Performance of Multi-Cell Massive MIMO Systems With Interference Decoding

Meysam Shahrbaf Motlagh, Subhajit Majhi, Patrick Mitran

arXiv:1911.11103v1 [cs.IT] 25 Nov 2019

Abstract—We consider a multi-cell massive MIMO system where a time-division duplex protocol is used to estimate the channel state information via uplink pilots. When maximum ratio combining (MRC) is used at the BSs, the re-use of pilots across cells causes the pilot contamination effect which yields interference components that do not vanish as the number of base-station (BS) antennas $M \to \infty$. When treating interference as noise (TIN), this phenomenon limits the performance of multi-cell massive MIMO systems. In this paper, we analyze more advanced schemes based on simultaneous unique decoding (SD) as well as simultaneous non-unique decoding (SND) of the interference that can provide unbounded rate as $M \to \infty$. We also establish a worst-case uncorrelated noise technique for multiple-access channels to derive achievable rate expressions for finite M. Furthermore, we study a much simpler subset of SND (called S-SND) which provides a lower bound to SND and achieves unbounded rate as $M \to \infty$, and also outperforms SD for finite M. For the special cases of two-cell and three-cell systems, using a maximum symmetric rate allocation policy we compare the performance of different interference decoding schemes with that of TIN. Finally, we numerically illustrate the improved performance of the proposed schemes.

Index Terms—Massive MIMO, pilot contamination, simultaneous unique/non-unique decoding (SD/SND), treating interference as noise (TIN).

I. INTRODUCTION

WIRELESS communication standards are rapidly evolving to deal with challenges such as the ever increasing number of users as well as the demand for higher data rates and energy efficiency. These challenges give rise to the need to incorporate new protocols and techniques in 5G cellular networks. For instance, the METIS 5G project has as overall technical goals to increase the typical user data rate in a mobile network by 10x to 100x, handle 1000x more mobile data traffic per unit area, and support 10x to 100x more connected devices, all by 2020 [1]. Also in light of the demand for increased energy efficiency, the GreenTouch initiative has, for example, aimed to reduce the net energy consumption in end-to-end communication networks by up to 98% by 2020, compared to 2010 [2]. In order to meet these goals, several promising solutions have been proposed for 5G, including massive multi-input multi-output antenna systems [3], cloud-RAN [4], ultra-densification [5] and millimeter wave communications [6], [7].

In a massive MIMO communication system, each BS utilizes a very large number of antennas, which allows for the simultaneous serving of several (single or multi-antenna) users, where the number of BS antennas is normally assumed to be significantly larger than the number of users. The introduction of massive MIMO technology dates back to the seminal

work of Marzetta in [8], where it was shown that, under the assumption of independent Rayleigh fading channels and when the number of BS antennas grows to infinity, the effects of small-scale fading, intra-cell interference and additive noise all vanish due to channel hardening effects. In theory, massive MIMO has several advantages including high energy efficiency [9], high spectral efficiency and increased capacity through the aggressive spatial multiplexing of many users [3], [10], all enabled by the use of simple linear precoding/decoding techniques [11]. With regards to energy efficiency in particular, the work of [9] has analytically shown that when perfect CSI is available at the BSs, the uplink transmit power of each user can be scaled inversely proportionally with the number of BS antennas, without any performance loss.

Successful implementation of massive MIMO in practice, however, depends heavily on the availability of knowledge of channel state information (CSI) at the BSs. Thanks to the use of time-division duplex (TDD) protocols as suggested in [8] as well as uplink and downlink channel reciprocity, the BSs are able to estimate downlink channels using uplink pilots and later employ these estimates for both precoding and decoding purposes. In reality, the length of the channel coherence time is finite, and therefore the number of available orthonormal pilot sequences is limited. Consequently, in a multi-cell system, the same set of orthonormal pilots is shared across multiple cells. This, in turn, degrades the channel estimation performance for a user in one cell in that it will be contaminated by the channels of users in other cells whose pilots are not orthonormal to the first user. More specifically, the pilot contamination (PC) effect becomes a source of interference that precludes the logarithmic growth of users' achievable rate with the number of BS antennas [8]. Hence, it is believed that PC constitutes a fundamental bottleneck in multi-cell massive MIMO systems [12]. Some notable exceptions to this belief are the works of [13]–[15] that will be discussed in the sequel.

A. Contributions

In this paper, we assume the same orthonormal pilots are used across multiple cells, and take a different view of the inter-cell interference caused by PC. More specifically, we show that when $M \to \infty$ the use of more sophisticated schemes such as decoding the PC interference, rather than simply treating it as noise, allows one to attain unbounded rates, even in the presence of the PC effect.

We summarize the major contributions of this paper as follows:

- Using the capacity region obtained by simultaneous unique decoding of the desired signal and PC interference

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

(i.e., employing SD as opposed to TIN), it is shown that when maximum ratio combining (MRC) is employed at BS, the per-user rate tends to infinity as $M \to \infty$.

- It is shown that when decoding interference due to the PC, reusing the same pilots across cells (as opposed to using different pilots) is preferable as it requires decoding significantly fewer interference terms.
- The benefits of using simultaneous non-unique decoding (SND) is investigated, which strictly contains regions SD and TIN. Moreover, a simplified subset of SND (S-SND) is studied, which is shown to be strictly larger than SD and also provides a lower bound to SND.
- A worst-case uncorrelated noise technique for multiple access channels (MAC) is established that for finite M yields achievable rate expressions over regions SD/SND/S-SND.
- The problem of maximum symmetric rate allocation (i.e., maximizing the minimum achievable rate) for TIN/SD/SND/S-SND is investigated. Some structural results are also presented for the two extreme regimes of high and low SINR. In particular, it is found that when the number of BS antennas is truly large the interference decoding schemes SD/SND achieve the same performance and also strictly outperform TIN.
- For the special case of a two-cell system and assuming a symmetric geometry, it is shown that for relatively small values of M, the PC interference is "weak" in that SND and TIN achieve the same rate and both of these strictly outperform SD and S-SND. Hence, one may choose TIN which is simpler to implement. Nevertheless, for large values of M (beyond a threshold), the PC interference becomes "strong" so that the interference decoding schemes SD/SND provide the same performance and both of these strictly outperform TIN. Hence, one may choose SD which is simpler to implement. Analytical conditions in terms of mutual information expressions under which these results hold are also found.
- For the special case of a three-cell system, it is numerically shown that the use of SND can provide a strictly better performance compared to *all the other schemes*.

One should note that the theoretical contributions in the first five items listed above as well as the analytical conditions for the two-cell case are valid regardless of the numerical results presented in Section V. It is only in Section V that specific values for the system parameters are chosen to numerically compare the performance of different schemes and hence validate the analytical findings of Sections III and IV.

B. Related Work

In order to tackle the PC problem, systematic solutions have been extensively studied in the literature; some attempt to alleviate the PC effects by reducing its impact on the system performance [16]–[19] whereas others, given that some assumptions and requirements are satisfied, suggest schemes that completely eliminate PC and provide unbounded achievable rates in the asymptotic regime [13]–[15], [20]. Specifically, in the framework of PC mitigation schemes, the use of time-shifted pilots is proposed in [17], [18], where in order to make

sure that non-orthonormal pilots do not overlap in time, the location of pilots in frames shift so that transmission in different cells is done at non-overlapping times. In addition to time-shifted pilots, power allocation algorithms are also proposed in [17] and have shown to provide significant gains. In [19], a multi-cell MMSE based precoding is investigated for downlink to minimize the sum of the mean square error of signals received at the users in the same cell and the mean square interference seen by users in other cells. Unfortunately, in all these techniques the PC effect is only partially suppressed and thus the achievable rates do not grow without bound as the number of BS antennas is increased.

In the line of works that construct asymptotically noise and interference free systems with infinite capacity, [13] considers a semi-blind channel estimation technique to separate the subspace of the desired user channels from the subspace of interfering channels due to PC. However, in order for complete elimination of the PC effect, this method requires that the channel coherence time goes to infinity. Unfortunately, this assumption is not true in practice. Other interesting works include [14], [20], where a large-scale fading decoding (LSFD) technique is used that eliminates PC interference with the help of a network controller resulting in achievable rates that scale as $O(\log M)$. Therein, BS cooperation is required for the exchange of large-scale fading coefficients between BSs and the network controller, which results in backhaul overhead. Different from [14], [20], the work of [15] uses a multi-cell MMSE precoding/combining technique and assumes that pilot-sharing users must have asymptotically linearly independent covariance matrices. This assumption, however, may not always be true and also requires the knowledge of channel covariance matrices at the BSs.

In this paper, we do not view the interference caused by PC as a fundamental limitation in a multi-cell system due to treating it as noise. This is because TIN is known to be suboptimal in some scenarios [21], [22], and hence the present work proposes more sophisticated schemes based on interference decoding.

The rest of this paper is organized as follows. In Section II, we present the system model and describe the PC problem. In Section III, to combat the PC problem we propose two interference decoding schemes, i.e., SD and SND, as well as an achievable subset of SND (S-SND). Achievable rate expressions based on the worst-case uncorrelated noise technique are also derived in this section. The problem of maximum symmetric rate allocation is studied in section IV along with some structural results for the extreme regimes of high and low SINR as well as for the special cases of two-cell and three-cell systems. In Section V, we demonstrate numerical results, and finally, section VI concludes this paper.

Notation: We use boldface upper and lower case symbols to represent matrices and vectors, respectively. An $M \times M$ identity matrix and an all-zero vector are denoted by $\boldsymbol{I}_M$ and $\boldsymbol{0}$, respectively. The superscripts $(.)^T$, $(.)^\dagger$, $(.)^*$, and $(.)^{-1}$ denote the transpose, Hermitian transpose, conjugate and inverse operations. The notation $\mathrm{diag}(\boldsymbol{v})$ represents a diagonal matrix with elements $v[1], v[2], ...$ of vector $\boldsymbol{v}$ along its main diagonal. The expressions $\mathbb{E}[.]$ and $\mathrm{var}[]$ are used to denote mean and
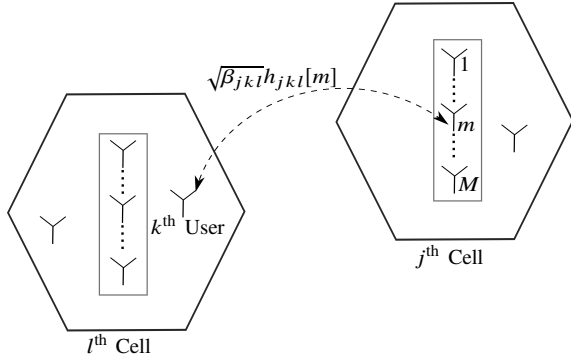
Fig. 1: System model showing the channel gain between the mth antenna of the BS in cell j and the kth user in cell l.

variance of a random variable, respectively, and $CN(\boldsymbol{m}, \boldsymbol{R})$ denotes the circular symmetric complex Gaussian distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{R}$.

II. PRELIMINARIES

A. System Model

We consider a multi-cell communication system with L cells, where each cell has a BS equipped with M antennas serving K ($M \gg K$) single antenna users. Assuming a flat-fading model, the channel between the M antennas of the BS in cell j and the users in cell l is described by

$$\boldsymbol{G}_{jl} = \boldsymbol{H}_{jl}\boldsymbol{D}_{jl}^{1/2}, \qquad (1)$$

where $\boldsymbol{H}_{jl} = [\boldsymbol{h}_{j1l}, \boldsymbol{h}_{j2l}, ..., \boldsymbol{h}_{jKl}] \in \mathbb{C}^{M \times K}$ is the channel matrix associated with the channel vectors $\boldsymbol{h}_{jkl} \in \mathbb{C}^{M \times 1}$ of small-scale fading coefficients between antennas of the BS in cell j and the kth user in cell l, and $\boldsymbol{D}_{jl} = \text{diag}(\beta_{j1l}, \beta_{j2l}, ..., \beta_{jKl})$ is the matrix of large-scale fading coefficients. One may rewrite (1) as

$$\boldsymbol{g}_{jkl} = \sqrt{\beta_{jkl}}\boldsymbol{h}_{jkl}, \qquad (2)$$

which explicitly shows that the large-scale fading coefficient β_{jkl}, which models shadowing and path loss and is assumed to be known at the BS, is constant with respect to the index m of the BS antenna (see Fig. 1). The latter follows since the distance between the BS and a user is much larger than the spacing between the antennas of the BS. We also assume a block fading model where the large-scale fading coefficients β_{jkl} are constant over many coherence time intervals T and known a *priori*, whereas small-scale fading coefficients $h_{jkl}[m], m = 1, ..., M$ are constant over one coherence interval, and drawn independently in each coherence interval with $\boldsymbol{h}_{jkl} \sim CN(\boldsymbol{0}, \boldsymbol{I}_M)$ (i.e., flat-fading model).

Furthermore, we consider TDD operation such that reciprocity holds between uplink and downlink channels. We take the frequency re-use factor to be one, i.e., the whole frequency band is used in one cell, and re-used in all the adjacent cells. This assumption, in particular, entails a worst-case inter-cell interference.

B. Uplink Data Transmission

We point out that the model used for uplink data transmission in this paper is similar to that of [11] with slight change of notation. During the uplink data transmission phase, the BS in cell j receives the baseband signal $\boldsymbol{y}_j \in \mathbb{C}^{M \times 1}$ given by

$$\boldsymbol{y}_j = \sum_{l=1}^{L} \sum_{k=1}^{K} \sqrt{\rho_{\mathrm{u}}}\boldsymbol{g}_{jkl}x_l[k] + \boldsymbol{n}_j, \qquad (3)$$

where $\boldsymbol{x}_l = [x_l[1], x_l[2], ..., x_l[k]]^T$ is the vector of transmit signals of the users in cell l, ρ_{u} is the average uplink transmit power of the users, and $\boldsymbol{n}_j \sim CN(\boldsymbol{0}, \boldsymbol{I}_M)$ is the additive Gaussian noise vector at the BS in cell j. Thus, ρ_{u} can be interpreted as the uplink transmit SNR of the users.

C. CSI Estimation at BS

Similar to the approach of [14] for CSI estimation, it is assumed that the same set of pilot sequences $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, ..., \boldsymbol{\psi}_K \in \mathbb{C}^{\tau \times 1}$ of length τ (usually $\tau \geq K$, however without essential loss of generality we assume $\tau = K$) are used in all cells and thus the channel estimate will be corrupted by the PC from the adjacent cells. Defining the pilot matrix $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, ..., \boldsymbol{\psi}_K]^T \in \mathbb{C}^{K \times K}$, we assume orthonormal pilots $\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger = I_K$.

During the uplink training phase of the TDD protocol, user $k = 1, 2, ..., K$ in each cell transmits the pilot sequence $\boldsymbol{\psi}_k$ to its BSs. The BS in cell j then finds the estimate $\hat{\boldsymbol{G}}_{jj}$ of the local channels $\boldsymbol{G}_{jj}$. More specifically, the BS in cell j receives the matrix $\boldsymbol{Y}_j^{\mathrm{p}} \in \mathbb{C}^{M \times K}$, i.e.,

$$\boldsymbol{Y}_j^{\mathrm{p}} = \sum_{l=1}^{L} \sqrt{\rho_{\mathrm{p}}}\boldsymbol{G}_{jl}\boldsymbol{\Psi} + \boldsymbol{Z}_j, \qquad (4)$$

where ρ_{p} is the average pilot transmission power, and $\boldsymbol{Z}_j$ is the AWGN at the BS with entries that are iid $CN(0, 1)$ random variables. Similar to the uplink data transmission, ρ_{p} can be interpreted as the pilot SNR. Generally, ρ_{p} is a function of the average transmit power of users ρ_{u} and the length of pilot sequences τ.

Multiplying $\boldsymbol{Y}_j^{\mathrm{p}}$ by $\boldsymbol{\Psi}^\dagger$, the kth column of the resulting matrix is

$$\boldsymbol{r}_{jk} = \sum_{l=1}^{L} \sqrt{\rho_{\mathrm{p}}}\boldsymbol{g}_{jkl} + \tilde{\boldsymbol{z}}_{jk}, \qquad (5)$$

where $\tilde{\boldsymbol{z}}_{jk} \sim CN(\boldsymbol{0}, \boldsymbol{I}_M)$. The MMSE estimate $\hat{\boldsymbol{g}}_{jkj}$ of $\boldsymbol{g}_{jkj}$ based on the observation $\boldsymbol{r}_{jk}$ is given by [23]

$$\hat{\boldsymbol{g}}_{jkj} = \mathbb{E}\left[\boldsymbol{g}_{jkj}\boldsymbol{r}_{jk}^\dagger\right]\mathbb{E}\left[\boldsymbol{r}_{jk}\boldsymbol{r}_{jk}^\dagger\right]^{-1}\boldsymbol{r}_{jk} \qquad (6)$$

$$= \sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\left(1 + \rho_{\mathrm{p}}\sum_{l=1}^{L}\beta_{jkl}\right)^{-1}\boldsymbol{r}_{jk} \qquad (7)$$

$$= \alpha_{jkj}\left(\sum_{l=1}^{L}\sqrt{\rho_{\mathrm{p}}}\boldsymbol{g}_{jkl} + \tilde{\boldsymbol{z}}_{jk}\right), \qquad (8)$$

where $\alpha_{jkj} := \frac{\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}}{1+\rho_{\mathrm{p}}\sum_{l=1}^{L}\beta_{jkl}}$. Due to the orthogonality property of MMSE estimation, one can decompose the channel $\boldsymbol{g}_{jkj}$ as $\boldsymbol{g}_{jkj} = \hat{\boldsymbol{g}}_{jkj} + \boldsymbol{\epsilon}_{jkj}$, where $\boldsymbol{\epsilon}_{jkj}$ is the estimation error. It is also known that $\boldsymbol{\epsilon}_{jkj}$ is uncorrelated (and thus independent due to the Gaussian assumption) from the estimate $\hat{\boldsymbol{g}}_{jkj}$ [14]. Therefore, it can be verified that $\hat{\boldsymbol{g}}_{jkj} \sim CN(\boldsymbol{0}, \sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}\boldsymbol{I}_M)$ and $\boldsymbol{\epsilon}_{jkj} \sim CN(\boldsymbol{0}, \beta_{jkj}(1 - \sqrt{\rho_{\mathrm{p}}}\alpha_{jkj})\boldsymbol{I}_M)$. Using (6), one can

see that the estimate $\hat{\boldsymbol{g}}_{jkl}$ can be written in terms of $\hat{\boldsymbol{g}}_{jkj}$ as follows

$$\hat{\boldsymbol{g}}_{jkl} = \left(\frac{\beta_{jkl}}{\beta_{jkj}}\right)\hat{\boldsymbol{g}}_{jkj}. \tag{9}$$

D. Treating Interference as Noise (TIN)

Assuming maximum ratio combining (MRC), from (3) the estimate of the ith user's signal in cell j is

$$\hat{y}_{ji} = \hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{y}_j \tag{10}$$

$$= \sum_{l=1}^{L} \sqrt{\rho_{\text{u}}}\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jil}x_l[i]$$

$$+ \sum_{l=1}^{L} \sum_{k=1,k\neq i}^{K} \sqrt{\rho_{\text{u}}}\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jkl}x_l[k] + \hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{n}_j. \tag{11}$$

Substituting (8) in (11), and applying the strong law of large numbers, as $M \to \infty$, the following is obtained

$$\frac{\hat{y}_{ji}}{M} \xrightarrow{\text{a.s.}} \sqrt{\rho_{\text{u}}\rho_{\text{p}}}\alpha_{jij}\left(\beta_{jij}x_j[i] + \sum_{l=1,l\neq j}^{L}\beta_{jil}x_l[i]\right), \tag{12}$$

where $\xrightarrow{\text{a.s.}}$ represents the almost sure convergence. Note that a channel hardening effect is observed in (12).

Assuming TIN in uplink, the BS in cell j only decodes the desired signal $x_j[i]$ and treats the remaining interfering signals $x_l[i]$, $l \neq j$ as noise. Thus, defining R_{ij} as the uplink rate of the ith user in cell j, any rate tuple $(R_{i1}, ..., R_{iL})$ is achievable if it satisfies the following set of inequalities

$$R_{ij} \leq I\left(\hat{y}_{ji}; x_j[i]\middle|\, \hat{\boldsymbol{g}}_{jij}\right), \quad \text{for} \quad j = 1, ..., L. \tag{13}$$

Based on (12), it is known that [8]

$$I\left(\hat{y}_{ji}; x_j[i]\middle|\, \hat{\boldsymbol{g}}_{jij}\right) \xrightarrow{\text{a.s.}} C\left(\frac{\beta_{jij}^2}{\sum_{l=1,l\neq j}^{L}\beta_{jil}^2}\right), \quad \text{as} \quad M \to \infty, \tag{14}$$

where $C(x) := \log(1+x)$ is the Shannon rate function.

Remark 1. As the number of BS antennas M in (12) becomes large, except for the terms associated with the pilot-sharing users, i.e., the interference caused by PC, the effects of interference and noise vanish.

Throughout the paper, it is assumed that the noisy channel estimates $\hat{\boldsymbol{g}}_{jij}$ are known locally at the BSs. Thus, from now on to simplify notation they will be omitted from the mutual information expressions.

III. DECODING THE PC INTERFERENCE

One can see that, as the number of BS antennas M becomes large enough, the expression of (14) converges to a constant independent of M and thus the benefits of increasing M saturate. In other words, treating interference due to PC as noise results in a fundamental limitation that constitutes a major bottleneck in overall performance of massive MIMO systems [8].

In this paper, as opposed to simply performing TIN, we consider more sophisticated schemes based on interference decoding. More specifically, we treat the PC interference terms as individual users (similar to a MAC) and thus try to decode them. As will be seen in the subsequent part, this change of perspective results in new achievable rate expressions that grow without bound as $M \to \infty$.

A. Simultaneous unique Decoding (SD)

Note that in the expression of the received signal after performing MRC in (12), the first term is the desired signal and the remaining non-vanishing terms are all inter-cell interference caused by users in other cells that are sharing the same pilot sequence, $\boldsymbol{\psi}_i, i = 1, \ldots, K$, as the ith user of cell j. Now, consider (11) which is the output of the jth BS after performing MRC. If $\hat{y}_{ji}$, for $j = 1, 2, ..., L$, $i = 1, ..., K$, are considered together, then these represent the output of K separate/non-interfering L-user interference channels (ICs), one such L-user IC for each pilot sequence $\boldsymbol{\psi}_i, i = 1, \ldots, K$: each L-user IC consists of L transmitters, i.e., i^{th} user of all cells that are using the same pilot sequence $\boldsymbol{\psi}_i$, and L receivers, i.e., the BSs. One should also note that at each of the L receivers of each IC, an asymptotically noise-free L-user MAC is observed (see (12)). For instance, in the noise-free L-user MAC of (12), by unique joint decoding of the users $[x_1[i], x_2[i], ..., x_L[i]]^T$, unbounded rates are obtained as $M \to \infty$.

Remark 2. Note that since large-scale fading coefficients from contaminating users are unknown at the BS, and also the effective channel gains in the MAC of (12) are functions of these coefficients, this MAC can be regarded as a compound MAC [24], where the channel gains from users to the receiver are unknown. It has been shown in [24] that the achievable rates of a compound MAC (i.e., a MAC with unknown channel gains) are the same as those of the standard MAC, where all channel gains are known. Therein, it has been shown that the lack of knowledge of channel gains at the receiver *does not affect* the achievable rates, i.e., the users' signal can still be successfully decoded.

Due to finite coherence time of wireless channels, the number of available orthonormal pilot sequences is smaller than KL for typical values of K and L. Thus, one way to address this issue is to re-use the *same* set of orthonormal pilots across all cells as described in Section II-C. However, an alternative approach to that of Section II-C is to use *different* sets of orthonormal pilots in different cells. To illustrate this alternative, assume that a single set of orthonormal pilots is picked for one cell, and different rotated versions of this set are used in all other cells. In particular, user k in cell l transmits the pilot sequence $\boldsymbol{\psi}_{kl}$ to its BS, where the entire pilot matrix used in cell l is denoted by $\boldsymbol{\Psi}_l$. As the sequences of other pilot matrices, $\boldsymbol{\Psi}_j$, $j \neq l \in \{1, 2, ..., L\}$, are rotated versions of sequences in $\boldsymbol{\Psi}_l$, they have non-zero inner product.

After transmission of all pilot sequences, the BS in cell j receives the matrix $\boldsymbol{Y}_j^{\text{p}} \in \mathbb{C}^{M \times K}$, given by

$$\boldsymbol{Y}_j^{\text{p}} = \sum_{l=1}^{L} \sqrt{\rho_{\text{p}}}\boldsymbol{G}_{jl}\boldsymbol{\Psi}_l + \boldsymbol{Z}_j. \tag{15}$$

Multiplying $\boldsymbol{Y}_j^{\text{p}}$ by $\boldsymbol{\Psi}_j^{\dagger}$, the kth column of the resulting matrix is

$$\boldsymbol{r}_{jk} = \sqrt{\rho_{\text{p}}}\boldsymbol{g}_{jkj} + \sum_{l=1,l\neq j}^{L} \sqrt{\rho_{\text{p}}}\boldsymbol{G}_{jl}\boldsymbol{\Psi}_l\boldsymbol{\psi}_{kj}^{\dagger} + \boldsymbol{q}_{jk}, \tag{16}$$

where $\boldsymbol{q}_{jk} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$. Therefore, the MMSE estimate $\hat{\boldsymbol{g}}_{jkj}$ of $\boldsymbol{g}_{jkj}$ based on the observation $\boldsymbol{r}_{jk}$ is

$$\hat{\boldsymbol{g}}_{jkj} = \mathbb{E}\left[\boldsymbol{g}_{jkj}\boldsymbol{r}_{jk}^{\dagger}\right]\mathbb{E}\left[\boldsymbol{r}_{jk}\boldsymbol{r}_{jk}^{\dagger}\right]^{-1}$$

$$\times \left(\sqrt{\rho_\mathrm{p}} \boldsymbol{g}_{jkj} + \sum_{l=1,l\neq j}^{L} \sqrt{\rho_\mathrm{p}} \boldsymbol{G}_{jl} \boldsymbol{\Psi}_l \boldsymbol{\psi}_{kj}^{\dagger} + \boldsymbol{q}_{jk} \right). \quad (17)$$

One can readily see from (17) that the channel estimate $\hat{\boldsymbol{g}}_{jkj}$ is now contaminated by the channel of all the users in other cells. Thus, after performing MRC and letting $M \rightarrow \infty$, the non-vanishing terms in (12) will include the signal of every user in every other cell. In turn, when employing interference decoding schemes, using the same set of pilots in different cells results in decoding L users, whereas using different sets of pilots in different cells, as explained above, results in decoding $K(L-1)+1$ users. As will be explained later in Remark 7, this alternative approach that requires decoding $K(L-1)+1$ users (instead of L users) degrades the performance of interference decoding schemes, as compared to the approach of Section II-C. Moreover, the complexity of jointly decoding $K(L-1)+1$ users is larger than that of decoding L users. Hence, when decoding the PC interference, using the same set of pilots in different cells (as opposed to different pilots) is preferable as it results in fewer interference terms to be decoded.

We now provide a detailed analysis of the achievable rates for finite values of M. Following the approach of [14], by adding and subtracting a term associated with the mean of the effective channel $\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil}$ in (10), the following is obtained over one coherence interval

$$\hat{y}_{ji} = \underbrace{\sqrt{\rho_\mathrm{u}} \sum_{l=1}^{L} \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil} \right] x_l[i]}_{\text{Desired signals}}$$

$$+ \underbrace{\sum_{l=1}^{L} \sqrt{\rho_\mathrm{u}} \left(\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil} \right] \right) x_l[i]}_{\text{Interference due to channel estimation error}}$$

$$+ \underbrace{\sum_{l=1}^{L} \sum_{k=1,k\neq i}^{K} \sqrt{\rho_\mathrm{u}} \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jkl} x_l[k]}_{\text{Interference caused by other users}} + \underbrace{\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{n}_j}_{\text{Noise}} \quad (18)$$

$$= \sum_{l=1}^{L} \gamma_{il} x_l[i] + z'_{jij}, \quad (19)$$

where $\gamma_{il} := \sqrt{\rho_\mathrm{u}} \mathbb{E}[\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil}]$ and z'_{jij} is the effective noise term.

The power of the desired signals in (19) is proportional to $|\gamma_{il}|^2$ and is thus proportional to M^2. Moreover, the power of the effective noise term z'_{jij} is proportional to M. Therefore, by unique joint decoding of the users' signals $\{x_l[i]\}_{l=1}^{L}$ in (19), the achievable rates of the corresponding MAC grow unboundedly as $M \rightarrow \infty$.

Note that the effective noise z'_{jij} in (19) contains the last three terms in (18) including the inner product $\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{n}_j$ of two Gaussian vectors, and hence is neither Gaussian nor independent of the users' signals. However, as will be shown in the sequel, it is uncorrelated from the users' signals. The following lemma lower bounds the mutual information terms defining the boundaries of the MAC region, using a Gaussian effective noise with the same power as that of z'_{jij} in (19).

Lemma 1. *Consider the L-user MAC given by $y = \sum_{i=1}^{L} x_i^G + z$, where the users' signals x_i^G, $i = 1, ..., L$ are independent with complex Gaussian distribution $x_i^G \sim \mathcal{CN}(0, P_i)$, and the additive noise z is a complex random variable with mean zero*

and variance σ_z^2. If z is uncorrelated from x_i^G, $i = 1, ..., L$, then

$$I\left(\boldsymbol{x}_\Omega^G; \ y^G \big| \boldsymbol{x}_{\Omega^c}^G \right) \leq I\left(\boldsymbol{x}_\Omega^G; \ y \big| \boldsymbol{x}_{\Omega^c}^G \right), \quad (20)$$

where $\boldsymbol{x}_\Omega^G$ is the vector with entries $x_i^G, i \in \Omega \subseteq \{1, 2, ..., L\}, \Omega \neq \emptyset, \Omega^c := \{1, 2, ..., L\} \setminus \Omega, y^G = \sum_{i=1}^{L} x_i^G + z^G$, and $z^G \sim \mathcal{CN}(0, \sigma_z^2)$.

Proof. See Appendix A. $\qquad\square$

Note that using Lemma 1 one can obtain an achievable lower bound on the capacity of a MAC with uncorrelated additive non-Gaussian noise by replacing the noise term with an independent zero mean Gaussian noise having the same variance. This is known as the worst-case uncorrelated noise result which has been previously established for the case of a point-to-point channel in the literature [25]. When the additive noise is independent of the users' signals, Gaussian noise has been proven to be the worst-case noise for point-to-point, MAC, degraded broadcast and MIMO channels [26]. However, the proof provided in Lemma 1 only requires the additive noise to be uncorrelated of the users' signals.

We now consider the MAC of (19) at BS j. Using the usual definitions as in [27], each message $m_l \in [1 : 2^{nR_{il}}]$, $l = 1, ..., L$ (distributed uniformly) is encoded into the codeword $\boldsymbol{x}_l^n[i](m_l)$ of length n which is generated iid $\mathcal{CN}(0, 1)$. Using SD and the standard random coding analysis as in [27], it can be shown that decoding error probability tends to zero as $n \rightarrow \infty$, i.e., the rate tuple $(R_{i1}, ..., R_{iL})$ is achievable, if

$$R_\Omega := \sum_{l \in \Omega} R_{il} \leq I\left(\hat{y}_{ji}; \ \boldsymbol{x}_\Omega \big| \boldsymbol{x}_{\Omega^c} \right), \quad (21)$$

where $S = \{1, 2, ..., L\}$, $\Omega \subseteq S$, and $\boldsymbol{x}_\Omega$ is the vector with entries $x_l[i], l \in \Omega$. Finally, to obtain the achievable region network-wide (at all BSs), one should take the intersection of achievable regions over all BSs. It should be pointed out that recent studies in the literature have proposed practical schemes using off-the-shelf LDPC codes that can achieve a performance very close to the theoretical SD [28], [29].

Note that $x_l[i]$ and $x_j[k]$ are independent for $(l, i) \neq (j, k)$, $\boldsymbol{g}_{jil}$ and $\boldsymbol{g}_{mkn}$ are independent for $(j, i, l) \neq (m, k, n)$, and also $\hat{\boldsymbol{g}}_{jil}$ and $\boldsymbol{\epsilon}_{jil}$ are uncorrelated. Therefore, for transmission over multiple coherence intervals all interference and noise terms in (18) are uncorrelated from the desired signal components. Thus applying Lemma 1, an achievable lower bound to $I(\hat{y}_{ji}; \ \boldsymbol{x}_\Omega \big| \boldsymbol{x}_{\Omega^c})$ in (21) is obtained, similar to the ones established in [14]. More specifically, replacing the effective noise z'_{jij} in (19) by an independent Gaussian noise with a variance equal to the sum of the variances of the interference and noise terms in (18), provides a lower bound in (21). This is formally presented in the following theorem.

Theorem 1. *Assuming $\boldsymbol{x}_l = [x_l[1], x_l[2], ..., x_l[K]]^T \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_K)$ for $l \in \{1, 2, ..., L\}$, the following set of lower bounds can be achieved for the MAC given in (19) at BS j*

$$I\left(\hat{y}_{ji}; \ \boldsymbol{x}_\Omega \big| \boldsymbol{x}_{\Omega^c} \right) \geq C\left(\frac{P_1}{P_2 + P_3 + P_4} \right) := C_{\mathrm{LB}}(\Omega), \quad (22)$$

which holds for $\forall \Omega \subseteq \{1, 2, ..., L\}$, where

$$P_1 = M^2 \sum_{l \in \Omega} \rho_\mathrm{p} \rho_\mathrm{u} \beta_{jil}^2 \alpha_{jij}^2, \quad (23)$$

$$P_2 = M\sqrt{\rho_\mathrm{p}}\beta_{jij}\alpha_{jij} \sum_{l=1}^{L} \rho_\mathrm{u}\beta_{jil}, \tag{24}$$

$$P_3 = M\sqrt{\rho_\mathrm{p}}\beta_{jij}\alpha_{jij} \sum_{l=1}^{L} \sum_{k=1,k\neq i}^{K} \rho_\mathrm{u}\beta_{jkl}, \tag{25}$$

$$P_4 = M\sqrt{\rho_\mathrm{p}}\beta_{jij}\alpha_{jij}. \tag{26}$$

Proof. See Appendix B. □

Remark 3. Note that P_1 is the power of the desired signal components associated with $\boldsymbol{x}_\Omega$, whereas P_2 is the power of the interference due to the channel estimation error, P_3 is the power of the interference of other users, and P_4 is the power of the noise.

Using (23)-(26), one can simplify the right hand side of (22) as

$$C_{\mathrm{LB}}(\Omega) = C\left(\frac{M\sum_{l\in\Omega}\sqrt{\rho_\mathrm{p}}\rho_\mathrm{u}\beta_{jil}\alpha_{jil}}{\sum_{l=1}^{L}\sum_{k=1}^{K}\rho_\mathrm{u}\beta_{jkl}+1} \right), \quad \forall\Omega \subseteq \{1,2,...,L\} \tag{27}$$

which follows from the fact that $\beta_{jil}\alpha_{jij} = \beta_{jij}\alpha_{jil}$. Therefore, it is clear that $I(\hat{y}_{ji}; \ \boldsymbol{x}_\Omega \mid \boldsymbol{x}_{\Omega^c}) \geq C(M \times \kappa)$, where κ is a function of ρ_u, ρ_p and large-scale fading coefficients, and is also constant. Thus, the uplink achievable rates in (22) grow as $O(\log M)$.

It is important to note that the BS j is only interested in correct decoding of $x_j[i]$ in uplink. Thus, incorrectly decoding $x_l[i]$, $l \neq j$, should not penalize the rates achievable at BS j. Furthermore, the power of received signal for the users located in distant cells is very small, and thus trying to decode signals of such users can reduce achieved rates considerably. As later illustrated in the paper, there exist scenarios where system performance is constrained by these distant cells, which motivates the need for more advanced decoding schemes.

B. Simultaneous Non-unique Decoding (SND)

In this part, we investigate the benefit of using SND and further show that it enlarges the region obtained by SD for finite M. The optimality of this decoding scheme for interference networks with point-to-point codes and time-sharing has been shown in [30]. Associated with the estimate of $x_l[i]$ in cell l, we consider an IC that consists of the L senders, i.e., the ith user in each of cell $l = 1, ..., L$, and the L BS receivers. In particular, the BS $j \in S = \{1,2,...,L\}$ simultaneously decodes the intended message $x_j[i]$ and the interference signals $x_l[i]$, $l \neq j$, where incorrect decoding of the interference signals does not incur any penalty. More precisely, BS j finds the unique message $\hat{m}_j$ such that $(\hat{\boldsymbol{x}}_j^n[i](\hat{m}_j), \hat{\boldsymbol{x}}_{S\setminus\{j\}}^n[i](m_{S\setminus\{j\}}), \hat{\boldsymbol{y}}_{ji}^n)$ is *jointly typical* for some $m_{S\setminus\{j\}}$, where $\hat{\boldsymbol{x}}_{S\setminus\{j\}}^n[i](m_{S\setminus\{j\}})$ is the tuple of all codewords $\hat{\boldsymbol{x}}_l^n[i](m_l)$ for $l \in S \setminus \{j\}$. For a comprehensive treatment of random code ensembles and joint typicality, we refer the reader to [27, Chap. 3].

It has been shown in [30] that, assuming point-to-point random code ensembles, the capacity region of the IC associated with the ith users across the L cells can be described by

$$C_i = \bigcap_{j\in S} \mathcal{R}_{ji}, \tag{28}$$

where $\mathcal{R}_{ji}$ is the rate region achievable at BS j given by

$$\mathcal{R}_{ji} = \bigcup_{\{j\}\subseteq\Omega\subseteq S} \mathcal{R}_{\mathrm{MAC}(\Omega,j)}^i, \tag{29}$$

and $\mathcal{R}_{\mathrm{MAC}(\Omega,j)}^i$ represents the achievable rate region obtained from unique joint decoding of the signals $x_l[i]$, $l \in \Omega$ at BS j. Note that Ω at BS j must contain the index of the desired signal $x_j[i]$.

The rate region $\mathcal{R}_{\mathrm{MAC}(\Omega,j)}^i$ has the following properties:

[P1] The region does not include the rates R_{il}, $l \in \Omega^c$, and is thus unbounded in these variables.

[P2] The signals $x_l[i]$, $l \in \Omega^c$, are treated as noise in the rate expressions defining the region.

One can readily see that $\mathcal{R}_{ji}$ strictly contains the MAC region at BS j. Therefore, the capacity region C_i in (28) (obtained from SND) is strictly larger than the intersection of the MAC regions (obtained from SD) at BSs $l = 1, ..., L$. Another important observation is that, due to [P2], $\mathcal{R}_{ji}$ also contains the TIN region (a similar observation was also made in [30] and [21]) and thus parts of region $\mathcal{R}_{ji}$'s boundary remain constant when $M \to \infty$. It is also worth mentioning that a low complexity technique, called sliding-window coded modulation (SWCM) has been recently proposed in the literature that can achieve a performance close to that of the theoretical SND, while outperforming TIN in the *strong* interference regime [31]–[34]. We will see in the next section that depending on the number of BS antennas and geometry of the cells, the use of SND automatically specifies the optimal subset of signals that should be jointly decoded while the remaining signals will be treated as noise.

Note that in SD, the decoder attempts to uniquely decode the message tuple of all users (i.e., the intended one as well as the interfering users), as in a MAC. While in SND the decoder attempts to decode only the intended message uniquely and the messages of interfering users non-uniquely. More specifically, for SND the decoder needs to perform jointly typical decoding of all possible message tuples that include the message of the intended user (i.e., the intended message only, all 2-message tuples containing the intended message, ..., all $(n-1)$-message tuples containing the intended message and the only n-message tuple) as in (29). Hence, the SND decoder is more complex than that of SD.

Remark 4. Note that there exists a *complexity-performance trade-off* between the two interference decoding schemes SND and SD, and also between SD/SND and TIN. As explained above, the SD scheme attempts to decode all users, including those that have weak interference. Hence, while SD requires *less complexity* than SND, as will be seen later in the paper, it achieves *worse rates* than SND. In contrast, SND is able to adaptively determine whether a user should be decoded or treated as noise based on the strength of the interference. Hence, even though SND has *more complexity* than SD, it achieves *larger rates* than SD. A similar trade-off exists between SD/SND and TIN. Specifically, while the proposed schemes of SD/SND have *more complexity* than TIN as they need to decode additional users, for sufficiently large number of antennas M, the rates achieved by TIN *saturate* to a fixed

value that does not increase with M. In contrast, the rates for SND/SD *increase* as $O(\log M)$, and hence as $M \to \infty$, *unbounded rates* are obtained.

Remark 5. Note that the successive interfernce cancelation (SIC) technique used in [35] is different from the SND/SD of this paper in the following manner: the work of [35] considers a setting in downlink where each user is served by all BSs through the reception of L independent data symbols from the L BSs. In particular, each user applies SIC to sequentially decode the L intended data symbols transmitted by the BSs, while treating all interfering signals, including pilot-sharing interfering signals, as noise, thus resulting in the rate saturation problem. This is in contrast to the approach proposed in this paper. As the BSs try to jointly decode (either uniquely or non-uniquely) the intended signal along with the signal coming from the pilot-sharing users, there is no rate saturation as M increases.

C. A Simplified Subset of SND (S-SND)

We now consider a simplified achievable region which is a subset of SND and also described in [26, Eq. (6.5)]. We refer to this region as S-SND, which is given by the following set of inequalities at BS j

$$\sum\nolimits_{l\in\Omega} R_{il} \le I\left(\hat{y}_{ji}; \, \boldsymbol{x}_\Omega \,\middle|\, \boldsymbol{x}_{\Omega^c}\right), \qquad (30)$$

for all Ω such that $\{j\} \subseteq \Omega \subseteq \{1, 2, ..., L\}$.

One can directly verify that region S-SND can be obtained from SD by removing all $2^{L-1} - 1$ inequalities in (21) that do not involve the rate R_{ij}. Hence, the region SD is strictly contained in S-SND. Furthermore, due to Theorem 1 it can be verified that the boundaries of S-SND in (30) grow as $O(\log M)$.

The motivation behind considering this region is as follows. It will be shown in the next section that, as opposed to SND, the S-SND region is in the form of a convex polytope which makes it tractable for computing the maximum symmetric rate allocation. Therefore, even though for large networks (e.g., more than 3 cells) the maximum symmetric rate of SND can not be computed in a computationally efficient way, it is feasible under S-SND. Furthermore, since S-SND is a subset of SND, it provides a lower bound to SND. As will be shown later, there are cases where S-SND strictly outperforms other schemes (e.g., it strictly outperforms SD in the low SINR regime). Thus, based on these findings we are able to draw conclusions regarding the performance of SND.

IV. Maximum Symmetric Rate Allocation

Considering (19), it is evident that users with relatively small effective channel gains γ_{il}, $l \ne i$, suffer from smaller rates compared to those users with stronger channels. Therefore, it is crucial to assure fairness among users when allocating resources in cellular networks. As such, we study the problem of maximum symmetric rate allocation policy (which is the same as maximizing the minimum achievable rate among all users) for various schemes. More specifically, we will compare the performance of all interference decoding schemes

SD/SND/S-SND with that of TIN based on the maximum symmetric rate they can offer. In what follows, the analysis is shown only for the ith (i is arbitrary) users across multiple cells that are employing the same pilots, since the same results hold for other sets of pilot-sharing users.

A rate allocation is said to be symmetric when all users are assigned the same rate. Thus, the maximum symmetric rate associated with BS j is obtained by $R_{\text{Sym}}^j = \max R$ such that the rate vector $\boldsymbol{R} = [R, R, ..., R]^T$ belongs to the achievable region at BS j. Therefore, the rate vector $[R_{\text{Sym}}^j, R_{\text{Sym}}^j, ..., R_{\text{Sym}}^j]^T$ must lie at the intersection of the diagonal ($R_{i1} = ... = R_{iL}$) with the boundary of the achievable region at BS j.

One can verify that the SD region described in (21) (achieved at BS j) can be represented as the intersection of a finite number of closed half-spaces and is also bounded. Hence, it is a convex polytope, denoted by $\mathscr{R}_j$, shown below

$$\mathscr{R}_j = \left\{ [R_{i1}, ..., R_{iL}]^T : \sum\nolimits_{l\in\Omega} R_{il} \le g_j(\Omega), \forall \Omega \subseteq \{1, 2, ..., L\} \right\}, \qquad (31)$$

where the function $g_j(\Omega)$ is the r.h.s of the inequality in (21). Similarly, it can be verified that the region S-SND in (30) is of the form (31) except now $g_j(\Omega) = \infty$ if $j \notin \Omega$, and is also a convex polytope.

The following lemma can be used to find the maximum possible value for the minimum entry of a vector $\boldsymbol{R}$, where $\boldsymbol{R} \in \mathscr{R}_j$.

Lemma 2. In the polytope $\mathscr{R}_j$, define

$$\pi = \max \quad \min_{i\in S} \quad R_i \qquad (32)$$

$$\text{subject to} \quad [R_1, ..., R_L]^T \in \mathscr{R}_j, \qquad (33)$$

where $S = \{1, 2, ..., L\}$. *Then,*

$$\pi = \min_{\Omega \subseteq \{1, 2, ..., L\}, \Omega \ne \emptyset} \frac{g_j(\Omega)}{|\Omega|}. \qquad (34)$$

Proof. Following the steps of [36], consider an arbitrary vector $\boldsymbol{R} \in \mathscr{R}_j$, and define $\delta = \min_i R_i$. Hence, for all $\Omega \ne \emptyset$, we have $\delta \le \sum_{i\in\Omega} R_i / |\Omega| \le g_j(\Omega)/|\Omega|$. Therefore, $\min_{\Omega \ne \emptyset} g_j(\Omega)/|\Omega|$ is an upper bound on $\min_i R_i$. Choosing $\boldsymbol{R} = (\pi_0, ..., \pi_0) \in \mathscr{R}_j$, where $\pi_0 = \min_{\Omega \ne \emptyset} g_j(\Omega)/|\Omega|$, the upper bound is thus achieved. $\qquad \square$

Thus, the maximum symmetric rate (which also maximizes the minimum rate due to Lemma 2) at BS j is

$$R_{\text{Sym}}^j = \min_{\Omega_j \subseteq \{1, 2, ..., L\}, \Omega_j \ne \emptyset} \frac{g_j(\Omega_j)}{|\Omega_j|}. \qquad (35)$$

Finally, to find the maximum symmetric rate network-wide one needs to compute $\min R_{\text{Sym}}^j$ for $j \in \{1, 2, ..., L\}$. In the following, we discuss how (35) can be solved over various regions.

SD: At BS j, the minimization over the SD region in (21) can be carried out by solving

$$\min_{\Omega_j} \quad \frac{I(\hat{y}_{ji}; \, \boldsymbol{x}_{\Omega_j} \mid \boldsymbol{x}_{\Omega_j^c})}{|\Omega_j|} \qquad (36)$$

$$\text{subject to} \quad \Omega_j \subseteq \{1, 2, ..., L\}. \qquad (37)$$

The inequality in (27) then allows one to find an achievable lower bound to the above problem by solving

$$[\mathcal{P}1] \quad \min_{\Omega_j} \quad \frac{1}{|\Omega_j|} \log\left(1 + \frac{M \sum_{l \in \Omega_j} \sqrt{\rho_\mathrm{p}} \rho_\mathrm{u} \beta_{jil} \alpha_{jil}}{\sum_{l=1}^{L} \sum_{k=1}^{K} \rho_\mathrm{u} \beta_{jkl} + 1}\right) \tag{38}$$

$$\text{subject to} \quad \Omega_j \subseteq \{1, 2, ..., L\}. \tag{39}$$

SND: It can be seen from (29) that the region achieved by SND at BS j can not in general be represented by the intersection of a finite number of half-spaces and thus does not fall in the category of convex polytopes. Hence, finding the maximum symmetric rate over this region in general does not appear to have a closed-form formulation as in (35). However, in order to provide insights into the benefits of using SND, below we investigate S-SND which provides a lower bound to SND. The special cases of two-cell and three-cell systems, for which analyzing SND is tractable, are also studied at the end of this section.

S-SND: Under S-SND, it is solved at BS j

$$[\mathcal{P}2] \quad \min_{\Omega_j} \quad \frac{1}{|\Omega_j|} \log\left(1 + \frac{M \sum_{l \in \Omega_j} \sqrt{\rho_\mathrm{p}} \rho_\mathrm{u} \beta_{jil} \alpha_{jil}}{\sum_{l=1}^{L} \sum_{k=1}^{K} \rho_\mathrm{u} \beta_{jkl} + 1}\right) \tag{40}$$

$$\text{subject to} \quad \{j\} \subseteq \Omega_j \subseteq \{1, 2, ..., L\}. \tag{41}$$

Note that even though $[\mathcal{P}1]$ and $[\mathcal{P}2]$ have the same objective function, following the discussion below (30) the solution Ω_j of $[\mathcal{P}2]$ must include the index j associated with the rate R_{ij}, and is thus not necessarily identical to that of $[\mathcal{P}1]$.

To tackle $[\mathcal{P}1]$ (or $[\mathcal{P}2]$), we first consider two extreme regimes of high and low SINR.

A. High SINR regime

In this regime, the values of M and L are such that

$$\log\left(1 + \frac{M \sum_{l \in \Omega_j} \sqrt{\rho_\mathrm{p}} \rho_\mathrm{u} \beta_{jil} \alpha_{jil}}{\sum_{l=1}^{L} \sum_{k=1}^{K} \rho_\mathrm{u} \beta_{jkl} + 1}\right) \simeq \log(M). \tag{42}$$

For instance, this approximation holds when the number of BS antennas M is truly large but finite while the number of cells L is fixed.

Thus, in this regime the minimization in both $[\mathcal{P}1]$ and $[\mathcal{P}2]$ is achieved by $\Omega_j^* = \{1, 2, ..., L\}$, and thereby the maximum symmetric rate at BS j is given by

$$R_{\mathrm{Sym}}^{\mathrm{SD},j} = R_{\mathrm{Sym}}^{\mathrm{S\text{-}SND},j} = \frac{I\left(\hat{y}_{ji}; \, x_1[i], x_2[i], ..., x_L[i]\right)}{L}, \tag{43}$$

which scales as $O(\log M)$. As discussed before, the performance of SND is at least as good as SD and S-SND, i.e., $R_{\mathrm{Sym}}^{\mathrm{SND},j} \geq R_{\mathrm{Sym}}^{\mathrm{SD},j} = R_{\mathrm{Sym}}^{\mathrm{S\text{-}SND},j}$. Thus, in the high SINR regime the maximum symmetric rate of SND occurs on one of it's region boundaries that scales as $O(\log M)$. In other words, from (29) the maximum symmetric rate achieved by SND in the high SINR regime belongs to the full MAC, i.e., $R_{\mathrm{Sym}}^{\mathrm{SND},j} \in \mathcal{R}_{\mathrm{MAC}(\{1,...,L\},j)}^{i}$. Therefore, in the high SINR regime one can upper bound $R_{\mathrm{Sym}}^{\mathrm{SND},j}$ by $R_{\mathrm{Sym}}^{\mathrm{SND},j} \leq$

$\frac{1}{L} I\left(\hat{y}_{ji}; \, x_1[i], x_2[i], ..., x_L[i]\right)$. Consequently, we obtain for the high SINR regime $R_{\mathrm{Sym}}^{\mathrm{SND},j} = R_{\mathrm{Sym}}^{\mathrm{SD},j} = R_{\mathrm{Sym}}^{\mathrm{S\text{-}SND},j}$. To find the allocation network-wide, denoted by R_{Sym}, one needs to calculate the smallest value of (43) across all cells, i.e.,

$$R_{\mathrm{Sym}} = \min_{j} \frac{I\left(\hat{y}_{ji}; \, x_1[i], x_2[i], ..., x_L[i]\right)}{L}, \tag{44}$$

which is the same for all interference decoding schemes. Therefore, compared to TIN we obtain

$$R_{\mathrm{Sym}} > \min_{j} I\left(\hat{y}_{ji}; \, x_j[i]\right) = R_{\mathrm{Sym}}^{\mathrm{TIN}}, \tag{45}$$

i.e., joint decoding of all signals $\{x_l[i]\}_{l=1}^{L}$ performs strictly better than decoding only the desired signal (e.g., $x_j[i]$ at BS j) while treating the interference signals (e.g., $\{x_l[i]\}_{l=1, l \neq j}^{L}$ at BS j) as noise (TIN).

Remark 6. Similar to the results of [14], [20], it is apparent from (44) that, for sufficiently large M, the proposed interference decoding schemes of this paper are also able to achieve rates that scale as $O(\log M)$. Note that the achievable rates of [14], [20] are higher than the ones reported in this paper due to an extra array processing at a centralized network controller, which results in a larger pre-log factor (e.g., 1 vs $1/L$ in the high SINR regime). However, such a centralized processing requires extra resources and hardware infrastructure to facilitate the BS cooperation at the network controller. In contrast, in this paper, all processing are performed locally at BSs without needing any cooperation.

Remark 7. Consider the alternative approach of using different pilots in different cells, as explained before (15). One should note that, for the regime of large but finite M, decoding all $K(L-1)+1$ number of interfering users at the current BS will generally produce a smaller symmetric rate than the approach of Section II-C which only decodes L users, due to the much smaller pre-log factor in the former case. For instance, in the regime of high SINR, using (42)-(44), the achieved maximum symmetric rate of the former case is $\approx 1/(K(L-1)+1)\log(M)$, whereas that of the latter case is $\approx 1/L\log(M)$. Hence, when decoding the PC interference, re-using orthonormal pilots cross all cells is preferred as, for finite M, it results in larger symmetric rate across the network.

Observation: In the high SINR regime, regardless of the cells geometry, all interference decoding schemes SD/SND/S-SND have identical performance and also strictly outperform TIN. In turn, one may choose to implement SD which has a simpler decoder.

B. Low SINR regime

In this regime, the values of M and L are such that

$$\log\left(1 + \frac{M \sum_{l \in \Omega_j} \sqrt{\rho_\mathrm{p}} \rho_\mathrm{u} \beta_{jil} \alpha_{jil}}{\sum_{l=1}^{L} \sum_{k=1}^{K} \rho_\mathrm{u} \beta_{jkl} + 1}\right) \simeq \frac{M \sum_{l \in \Omega_j} \sqrt{\rho_\mathrm{p}} \rho_\mathrm{u} \beta_{jil} \alpha_{jil}}{\sum_{l=1}^{L} \sum_{k=1}^{K} \rho_\mathrm{u} \beta_{jkl} + 1}. \tag{46}$$

For instance, this approximation holds when M and L are small such that the product of the number of BS antennas M and the ratio $\frac{\sum_{l \in \Omega_j} \sqrt{\rho_\mathrm{p}} \rho_\mathrm{u} \beta_{jil} \alpha_{jil}}{\sum_{l=1}^{L} \sum_{k=1}^{K} \rho_\mathrm{u} \beta_{jkl} + 1}$ becomes small.

Since $\alpha_{jil} = \frac{\sqrt{\rho_p}\beta_{jil}}{1+\rho_p\sum_{l_1=1}^{L}\beta_{jil_1}}$, provided that (46) holds one can see that $[\mathcal{P}1]$ (or $[\mathcal{P}2]$) has the same minimizer Ω_j^* as the problem below

$$[\mathcal{P}3] \qquad \min_{\Omega_j} \quad \frac{\sum_{l\in\Omega_j}\beta_{jil}^2}{|\Omega_j|}, \qquad (47)$$

subject to (39) for SD (or (41) for S-SND). To solve this problem, we first construct the sorted vector

$$\boldsymbol{\pi}_{ji} = \left[\beta_{ji(1)}^2, \beta_{ji(2)}^2, ..., \beta_{ji(L)}^2\right]^T, \qquad (48)$$

whose entries are $\beta_{ji1}, \beta_{ji2}, ..., \beta_{jiL}$ sorted in non-decreasing order, i.e., $\beta_{ji(1)}^2 \leq \beta_{ji(2)}^2 \leq ... \leq \beta_{ji(L)}^2$. Thus, considering a simple distance-based pathloss model for large-scale fading coefficients and assuming that users are associated to the nearest BS, $\beta_{ji(1)}$ and $\beta_{ji(L)}$ correspond to the furthest and closest users to BS j, respectively and $\beta_{ji(L)} = \beta_{jij}$. Further note that the objective function of $[\mathcal{P}3]$ is an averaging operation over a subset of the entries of (48). We now study the solution of $[\mathcal{P}3]$ under SD and S-SND in the following.

SD: Subject to (39), the average over any subset of the entries of (48) is always at least as large as the smallest entry, $\beta_{ji(1)}^2$, with equality when the average is only over the smallest entry. Therefore, Ω_j^* is the index of the BS located at the furthest distance from the BS j. In other words, in the low SINR regime the performance of SD at BS j is (unsurprisingly) limited by the rate of the furthest user from BS j, i.e., $R_{\text{Sym}}^{\text{SD},j} = I(\hat{y}_{ji}; x_{\Omega_j^*}[i] \mid \boldsymbol{x}_{S\setminus\Omega_j^*})$

S-SND: First, note that since Ω_j^* for S-SND at BS j contains index j, the averaging operation in $[\mathcal{P}3]$ must include the largest entry of (48), i.e., $\beta_{ji(L)}^2 = \beta_{jij}^2$. Moreover, by including any other entry of (48) to the averaging operation, its value decreases. Therefore, to find Ω_j^* one should start with the initial value $b^1 = (\beta_{ji(L)}^2 + \beta_{ji(1)}^2)/2$, and then repeatedly add the next smallest entry of the vector $\boldsymbol{\pi}_{ji}$ to the averaging operation. This process terminates when the average becomes larger than its value from the previous iteration. The following algorithm computes Ω_j^* for S-SND.

Algorithm 1. *Set $q = 1$, $b^q = (\beta_{ji(L)}^2 + \beta_{ji(1)}^2)/2$.*

(1) *Set $q = q + 1$, and compute $b^q = \frac{\beta_{ji(L)}^2 + \sum_{l=1,l\neq L}^{q}\beta_{ji(l)}^2}{q+1}$.*

(2) *If $b^q \geq b^{q-1}$, then stop and output $\Omega_j^* = \{j\} \cup \left\{l : \beta_{jil}^2 \in \{\boldsymbol{\pi}_{ji}[1:q-1]\}\right\}$, where $\boldsymbol{\pi}_{ji}[1:q-1]$ denotes the the first $q-1$ entries of $\boldsymbol{\pi}_{ji}$. Otherwise, go to step 1.*

Therefore, in the low SINR regime, by construction the allocation

$$R_{\text{Sym}}^{\text{S-SND},j} = \frac{I(\hat{y}_{ji}; \boldsymbol{x}_{\Omega_j^*} \mid \boldsymbol{x}_{S\setminus\Omega_j^*})}{|\Omega_j^*|}, \qquad (49)$$

where $\Omega_j^* \subseteq S = \{1, 2, ..., L\}$, is strictly larger than that of SD, i.e., in the low SINR regime S-SND strictly outperforms SD. Even though the performance of SND in the low SINR regime appears intractable, as previously pointed out S-SND

provides a lower bound to SND. Hence, network-wide we have $R_{\text{Sym}}^{\text{SND}} \geq R_{\text{Sym}}^{\text{S-SND}} > R_{\text{Sym}}^{\text{SD}}$.

C. General SINR

Now, consider the problem of determining the maximum symmetric rate of SD ($[\mathcal{P}1]$) or S-SND ($[\mathcal{P}2]$) in general, where approximations of high and low SINR are no longer assumed. In Appendix C, efficient methods to numerically compute the maximum symmetric rate of SD and S-SND are presented.

Since it is difficult to comment on the performance of maximum symmetric rate for SND in general due to the structure of the SND region, we next study two special cases of two-cell and three-cell systems which are analytically tractable. For the two-cell system, we find conditions under which either TIN/SND is optimal or interference decoding schemes SD/SND/S-SND are all optimal. Whereas, for the three-cell system we will briefly illustrate examples where SND outperforms *all the other schemes*.

1) Two-cell system: We now consider a cellular system consisting of only two cells, and denote the indices of the cells by $j = 1, 2$. Associated with the ith user, $i = 1, 2, ..., K$, the rate regions achieved at BS 1 are given as below.

SD: From (21) we obtain

$$R_{i1} \leq I\left(\hat{y}_{1i}; x_1[i] \mid x_2[i]\right) \qquad (50)$$
$$R_{i2} \leq I\left(\hat{y}_{1i}; x_2[i] \mid x_1[i]\right) \qquad (51)$$
$$R_{i1} + R_{i2} \leq I\left(\hat{y}_{1i}; x_1[i], x_2[i]\right). \qquad (52)$$

SND: From (29) we obtain

$$R_{i1} \leq I\left(\hat{y}_{1i}; x_1[i] \mid x_2[i]\right) \qquad (53)$$

$$R_{i1} + \min\left\{R_{i2}, I\left(\hat{y}_{1i}; x_2[i] \mid x_1[i]\right)\right\} \leq I\left(\hat{y}_{1i}; x_1[i], x_2[i]\right). \qquad (54)$$

S-SND: From (30) we obtain

$$R_{i1} \leq I\left(\hat{y}_{1i}; x_1[i] \mid x_2[i]\right) \qquad (55)$$
$$R_{i1} + R_{i2} \leq I\left(\hat{y}_{1i}; x_1[i], x_2[i]\right). \qquad (56)$$

Remark 8. One can similarly obtain the rate regions at BS 2 by replacing $\hat{y}_{1i}$ with $\hat{y}_{2i}$ and swapping appropriate indices in (50)-(56).

An interesting observation for a two-cell system is that the region for SND is the union of the SD/S-SND region and the TIN region. We now aim to investigate the performance of different schemes with maximum symmetric allocation. For the two-cell system, we first define the following cases:

Case (i): In this case, we have

$$I\left(\hat{y}_{1i}; x_2[i] \mid x_1[i]\right) < I\left(\hat{y}_{1i}; x_1[i]\right). \qquad (57)$$

Case (ii): In this case, we have

$$\frac{1}{2}I\left(\hat{y}_{1i}; x_1[i], x_2[i]\right)$$
$$\leq \min\left\{I\left(\hat{y}_{1i}; x_1[i] \mid x_2[i]\right), I\left(\hat{y}_{1i}; x_2[i] \mid x_1[i]\right)\right\}. \qquad (58)$$

Case (iii): In this case, we have

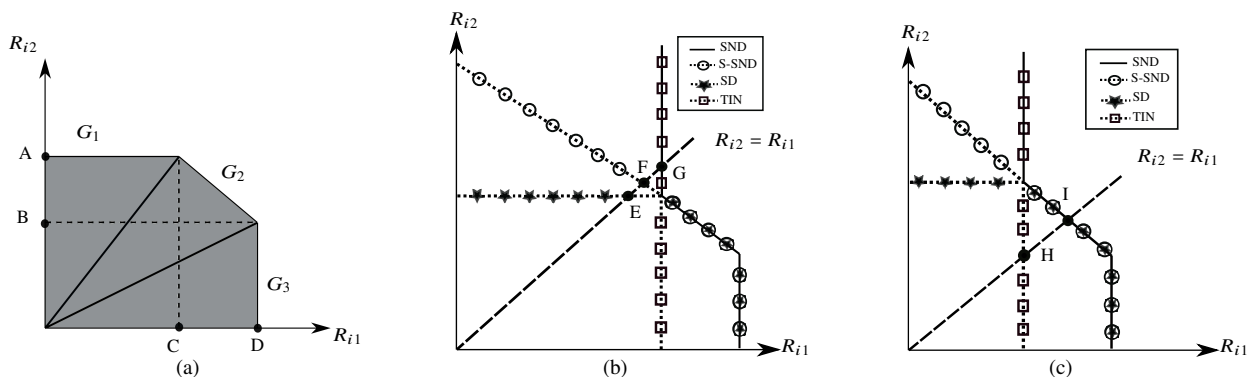$$I\left(\hat{y}_{1i}; x_1[i] \mid x_2[i]\right) < I\left(\hat{y}_{1i}; x_2[i]\right). \qquad (59)$$

Fig. 2: (a) Illustration of the SD region at BS 1 representing the 3 sub-regions G_1, G_2 and G_3 over which the diagonal $R_{i2} = R_{i1}$ will intersect a particular facet of the rate region, (b) Illustration of the rate regions achieved under TIN/SND/S-SND/SD at BS 1 for case (i): the diagonal $R_{i2} = R_{i1}$ intersects SD at point E, S-SND at point F, and SND/TIN at point G, resulting in (64), (c) Illustration of the rate regions achieved under TIN/SND/S-SND/SD at BS 1 for case (ii): the diagonal $R_{i2} = R_{i1}$ intersects TIN at point H, and SND/S-SND/SD at point I, resulting in (65).

From the perspective of the maximum symmetric rate, cases (i)-(iii) refer to conditions (in terms of mutual information) under which the diagonal $R_{i2} = R_{i1}$ intersects a particular facet of the rate region.

More specifically, consider the rate region achieved by SD at BS 1 depicted in Fig. 2a, where the entire region is divided into 3 sub-regions G_1, G_2 and G_3. Also, from (50)-(52), note that the corner points are given by $(C, A) = \left(I\left(\hat{y}_{1i}; x_1[i]\right), I\left(\hat{y}_{1i}; x_2[i] \big| x_1[i]\right)\right)$ and $(D, B) = \left(I\left(\hat{y}_{1i}; x_1[i] \big| x_2[i]\right), I\left(\hat{y}_{1i}; x_2[i]\right)\right)$. Now, the conditions under which the diagonal $R_{i2} = R_{i1}$ lies in sub-regions G_1, G_2 or G_3, are equivalent to the conditions of the three cases of (57)-(59) as follows: the diagonal $R_{i2} = R_{i1}$ lies in G_1, i.e., case (i) is true, iff $C > A$; the diagonal $R_{i2} = R_{i1}$ lies in G_2, i.e., case (ii) is true, iff $C \le A$ and $B \le D$; the diagonal $R_{i2} = R_{i1}$ lies in G_3, i.e., case (iii) is true, iff $B > D$. Specifically, the conditions for case (i) in (57) and case (iii) in (59) are exactly those given by $C > A$ and $B > D$, respectively.

For case (ii), note that one can also write

$$I(\hat{y}_{1i}; x_1[i], x_2[i]) = I(\hat{y}_{1i}; x_1[i]) + I(\hat{y}_{1i}; x_2[i] \mid x_1[i]) \tag{60}$$

$$= I(\hat{y}_{1i}; x_2[i]) + I(\hat{y}_{1i}; x_1[i] \mid x_2[i]). \tag{61}$$

Hence, in case (ii) where we have $C \le A$ and $B \le D$, by replacing C and B with their respective identity from (60) and (61), we reach the following conditions

$$\frac{1}{2} I(\hat{y}_{1i}; x_1[i], x_2[i]) \le I(\hat{y}_{1i}; x_2[i] \mid x_1[i]) \tag{62}$$

$$\frac{1}{2} I(\hat{y}_{1i}; x_1[i], x_2[i]) \le I(\hat{y}_{1i}; x_1[i] \mid x_2[i]), \tag{63}$$

resulting in (58).

Remark 9. If the worst-case uncorrelated noise bound in (27) is substituted for the mutual information expressions in (57)-(59), case (iii) can never happen as the effects of small-scale fading vanish in (27) and thus the received power of $x_2[i]$ at BS 1 can not be larger than that of $x_1[i]$ at BS 1. Hence, case (i) and case (ii) can be viewed as two complimentary and exhaustive conditions for a two-cell system at BS 1.

Note that the bounds of (27) differ from the mutual expressions in (59) due to two factors: (a) the expressions in (59) depend on the specific fading gains, and (b) the effective noise is not necessarily Gaussian. However, in the limit of large M the channel hardening of (12) minimizes the effects of (a). Moreover, due to the channel hardening of (12) as well as the assumption of Gaussian signaling in Theorem 1, the interference terms (effective noise) in (11) are asymptotically Gaussian.

The performance comparison of various schemes at BS 1 is summarized in the following corollary.

Corollary 1. *If the condition of case (i) in (57) holds at BS 1, then*

$$R_{Sym}^{SD,1} < R_{Sym}^{S\text{-}SND,1} < R_{Sym}^{SND,1} = R_{Sym}^{TIN,1}, \tag{64}$$

otherwise, if the condition of case (ii) in (58) holds at BS 1, then

$$R_{Sym}^{TIN,1} \le R_{Sym}^{SD,1} = R_{Sym}^{SND,1} = R_{Sym}^{S\text{-}SND,1}, \tag{65}$$

with strict equality in (65) if and only if (58) holds with strict equality.

Proof. See Appendix D. □

Fig. 2 illustrates an example of this corollary. Sub-figure (a) represents case (i) and its consequence in (64), whereas sub-figure (b) represents case (ii) and its consequence in (65).

To comment on the performance of various schemes over both cells, we consider a symmetric setting which is easy to analyze, and provides insights into the benefits of employing interference decoding schemes.

We define the symmetric setting as a scenario, where the MACs at both BS 1 and 2 are identical. Therefore, if case (i) is active at BS 1, it is also active at BS 2, and the resulting rates are equal at both BSs. Following Remark 8 it is thus obtained network-wide that

$$R_{Sym}^{SD} < R_{Sym}^{S\text{-}SND} < R_{Sym}^{SND} = R_{Sym}^{TIN}. \tag{66}$$

Observation: Both SND and TIN achieve the same performance and strictly outperform SD and S-SND. Thus, TIN may be the better choice of strategy in practice due to its simplicity.
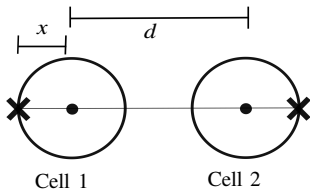
Fig. 3: An example of symmetric geometry with circular cells at a fixed distance of d from each other, where users are located at a distance x from the their BSs located at the center of the cells. The position of users is denoted by '$\times$'.

Similarly, if case (ii) is active with strict inequality at BS 1, it is also active with strict inequality at BS 2, and the resulting rates are equal at both BSs. Following Remark 8 it is thus obtained network-wide that

$$R_{\text{Sym}}^{\text{TIN}} < R_{\text{Sym}}^{\text{SD}} = R_{\text{Sym}}^{\text{SND}} = R_{\text{Sym}}^{\text{S-SND}}. \tag{67}$$

Observation: The interference decoding schemes SD/SND/S-SND achieve the same performance and strictly outperform TIN. Thus, SD (joint decoding of both users) may be the simplest one to implement in practice. Practical examples of these cases will be demonstrated in the numerical results section.

Consider, for instance, a setup where all users are located at a distance x from the corresponding BS as in Fig. 3. With respect to the lower bound in (27), this setup is symmetric as the effects of small-scale fading vanish in (27). Note that in a two-cell system, irrespective of whether the setting is symmetric or not, under no circumstances does SND strictly outperform all the other schemes. We next illustrate scenarios for a three-cell system, where SND can strictly outperform *all the other schemes*.

2) Three-cell system: Now, consider a cellular system consisting of only three cells, where the indices of the cells are denoted by $j = 1, 2, 3$. In this case, the rate regions under SD/S-SND can be obtained by a straightforward extension of (50)-(52) and (55)-(56) to the three-cell system, thus omitted for brevity. Moreover for SND, the rate region associated with the i-th user, $i = 1, 2, ..., K$, at BS 1 can be found using (29) as follows:

$$R_{i1} \le I\left(\hat{y}_{1i};\ x_1[i] \,\big|\, x_2[i], x_3[i]\right) \tag{68}$$

$$R_{i1} + \min\left\{I\left(\hat{y}_{1i};\ x_2[i] \,\big|\, x_1[i], x_3[i]\right),\ R_{i2}\right\}$$
$$\le I\left(\hat{y}_{1i};\ x_1[i], x_2[i] \,\big|\, x_3[i]\right) \tag{69}$$

$$R_{i1} + \min\left\{I\left(\hat{y}_{1i};\ x_3[i] \,\big|\, x_1[i], x_2[i]\right),\ R_{i3}\right\}$$
$$\le I\left(\hat{y}_{1i};\ x_1[i], x_3[i] \,\big|\, x_2[i]\right) \tag{70}$$

$$R_{i1} + \min\left\{I\left(\hat{y}_{1i};\ x_2[i], x_3[i] \,\big|\, x_1[i]\right),\right.$$
$$R_{i2} + I\left(\hat{y}_{1i};\ x_3[i] \,\big|\, x_1[i], x_2[i]\right),$$
$$R_{i3} + I\left(\hat{y}_{1i};\ x_2[i] \,\big|\, x_1[i], x_3[i]\right),$$
$$\left. R_{i2} + R_{i3}\right\}$$
$$\le I\left(\hat{y}_{1i};\ x_1[i], x_2[i], x_3[i]\right). \tag{71}$$

An example of this region is plotted in Fig. 4, where the dashed lines indicate that the region at BS 1 is unbounded in variables R_{i2} and R_{i3}, which is in agreement with property [P1] of the achievable region. Also, note that following Remark 8
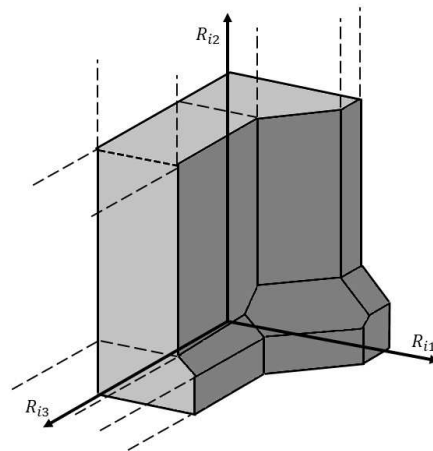

Fig. 4: An example of the rate region obtained by SND at BS 1.

the regions corresponding to BSs 2 and 3 can be similarly found. By comparing (68)-(71) with the achievable regions of SD and S-SND, it is noted that there are four faces in Fig. 4 that are only achieved by SND and not by any other schemes. More precisely at BS 1, it is possible for R_{Sym}^1 to achieve one of the rates, $I(\hat{y}_{1i};\ x_1[i] \,\big|\, x_3[i])$, $I(\hat{y}_{1i};\ x_1[i] \,\big|\, x_2[i])$, $\frac{1}{2}I(\hat{y}_{1i};\ x_1[i], x_2[i])$ or $\frac{1}{2}I(\hat{y}_{1i};\ x_1[i], x_3[i])$. Note that the first rate $I(\hat{y}_{1i};\ x_1[i] \,\big|\, x_3[i])$ can be interpreted as the maximum rate of the i-th user of cell 1, while treating the i-th user in cell 2 as noise. The second rate $I(\hat{y}_{1i};\ x_1[i] \,\big|\, x_2[i])$ can be interpreted similarly. Moreover, the rate $\frac{1}{2}I(\hat{y}_{1i};\ x_1[i], x_2[i])$ can be interpreted as the maximum symmetric rate achieved by joint decoding of the i-th users of cells 1 and 2, while treating the i-th user of cell 3 as noise. The fourth rate $\frac{1}{2}I(\hat{y}_{1i};\ x_1[i], x_3[i])$ can be interpreted similarly. Therefore, neither SD/S-SND nor TIN can provide these rates, in which case it is conceivable that SND could strictly outperform *all the other schemes*. More discussion will be provided in the numerical results section.

V. NUMERICAL RESULTS

To illustrate the performance of the two-cell system under both cases of (i) and (ii), we consider two different scenarios. In scenario (a), we assume that the cell radius and the distance of BSs are fixed, while the number of antennas M varies. In scenario (b), we assume that M and the distance of BSs are fixed, while the cell radius varies. In both scenarios for the geometry setting shown in Fig. 3, we compare the performance of various schemes TIN/SD/SND based on their achieved maximum symmetric rate using the bounds in (27). In particular, for scenario (a), we numerically quantify a threshold on M at which the transition from case (i) to case (ii) is observed. Analogously, for scenario (b), we numerically quantify a threshold on the distance x at which the transition from case (i) to case (ii) is observed. It should be pointed out that the results of this section are
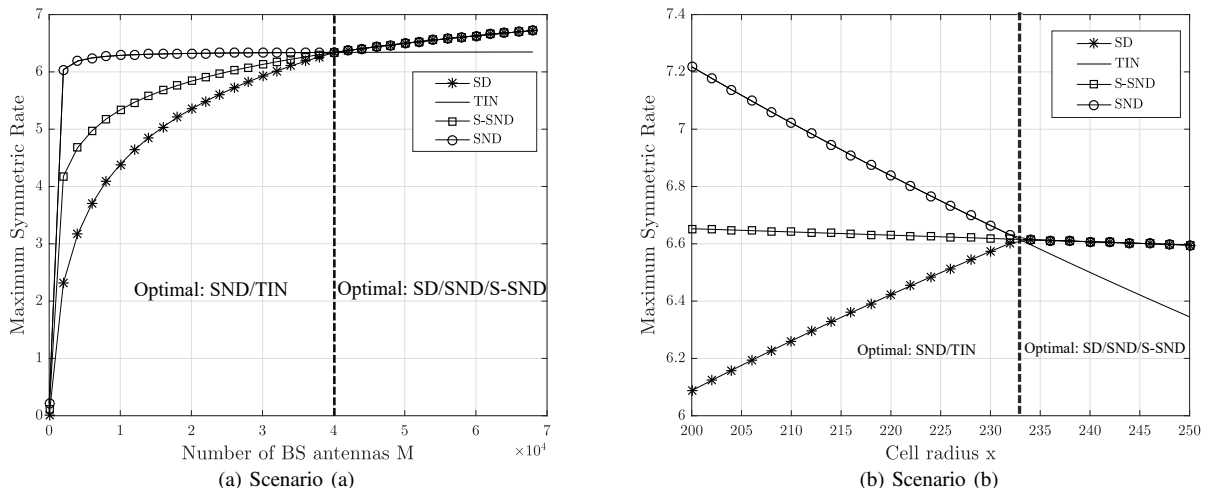
Fig. 5: Achieved maximum symmetric rate across two cells, R_{Sym}, over regions of TIN/SD/SND/S-SND: (a) for fixed values of $d = 2x$ and $x = 400$ m, as M is increased; (b) for fixed values of $d = 500$ m and $M = 5\times10^4$, as x is increased.

numerical examples presented only for the sake of illustration that validate the analytical findings of the previous sections. Hence, the identified thresholds depend on specific choices of system parameters. Moreover, for the large-scale fading coefficients β_{jil} we use a distance-based path loss model similar to [11] and neglect shadowing, i.e, $\beta_{jil} = (d_0/d_{jil})^{\alpha}$, where d_{jil} is the distance of the ith user in cell l from the BS j, α is the path loss exponent and d_0 is a normalization constant. For both scenarios (a) and (b), we take $K = 4$, $\alpha = 2$, $\rho_{\text{u}} = 30$, $\rho_{\text{p}} = 120$, and $d_0 = 100$ m. Achieved maximum symmetric rate of various schemes for scenario (a) are shown in Fig. 5a, where $x = 400$ m and $d = 2x$. It can be observed that for $M < 4 \times 10^4$, condition of case (i) in (57) is active; thus, SND and TIN have the same performance and strictly outperform SD/S-SND, i.e., $R_{\text{Sym}}^{\text{SD}} < R_{\text{Sym}}^{\text{S-SND}} < R_{\text{Sym}}^{\text{SND}} = R_{\text{Sym}}^{\text{TIN}}$. In other words, for $M < 4 \times 10^4$, to achieve the optimum performance each BS should only decode the signal of its own user while treating the signal of PC interference as noise. On the other hand, when $M > 4 \times 10^4$, the condition of case (ii) in (58) is active; thus, interference decoding schemes are all optimal, i.e., $R_{\text{Sym}}^{\text{TIN}} < R_{\text{Sym}}^{\text{SD}} = R_{\text{Sym}}^{\text{S-SND}} = R_{\text{Sym}}^{\text{SND}}$. Consequently, for significantly large values of M, to achieve the optimum performance each BS should jointly decode both the signal of its own user as well as that of the PC interference. This observation also matches with the consequence of the high SINR regime for truly large M in (45). Also, notice that there does not exist any range of M for which SND is strictly optimal. These observations are all in agreement with the analysis performed in subsection IV-C1.

Next consider scenario (b) where the BSs are at a distance of $d = 500$ m, $M = 5 \times 10^4$, and the cell radius x varies from 200 m to 250 m. The maximum symmetric rate for various schemes in this scenario, shown in Fig. 5b, illustrates that when approximately $x < 233$ m, the condition of case (i) in (57) is active, and thus using TIN is optimal. One implication of this observation is that for a fixed M, there
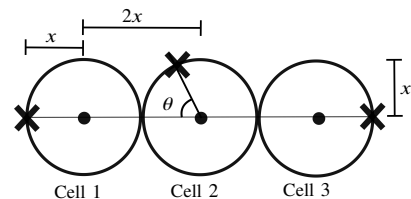

Fig. 6: Illustration of a three-cell system with circular cells, where all users of the left and the right cells are located on the cell edge at the farthest distance from the BSs located at the center of the cells, whereas the position of users on the edge of the middle cell is changing over $0° \leq \theta \leq 180°$. The position of users is denoted by '×'.

exists a threshold on cell radius such that if x is smaller than this threshold (i.e., 233 m in this example), interfering users located in the other cell are far away from the BS of the current cell, and hence treating interfering users as noise is optimal. On the other hand, if x is above the threshold, the condition of case (ii) in (58) is active: the interfering users are now close to the current BS, and thus interference decoding schemes achieve the optimal performance. We now extend the cell configuration model considered in Fig. 3, to the case of three cells. Particularly, we consider two scenarios with three circular cells based on Fig. 6 with $x = 400$ m, where the position of users in the left and the right cells is fixed on the cell edge at the maximum distance from the BSs located at the center, whereas those of the middle cell are determined based on the angle θ. More specifically, in scenario (a), it is assumed that the users of the middle cell are located on the cell edge at $\theta = 90°$, while in scenario (b), it is assumed that the position of users on the edge of the middle cell are swept over $0° \leq \theta \leq 180°$. Note that while the setting of Fig. 6 is impractical as all users will not be located at a single point at cell edge (i.e, at farthest distance from their BSs) in practice, nevertheless it does provide a conservative and
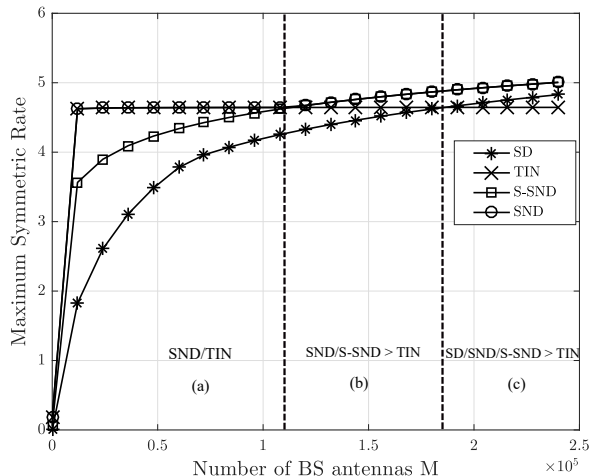
Fig. 7: Achieved maximum symmetric rate across three cells, R_{Sym}, over all four regions TIN/SD/SND/S-SND. The x-axis is divided into three intervals: (a) SND and TIN have identical performance and strictly outperform SD/S-SND, (b) SND/S-SND strictly outperform TIN/SD, (c) SD/SND/S-SND strictly outperform TIN.

somewhat pessimistic estimate of the user rates. In practice, users will typically be distributed more uniformly in the cell and some users will thus be located closer to their BSs. Hence, rate for users located closer to the BS will be higher than those located at cell edge, and thus the overall rate will be potentially higher.

The maximum symmetric rates of scenario (a) for various schemes and for different values of M are illustrated in Fig. 7, where the parameters of the setup are the same as before. It is observed from Fig. 7 that, even though for $M < 1.1 \times 10^5$ the performance of SND and TIN are identical and strictly better than SD/S-SND (and one can thus simply use TIN), for $M > 1.1 \times 10^5$ the PC interference is "strong" in that decoding it, as opposed to treating it as noise, produces better rates, i.e., SND/S-SND strictly outperform *all the other schemes* (and one should thus only use SND). In other words, for $1.1 \times 10^5 < M < 5 \times 10^5$ (the upper bound on M is not shown in the figure), the optimum performance is achieved only by SND and not by any other scheme. However, for $M > 5 \times 10^5$ all interference decoding schemes achieve the same rate as expected, while outperforming TIN. Finally, observe that for (approximately) $M > 1.85 \times 10^5$ SD outperforms TIN as well.

Finally, while keeping all other parameters the same as before, we consider scenario (b) where the position of users in the middle cell changes with θ, and plot the resulting maximum symmetric rates for various schemes against $0° \leq \theta \leq 180°$ with different values of BS antennas as illustrated in Fig. 8: (a) $M = 10^3$, (b) $M = 10^4$, (c) $M = 5 \times 10^4$, and (d) $M = 10^5$. We focus on the rates for $0° \leq \theta \leq 180°$ as, due to symmetry, the rates for $180° < \theta < 360°$ are identical to those for $0° \leq \theta \leq 180°$. Also, recall that the SND region, which contains TIN, S-SND and SD regions as special cases, as explained in subsection III-B, is always optimal.

Observe that in Fig. 8a for $M = 10^3$, the characteristics of rates can be classified into 3 regimes of θ: regime-1

where θ is smaller than $90°$ and SND outperforms TIN (i.e., $\theta \leq 66°$ in the setting of Fig. 8a), regime-2 where θ is close to $90°$ and SND and TIN have the same performance (i.e., $66° < \theta \leq 114°$), and regime-3 where θ is larger than $90°$ and SND outperforms TIN again (i.e., $\theta \geq 114°$). In regime-1, we have $R_{\text{Sym}}^{\text{SND}} > R_{\text{Sym}}^{\text{S-SND}} > R_{\text{Sym}}^{\text{TIN}} > R_{\text{Sym}}^{\text{SD}}$, as explained below. Note that θ captures the distance between users and BSs in different cells, and for θ in regime-1, users in the middle cell are much closer to BS 1 and farther away from BS 3. Therefore, users in the middle cell creates "strong" PC interference at BS 1 and "weak" PC interference at BS 3; hence, SND/S-SND outperforms all other schemes as it allows for users of the middle cell to be decoded at BS 1 and to be treated as noise at BS 3. In comparison, TIN provides poor performance as treating "strong" users from the middle cell as noise at BS 1 drastically reduces the rates, whereas SD performs poorly as it requires decoding of "weak" PC interference from users in the middle cell at BS 3. In the complementary setting of regime-3, the same principles apply with the roles of BS 1 exchanged with BS 3.

In contrast, for θ in regime-2, we have $R_{\text{Sym}}^{\text{SND}} = R_{\text{Sym}}^{\text{TIN}} > R_{\text{Sym}}^{\text{S-SND}} > R_{\text{Sym}}^{\text{SD}}$. Here, users in the middle cell are somewhat far from both BS 1 and BS 2, and thus the resulting PC interference becomes "weak" at both BS 1 and BS 2. Hence, performing TIN at both BSs is optimal and provides identical performance to that of SND while outperforming S-SND and SD.

As M is increased, the decoding rates continue to follow the same trend, but regime-2 shrinks to a small set of angles near $90°$, while the two other regimes expand, as illustrated in Fig. 8a to Fig. 8d. As M increases, users in the middle cell produce progressively stronger interference at BS 1 and BS 3, and thus using TIN at BS 1 and BS 3 to treat these users as noise results in poor rates except for a small set of θ near $90°$. These trends are similar to those in Fig. 7, where θ was fixed at $\theta = 90°$. These trends also confirm that when considering more than two cells, depending on the geometry and the parameters of the setting (e.g., M, cell radius, etc), there exist scenarios where SND performs strictly better than TIN/SD/S-SND.

VI. Conclusion

In this paper, to address the PC problem in the uplink of a multi-cell massive MIMO system it was proposed to decode the interference caused by PC rather than treating it as noise. In particular, when MRC is used at BS, it was shown that by coding over multiple coherence intervals and decoding the PC interference, the per-user rates tend to infinity as $M \to \infty$. Moreover, it was shown that when decoding the interference, using the same pilots across all cells (as opposed to using different pilots) is preferred as it results in decoding significantly fewer interference terms at each BS. A worst-case uncorrelated noise technique was also established for multiple access channels, from which achievable rates under two interference decoding schemes SD/SND were found for finite M. Comparing the performance of different schemes based on their maximum symmetric rate, structural results
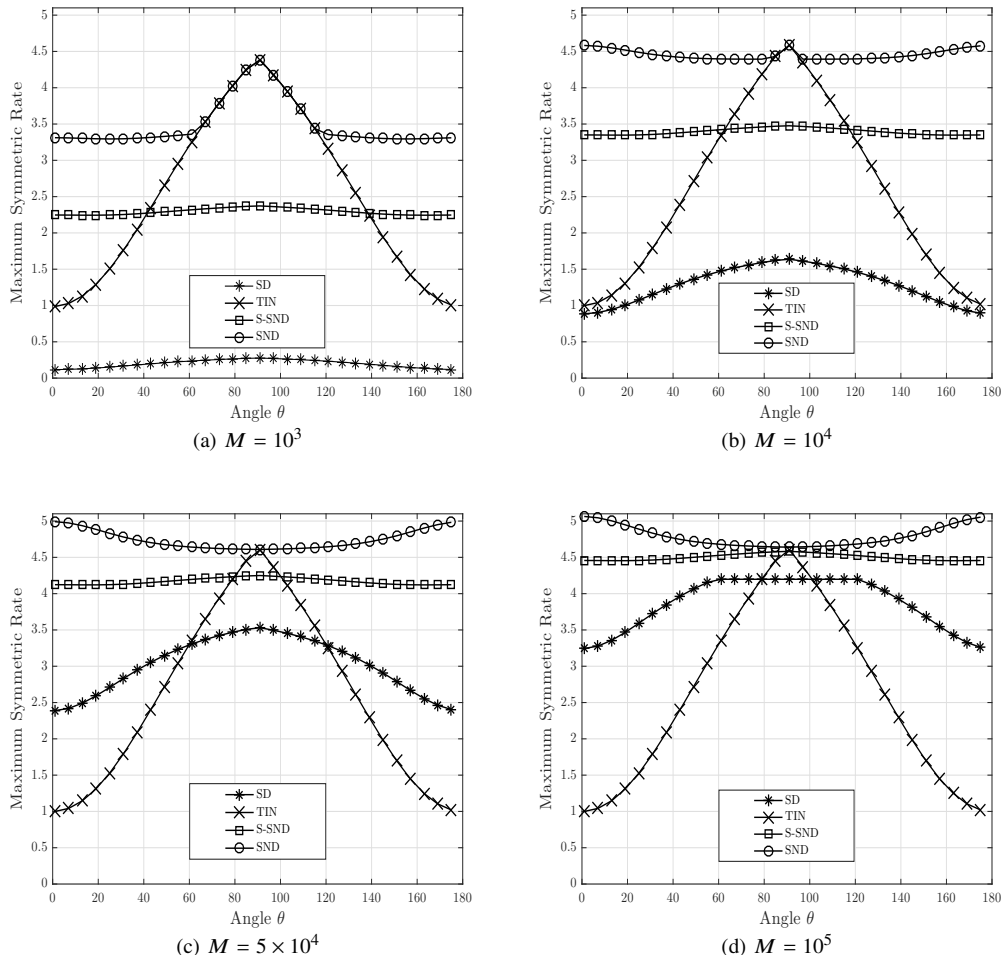
Fig. 8: Achieved maximum symmetric rates across three cells, R_{Sym}, over all four regions TIN/SD/SND/S-SND for $0° \leq \theta \leq 180°$, (a) $M = 10^3$, (b) $M = 10^4$, (c) $M = 5 \times 10^4$, (d) $M = 10^5$.

were found for the extreme regimes of high and low SINR. Specifically, for the high SINR regime when M is truly large, it was found that all interference decoding schemes achieve the maximum sum-rate and thus strictly outperform TIN.

The special cases of two-cell and three-cell systems were also studied. In the case of a two-cell system with symmetric geometry, conditions were found under which either SND and TIN achieve the same performance and are optimal (one should thus treat interference as noise), or all interference decoding schemes achieve the same performance and are optimal (one should thus jointly decode the desired signal as well as the PC interference). Furthermore, the analytical findings were numerically validated by quantifying a threshold on M (or on the cell radius), where TIN/SND was shown to be optimal below this threshold. On the other hand, beyond this threshold it was observed that only the interference decoding schemes achieve optimum performance.

Also in the case of a three-cell system, it was numerically shown that there exists a range of M for which the optimum performance is achieved only by SND and not by any other scheme. Hence, it was concluded that for large enough M when there are more than two cells (which is indeed true in

practice), SND can strictly outperform *all the other schemes*.

One possible future extension is to consider the downlink counterpart of this problem using well-known linear precoding techniques such as maximum ratio transmission (MRT), zero forcing (ZF), etc., where each BS simultaneously serves K users inside its cell. Specifically, after performing an arbitrary precoding technique at all BSs, K non-interfering L-user ICs will be obtained, whereby with simultaneous unique/non-unique decoding (SD/SND) of the intended signal along with the PC interference at each user, one can find achievable rates similar to (27) that scale as $O(\log M)$. Another possible future extension is to consider a correlated Rayleigh fading channel that will lead to a non-diagonal channel covariance matrix. This change of channel model will change the MMSE estimate of the channel vector in (6)-(8) and consequently the distribution of the channel estimate and the estimation error, and thus the power of different terms in P_1-P_4 in (23)-(26). One should note that, even though the updated expressions of (23)-(26) result in a new rate lower bound, it would still grow as $O \log(M)$ (similar to (27)), and thereby the final conclusions will remain the same.

APPENDIX A

Without loss of generality assume that $\Omega = \{1, 2, ..., l\}$ and thus $\Omega^c = \{l + 1, ..., L\}$. We start by expanding the r.h.s of (20) as follows

$$I\left(\boldsymbol{x}_\Omega^G; \, y | \boldsymbol{x}_{\Omega^c}^G\right) = h\left(\boldsymbol{x}_\Omega^G | \boldsymbol{x}_{\Omega^c}^G\right) - h\left(\boldsymbol{x}_\Omega^G | y, \boldsymbol{x}_{\Omega^c}^G\right) \quad (72)$$

$$\overset{(a)}{=} h\left(\boldsymbol{x}_\Omega^G\right) - h\left(\boldsymbol{x}_\Omega^G | y, \boldsymbol{x}_{\Omega^c}^G\right) \quad (73)$$

$$\overset{(b)}{=} \log\left((\pi e)^l \Pi_{i=1}^l P_i\right) - h\left(\boldsymbol{x}_\Omega^G | y, \boldsymbol{x}_{\Omega^c}^G\right), \quad (74)$$

where (a) is because the entries of $\boldsymbol{x}_\Omega^G$ and $\boldsymbol{x}_{\Omega^c}^G$ are independent, and (b) follows from the entropy of a complex Gaussian vector with independent entries. Also, using the chain rule one can write

$$h\left(\boldsymbol{x}_\Omega^G | y, \boldsymbol{x}_{\Omega^c}^G\right) \quad (75)$$

$$= \sum_{i \in \Omega} h\left(x_i^G | x_1^G, ..., x_{i-1}^G, y, \boldsymbol{x}_{\Omega^c}^G\right)$$

$$= \sum_{i \in \Omega} h\left(x_i^G - \alpha_i\left(y - \sum_{j=1}^{i-1} x_j^G - \sum_{k \in \Omega^c} x_k^G\right) \Big| x_1^G, ..., x_{i-1}^G, y, \boldsymbol{x}_{\Omega^c}^G\right),$$

where α_i is any constant. Defining $\tilde{y}_i = y - \sum_{j=1}^{i-1} x_j^G - \sum_{k \in \Omega^c} x_k^G$, we obtain

$$\sum_{i \in \Omega} h\left(x_i^G - \alpha_i \tilde{y}_i | x_1^G, ..., x_{i-1}^G, y, \boldsymbol{x}_{\Omega^c}^G\right) \quad (76)$$

$$\overset{(c)}{\leq} \sum_{i \in \Omega} h\left(x_i^G - \alpha_i \tilde{y}_i\right)$$

$$\overset{(d)}{\leq} \sum_{i \in \Omega} \log\left((\pi e) \text{var}\left[x_i^G - \alpha_i \tilde{y}_i\right]\right), \quad (77)$$

where (c) is due to the fact that conditioning reduces the entropy and (d) follows as Gaussian distributions maximize entropy. To obtain the tightest upper bound, one should minimize $\text{var}[x_i^G - \alpha_i \tilde{y}_i]$, i.e., $\alpha_i \tilde{y}_i$ must be the LMMSE estimate of x_i^G. More precisely, one can choose $\alpha_i = \mathbb{E}[\tilde{y}_i^* \tilde{y}_i]^{-1} \mathbb{E}[x_i^G \tilde{y}_i^*] = P_i/(\sum_{j=i}^l P_j + \sigma_z^2)$, where the second equality follows since z is uncorrelated from the users' signals. Thus

$$\text{var}\left[x_i^G - \alpha_i \tilde{y}_i\right] = \frac{P_i\left(\sum_{j=i+1}^l P_j + \sigma_z^2\right)}{\sum_{j=i}^l P_j + \sigma_z^2}, \quad (78)$$

and therefore we obtain

$$h\left(\boldsymbol{x}_\Omega^G | y, \boldsymbol{x}_{\Omega^c}^G\right) \leq \sum_{i \in \Omega} \log\left((\pi e) \frac{P_i\left(\sum_{j=i+1}^l P_j + \sigma_z^2\right)}{\sum_{j=i}^l P_j + \sigma_z^2}\right) \quad (79)$$

$$= \log\left((\pi e)^l \frac{(\Pi_{i \in \Omega} P_i) \sigma_z^2}{\sum_{i \in \Omega} P_i + \sigma_z^2}\right). \quad (80)$$

Hence, from (74) the following lower bound is obtained

$$I\left(\boldsymbol{x}_\Omega^G; \, y | \boldsymbol{x}_{\Omega^c}^G\right)$$

$$\geq \log\left((\pi e)^l \Pi_{i \in \Omega} P_i\right) - \log\left((\pi e)^l \frac{(\Pi_{i \in \Omega} P_i) \sigma_z^2}{\sum_{i \in \Omega} P_i + \sigma_z^2}\right) \quad (81)$$

$$= \log\left(1 + \frac{\sum_{i \in \Omega} P_i}{\sigma_z^2}\right) = I\left(\boldsymbol{x}_\Omega^G; \, y^G | \boldsymbol{x}_{\Omega^c}^G\right). \quad (82)$$

APPENDIX B

We start by computing the power of the desired signals, P_1. Note that

$$P_1 = \sum_{l \in \Omega} \rho_{\text{u}} \left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^\dagger \boldsymbol{g}_{jil}\right]\right|^2 = \sum_{l \in \Omega} \rho_{\text{u}} \left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^\dagger \left(\hat{\boldsymbol{g}}_{jil} + \boldsymbol{e}_{jil}\right)\right]\right|^2 \quad (83)$$

$$\overset{(a)}{=} \sum_{l \in \Omega} \rho_{\text{u}} \left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^\dagger \hat{\boldsymbol{g}}_{jil}\right]\right|^2 \quad (84)$$

$$= \sum_{l \in \Omega} M^2 \rho_{\text{u}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2 \frac{\rho_{\text{p}}^2 \beta_{jij}^4}{\left(1 + \rho_{\text{p}} \sum_{l_1=1}^L \beta_{jil_1}\right)^2} \quad (85)$$

$$= M^2 \sum_{l \in \Omega} \rho_{\text{p}} \rho_{\text{u}} \beta_{jil}^2 \alpha_{jij}^2, \quad (86)$$

where (a) follows from the fact that $\hat{\boldsymbol{g}}_{jij}$ and $\hat{\boldsymbol{e}}_{jij}$ are independent. Note that as explained earlier, all terms in the effective noise are uncorrelated; thus $\text{var}[z'_{jij}] = P_2 + P_3 + P_4$.

For the power of interference due to the channel estimation error, P_2, we write

$$P_2 = \sum_{l=1}^L \rho_{\text{u}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^\dagger \boldsymbol{g}_{jil} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^\dagger \boldsymbol{g}_{jil}\right]\right|^2\right] \quad (87)$$

$$= \sum_{l=1}^L \rho_{\text{u}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^\dagger \hat{\boldsymbol{g}}_{jil} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^\dagger \hat{\boldsymbol{g}}_{jil}\right]\right|^2\right] + \sum_{l=1}^L \rho_{\text{u}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^\dagger \boldsymbol{e}_{jil}\right|^2\right]. \quad (88)$$

For the first term in (88) we obtain

$$\sum_{l=1}^L \rho_{\text{u}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^\dagger \hat{\boldsymbol{g}}_{jil} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^\dagger \hat{\boldsymbol{g}}_{jil}\right]\right|^2\right]$$

$$= \sum_{l=1}^L \rho_{\text{u}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2 \text{var}\left[\hat{\boldsymbol{g}}_{jij}^\dagger \hat{\boldsymbol{g}}_{jij}\right] \quad (89)$$

$$= \sum_{l=1}^L M \rho_{\text{u}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2 \frac{\rho_{\text{p}}^2 \beta_{jij}^4}{\left(1 + \rho_{\text{p}} \sum_{l_1=1}^L \beta_{jil_1}\right)^2}. \quad (90)$$

Similarly, for the second term in (88) we obtain

$$\sum_{l=1}^L \rho_{\text{u}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^\dagger \boldsymbol{e}_{jil}\right|^2\right]$$

$$= \sum_{l=1}^L \rho_{\text{u}} \mathbb{E}\left[\text{tr}\left(\hat{\boldsymbol{g}}_{jij}^\dagger \boldsymbol{e}_{jil} \boldsymbol{e}_{jil}^\dagger \hat{\boldsymbol{g}}_{jij}\right)\right] \quad (91)$$

$$= M \rho_{\text{u}} \left(\frac{\rho_{\text{p}} \beta_{jij}^2}{1 + \rho_{\text{p}} \sum_{l_1=1}^L \beta_{jil_1}}\right) \sum_{l=1}^L \left(\beta_{jil} - \frac{\rho_{\text{p}} \beta_{jil}^2}{1 + \rho_{\text{p}} \sum_{l_1=1}^L \beta_{jil_1}}\right). \quad (92)$$

Therefore, using (90) and (92), one can verify that

$$P_2 = \sum_{l=1}^L M \rho_{\text{u}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2 \left(\frac{\rho_{\text{p}}^2 \beta_{jij}^4}{\left(1 + \rho_{\text{p}} \sum_{l_1=1}^L \beta_{jil_1}\right)^2}\right) \quad (93)$$

$$+ M \rho_{\text{u}} \left(\frac{\rho_{\text{p}} \beta_{jij}^2}{1 + \rho_{\text{p}} \sum_{l_1=1}^L \beta_{jil_1}}\right) \sum_{l=1}^L \left(\beta_{jil} - \frac{\rho_{\text{p}} \beta_{jil}^2}{1 + \rho_{\text{p}} \sum_{l_1=1}^L \beta_{jil_1}}\right)$$

$$= M\sqrt{\rho_{\mathrm{p}}}\beta_{jij}\alpha_{jij}\sum_{l=1}^{L}\rho_{\mathrm{u}}\beta_{jil}. \tag{94}$$

For the power of the interference of other users, P_3, we write

$$P_3 = \sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}\rho_{\mathrm{u}}\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jkl}\right|^2\right] \tag{95}$$

$$= \sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}\rho_{\mathrm{u}}\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jkl}\right|^2\right] + \sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}\rho_{\mathrm{u}}\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{e}_{jkl}\right|^2\right] \tag{96}$$

$$= \sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}M\rho_{\mathrm{u}}\left(\frac{\rho_{\mathrm{p}}\beta_{jij}^2}{1+\rho_{\mathrm{p}}\sum_{l_1=1}^{L}\beta_{jil_1}}\right)\left(\frac{\rho_{\mathrm{p}}\beta_{jkl}^2}{1+\rho_{\mathrm{p}}\sum_{l_2=1}^{L}\beta_{jkl_2}}\right)$$
$$+ \sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}M\rho_{\mathrm{u}}\left(\frac{\rho_{\mathrm{p}}\beta_{jij}^2}{1+\rho_{\mathrm{p}}\sum_{l_1=1}^{L}\beta_{jil_1}}\right)$$
$$\times\left(\beta_{jkl}-\frac{\rho_{\mathrm{p}}\beta_{jkl}^2}{1+\rho_{\mathrm{p}}\sum_{l_2=1}^{L}\beta_{jkl_2}}\right) \tag{97}$$

$$= M\sqrt{\rho_{\mathrm{p}}}\beta_{jij}\alpha_{jij}\sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}\rho_{\mathrm{u}}\beta_{jkl}. \tag{98}$$

Finally, for the power of the noise, P_4, we obtain

$$P_4 = \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{n}_j\right|^2\right] = \mathrm{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right]\mathbb{E}\left[\boldsymbol{n}_j\boldsymbol{n}_j^{\dagger}\right]\right) \tag{99}$$

$$= M\frac{\rho_{\mathrm{p}}\beta_{jij}^2}{1+\rho_{\mathrm{p}}\sum_{l_1=1}^{L}\beta_{jil_1}} = M\sqrt{\rho_{\mathrm{p}}}\beta_{jij}\alpha_{jij}. \tag{100}$$

APPENDIX C

We now propose efficient methods to compute the maximum symmetric rate of SD ($[\mathcal{P}1]$) or S-SND ($[\mathcal{P}2]$), where the assumption of high and low SINR are no longer made.

SD: One may first compare all $2^L - 1$ (total number of non-empty subsets of $S = \{1, 2, ..., L\}$) different rates with a computational complexity that is exponential in the number of cells. However, due to the structure of $[\mathcal{P}1]$ we can find the minimum by comparing only L rates at each BS. Thus network-wide, this task can be done in $O(L^2)$ time as there are in total L cells in the network. For notational brevity define $s_q = \sum_{l=1}^{q}\beta_{ji(l)}^2$ as the sum of the first q entries of $\boldsymbol{\pi}_{ji}$ and also

$$\mu_{ji} = \frac{M\rho_{\mathrm{p}}\rho_{\mathrm{u}}}{\left(\sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\mathrm{u}}\beta_{jkl}+1\right)\left(1+\rho_{\mathrm{p}}\sum_{l=1}^{L}\beta_{jil}\right)}. \tag{101}$$

One can see that due to the structure of the entries of $\boldsymbol{\pi}_{ji}$, if $|\Omega_j^*| = q$ for some integer $1 \leq q \leq L$, the minimum of the objective function in $[\mathcal{P}1]$ is $1/q \, \log\left(1 + \mu_{ji}s_q\right)$. Hence, define

$$v_q = \log\left(1 + \mu_{ji}s_q\right) / q, \quad \forall q. \tag{102}$$

Then, $[\mathcal{P}1]$ reduces to $\min_q v_q$, which can be calculated by comparing L values.

S-SND: Similarly when S-SND is used at BS j, define

$$c_q = \begin{cases} \log\left(1 + \mu_{ji}\beta_{jij}^2\right), & q = 1 \\ \frac{1}{q}\log\left(1 + \mu_{ji}\left(\beta_{jij}^2 + \sum_{l=1,l\neq j}^{q}\beta_{ji(l)}^2\right)\right), & \text{otherwise.} \end{cases} \tag{103}$$

Therefore, $[\mathcal{P}2]$ reduces to $\min_q c_q$, which can also be calculated by comparing L values.

APPENDIX D

Using (50)-(56), it can be verified that in case (i) we have $R_{\mathrm{Sym}}^{\mathrm{SD},1} = I(\hat{y}_{1i}; x_2[i] \mid x_1[i])$, $R_{\mathrm{Sym}}^{\mathrm{S\text{-}SND},1} = \frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i])$, and $R_{\mathrm{Sym}}^{\mathrm{SND},1} = R_{\mathrm{Sym}}^{\mathrm{TIN},1} = I(\hat{y}_{1i}; x_1[i])$.

Furthermore, using (60), one can rewrite (57) as $I(\hat{y}_{1i}; x_2[i] \mid x_1[i]) < \frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i])$, and conclude that $R_{\mathrm{Sym}}^{\mathrm{SD},1} < R_{\mathrm{Sym}}^{\mathrm{S\text{-}SND},1}$. Moreover, using (60), one can rewrite (57) as $\frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i]) < I(\hat{y}_{1i}; x_1[i])$, and conclude that $R_{\mathrm{Sym}}^{\mathrm{S\text{-}SND},1} < R_{\mathrm{Sym}}^{\mathrm{SND},1} = R_{\mathrm{Sym}}^{\mathrm{TIN},1}$. This is illustrated in Fig. 2b. Similarly, using (50)-(56), it can be verified that in case (ii) we have $R_{\mathrm{Sym}}^{\mathrm{TIN},1} = I(\hat{y}_{1i}; x_1[i])$, and $R_{\mathrm{Sym}}^{\mathrm{SD},1} = R_{\mathrm{Sym}}^{\mathrm{SND},1} = R_{\mathrm{Sym}}^{\mathrm{S\text{-}SND},1} = \frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i])$. Also, using (60)-(61), one can rewrite (58) as

$\max\{I(\hat{y}_{1i}; x_1[i]), I(\hat{y}_{1i}; x_2[i])\} \leq \frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i])$. Hence, when this condition holds it yields (65), which is also illustrated in Fig. 2c.

REFERENCES

[1] "Greentouch final results from green meter research study: A greentouch white paper," tech. rep., GreenTouch Foundation, 2015.

[2] D. Aziz, K. Kusume, O. Queseth, H. Tullberg, F. Mikael, M. Schellmann, M. Uusitalo, and M. Maternia, "Mobile and wireless communications enablers for the twenty-twenty information society (METIS)," tech. rep., -, 2015.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[4] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 29, no. 2, pp. 6–14, 2015.

[5] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, 2014.

[6] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, 2015.

[7] S. Majhi and P. Mitran, "On the capacity and the optimal sum-rate of a class of dual-band interference channels," *Entropy*, vol. 19, no. 9, 2017.

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

[9] 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, 2013.

[10] E. Björnson, E. G. Larsson, and M. Debbah, "Optimizing multi-cell massive MIMO for spectral efficiency: How many users should be scheduled?," in *2014 IEEE Global Conf. Signal and Inf. Process. (GlobalSIP)*, pp. 612–616, IEEE, 2014.

[11] 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, 2013.

[12] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO 5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, 2016.

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

[14] A. Adhikary, A. Ashikhmin, and T. L. Marzetta, "Uplink interference reduction in large-scale antenna systems," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2194–2206, 2017.

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

[16] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving massive MIMO spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, 2012.

[17] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, 2013.

[18] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *2010 IEEE Int. Conf. Commun. (ICC)*, pp. 1–5, IEEE, 2010.

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

[20] T. Van Chien, C. Mollén, and E. Björnson, "Large-scale-fading decoding in cellular massive MIMO systems with spatially correlated channels," *IEEE Trans. Commun.*, 2018.

[21] B. Bandemer and A. El Gamal, "Interference decoding for deterministic channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2966–2975, 2011.

[22] F. Baccelli, A. El Gamal, and N. David, "Interference networks with point-to-point codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2582–2596, 2011.

[23] S. K. Sengijpta, "Fundamentals of statistical signal processing: Estimation theory," 1995.

[24] A. Sabharwal, D. Dash, and S. Diggavi, "Compound Gaussian multiple access channels with noisy feedback," in *2008 Annu. Allerton Conf. Commun., Control, and Comput.*, pp. 887–894, IEEE, 2008.

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

[26] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[27] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[28] J. Zhu and M. Gastpar, "Gaussian multiple access via compute-and-forward," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2678–2695, 2017.

[29] E. Sula, J. Zhu, A. Pastore, S. H. Lim, and M. Gastpar, "Compute–forward multiple access (CFMA): Practical implementations," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1133–1147, 2019.

[30] B. Bandemer, A. El Gamal, and Y.-H. Kim, "Optimal achievable rates for interference networks with random codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6536–6549, 2015.

[31] L. Wang, E. Şaşoğlu, and Y.-H. Kim, "Sliding-window superposition coding for interference networks," in *2014 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2749–2753, IEEE, 2014.

[32] H. Park, Y.-H. Kim, and L. Wang, "Interference management via sliding-window superposition coding," in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 972–976, IEEE, 2014.

[33] K. T. Kim, S.-K. Ahn, Y.-H. Kim, H. Park, L. Wang, C.-Y. Chen, and J. Park, "Adaptive sliding-window coded modulation in cellular networks," in *2015 IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–7, IEEE, 2015.

[34] K. T. Kim, S.-K. Ahn, Y.-S. Kim, J. Park, C.-Y. Chen, and Y.-H. Kim, "Interference management via sliding-window coded modulation for 5G cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 82–89, 2016.

[35] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6384–6399, 2016.

[36] S. Fujishige, *Submodular functions and optimization*, vol. 58. Elsevier, 2005.