}$ is the null space of the channels from BS_b to all the MSs in $\mathcal{S}_{b_m}$, and $\mathcal{S}_{b_m}$ is the set of the users located in cell b_m except MS_m.

The received signal at MS_m is

$$\begin{aligned} y_m &= \mathbf{g}_{mb_m}^H\mathbf{w}_m^{NC}x_m + \sum_{j\in\mathcal{S}_{b_m}}\mathbf{g}_{mb_m}^H\mathbf{w}_j^{NC}x_j \\ &\quad + \sum_{k\notin\mathcal{S}_{b_m},k\neq m}\mathbf{g}_{mb_k}^H\mathbf{w}_k^{NC}x_k + z_m, \end{aligned} \tag{9}$$

where in the right hand side the second term is the MUI for MS_m, and the third term is the ICI generated by BS_{b_k} when it transmits to MS_k, $k \notin \mathcal{S}_{b_m}$ and $k \neq m$.

III. Downlink Transmission Performance Analysis

In this section, we derive the average per-user data rates of CoMP-JP, CoMP-CB and Non-CoMP systems, respectively. From which we can show the different rate gain of channel prediction over channel estimation for CoMP and Non-CoMP.

A. Average Per-user Rate of CoMP-JP Systems

From the receive signal model in (5), the average rate of MS_m achieved by the joint ZF precoding of CoMP-JP is

$$\bar{R}_m^{JP} = \mathbb{E}\{\log_2(1+\frac{|\mathbf{g}_m^H\mathbf{w}_m^{JP}|^2}{\sigma^2+\sum_{j\neq m}|\mathbf{g}_m^H\mathbf{w}_j^{JP}|^2})\}. \quad (10)$$

In order to find an explicit expressions of the average per-user rate with respect to the NMSE of the channel prediction/estimation, we resort to large system analysis where the number of transmit antennas at each BS approaches infinity. We begin with deriving the power of the signal and interference.

○ Signal/MUI Power of Centralized CoMP-JP

For the *centralized CoMP-JP* systems, the powers of the desired signal and MUI are shown as follows.

Prop. 1: When $N_t \to \infty$, the random variable $|\mathbf{g}_m^H\mathbf{w}_{c,m}^{JP}|^2/N_t$ converges to a deterministic variable, i.e.,

$$\lim_{N_t\to\infty}\frac{|\mathbf{g}_m^H\mathbf{w}_{c,m}^{JP}|^2}{N_t} = p_m^{JP}\sum_{b=1}^{B}\alpha_{mb}(1-\bar{\varepsilon}_{mb}). \quad (11)$$

Proof: See Appendix A.

It indicates that the signal power increases linearly with the number of antennas.

Prop. 2: When $N_t \to \infty$, the power of the interference caused by MS_j to MS_m, $|\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}|^2$, has exponential distribution with parameter $1/\lambda_{c,mj}^{JP}$, whose *probability density function* (PDF) is

$$f_{mj}^{JP}(x) = \frac{1}{\lambda_{c,mj}^{JP}}\exp(-\frac{x}{\lambda_{c,mj}^{JP}}), x\geq 0, \quad (12)$$

where

$$\lambda_{c,mj}^{JP} = p_j^{JP}\sum_{b=1}^{B}\alpha_{mb}\bar{\varepsilon}_{mb}\beta_{jb}^{c} \quad (13)$$

is the average MUI power induced by MS_j to the desired user, MS_m, and

$$\beta_{jb}^{c} = \frac{\alpha_{jb}(1-\bar{\varepsilon}_{jb})}{\sum_{l=1}^{B}\alpha_{jl}(1-\bar{\varepsilon}_{jl})} \quad (14)$$

depends on the location of MS_j. Moreover, the MUI respectively caused by MS_j and MS_k, $\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}$ and $\mathbf{g}_m^H\mathbf{w}_{c,k}^{JP}$, are mutually independent.

Proof: See Appendix B.

○ Signal/MUI Power of Decentralized CoMP-JP

For the *decentralized CoMP-JP* systems, we can also prove two propositions.

Prop. 3: When $N_t \to \infty$, the random variable $|\mathbf{g}_m^H\mathbf{w}_{d,m}^{JP}|^2/N_t$ converges to a deterministic variable, i.e.,

$$\lim_{N_t\to\infty}\frac{|\mathbf{g}_m^H\mathbf{w}_{d,m}^{JP}|^2}{N_t} = \big(\sum_{b=1}^{B}\sqrt{p_m^{(b)}\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)})}\big)^2, \quad (15)$$

where $\bar{\varepsilon}_{mb}^{(b)}$ is the NMSE of $\hat{\mathbf{g}}_{mb}$ at BS_b.

Proof: See Appendix C.

Prop. 4: When $N_t \to \infty$, the power of the interference caused by MS_j to MS_m, $|\mathbf{g}_m^H\mathbf{w}_{d,j}^{JP}|^2$, has exponential distribution with parameter $1/\lambda_{d,mj}^{JP}$, whose PDF is (12) and

$$\lambda_{d,mj}^{JP} = \sum_{b=1}^{B}p_j^{(b)}\alpha_{mb}\bar{\varepsilon}_{mb}^{(b)}. \quad (16)$$

Proof: See Appendix D.

From (15) and (16), we can see that the signal and MUI powers in the decentralized CoMP-JP system depend on the allocated powers from multiple BSs to each MS, $p_m^{(b)}, b = 1,\cdots,M$. In the following, we show that the decentralized CoMP-JP will achieve the same performance as the centralized CoMP-JP if the power allocation is optimized as follows with the same total transmit power to each MS.

Consider a "selfish" power allocation problem that maximizes the signal power of MS_m under the total power constraint for the MS without considering the interference to the other MSs, which is

$$\max_{p_m^{(b)}} \quad \big(\sum_{b=1}^{B}\sqrt{p_m^{(b)}}\cdot\sqrt{\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)})}\big)^2 \quad (17a)$$

$$s.t. \quad \sum_{b=1}^{B}p_m^{(b)} = p_m^{JP}. \quad (17b)$$

From Cauchy-Schwarz inequality, the optimal solution of the problem is[5]

$$p_m^{(b)} = \frac{p_m^{JP}}{\sum_{j=1}^{B}\alpha_{mj}(1-\bar{\varepsilon}_{mj}^{(j)})}\cdot\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)}). \quad (18)$$

Substituting (18) into (15) and (16), we obtain

$$\begin{aligned}&\lim_{N_t\to\infty}\frac{|\mathbf{g}_m^H\mathbf{w}_{d,m}^{JP}|^2}{N_t}\\ =\ &\big(\sum_{b=1}^{B}\sqrt{\frac{p_m^{JP}\cdot\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)})}{\sum_{j=1}^{B}\alpha_{mj}(1-\bar{\varepsilon}_{mj}^{(j)})}\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)})}\big)^2\\ =\ &p_m^{JP}\sum_{b=1}^{B}\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)})\end{aligned}$$

and

$$\lambda_{d,mj}^{JP} = p_j^{JP}\sum_{b=1}^{B}\alpha_{mb}\bar{\varepsilon}_{mb}\beta_{jb}^{d},$$

where $\beta_{jb}^{d} = \frac{\alpha_{jb}(1-\bar{\varepsilon}_{jb}^{(b)})}{\sum_{l=1}^{B}\alpha_{jl}(1-\bar{\varepsilon}_{jl}^{(l)})}$.

Now the signal and MUI powers in the decentralized CoMP-JP system have the same expressions as the centralized system. This is valid when the number of antennas goes to infinity, the number of MSs is fixed, and the optimal power allocation in (18) is applied. Consequently, we do not distinguish the centralized and decentralized CoMP-JP in the following analysis.

[5]It is worth to note that such an optimal power allocation is hard to be implemented in practice, because each BS, say BS_b, has to know the NMSE of the estimated/predicted channels in other BSs, $\bar{\varepsilon}_{mj}^{(j)}$. The reason to present the optimal power allocation is to help understand the decentralized CoMP-JP, and to unify the analysis for centralized and decentralized CoMP-JP systems.

◦ **Asymptotic Average Per-user Rate of CoMP-JP**

We can see that the signal power grows linearly with the number of antennas, but the interference power is not. It implies that when $N_t \to \infty$, the signal to interference plus noise ratio (SINR) of MS_m is much higher than 1, and the term "1" inside the log function in (10) can be ignored. Then the asymptotic average achievable data rate of MS_m served by CoMP-JP can be obtained as follows

$$\begin{aligned} R_m^{JP_A} &\triangleq \lim_{N_t\to\infty} \bar{R}_m^{JP} \\ &= \lim_{N_t\to\infty} \mathbb{E}\{\log_2(\frac{|\mathbf{g}_m^H \mathbf{w}_m^{JP}|^2}{\sigma^2 + \sum_{j\neq m} |\mathbf{g}_m^H \mathbf{w}_j^{JP}|^2})\} \\ &= \lim_{N_t\to\infty} \mathbb{E}\{\log_2(|\mathbf{g}_m^H \mathbf{w}_m^{JP}|^2\} \\ &\quad - \lim_{N_t\to\infty} \mathbb{E}\{\log_2(\sigma^2 + \sum_{j\neq m} |\mathbf{g}_m^H \mathbf{w}_j^{JP}|^2)\} \\ &\overset{(a)}{=} \underbrace{\lim_{N_t\to\infty} \mathbb{E}\{\log_2(|\mathbf{g}_m^H \mathbf{w}_m^{JP}|^2\}}_{S_m^{JP}\text{:contributed by signal}} \\ &\quad \underbrace{- \mathbb{E}\{\log_2(\sigma^2 + \sum_{j\neq m} \lambda_{mj}^{JP}\xi_j)\}}_{I_m^{JP}\text{:contributed by MUI and noise}}, \end{aligned} \tag{19}$$

where $\xi_j, j \neq m$ are i.i.d. random vector of exponential distribution with parameter 1, and the equality (a) is because $|\mathbf{g}_m^H \mathbf{w}_j^{JP}|^2$ and $\lambda_{mj}^{JP}\xi_j$ have the same distribution.

The term I_m^{JP} in (19) is an increasing function of λ_{mj}^{JP}, because its first order derivative $\frac{\mathrm{d}I_m^{JP}}{\mathrm{d}\lambda_{mj}^{JP}} = \lim_{N_t\to\infty} \mathbb{E}\{\frac{\xi_j}{\ln 2(\sigma^2+\sum_{j\neq m}\lambda_{mj}^{JP}\xi_j)}\} > 0$.

According to the result in [21], we can obtain the cumulative distribution function (CDF) of $\sum_{j\neq m} \lambda_{mj}^{JP}\xi_j$ as

$$F(x) = 1 - \sum_{j\neq m} \gamma_j^{JP} \exp(-\frac{x}{\lambda_{mj}^{JP}}), x \geq 0, \tag{20}$$

where $\gamma_j^{JP} = 1/(\prod_{l\neq j,m}(1-\lambda_{ml}^{JP}/\lambda_{mj}^{JP}))$. Because $F(0) = 0$, we have $\sum_{j\neq m} \gamma_j^{JP} = 1$. With this CDF, the second term of (19) can be derived as

$$\begin{aligned} I_m^{JP} &= \lim_{N_t\to\infty} \mathbb{E}\{\log_2(\sigma^2 + \sum_{j\neq m} |\mathbf{g}_m^H \mathbf{w}_j^{JP}|^2)\} \\ &= \int_0^\infty \log_2(\sigma^2 + x) dF(x) \\ &= \log_2(\sigma^2 + x)(F(x) - 1)\Big|_0^\infty - \int_0^\infty \frac{F(x) - 1}{\ln 2(\sigma^2 + x)} dx \\ &= \log_2 \sigma^2 + \sum_{k\neq m} \gamma_k^{JP} \int_0^\infty \frac{\exp(-\frac{x}{\lambda_{mk}^{JP}})}{\ln 2(\sigma^2 + x)} dx \\ &= \log_2 \sigma^2 - \frac{1}{\ln 2} \sum_{k\neq m} \gamma_k^{JP} \exp(\frac{\sigma^2}{\lambda_{mk}^{JP}}) Ei(-\frac{\sigma^2}{\lambda_{mk}^{JP}}), \end{aligned} \tag{21}$$

where the last equality is obtained by using the formula $\int_0^\infty \frac{\exp(-\mu x)}{x+\beta} dx = -\exp(\beta\mu)Ei(-\beta\mu)$, $Ei(x) = \int_{-\infty}^x \frac{\exp(t)}{t} dt$ is the exponential integral function, and $Ei(x) < 0$ when $x < 0$.

With Prop. 1 and (21), the asymptotic average per-user rate of the CoMP-JP system shown in (19) becomes

$$\begin{aligned} R_m^{JP_A} &= \underbrace{\log_2(N_t p_m^{JP} \sum_{b=1}^B \alpha_{mb}(1 - \bar{\varepsilon}_{mb}))}_{S_m^{JP}\text{:contributed by signal}} \\ &\quad + \underbrace{\frac{1}{\ln 2} \sum_{k\neq m} \gamma_k^{JP} \exp(\frac{\sigma^2}{\lambda_{mk}^{JP}}) Ei(-\frac{\sigma^2}{\lambda_{mk}^{JP}}) - \log_2 \sigma^2}_{I_m^{JP}\text{:contributed by MUI and noise}} \end{aligned} \tag{22}$$

As will be shown in simulations later, the value of $R_m^{JP_A}$ is close to the value of $\bar{R}_m^{JP}$ even with finite N_t.

◦ **Impact of the Imperfect CSI**

In (22), S_m^{JP} depends on the average received signal power and the NMSE of the predicted/estimated global channel of MS_m. The imperfect CSI leads to a rate loss from reducing the received signal power. $I_m^{JP} < 0$ is the rate loss led by the MUI and the noise, which is a complicate function of the interference power caused by the co-scheduled MSs, λ_{mk}^{JP}. Because I_m^{JP} is an increasing function of each value of λ_{mj}^{JP}, we analyze the parameter λ_{mj}^{JP} in the following.

From λ_{mj}^{JP} in (13) we know that it depends on both α_{mb} and β_{jb}, i.e., the MUI of MS_m depends on its own location as well as those of its co-scheduled MSs. In the sequel, we briefly analyze the impact of the location of MSs on the performance of MS_m.

1) Location of the desired MS: When MS_m moves from cell edge to its master BS, BS_{b_m}, the large scale fading gain of its local channel α_{mb_m} will increase, while those of its cross channels $\alpha_{mb}, b \neq b_m$ will decrease. Since $\alpha_{mb_m} \gg \alpha_{mb}, b \neq b_m$ when MS_m is in cell center, we can see from (22) that the values of S_m^{JP} and I_m^{JP} will be dominated by the NMSE of the local channel prediction/estimation The NMSE of channel prediction $\bar{\varepsilon}_{mb_m} = \bar{\varepsilon}_{mb_m}^{CP}$ and the NMSE of channel estimation $\bar{\varepsilon}_{mb_m} = \bar{\varepsilon}_{mb_m}^{CE}$. In other words, given the location of the co-scheduled MSs, the reduction of NMSE of local channel prediction/estimation of the desired MS, $\bar{\varepsilon}_{mb_m}$, will improve the performance more significantly when the MS is close to the cell center.

2) Location of the co-scheduled MSs: If the desired MS_m is located at the "exact cell edge", where $\alpha_{m1} = \cdots = \alpha_{mB} = \alpha_m$, we have $\bar{\varepsilon}_{m1} = \cdots = \bar{\varepsilon}_{mB} = \bar{\varepsilon}_m$. In this case the interference experienced at MS_m does not depend on the large scale fading gains of its co-scheduled MSs, because $\sum_{b=1}^B \beta_{kb} = 1$ and from (13) we have

$$\lambda_{mj}^{JP} = p_j^{JP} \alpha_m \bar{\varepsilon}_m \sum_{b=1}^B \beta_{jb} = p_j^{JP} \alpha_m \bar{\varepsilon}_m. \tag{23}$$

When MS_m is located at any other positions, the value of λ_m^{JP} depends on β_{kb}, i.e., the MUI of the desired MS will largely depend on the location of its co-scheduled users.

When a co-scheduled user MS_j is located close to its master BS, BS_{b_j}, its local channel gain α_{jb_j} dominates β_{jb} as shown in (14), thus $\beta_{jb_j} \approx 1$ and $\beta_{jb} \approx 0, b \neq b_j$. Then from (13) the power of the interference at MS_m generated from the co-scheduled user MS_j is

$$\lambda_{mj}^{center} = p_j^{JP} \alpha_{mb_j} \bar{\varepsilon}_{mb_j}, \tag{24}$$

which only depends on the large scale fading gain of the cross channel between MS_m and BS_{b_j}.

When MS_j is located at the "exact cell edge" where $\alpha_{j1} = \cdots = \alpha_{jB}$, then $\beta_{j1} = \cdots = \beta_{jB} = \frac{1}{B}$. The interference power at MS_m caused from MS_j becomes

$$\lambda_{mj}^{edge} = \frac{p_j^{JP}}{B} \sum_{b=1}^{B} \alpha_{mb} \bar{\varepsilon}_{mb}, \tag{25}$$

which depends on the overall average receive channel power of MS_m.

B. Average Per-user Rate of CoMP-CB Systems

From (7), the average rate of MS_m achieved by the individual ZF precoding of CoMP-CB is

$$\bar{R}_m^{CB} = \mathbb{E}\{\log_2(1 + \frac{|\mathbf{g}_{mb_m}^H \mathbf{w}_m^{CB}|^2}{\sigma^2 + \sum_{j \neq m} |\mathbf{g}_{mb_j}^H \mathbf{w}_j^{CB}|^2})\}. \tag{26}$$

Analogous to the analysis for CoMP-JP systems, we start by deriving the powers of the desired signal and interference.

Prop. 5: When $N_t \to \infty$, the random variable $|\mathbf{g}_{mb_m}^H \mathbf{w}_m^{CB}|^2/N_t$ converges to a deterministic variable, i.e.,

$$\lim_{N_t \to \infty} \frac{|\mathbf{g}_{mb_m}^H \mathbf{w}_m^{CB}|^2}{N_t} = p_m^{CB} \alpha_{mb_m} (1 - \bar{\varepsilon}_{mb_m}). \tag{27}$$

Prop. 6: When $N_t \to \infty$, the interference power from MS_j to MS_m, $|\mathbf{g}_{mb_j}^H \mathbf{w}_j^{CB}|^2$, has an exponential distribution with parameter $1/\lambda_{mj}^{CB}$, whose PDF is

$$f_{mj}^{CB}(x) = \frac{1}{\lambda_{mj}^{CB}} \exp(-\frac{x}{\lambda_{mj}^{CB}}), x \geq 0, \tag{28}$$

where $\lambda_{mj}^{CB} = p_j^{CB} \alpha_{mb_j} \bar{\varepsilon}_{mb_j}$ is the average interference power.

The proofs for these two propositions are similar to those for Prop.1 and Prop. 2, therefore are omitted for conciseness.

In contrast to λ_{mj}^{JP} in (13), the average interference power experienced at the desired user MS_m from its co-scheduled user MS_j only depends on the NMSE of the cross channel of MS_j, $\hat{\mathbf{g}}_{mb_j}$. This is consistent with the intuition, because BS_j only employs $\hat{\mathbf{g}}_{mb_j}$ for computing its individual precoding to avoid creating interference to MS_m. Moreover, from (24) we can see that the average MUI caused by MS_j in CoMP-CB systems will equal to that in CoMP-JP systems if MS_j is located at cell center and $p_j^{CB} = p_j^{JP}$.

When the SINR of MS_m is much higher than 1, the term "1" inside the log function in (26) can be ignored. Then, by substituting (27) into (26), using (28) and after some regular manipulations as we derive $R_m^{JP_A}$, the asymptotic average achievable rate of MS_m served by CoMP-CB systems can be derived as follows,

$$R_m^{CB_A} \triangleq \underbrace{\log_2(N_t p_m \alpha_{mb_m}(1 - \bar{\varepsilon}_{mb_m}))}_{S_m^{CB}\text{:contributed by signal}} + \underbrace{\frac{1}{\ln 2} \sum_{k \neq m} \gamma_k^{CB} \exp(\frac{\sigma^2}{\lambda_{mk}^{CB}}) Ei(-\frac{\sigma^2}{\lambda_{mk}^{CB}}) - \log_2 \sigma^2}_{I_m^{CB}\text{:contributed by MUI}}, \tag{29}$$

where $\gamma_k^{CB} = 1/(\prod_{l \neq k,m}(1 - \lambda_{ml}^{CB}/\lambda_{mk}^{CB}))$.

C. Average Per-user Rate of Non-CoMP Systems

From (9), the average data rate achieved by MS_m by the independent ZF precoding in Non-CoMP is

$$\bar{R}_m^{NC} = \mathbb{E}\{\log_2(1 + \frac{|\mathbf{g}_{mb_m}^H \mathbf{w}_m^{NC}|^2}{\sigma^2 + \sum_{j \in \mathcal{S}_{b_m}} |\mathbf{g}_{mb_m}^H \mathbf{w}_j^{NC}|^2 + \sum_{k \notin \mathcal{S}_{b_m}, k \neq m} |\mathbf{g}_{mb_k}^H \mathbf{w}_k^{NC}|^2}\}. \tag{30}$$

The powers of the desired signal and interference are shown in the following two propositions.

Prop. 7: When $N_t \to \infty$, the random variable $|\mathbf{g}_{mb_m}^H \mathbf{w}_m^{NC}|^2/N_t$ converges to a deterministic variable, i.e.,

$$\lim_{N_t \to \infty} \frac{|\mathbf{g}_{mb_m}^H \mathbf{w}_m^{NC}|^2}{N_t} = p_m^{NC} \alpha_{mb_m} (1 - \bar{\varepsilon}_{mb_m}), \tag{31}$$

and the MUI, $|\mathbf{g}_{mb_j}^H \mathbf{w}_j^{NC}|^2, j \in \mathcal{S}_{b_m}$, is subject to exponential distribution with parameter $\lambda_{MUI,j}^{NC}$, where $\lambda_{MUI,j}^{NC} = p_j^{NC} \alpha_{mb_m} \bar{\varepsilon}_{mb_m}$ is the average MUI power caused by MS_j who is located in the same cell with MS_m.

Prop. 8: When $N_t \to \infty$, the ICI power experienced by the desired user MS_m, $|\mathbf{g}_{mb_k}^H \mathbf{w}_k^{NC}|^2$, $k \notin \mathcal{S}_{b_m}$ and $k \neq m$, is subject to an exponential distribution with parameter $\lambda_{ICI,k}^{NC}$, where $\lambda_{ICI,k}^{NC} = p_k^{NC} \alpha_{mb_k}$ is the average ICI power caused by MS_k who is not located in the same cell with MS_m.

The proofs for these two propositions are similar to those for Prop.1 and Prop. 2, therefore are omitted for conciseness.

Comparing Prop. 7 with Prop. 5, we can see that the asymptomatic signal power of Non-CoMP is identical to that of CoMP-CB, and $\lambda_{MUI,j}^{NC}$ will equal to λ_{mj}^{CB} if $p_j^{NC} = p_j^{CB}$. It is shown from Prop. 8 that $\lambda_{ICI,k}^{NC}$ does not depend on the NMSE of predicted/estimated cross channel. This is because in Non-CoMP systems each BS does not predict/estimate the channels to the MSs located in other cells.

Analogous to the derivations for CoMP-JP, after some regular manipulations we obtain the interference and noise term in the average rate in Non-CoMP systems as

$$I_m^{NC} = -\log \sigma^2 + \frac{1}{\ln 2} \sum_{k \in \mathcal{S}_{b_m}} \gamma_k^{NC} \exp(\frac{\sigma^2}{\lambda_{mk}^{NC}}) Ei(-\frac{\sigma^2}{\lambda_{mk}^{NC}}) + \frac{1}{\ln 2} \sum_{k \notin \mathcal{S}_{b_m}, k \neq m} \gamma_k^{NC} \exp(\frac{\sigma^2}{\lambda_{mk}^{NC}}) Ei(-\frac{\sigma^2}{\lambda_{mk}^{NC}}), \tag{32}$$

where $\gamma_k^{NC} = 1/(\prod_{l \neq k,m}(1 - \lambda_{ml}^{NC}/\lambda_{mk}^{NC}))$, $\lambda_{mj}^{NC} = \lambda_{MUI,j}^{NC} = p_j^{NC} \alpha_{mb_j} \bar{\varepsilon}_{mb_j}$ is the average MUI when $j \in \mathcal{S}_{b_m}$, and $\lambda_{mk}^{NC} = \lambda_{ICI,k}^{NC} = p_k^{NC} \alpha_{mb_k}$ is the average ICI when $k \notin \mathcal{S}_{b_m}$ and $k \neq m$.

Considering that $\lambda_{MUI,j}^{NC} = p_j^{NC} \alpha_{mb_m} \bar{\varepsilon}_{mb_m} \ll \lambda_{ICI,k}^{NC} = p_k^{NC} \alpha_{mb_k}$ due to the using of the independent ZF precoding, we have

$$\gamma_k^{NC} = 1/(\prod_{l \neq k,m} (1 - \lambda_{ml}^{NC}/\lambda_{mk}^{NC})) \approx 0, \quad k \in \mathcal{S}_{b_m}. \tag{33}$$

Substituting (33) into (32), we have

$$I_m^{NC} \approx \frac{1}{\ln 2} \sum_{k \notin \mathcal{S}_{b_m}, k \neq m} \gamma_k^{NC} \exp(\frac{\sigma^2}{\lambda_{mk}^{NC}}) Ei(-\frac{\sigma^2}{\lambda_{mk}^{NC}}) - \log \sigma^2, \tag{34}$$

which reflects the impact of the ICI and noise.

Again, assuming that the SINR experienced at MS_m is much larger than 1, we obtain an approximated asymptotic average data rate of MS_m in Non-CoMP as follows

$$R_m^{NC_A} \triangleq \underbrace{\log_2(N_t p_m \alpha_{mb_m}(1-\bar{\varepsilon}_{mb_m}))}_{S_m^{NC}\text{:contributed by signal}} + I_m^{NC}. \quad (35)$$

Note that in Non-CoMP systems, the NMSE of the channel prediction/estimation only affects the average per-user rate loss induced by the power reduction of the desired signal.

D. Rate Gain of Channel Prediction Over Channel Estimation

For mathematical tractability, we rewrite the exponential integral function as

$$Ei(x) = \ln|x| + \sum_{k=1}^{\infty} \frac{x^k}{k \cdot k!} + C, x \neq 0, \quad (36)$$

where $C = 0.5772$ is Euler-Mascheroni constant.

Substituting (13) and (36) into (22), the asymptotic average per-user rate of CoMP-JP can be obtained as follows

$$\begin{aligned} & R_m^{JP_A} \\ = \; & \log_2(N_t p_m^{JP} \sum_{b=1}^{B} \alpha_{mb}(1-\bar{\varepsilon}_{mb})) - \log_2 \sigma^2 + \frac{1}{\ln 2} \\ & \sum_{k \neq m} \gamma_k^{JP} \exp(\frac{\sigma^2}{\lambda_{mk}^{JP}})(\ln \frac{\sigma^2}{\lambda_{mk}^{JP}} + \sum_{n=1}^{\infty} \frac{(-\sigma^2/)^n}{n \cdot n!} + C) \\ = \; & \log_2(N_t p_m^{JP} \sum_{b=1}^{B} \alpha_{mb}(1-\bar{\varepsilon}_{mb})) - \log_2 \sigma^2 \\ & + \frac{1}{\ln 2} \sum_{k \neq m} \gamma_k^{JP} (\ln \frac{\sigma^2}{\lambda_{mk}^{JP}} + C) + \frac{1}{\ln 2} \sum_{k \neq m} \gamma_k^{JP} \\ & (\sum_{n=1}^{\infty} \frac{(-\sigma^2/\lambda_{mk}^{JP})^n}{n \cdot n!} + (\exp(\frac{\sigma^2}{\lambda_{mk}^{JP}}) - 1) Ei(-\frac{\sigma^2}{\lambda_{mk}^{JP}})) \\ \overset{(a)}{=} \; & \log_2(N_t p_m^{JP} \sum_{b=1}^{B} \alpha_{mb}(1-\bar{\varepsilon}_{mb})) - N_m^{JP} \\ & \underbrace{- \sum_{k \neq m} \gamma_k^{JP} \log_2(p_k^{JP} \sum_{b=1}^{B} \alpha_{mb} \bar{\varepsilon}_{mb} \beta_{kb})}_{I_{JP}^{UB}} + \frac{C}{\ln 2}, \end{aligned} \quad (37)$$

where the equality (a) holds because $\sum_{j \neq m} \gamma_j^{JP} = 1$ as shown below (20), and $N_m^{JP} \triangleq \frac{1}{\ln 2} \sum_{k \neq m} \gamma_k^{JP} (\frac{\sigma^2}{\lambda_{mk}^{JP}} \sum_{n=1}^{\infty} \frac{(-\sigma^2/\lambda_{mk}^{JP})^{n-1}}{n \cdot n!} - (\exp(\frac{\sigma^2}{\lambda_{mk}^{JP}}) - 1) Ei(-\frac{\sigma^2}{\lambda_{mk}^{JP}}))$.

With similar derivation, we can derive the asymptotic average per-user rates of CoMP-CB and Non-CoMP as follows

$$\begin{aligned} R_m^{CB_A} &= \log_2(N_t p_m^{CB} \alpha_{mb_m}(1-\bar{\varepsilon}_{mb_m})) + \frac{C}{\ln 2} \\ & \underbrace{- \sum_{k \neq m} \gamma_k^{CB} \log_2(p_k^{CB} \alpha_{mb_k} \bar{\varepsilon}_{mb_k})}_{I_{CB}^{UB}} - N_m^{CB} \end{aligned} \quad (38)$$

$$\begin{aligned} R_m^{NC_A} &= \log_2(N_t p_m^{NC} \alpha_{mb_m}(1-\bar{\varepsilon}_{mb_m})) + \frac{C}{\ln 2} \\ & \underbrace{- \sum_{k \notin \mathcal{S}_{b_m}} \gamma_k^{NC} \log_2(p_k^{NC} \alpha_{mb_k})}_{I_{ICI}^{UB}} - N_m^{NC}, \end{aligned} \quad (39)$$

where $N_m^{CB} \triangleq \sum_{k \neq m} \gamma_k^{CB} (\frac{\sigma^2}{\lambda_{mk}^{CB}} \sum_{n=1}^{\infty} \frac{(-\sigma^2/\lambda_{mk}^{CB})^{n-1}}{n \cdot n!} - (\exp(\frac{\sigma^2}{\lambda_{mk}^{CB}}) - 1) Ei(-\frac{\sigma^2}{\lambda_{mk}^{CB}}))$, and $N_m^{NC} \triangleq \sum_{k \neq m} \gamma_k^{NC} (\frac{\sigma^2}{\lambda_{mk}^{NC}} \sum_{n=1}^{\infty} \frac{(-\sigma^2/\lambda_{mk}^{NC})^{n-1}}{n \cdot n!} - (\exp(\frac{\sigma^2}{\lambda_{mk}^{NC}}) - 1) Ei(-\frac{\sigma^2}{\lambda_{mk}^{NC}}))$.

Because when $N_t \to \infty$, the asymptotic average rates are dominated by the interference terms and the noise terms are negligible, we have $N_m^{JP} \approx 0$, $N_m^{CB} \approx 0$ and $N_m^{NC} \approx 0$.

In the sequel, we analyze the performance gain of using channel prediction over using channel estimation in the considered systems. To obtain explicit expressions for ease of comparison, we consider that the desired MS_m is located at the "exact cell edge", i.e., $\alpha_{m1} = \cdots = \alpha_{mB} = \alpha_m$. This is the worst case, because the interference caused by the outdated CSI is most severe for the MS. Then, we have $\bar{\varepsilon}_{m1}^{CE} = \cdots = \bar{\varepsilon}_{mB}^{CE} = \bar{\varepsilon}_m^{CE}$ and $\bar{\varepsilon}_{m1}^{CP} = \cdots = \bar{\varepsilon}_{mB}^{CP} = \bar{\varepsilon}_m^{CP}$. From (23), we obtain

$$\frac{(\lambda_{mj}^{JP})^{CE}}{(\lambda_{mk}^{JP})^{CE}} = \frac{p_j \alpha_m \bar{\varepsilon}_m^{CE}}{p_k \alpha_m \bar{\varepsilon}_m^{CE}} = \frac{p_j}{p_k} = \frac{(\lambda_{mj}^{JP})^{CP}}{(\lambda_{mk}^{JP})^{CP}}. \quad (40)$$

From the definition of γ_j^{JP} in (20), we have

$$\begin{aligned} (\gamma_j^{JP})^{CE} &= 1/(\prod_{l \neq j,m} (1 - (\lambda_{ml}^{JP})^{CE}/(\lambda_{mj}^{JP})^{CE})) \\ &= (\gamma_j^{JP})^{CP}. \end{aligned} \quad (41)$$

For notational simplicity, in the following we omit the superscript in $(\gamma_j^{JP})^{CE}$ and $(\gamma_j^{JP})^{CP}$, and write them as γ_j^{JP}.

For CoMP-JP systems, the gain of the average per-user rate with channel prediction over channel estimation can be derived by considering $N_m^{JP} \approx 0$ as follows

$$\begin{aligned} \Delta R_m^{JP} &= R_m^{JP_A}(\bar{\varepsilon}_m^{CP}) - R_m^{JP_A}(\bar{\varepsilon}_m^{CE}) \\ &\approx \underbrace{\log_2(\frac{N_t p_m^{JP} \sum_{b=1}^{B} \alpha_m (1-\bar{\varepsilon}_m^{CP})}{N_t p_m^{JP} \sum_{b=1}^{B} \alpha_m (1-\bar{\varepsilon}_m^{CE})})}_{\text{contribution from signal}} \\ &+ \underbrace{\sum_{k \neq m} \gamma_k^{JP} \log_2(\frac{p_k^{JP} \alpha_m \bar{\varepsilon}_m^{CE}}{p_k^{JP} \alpha_m \bar{\varepsilon}_m^{CP}})}_{\text{contribution from interference}} \\ &= \log_2(\frac{1-\bar{\varepsilon}_m^{CP}}{1-\bar{\varepsilon}_m^{CE}}) + \sum_{k \neq m} \gamma_k^{JP} \log_2(\frac{\bar{\varepsilon}_m^{CE}}{\bar{\varepsilon}_m^{CP}}) \\ &= \log_2(\frac{1-\bar{\varepsilon}_m^{CP}}{1-\bar{\varepsilon}_m^{CE}}) + \log_2(\frac{\bar{\varepsilon}_m^{CE}}{\bar{\varepsilon}_m^{CP}}). \end{aligned} \quad (42)$$

For CoMP-CB systems, the gain of the average per-user rate with channel prediction can be similarly derived by considering $N_m^{CB} \approx 0$ as

$$\begin{aligned}\Delta R_m^{CB} &= R_m^{CB_A}(\bar{\varepsilon}_m^{CP}) - R_m^{CB_A}(\bar{\varepsilon}_m^{CE}) \\ &\approx \underbrace{\log_2\left(\frac{N_t p_m^{CB}\alpha_m(1-\bar{\varepsilon}_m^{CP})}{N_t p_m^{CB}\alpha_m(1-\bar{\varepsilon}_m^{CE})}\right)}_{\text{contribution from signal}} \\ &+ \underbrace{\sum_{k\neq m}\gamma_k^{CB}\log_2\left(\frac{p_k^{CB}\alpha_m\bar{\varepsilon}_m^{CE}}{p_k^{CB}\alpha_m\bar{\varepsilon}_m^{CP}}\right)}_{\text{contribution from interference}} \\ &= \log_2\left(\frac{1-\bar{\varepsilon}_m^{CP}}{1-\bar{\varepsilon}_m^{CE}}\right) + \log_2\left(\frac{\bar{\varepsilon}_m^{CE}}{\bar{\varepsilon}_m^{CP}}\right) \\ &= \Delta R_m^{JP}. \end{aligned} \quad (43)$$

For Non-CoMP systems, the gain can be derived by considering $N_m^{NC} \approx 0$ as

$$\begin{aligned}\Delta R_m^{NC} &= R_m^{NC_A}(\bar{\varepsilon}_m^{CP}) - R_m^{NC_A}(\bar{\varepsilon}_m^{CE}) \\ &\approx \underbrace{\log_2\left(\frac{N_t p_m^{NC}\alpha_{mb_m}(1-\bar{\varepsilon}_m^{CP})}{N_t p_m^{NC}\alpha_{mb_m}(1-\bar{\varepsilon}_m^{CE})}\right)}_{\text{contribution from signal}} \\ &= \log_2\left(\frac{1-\bar{\varepsilon}_m^{CP}}{1-\bar{\varepsilon}_m^{CE}}\right). \end{aligned} \quad (44)$$

Comparing (42), (43) with (44), we can see that ΔR_m^{NC} equals to the first terms in ΔR_m^{JP} and ΔR_m^{CB}. Since $\bar{\varepsilon}_{mb}^{CP} < \bar{\varepsilon}_{mb}^{CE}$ in time varying channels, it implies that CoMP systems benefit more significantly from channel prediction than Non-CoMP systems.

This can be explained as follows. For Non-CoMP systems, we can observe from (35) or (39) that using channel prediction can retrieve the power loss of the desired signal, but cannot retrive the rate loss induced by the ICI. For CoMP systems, on the contrary, using channel prediction retrieves the rate loss caused both by the power reduction on the desired signal and by the MUI, as shown in (37) and (38). Since interference is more detrimental than power loss in cellular systems, using channel prediction can improve the performance of CoMP systems remarkably.

Though these results are obtained from asymptotic analysis when the MSs are located at the exact cell-edge, the simulations later will show that the conclusion is still valid for general cases with finite number of antennas.

IV. Simulation and Numerical Results

In this section, we first verify our previous analysis via simulations, and then show the gain of channel prediction for CoMP systems with large backhaul latency by comparing with Non-CoMP systems.

A. Simulation and Parameters Setup

Consider CoMP systems with a BS-to-BS distance of 500 m. Each BS transmits with a maximal power of 40 W and with 10 MHz bandwidth. Each MS transmits with a maximal power of 0.2 W with 180 kHz bandwidth. The carrier frequency is 2 GHz. The path loss exponent is 3.76, the average power loss at the reference distance of 1 m is 36.3 dB, the minimum distance between the MSs and each BS is 35 m, and 8 dB shadowing is considered. This simulation setup is based on [22].

We consider a realistic time-varying channel model, Jakes Model, which is widely applied in various standardization organizations. Its temporal correlation function is $R_h(\tau) = J_0(2\pi f_d \tau)$, where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, and f_d is the Doppler spread.

When simulating the average per-user rate achieved by the considered precoders, we use Wiener filter with finite coefficients of Q because in practical systems the number of training symbols is limited, where $Q = 5$. Specifically, each BS employs five received uplink training symbols respectively in the successive uplink frames to predict the downlink channel [11], and uses five received training symbols in each uplink frame to estimate the channel for next downlink frame. The speed of the MSs is 3 km/h or 10 km/h. Each MS is located in a cell edge region, where the ratio of the large scale fading gain of its local channel to the sum of those of the cross channels, i.e., $\alpha_{mb_m}/\sum_{b\neq b_m}\alpha_{mb}$ for MS_m, is less than a certain value.

Unless otherwise specified, we consider the following setting. The cell edge SNR is 10 dB, which is the average downlink receive SNR at the MS with 250 m distance from each BS. We consider i.i.d. Rayleigh fading channels, and the simulation results are obtained from 100 small scale channel realizations. We consider equal power allocation among the users. The backhaul latency $T = 7.5$ ms, and the delay between the locally estimated channels and the realistic channels during downlink transmission $\tau = 5$ ms.

Remind that when the channel estimation method is applied, in the decentralized CoMP-JP (with legend "Dec-JP") with $\mathbf{w}_{mb}^{(b)}$ in (4) and optimal power allocation in (18), CoMP-CB (with legend "CB") and Non-CoMP systems, the outdated CSI is caused by the delay between training and transmission, and in centralized CoMP-JP (with legend "Cen-JP") systems, the outdated CSI is caused by both the backhaul latency and delay between training and transmission. The overall CSI delays for these systems are listed in Table I.

TABLE I: The overall CSI delays in different systems

	Cen-JP	Dec-JP	CB	Non-CoMP
Δ	$2T+\tau$	τ	τ	τ

The NMSE of the channel prediction and channel estimation for all the numerical results with various values of Δ is computed from [11], where $Q = 5$. Because the NMSE of the predicted channels for the centralized and decentralized CoMP-JP systems differ, their average per-user rates are different, although in Section III A we have used a unified form to express their asymptotic average per-user rates.

B. Validation of the Analytical Analysis

To validate the asymptotic results in Section III derived for $N_t \to \infty$, Fig. 2 shows the simulated average per-user rate and the numerical results obtained from (22), (29) and (35) with finite N_t. To validate the impact of ignoring the noise terms on ΔR_m^{JP}, ΔR_m^{CB} and ΔR_m^{NC}, Table II shows the simulated per-user average rate and the numerical results obtained from (37), (38) and (39) with $N_m^{JP} = 0$, $N_m^{CB} = 0$ and $N_m^{NC} = 0$. We

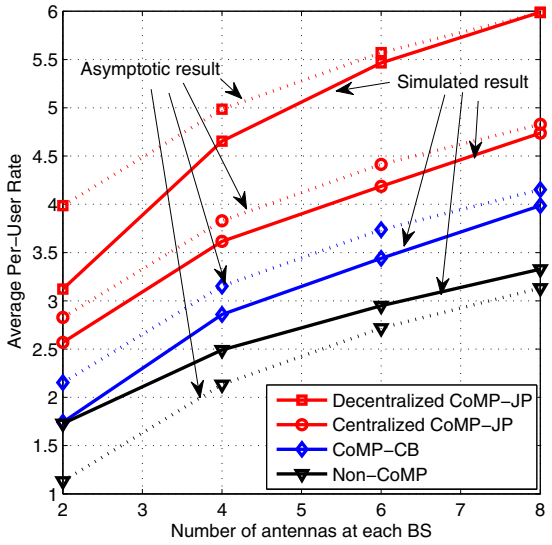


Fig. 2: Simulated average per-user rate and numerical results of the asymptotic average per-user rate of CoMP-JP, CoMP-CB and Non-CoMP systems. $B = M = 2$.

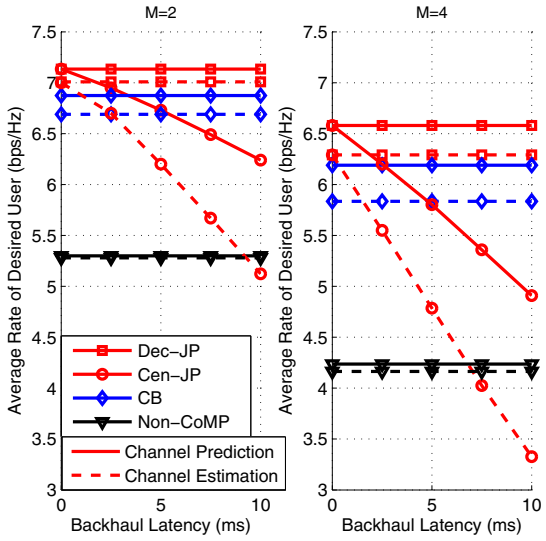


Fig. 3: Numerical results of the asymptotic average per-user rate when channel prediction or channel estimation is used. $B = 2$, $M = 2, 4$ and $N_t = 8$.

consider two MSs located on the line connecting the two BSs, the speed of each MS is 3 km/h. The distance between each MS and its local BS is 240 m, which corresponds to a 1 dB cell edge region. From the figure we can see that the asymptotic rates of both centralized and decentralized CoMP-JP systems converge to the simulation result faster than CoMP-CB and Non-CoMP. This is because the CU of CoMP-JP systems is actually a "super BS" with BN_t antennas. Both the figure and the table show that the numerical results are close to the simulated results. The results for the MSs with other speeds such as with 10 km/h are similar, which are not shown for conciseness.

TABLE II: Simulation results and Numerical results of the approximations by ignoring noise. Cell edge SNR= 30 dB, $B = M = 2$ and $N_t = 8$.

Average Per-User Rate	Cen-JP	CB	Non-CoMP
Approximations (bps/Hz)	5.71	5.11	3.81
Simulated results (bps/Hz)	5.62	4.91	3.63

C. Numerical Results of Rate Gain of Channel Prediction Over Channel Estimation

In the sequel we respectively show the asymptotic average rate of the desired MS versus the backhaul latency T (where τ is the same as in Section IV.A) and $f_d\Delta$ when channel prediction or channel estimation is employed, which are numerically obtained from (22), (29) and (35). The desired user is located in cell edge, whose distance to its local BS is 200 m.

Figure 3 is the average rate versus T. The speed of each MS is 3 km/h. The distances from the co-scheduled user(s) to their local BSs are 240 m. The rate gain of the centralized CoMP-JP system is significant compared to others because the CSI delay in this kind of system is the largest. Moreover, the performance gain increases with the number of MSs M in each system, because the the MUI power caused by the imperfect CSI increases with M. By contrast, the Non-CoMP system almost has no performance gain with the channel prediction.

Figure 4 is the average rate versus $f_d\Delta$. To differentiates the impact of the backhaul latency from the uplink training delay, we set $T = 0$ ms, which reflects a perfect backhaul with no latency. Then all the systems have identical CSI delay $\Delta = \tau$, and the centralized and decentralized CoMP-JP systems achieve the same performance as we analyzed in Section III A. To illustrate the impact of the location of the co-scheduled MSs, the distance between the co-scheduled user and its local BS is 50 m (cell center co-scheduled MS) or 240 m (cell edge co-scheduled MS).

When the values of $f_d\Delta$ are very low, e.g., equals zero, there is no performance gain of channel prediction over the channel estimation, as expected. In this scenario of static channel, all CoMP systems is superior to Non-CoMP system significantly.

When the values of $f_d\Delta$ are high, for cell edge user CoMP-JP slightly outperforms Non-CoMP when using channel prediction, but is inferior to Non-CoMP when using channel estimation. This is because without backhaul latency the only difference of CoMP-JP with Non-CoMP is the channel statistics. When all the users are located in cell edge, their CoMP channels are approximately i.i.d., which are similar to Non-CoMP channels. When the co-scheduled MS is located in cell center, CoMP-JP always outperforms Non-CoMP. This can be explained by the MUI power analysis in (24).

Comparing Figs. 3 and 4, we can see that when the CSI delay is also led by the backhaul latency, the gain of channel prediction over channel estimation is distinct from those with perfect backhaul, which largely depends on the specific manner of the BS cooperation.

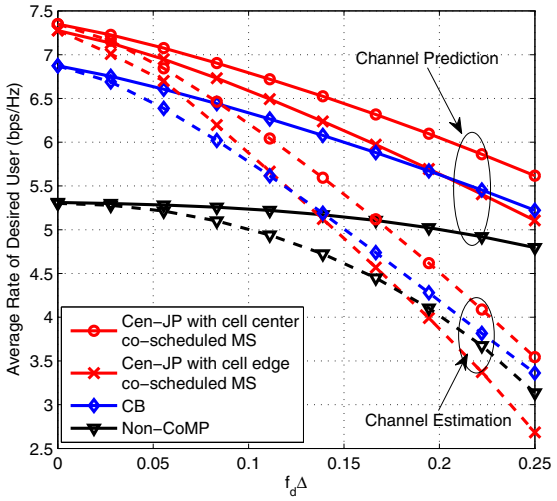
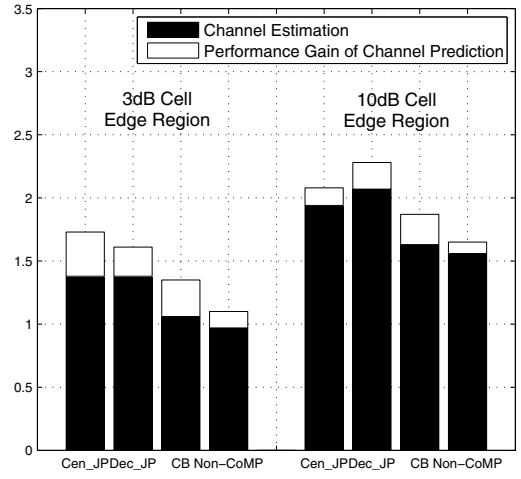
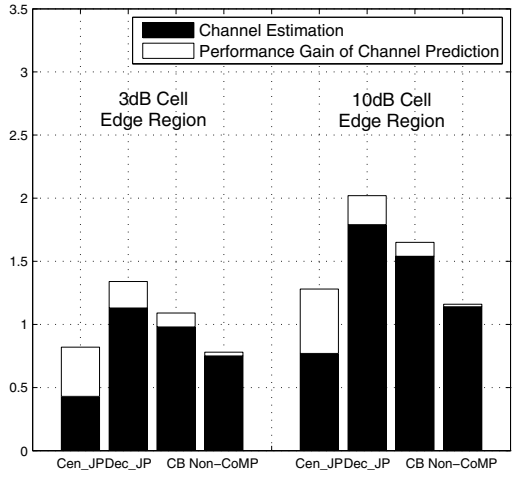


Fig. 4: Numerical results of the asymptotic average per-user rate when channel prediction or channel estimation is used. $B = M = 2$ and $N_t = 8$.



(a) MS's speed of 3 km/h



(b) MS's speed of 10 km/h

Fig. 5: Simulated average per-user rates. $B = 3$, $M = 6$ and $N_t = 8$. The height of the black bar is the average per-user rate with channel estimation, and the overall height of the black and white bars in each case is the average per-user rate with channel prediction.

D. Simulation of the Rate Gain of Channel Prediction

Finally, we simulate a more realistic scenario. We consider 19 hexagonal cells with the wrap around topology in order to eliminate boundary effects, where each of the three neighboring BSs is cooperative without overlapping. 10 candidate MSs are randomly placed in each cell within a 3 or 10 dB cell edge region. Considering that the user scheduling for CoMP-CB and for decentralized CoMP-JP are on-going research topics, we only study the impact of the imperfect CSI on precoding, i.e., on beamforming and power allocation. Therefore, we consider that each BS selects two MSs with a Round-Robin scheduling. PBPC is considered for CoMP systems. For the centralized and decentralized CoMP-JP, the precoders are respectively obtained from (2), (3) and (4), which consist of the joint ZFBF together with a power allocation that maximizes the sum rate of the six MSs under PBPC [23]. For CoMP-CB, the power allocation in the individual ZF precoder is obtained by maximizing the sum rate of the two MSs in each cell under PBPC. All the results are obtained from 1000 random places for each MS, and for different places the MSs undergo independent Rayleigh fading channels. The simulation results are shown in Fig. 5. We can see that the previous conclusions are still valid. Comparing the results for the cases where the users are located in a 3 and 10 dB cell edge region, we can find that the performance gain of channel prediction is more pronounced when the channels are more asymmetric, i.e., the large scale channel gains are more different.

The following observations can be obtained. (1) CoMP systems always outperform Non-CoMP systems when channel prediction is applied, which is not true when channel estimation is applied. (2) The performance gains of channel prediction over channel estimation for CoMP systems are much larger than the gain for Non-CoMP systems, which depends on the speed and location of the MSs. (3) When channel prediction is applied, the centralized CoMP-JP achieves the highest average per-user rate only when the MSs are with low speed and located in the 3 dB cell edge region, while the decentralized CoMP-JP will perform the best if the MSs are with more asymmetric channels. Comparing the decentralized CoMP-JP with CoMP-CB which employ the same CSI with the same delay, it is interesting to see that the decentralized CoMP-JP is superior to CoMP-CB.

Note that we have assumed i.i.d. small scale fading channels in previous analysis. In existing systems, the channels from each BS to each user are usually spatially correlated [22]. Moreover, the temporal correlation usually depends on the spatial correlation, which can be exploited to predict the MIMO channels with shorter channel coherent time [12]. This implies that a much higher user speed than 10 km/h can be supported for CoMP in prevalent 3GPP systems.

V. Conclusion

In this paper, we investigated the performance gain of channel prediction over channel estimation for centralized and decentralized CoMP-JP, CoMP-CB and Non-CoMP systems. Analytical and simulation results showed that CoMP systems benefit more significantly from channel prediction than Non-CoMP systems, where the performance gain largely depends on the way of cooperation and the locations of the users. When channel estimation is used for the users with higher speed, the centralized CoMP-JP will perform the worst and be even inferior to Non-CoMP. When channel prediction is used, CoMP systems always outperform Non-CoMP systems. In general, the performance of CoMP systems will be improved remarkably even with large backhaul latency whenever the channels are predictable. Considering that CoMP systems are expected to serve low mobility users, our results suggest that the current backhaul may be sufficient for CoMP in terms of the latency. This is very attractive since upgrading the deployed backhaul is of high cost.

Appendix A
Proof of Prop.1

The global channel vector of MS$_m$ can be expressed by its estimation or prediction as follows,

$$\mathbf{g}_m(t) = \hat{\mathbf{g}}_m(t) + \mathbf{e}_m(t), \tag{A.1}$$

where $\hat{\mathbf{g}}_m(t)$ is uncorrelated with $\mathbf{e}_m(t)$, for channel estimation $\hat{\mathbf{g}}_m(t) = \rho\hat{\mathbf{g}}_m(t-q)$, ρ is the temporal correlation between $\mathbf{g}(t)$ and $\mathbf{g}(t-q)$ and q is the delay between uplink training and downlink transmission. The elements of $\mathbf{e}_m(t)$ are zero mean and $\mathbb{E}\{\mathbf{e}_m(t)\mathbf{e}_m^H(t)\} = \mathrm{diag}(\varepsilon_{m1}\mathbf{I}\cdots\varepsilon_{mB}\mathbf{I})$.

For notational simplicity, we omit the time variable of the time-varying channels. Further considering the precoder of MS$_m$ shown in (2), we can obtain

$$\lim_{N_t\to\infty}\frac{|\mathbf{g}_m^H\mathbf{w}_{c,m}^{JP}|^2}{N_t} = \lim_{N_t\to\infty}\frac{p_m^{JP}|(\hat{\mathbf{g}}_m^H+\mathbf{e}_m^H)\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m|^2}{N_t\|\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m\|^2}. \tag{A.2}$$

Noticing that $\hat{\mathbf{Q}}_{c,m}$ is the null space spanned by the global channel vectors of all the MSs except MS$_m$, and $(\hat{\mathbf{Q}}_{c,m})^H\cdot\hat{\mathbf{Q}}_{c,m} = \hat{\mathbf{Q}}_{c,m}$, we can rewrite (A.2) as

$$\begin{aligned}\lim_{N_t\to\infty}\frac{|\mathbf{g}_m^H\mathbf{w}_{c,m}^{JP}|^2}{N_t} &= \lim_{N_t\to\infty}\frac{p_m^{JP}}{N_t}(\hat{\mathbf{g}}_m^H\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m \\ &+\mathbf{e}_m^H\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m+\hat{\mathbf{g}}_m^H\hat{\mathbf{Q}}_{c,m}\mathbf{e}_m+\frac{|\mathbf{e}_m^H\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m|^2}{\hat{\mathbf{g}}_m^H\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m}).\end{aligned} \tag{A.3}$$

Considering (A.1) and the fact that $\mathbf{g}_m$ is the global channel of MS$_m$ that is independent from the global channels of other users, from the law of large numbers we obtain

$$\lim_{N_t\to\infty}\frac{\hat{\mathbf{g}}_m^H\hat{\mathbf{g}}_j}{N_t} = \begin{cases}\sum_{b=1}^B(\alpha_{mb}-\varepsilon_{mb}), & \text{if } m=j,\\ 0, & \text{if } m\neq j,\end{cases} \tag{A.4}$$

$$\lim_{N_t\to\infty}\frac{\hat{\mathbf{e}}_m^H\hat{\mathbf{g}}_j}{N_t} = 0. \tag{A.5}$$

From the definition of $\mathbf{Q}_{c,m}$ we can rewrite it as

$$\mathbf{Q}_{c,m} = \mathbf{I} - \frac{1}{N_t}\bar{\mathbf{G}}_{c,m}^H(\frac{1}{N_t}\bar{\mathbf{G}}_{c,m}\bar{\mathbf{G}}_{c,m}^H)^{-1}\bar{\mathbf{G}}_{c,m}, \tag{A.6}$$

where $\bar{\mathbf{G}}_{c,m} = [\hat{\mathbf{g}}_1\cdots\hat{\mathbf{g}}_{m-1}\hat{\mathbf{g}}_{m+1}\cdots\hat{\mathbf{g}}_M]^H$.

From (A.4), we have

$$\begin{aligned}&\lim_{N_t\to\infty}\frac{1}{N_t}\bar{\mathbf{G}}_{c,m}\bar{\mathbf{G}}_{c,m}^H \\ &= \begin{bmatrix}\sum_{b=1}^B(\alpha_{1b}-\varepsilon_{1b}) & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sum_{b=1}^B(\alpha_{Mb}-\varepsilon_{Mb})\end{bmatrix},\end{aligned} \tag{A.7}$$

and then from (A.6) we obtain

$$\lim_{N_t\to\infty}\mathbf{Q}_{c,m} = \mathbf{I} - \frac{1}{N_t}\sum_{j\neq m}\frac{\hat{\mathbf{g}}_j\hat{\mathbf{g}}_j^H}{\sum_{b=1}^B(\alpha_{jb}-\varepsilon_{jb})}. \tag{A.8}$$

Substituting (A.4) and (A.8) into (A.3), the first term without p_m^{JP} becomes

$$\begin{aligned}&\lim_{N_t\to\infty}\frac{\hat{\mathbf{g}}_m^H\mathbf{Q}_{c,m}\hat{\mathbf{g}}_m}{N_t} \\ &= \lim_{N_t\to\infty}(\frac{\hat{\mathbf{g}}_m^H\hat{\mathbf{g}}_m}{N_t} - \sum_{j\neq m}\frac{1}{\sum_{b=1}^B(\alpha_{jb}-\varepsilon_{jb})}\frac{\hat{\mathbf{g}}_m^H\hat{\mathbf{g}}_j}{N_t}\frac{\hat{\mathbf{g}}_j^H\hat{\mathbf{g}}_m}{N_t}) \\ &= \sum_{b=1}^B\alpha_{mb}(1-\bar{\varepsilon}_{mb}).\end{aligned} \tag{A.9}$$

By using (A.5), the second term without p_m^{JP} becomes

$$\begin{aligned}&\lim_{N_t\to\infty}\frac{\mathbf{e}_m^H\mathbf{Q}_{c,m}\hat{\mathbf{g}}_m}{N_t} \\ &= \lim_{N_t\to\infty}(\frac{\mathbf{e}_m^H\hat{\mathbf{g}}_m}{N_t} - \sum_{j\neq m}\frac{1}{\sum_{b=1}^B(\alpha_{jb}-\varepsilon_{jb})}\frac{\mathbf{e}_m^H\hat{\mathbf{g}}_j}{N_t}\frac{\hat{\mathbf{g}}_j^H\hat{\mathbf{g}}_m}{N_t}) \\ &= 0.\end{aligned} \tag{A.10}$$

With similar proofs, the third and fourth terms of (A.3) become

$$\lim_{N_t\to\infty}\frac{\hat{\mathbf{g}}_m^H\hat{\mathbf{Q}}_{c,m}\mathbf{e}_m}{N_t} = \lim_{N_t\to\infty}\frac{1}{N_t}\frac{|\mathbf{e}_m^H\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m|^2}{\hat{\mathbf{g}}_m^H\hat{\mathbf{Q}}_{c,m}\hat{\mathbf{g}}_m} = 0. \tag{A.11}$$

Therefore,

$$\begin{aligned}\lim_{N_t\to\infty}\frac{|\mathbf{g}_m^H\mathbf{w}_{c,m}^{JP}|^2}{N_t} &= \lim_{N_t\to\infty}\frac{p_m^{JP}\hat{\mathbf{g}}_m^H\mathbf{Q}_{c,m}\hat{\mathbf{g}}_m}{N_t} \\ &= p_m^{JP}\sum_{b=1}^B\alpha_{mb}(1-\bar{\varepsilon}_{mb}).\end{aligned} \tag{A.12}$$

Appendix B
Proof of Prop.2

Considering the precoder of MS$_m$ shown in (2), we obtain

$$\lim_{N_t\to\infty}\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP} = \lim_{N_t\to\infty}\frac{\sqrt{p_j^{JP}}\mathbf{g}_m^H\hat{\mathbf{Q}}_{c,j}\hat{\mathbf{g}}_j}{\|\hat{\mathbf{Q}}_{c,j}\hat{\mathbf{g}}_j\|}. \tag{B.1}$$

Substituting (A.8) into (B.1), we have

$$\begin{aligned}&\lim_{N_t\to\infty}\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP} \\ &= \lim_{N_t\to\infty}\frac{\sqrt{p_j^{JP}}}{\|\hat{\mathbf{Q}}_{c,j}\hat{\mathbf{g}}_j\|}(\mathbf{g}_m^H\hat{\mathbf{g}}_j - \sum_{l\neq j}\frac{\mathbf{g}_m^H\hat{\mathbf{g}}_l}{N_t}\frac{\hat{\mathbf{g}}_l^H\hat{\mathbf{g}}_j}{\sum_{b=1}^B(\alpha_{lb}-\varepsilon_{lb})}) \\ &= \lim_{N_t\to\infty}\frac{\sqrt{p_j^{JP}}}{\|\hat{\mathbf{Q}}_{c,j}\hat{\mathbf{g}}_j\|}(\mathbf{g}_m^H\hat{\mathbf{g}}_j - \frac{\mathbf{g}_m^H\hat{\mathbf{g}}_m}{N_t}\frac{\hat{\mathbf{g}}_m^H\hat{\mathbf{g}}_j}{\sum_{b=1}^B(\alpha_{mb}-\varepsilon_{mb})} \\ &\quad - \sum_{l\neq m,j}\frac{\mathbf{g}_m^H\hat{\mathbf{g}}_l}{N_t}\frac{\hat{\mathbf{g}}_l^H\hat{\mathbf{g}}_j}{\sum_{b=1}^B(\alpha_{lb}-\varepsilon_{lb})})).\end{aligned} \tag{B.2}$$

After substituting (A.4) into the second and third terms, (B.2) becomes

$$\begin{aligned}\lim_{N_t\to\infty}\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP} &= \lim_{N_t\to\infty}\frac{\sqrt{p_j^{JP}}}{\|\hat{\mathbf{Q}}_{c,j}\hat{\mathbf{g}}_j\|}(\mathbf{g}_m^H\hat{\mathbf{g}}_j-\hat{\mathbf{g}}_m^H\hat{\mathbf{g}}_j)\\ &= \lim_{N_t\to\infty}\frac{\sqrt{p_j^{JP}}}{\|\hat{\mathbf{Q}}_{c,j}\hat{\mathbf{g}}_j\|}\mathbf{e}_m^H\hat{\mathbf{g}}_j\\ &= \lim_{N_t\to\infty}\frac{\sqrt{p_j^{JP}/N_t}\mathbf{e}_m^H\hat{\mathbf{g}}_j}{\sqrt{\hat{\mathbf{g}}_j^H\hat{\mathbf{Q}}_{c,j}\hat{\mathbf{g}}_j/N_t}}. \qquad \text{(B.3)}\end{aligned}$$

After substituting (A.9) into the denominator, (B.3) turns into

$$\lim_{N_t\to\infty}\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}=\lim_{N_t\to\infty}\sqrt{\frac{p_j^{JP}}{\sum_{b=1}^B\alpha_{jb}(1-\bar{\varepsilon}_{jb})}}\cdot\frac{\mathbf{e}_m^H\hat{\mathbf{g}}_j}{\sqrt{N_t}}. \qquad \text{(B.4)}$$

From the central limit theorem, the random variable $\mathbf{e}_m^H\hat{\mathbf{g}}_j/\sqrt{N_t}$ converges in distribution to a complex Gaussian random variable with zero mean, and the variance is

$$\sum_{b=1}^B\alpha_{mb}\bar{\varepsilon}_{mb}\alpha_{jb}(1-\bar{\varepsilon}_{jb}). \qquad \text{(B.5)}$$

From (B.4) and (B.5), the variance of the random variable $\lim_{N_t\to\infty}\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}$ can be obtained as

$$\lambda_{c,mj}^{JP}=p_j^{JP}\frac{\sum_{b=1}^B\alpha_{mb}\bar{\varepsilon}_{mb}\alpha_{jb}(1-\bar{\varepsilon}_{jb})}{\sum_{b=1}^B\alpha_{jb}(1-\bar{\varepsilon}_{jb})}. \qquad \text{(B.6)}$$

Denote $\beta_{jb}^c=\frac{\alpha_{jb}(1-\bar{\varepsilon}_{jb})}{\sum_{b=1}^B\alpha_{jb}(1-\bar{\varepsilon}_{jb})}$, then (B.6) can be rewritten as

$$\lambda_{c,mj}^{JP}=p_j^{JP}\sum_{b=1}^B\alpha_{mb}\bar{\varepsilon}_{mb}\beta_{jb}^c. \qquad \text{(B.7)}$$

Therefore, $\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}$ converges in distribution to an exponential distribution with the parameter $1/\lambda_{c,mj}^{JP}$.

In the following, we prove that $\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}$ and $\mathbf{g}_m^H\mathbf{w}_{c,k}^{JP}$ are uncorrelated by proving that $\mathbb{E}\{\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}\mathbf{g}_m^H\mathbf{w}_{c,k}^{JP}\}=\mathbb{E}\{\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}\}\mathbb{E}\{\mathbf{g}_m^H\mathbf{w}_{c,k}^{JP}\}$. On one hand,

$$\mathbb{E}\{\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}\}\mathbb{E}\{\mathbf{g}_m^H\mathbf{w}_{c,k}^{JP}\}=0. \qquad \text{(B.8)}$$

On the other hand, the global channel vector of MS_k, i.e., $\mathbf{g}_k$, is independent from $\mathbf{g}_j$ and $\mathbf{e}_m$. Then, we have

$$\begin{aligned}&\mathbb{E}\{\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}\mathbf{g}_m^H\mathbf{w}_{c,k}^{JP}\}\\ =\;&\mathbb{E}\{\frac{\sqrt{p_j^{JP}}\mathbf{e}_m^H\hat{\mathbf{g}}_j}{\sqrt{N_t\sum_{b=1}^B(\alpha_{jb}-\varepsilon_{jb})}}\frac{\sqrt{p_k^{JP}}\mathbf{e}_m^H\hat{\mathbf{g}}_k}{\sqrt{N_t\sum_{b=1}^B(\alpha_{kb}-\varepsilon_{kb})}}\}\\ =\;&\frac{\sqrt{p_j^{JP}p_k^{JP}}\mathbb{E}\{\mathbf{e}_m^H\hat{\mathbf{g}}_j\mathbf{e}_m^H\}\mathbb{E}\{\hat{\mathbf{g}}_k\}}{N_t\sqrt{\sum_{b=1}^B(\alpha_{jb}-\varepsilon_{jb})\sum_{b=1}^B(\alpha_{kb}-\varepsilon_{kb})}}\\ =\;&\frac{\sqrt{p_j^{JP}p_k^{JP}}\mathbb{E}\{\mathbf{e}_m^H\hat{\mathbf{g}}_j\mathbf{e}_m^H\}\cdot\mathbf{0}}{N_t\sqrt{\sum_{b=1}^B(\alpha_{jb}-\varepsilon_{jb})\sum_{b=1}^B(\alpha_{kb}-\varepsilon_{kb})}}\\ =\;&0. \qquad \text{(B.9)}\end{aligned}$$

From (B.8) and (B.9) we know that $\mathbf{g}_m^H\mathbf{w}_{c,j}^{JP}$ and $\mathbf{g}_m^H\mathbf{w}_{c,k}^{JP}$ are uncorrelated. Since both of them are complex Gaussian random variables, they are independent.

APPENDIX C
PROOF OF PROP.3

From (3), the precoding vector for MS_m is given by $\mathbf{w}_{d,m}^{JP}=[\mathbf{w}_{m1}^{(1)H}\cdots\mathbf{w}_{mB}^{(B)H}]^H$. Therefore, we have

$$\begin{aligned}&\lim_{N_t\to\infty}\frac{|\mathbf{g}_m^H\mathbf{w}_{d,m}^{JP}|^2}{N_t}\\ =\;&\lim_{N_t\to\infty}\frac{|\sum_{b=1}^B\mathbf{g}_{mb}^H\mathbf{w}_{mb}^{(b)}|^2}{N_t}\\ =\;&\lim_{N_t\to\infty}\Big|\sum_{b=1}^B\frac{\sqrt{p_m^{(b)}}\mathbf{g}_{mb}^H\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}}{\sqrt{N_t}\|\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}\|}\Big|^2\\ =\;&\lim_{N_t\to\infty}\Big|\sum_{b=1}^B\frac{\sqrt{p_m^{(b)}}\hat{\mathbf{g}}_{mb}^{(b)H}\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}}{\sqrt{N_t}\|\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}\|}\\ &+\sum_{b=1}^B\frac{\sqrt{p_m^{(b)}}\mathbf{e}_{mb}^{(b)H}\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}}{\sqrt{N_t}\|\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}\|}\Big|^2\\ \triangleq\;&\lim_{N_t\to\infty}|\sum_{b=1}^BA_b+\sum_{b=1}^BB_b|^2. \qquad \text{(C.1)}\end{aligned}$$

With a similar derivation as for the proof of Prop.1, we obtain

$$\lim_{N_t\to\infty}|A_b|^2=\lim_{N_t\to\infty}\frac{p_m^{(b)}\hat{\mathbf{g}}_{mb}^{(b)H}\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}}{N_t}=p_m^{(b)}\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)}),$$

and

$$\begin{aligned}\lim_{N_t\to\infty}|B_b|^2 &= \lim_{N_t\to\infty}\frac{p_m^{(b)}|\mathbf{e}_{mb}^{(b)H}\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}|^2}{N_t\hat{\mathbf{g}}_{mb}^{(b)H}\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}}\\ &= \lim_{N_t\to\infty}\frac{\sqrt{p_m^{(b)}}\mathbf{e}_{mb}^{(b)H}\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}}{\sqrt{N_t}\|\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}\|}=0. \qquad \text{(C.2)}\end{aligned}$$

Because $\lim_{N_t\to\infty}A_b=\lim_{N_t\to\infty}\frac{\sqrt{p_m^{(b)}}}{\sqrt{N_t}}\|\hat{\mathbf{Q}}_{mb}\hat{\mathbf{g}}_{mb}^{(b)}\|$, which should be a real number, we have

$$\lim_{N_t\to\infty}\sqrt{A_b}=\sqrt{p_m^{(b)}\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)})}. \qquad \text{(C.3)}$$

Substituting (C.2) and (C.3) into (C.1), we obtain

$$\lim_{N_t\to\infty}\frac{|\mathbf{g}_m^H\mathbf{w}_{d,m}^{JP}|^2}{N_t}=\Big(\sum_{b=1}^B\sqrt{p_m^{(b)}\alpha_{mb}(1-\bar{\varepsilon}_{mb}^{(b)})}\Big)^2. \qquad \text{(C.4)}$$

APPENDIX D
PROOF OF PROP.4

Analogous to the proof of Prop. 3, substituting the precoding vector for MS_m into $\mathbf{g}_m^H\mathbf{w}_{d,j}^{JP}$, we have

$$\mathbf{g}_m^H\mathbf{w}_{d,j}^{JP}=\sum_{b=1}^B\mathbf{g}_{mb}^H\mathbf{w}_{jb}^{(b)}. \qquad \text{(D.1)}$$

With a similar derivation as for the proof of Prop. 2, we obtain that $\mathbf{g}_{mb}^H\mathbf{w}_{jb}^{(b)}$ converges in distribution to a complex Gaussian random variable with zero mean and variance as $p_j^{(b)}\alpha_{mb}\bar{\varepsilon}_{mb}^{(b)}$. Therefore, $\mathbf{g}_m^H\mathbf{w}_{d,j}^{JP}$ converges in distribution to a complex Gaussian random variable with zero mean and variance as

$$\lambda_{d,mj}^{JP}=\sum_{b=1}^Bp_j^{(b)}\alpha_{mb}\bar{\varepsilon}_{mb}^{(b)}.$$

As a result, $|\mathbf{g}_m^H \mathbf{w}_{d,j}^{JP}|^2$ converges in distribution to an exponential distribution with parameter $1/\lambda_{d,mj}^{JP}$.

REFERENCES

[1] 3GPP Long Term Evolution (LTE), "Coordinated multi-point operation for LTE," TSG RAN TR 36.819 v1.1.0, Aug. 2011.

[2] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun. Mag.*, vol. 13, pp. 56–61, Aug. 2006.

[3] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1–29, Dec. 2010.

[4] J. Zhang and J. Andrews, "Adaptive spatial intercell interference cancellation in multicell wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, pp. 1455–1468, Dec. 2010.

[5] X. Hou, C. Yang, and B. K. Lau, "Impact of non-orthogonal training on performance of downlink base station cooperative transmission," *IEEE Trans. Veh. Tech.*, vol. 60, pp. 4633–4639, Sept. 2011.

[6] X. Hou and C. Yang, "How much feedback overhead is required for base station cooperative transmission to outperform non-cooperative transmission?" in *Proc. 2011 IEEE ICASSP*.

[7] 3GPP TSG RAN and TS 36.420 v9.0.0, "Evolved universal terrestrial radio access network (e-utran): X2 general aspects and principles," Sept. 2009.

[8] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: concepts performance, and field trial results," *IEEE Wireless Commun. Mag.*, vol. 49, pp. 102–111, Feb 2011.

[9] M. Dong and L. Tong, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, pp. 1403–1418, May 2004.

[10] C. Shin, J. G. Andrews, and E. J. Powers, "An efficient design of doubly selective channel estimation for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 585–588, Sept. 2007.

[11] S. Han, Y. Tian, and C. Yang, "User-specified training symbol placement for channel prediction in TDD MIMO systems," *IEEE Trans. Veh. Tech.*, vol. 60, pp. 2837–2843, July 2011.

[12] T. Svantesson and A. L. Swindlehurst, "A performance bound for prediction of MIMO channels," *IEEE Trans. Signal Process.*, vol. 54, pp. 520–529, Feb. 2006.

[13] H. Huh, A. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, pp. 2911–2934, May 2012.

[14] J. Zhang, C. Wen, S. Jin, and X. Gao, "A large system analysis of cooperative multicell downlink system with imperfect CSIT," in *Proc. 2012 IEEE ICC*.

[15] J. Hoydis, S. 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, pp. 160–171, Feb. 2013.

[16] H. L. Trees, *Detection, Estimation, and Modulation Theory*. Wiley, 2002.

[17] L. Su, C. Yang, G. Wang, and M. Lei, "Retrieving channel reciprocity for coordinated multi-point transmission," in *Proc. 2013 IEEE WCNC*.

[18] D. Su, X. Hou, and C. Yang, "Quantization based on per-cell codebook in cooperative multi-cell systems," in *Proc. 2011 IEEE WCNC*.

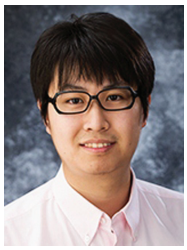
[19] A. Wiesel, Y. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4409–4418, Sep. 2008.

[20] J. Zhang, R. Chen, J. Andrews, A. Ghosh, and R. W. Heath Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

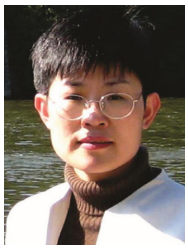
[21] U. Kamps, "Characterizations of the exponential distribution by weighted sums of i.i.d. random variables," *Statistic. Papers*, vol. 31, pp. 233–237, 1990.

[22] 3GPP TSG RAN and TR 25.814 v7.1.0, "Physical layer aspects for evolved UTRA," Sept. 2006.

[23] F. Boccardi and H. Huang, "Optimum power allocation for the MIMO-BC zero-forcing precoder with per-antenna power constraints," in *Proc. 2006 Conf. Inform. Sciences and Systems*.



Liyan Su received his B. Eng degree in School of Advanced Engineering from Beihang University (BUAA), Beijing, China in 2010. He is now pursuing Ph.D degree in School of Electronics and Information Engineering from BUAA. His research interests lie in the area of energy efficient and spectral efficient multi-cell MIMO systems.



Chenyang Yang (SM'08) received the M.S.E and Ph.D. degrees in electrical engineering from Beihang University (formerly Beijing University of Aeronautics and Astronautics), Beijing, China, in 1989 and 1997, respectively. She is currently a Full Professor with the School of Electronics and Information Engineering, Beihang University. She has published various papers and filed many patents in the fields of signal processing and wireless communications. Her recent research interests include signal processing in network MIMO, cooperative communication, energy efficient transmission and interference management. Prof. Yang was the Chair of the IEEE Communications Society Beijing chapter from 2008 to 2012. She has served as Technical Program Committee Member for many IEEE conferences, such as the IEEE International Conference on Communications and the IEEE Global Telecommunications Conference. She currently serves as an Associate editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor-in-Chief of the *Chinese Journal of Communications*, and an Associate editor-in-chief of the *Chinese Journal of Signal Processing*. She was nominated as an Outstanding Young Professor of Beijing in 1995 and was supported by the First Teaching and Research Award Program for Outstanding Young Teachers of Higher Education Institutions by Ministry of Education (P.R.C. "TRAPOYT") during 1999-2004.

Shengqian Han (S'05-M'12) received his B.S. degree in communication engineering and Ph.D. degree in signal and information processing from Beihang University (formerly Beijing University of Aeronautics and Astronautics), Beijing, China, in 2004 and 2010 respectively. From 2010 to 2012, he held a postdoctoral research position with the School of Electronics and Information Engineering, Beihang University. He is currently a Lecturer of the same school. His research interests are multiple antenna techniques, cooperative communication and energy efficient transmission in the areas of wireless communications and signal processing.