Optimal SIC Ordering and Power Allocation in Downlink Multi-Cell NOMA Systems

Sepehr Rezvani, Student Member, IEEE, Eduard Axel Jorswieck, Fellow, IEEE, Nader Mokari Yamchi, Senior Member, IEEE, and Mohammad Reza Javan, Senior Member, IEEE

Abstract—In this work, we propose a globally optimal joint successive interference cancellation (SIC) ordering and power allocation (JSPA) algorithm for the sum-rate maximization problem in downlink multi-cell non-orthogonal multiple access (NOMA) systems. The proposed algorithm is based on the exploration of base stations (BSs) power consumption, and closed-form of optimal powers obtained for each cell. Although the optimal JSPA algorithm scales well with larger number of users, it is still exponential in the number of cells. For any suboptimal decoding order, we propose a low-complexity near-optimal joint rate and power allocation (JRPA) strategy in which the complete rate region of users is exploited. Furthermore, we design a near-optimal semi-centralized JSPA framework for a two-tier heterogeneous network such that it scales well with larger number of small-BSs and users. Numerical results show that JRPA highly outperforms the case that the users are enforced to achieve their channel capacity by imposing the well-known SIC necessary condition on power allocation. Moreover, the proposed semi-centralized JSPA framework significantly outperforms the fully distributed framework, where all the BSs operate in their maximum power budget. Therefore, the centralized JRPA and semi-centralized JSPA algorithms with near-optimal performances are good choices for larger number of cells and users.

Index Terms—Multi-cell, NOMA, successive interference cancellation, optimal SIC ordering, power allocation.

I. INTRODUCTION

A. Concept of NOMA

It is shown that the channel capacity of degraded broadcast channels (BCs) can be achieved by performing linear superposition coding (SC) in power domain at the transmitter side combined with coherent mutliuser detection algorithms, such as successive interference cancellation (SIC), at the receivers side [1]–[3]. The SC-SIC technique, also called power-domain non-orthogonal multiple access (NOMA), is considered as a candidate radio access technique for the fifth generation (5G) wireless networks and beyond [4]–[6]. In information theory, the main purpose of SC-SIC is reducing the prohibitive complexity of dirty paper coding (DPC) to attain the capacity region of degraded BCs.

B. Single-Cell NOMA

It is well-known that the downlink single-input single-output (SISO) Gaussian BCs are degraded [1]–[3]. Hence, NOMA with channel-to-noise ratio (CNR)-based decoding order is capacity-achieving in SISO Gaussian BCs meaning that any rate region is a subset of the rate region of NOMA with CNR-based decoding order [1]–[3]. The superiority of single-cell NOMA over single-cell OMA is also well-known in information theory [6]–[8].

In single-cell NOMA, it is verified that the power allocation optimization is necessary to achieve the maximum users sum-rate [5]–[10]. From the optimization perspective, the optimal (CNR-based) decoding order is independent from the power allocation, so is robust and straightforward. Moreover, it is shown that the Hessian of sum-rate function under the CNR-based decoding order is negative definite, so the sum-rate function is strictly concave in powers [11], [12]. In this way, the sum-rate maximization problem in downlink single-cell NOMA is convex.

There are some research studies on finding the closed-form of optimal powers for sum-rate maximization problem of $M$-user single-cell NOMA [11], [12], or its special 2-user case [13]. However, the analysis in [11]–[13] is based on some additional constraints on power allocation among multiplexed users to guarantee successful SIC. In information theory, it is proved that in SISO Gaussian BCs which are degraded, each user can achieve its channel capacity after SIC independent from power allocation among multiplexed users [2]. Hence, the SIC at users in SISO Gaussian BCs can be successfully proceed independent from power allocation among multiplexed users [10]. To this end, the Karush-Kuhn-Tucker (KKT) optimality conditions analysis as well as finding

1In this work, the term ‘NOMA’ is referred to power-domain NOMA.

2The feasible region of the general power allocation problem in single-cell NOMA under the minimum rate constraints is affine, so is convex [11], [12].
the closed-form of optimal powers for the general $M$-user single-cell NOMA are still open problems.

C. Multi-Cell NOMA With Single-Cell Processing

Unfortunately, the capacity-achieving schemes are still unknown in downlink multi-cell networks, since the capacity region of the two-user downlink interference channel is still unknown in general [1], [2], [7]. For the case that the information of each user is available at only one transmitter, i.e., multi-cell networks with single-cell processing, the signals from neighboring cells are fully treated as additive white Gaussian noise (AWGN) at each user, also called inter-cell interference (ICI). Inspired by the degradation of SISO Gaussian BCs, SC-SIC achieves the channel capacity of each cell in multi-cell systems with single-cell processing [1], [2]. In this system, ‘ICI+AWGN’ can be viewed as equivalent noise power at the users. Therefore, NOMA with channel-to-interference-plus-noise ratio (CINR)-based decoding order is capacity-achieving in each cell [1], [2]. In contrast to single-cell NOMA, finding optimal decoding order in multi-cell NOMA is challenging, because of the impact of ICI on the CINR of multiplexed users. The ICI at users in each cell is affected by the total power consumption of each neighboring (interfering) base station (BS). Therefore, the optimal SIC decoding order in each cell depends on the optimal power control among interfering cells. In this way, the optimal joint SIC ordering and power allocation (JSPA) problem in downlink multi-cell NOMA consists of two components as 1) Power control among interfering cells; 2) optimal joint SIC ordering and power allocation among multiplexed users within each cell. Note that the optimal SIC ordering in case 2 follows the CINR of users known from information theory [1], [2].

It is verified that under the CINR-based decoding order, the ICI in the centralized total power minimization problem verifies the basic properties of the standard interference function [14]–[17]. Hence, the optimal JSPA can be obtained by using the well-known Yates power control framework [18]. In other words, the globally optimal JSPA for the total power minimization problem in multi-cell NOMA can be found in an iterative distributed manner with a fast convergence speed. However, the ICI in the centralized sum-rate maximization problem does not verify the basic properties of the standard interference function. As a result, Yates power control framework does not guarantee any global optimality for the sum-rate maximization problem [16]. It is shown that the sum-rate function in multi-cell NOMA is nonconcave in powers, due to existing ICI, which makes the centralized sum-rate maximization problem nonconvex and strongly NP-hard [16], [19], [20]. The best existing solution for solving the power allocation problem is monotonic optimization which is still approximately exponential in the number of users [19], [21]. The JSPA needs to examine the monotonic-based power allocation $(M)_B^B$ times, where $(M)_B^B$ is the total number of possible decoding orders in $B$ cells each having $M$ users. Therefore, the joint optimization via the monotonic-based power allocation [21] is basically impractical even at lower number of BSs and users. Generally speaking, the JSPA sum-rate maximization problem in multi-cell NOMA falls into the following two schemes:

1) **Optimal SIC ordering (JSPA scheme):** In this scheme, we dynamically update the decoding order of users during power control among the interfering cells. This scheme achieves the globally optimal JSPA, however, it is not yet addressed in the literature.

2) **Fixed SIC ordering:** In this scheme, we consider a fixed decoding order (for instance, CNR-based decoding order) during power control among the cells, and find optimal power allocation. This scheme is suboptimal if and only if the fixed decoding order is not guaranteed to be optimal at any feasible region of power allocation. For any fixed, and possibly suboptimal, decoding order, there are two schemes in which we can guarantee successful SIC at users:

1) **Fixed-rate-region power allocation (FRPA):** In this scheme, we limit the power consumption of interfering cells such that the CINR of user with higher decoding order remains larger than that of user with lower decoding order. In other words, we impose an additional constraint on the power consumption of interfering cells, known as SIC necessary condition [19], [20], such that the fixed decoding order remains optimal. In this scheme, each user achieves its channel capacity for decoding its own signal after SIC. The monotonic optimization is used in [19] to find the optimal power allocation in the FRPA scheme. Due to the exponential complexity of monotonic optimization in the number of BSs and users, a sequential programming is proposed in [19], [20] to achieve a near-optimal solution with polynomial time complexity.

2) **Joint rate and power allocation (JRPA):** The so-called SIC necessary condition [19], [20] is indeed not a necessary condition for successful SIC when the decoding order is not guaranteed to be optimal. The SIC necessary condition in FRPA imposes additional limitations on the power consumption of neighboring cells to only guarantee that each user achieves its channel capacity for decoding its desired signal. Therefore, FRPA may degrade the rate region of users depending on how much the so-called SIC necessary condition limits the feasible region of BSs power consumption. In JRPA, the SIC necessary condition is removed, while users are allowed to operate in lower rates than their channel capacity to ensure successful SIC. Therefore, we consider the complete rate region of users, known from information theory [1], [2]. It is obvious that the rate region of FRPA is a subset of the rate region of JRPA. The power allocation problem in the JRPA scheme is not yet addressed in the literature. Therefore, finding an efficient algorithm for the JRPA scheme is still an open problem in multi-cell NOMA.

The work in [16] considers the FRPA scheme with CNR-based decoding order, while SIC necessary condition is not applied, which may lead to an infeasible solution. Moreover, the ICI management with efficient power control among the cells...
is not addressed in [22], since it is assumed that all the BSs operate in their maximum power budget. Subsequently, the challenges of finding optimal decoding order in multi-cell NOMA are not addressed in [22]. We show that the power allocation problem in [22] falls into our distributed resource allocation framework, where all the BSs operate in their maximum power budgets. In addition to the above-mentioned open problems in multi-cell NOMA, it is important to know the performance gap between optimal and CNR-based decoding orders, since adopting the CNR-based decoding order is more challenging. In other words, the performance gap between JSPA and FRPA/JRPA is still unknown. The latter performance gap depends on the suboptimality of CNR-based decoding order. Therefore, another interesting topic is how to ensure that the CNR-based decoding order is optimal independent from ICI? Moreover, for the CNR-based decoding order, it is still unknown how much is the performance gap between FRPA and JRPA?

D. Our Contributions

In this work, we address the above-mentioned open problems for the sum-rate maximization problem in the downlink of general single-carrier multi-cell NOMA system with single-cell processing. Our main contributions are presented as follows:

• We prove that at the optimal point, only the NOMA cluster-head user, which has the highest CNR (thus highest decoding order), deserves additional power, while the users with lower decoding order get power to only maintain their individual minimal rate demands. Subsequently, we obtain the closed-form expression of optimal powers among multiplexed users within each cell for any given feasible power consumption of BSs. The optimal value is also formulated in closed form. The analysis shows that the optimal power coefficients among multiplexed users are highly insensitive to the users channel gain, specifically for the high signal-to-interference-plus-noise ratio (SINR) regions.

• We propose a globally optimal JSPA algorithm for maximizing users sum-rate under the individual minimum rate demand of users. This algorithm utilizes both the exploration of BSs power consumption and closed-form of optimal powers for single-cell NOMA, so is a good benchmark for larger number of users while small number of BSs.

• The optimal JSPA algorithm is modified to find the globally optimal solution of the FRPA problem, resulting in dramatically reducing the complexity of monotonic optimization which is still exponential in the number of users.

• We propose a suboptimal sequential programming-based JRPA algorithm which considers the complete achievable rate region of users when the (potentially suboptimal) decoding order is prefixed. The convergence speed and performance of this algorithm for different initialization methods are investigated.

• We propose a sufficient condition which determines the optimality of CNR-based decoding order for a user pair independent from ICI. The applications of this theorem are listed in the paper.

• We propose a semi-centralized framework for a two-tier heterogeneous network (HetNet) consisting of multiple femto BSs (FBSs) underlying a single macro BS (MBS), where the ICI from MBS to the FBS users is managed efficiently.

The complete source code of the simulations including a user guide is available in [23]. Due to the space limitations, the additional notes are moved to the supplementary material in [24].

E. Paper Organization

The rest of this paper is presented as follows. Section II describes the general multi-cell NOMA system, and formulates the main optimization problem for maximizing users sum-rate. The solution algorithms are presented in Section III. Numerical results are provided in Section IV. Our conclusions and future research directions are presented in Section V.

II. GENERAL DOWNLINK MULTI-CELL NOMA SYSTEM

Consider the downlink transmission of a multiuser single-carrier multi-cell NOMA system. The set of single-antenna BSs, and users served by BS $b$ are indicated by $B$ and $U_b$, respectively. According to the NOMA protocol, the users associated to the same transmitter form a NOMA cluster. The signal of users associated to other transmitters (known as ICI) is fully treated as AWGN at users within the NOMA cluster. Hence, we consider a single NOMA cluster at each cell $b$ [14] including $|U_b|$ users, where $|.|$ is the cardinality of a finite set. The term $k \rightarrow i$ indicates that user $k$ has a higher decoding order than user $i$ such that user $k$ is scheduled (and enforced) to decode and cancel the whole signal of user $i$, while the whole signal of user $k$ is treated as noise at user $i$. For instance, assume that each cell $b$ serves $M$ users, i.e., $|U_b| = M$. Generally, there are $M!$ possible decoding orders for users within the $M$-order NOMA cluster. Without loss of generality, let $k \rightarrow i$ if $k > i$, i.e., the SIC of NOMA in each cell $b$ follows $M \rightarrow M - 1 \rightarrow \ldots \rightarrow 1$. As shown in Fig. 1 in [14], in this SIC decoding order, the signal of each user $i$ will be decoded prior to user $k > i$. In general, each user $i$ first decodes and cancels the signal of users $1, \ldots, i - 1$. Then, it decodes its desired signal such that the signal of users $i + 1, \ldots, M$ is treated as noise at user $i$ [6]. In this regard, in each cell, the NOMA cluster-head user $M$ does not experience any intra-NOMA interference (INI).

Let $s_{b,i} \sim CN(0, 1)$ be the desired signal of user $i \in U_b$. Denoted by $\lambda_{b,i,k} \in \{0, 1\}$, the binary decoding decision indicator, where $\lambda_{b,i,k} = 1$ if user $k \in U_b$ is scheduled to decode (and cancel when $k \neq i$) $s_{b,i}$, and otherwise, $\lambda_{b,i,k} = 0$. Since for each user pair within a NOMA cluster, only one user can decode and cancel the signal of other user, we have $\lambda_{b,i,k} + \lambda_{b,k,i} = 1, \forall b \in B, i, k \in U_b, k \neq i$. Moreover, the signal of each user should be decoded at that user, meaning that $\lambda_{b,i,i} = 1, \forall b \in B, i \in U_b$. Due to the
transitive nature of SIC ordering, if \( \lambda_{b,i,k} = 1 \) and \( \lambda_{b,k,h} = 1 \), then we should have \( \lambda_{b,i,h} = 1 \). In other words, we have \( \lambda_{b,i,k} \lambda_{b,k,h} \leq \lambda_{b,i,h}, \quad \forall b \in B, i, k, h \in U_b \). According to the SIC protocol, \( \Phi_{b,i} \) should be decoded at user \( i \in U_b \) as well as all the users in \( \Phi_{b,i} \) for convenience, let \( \Phi_{b,i} = \{ k \in U_b \setminus \{ i \} \mid \lambda_{b,i,k} = 1 \} \). Therefore, in the SIC of NOMA, each user \( i \) first decodes and subtracts each signal \( s_{b,j} \), \( \forall j \in U_b \setminus \{ i \} \cup \Phi_{b,i} \), then it decodes its desired signal \( s_{b,i} \) such that the signal of users in \( \Phi_{b,i} \) is treated as noise (called INI). According to the SIC protocol, the signal of user \( i \) will be decoded prior to user \( k \) if \( |\Phi_{b,i}| > |\Phi_{b,k}| \). According to the above, the SIC decoding order among users can be determined by finding \( \lambda_{b,i,k} \). Moreover, let us denote the matrix of all the decoding indicators by \( \Phi_{b,i} \). Without loss of generality, assume that \( s_{b,i} \in |U| \) \( s_{b,\lambda} \geq \lambda \). Actually, \( \lambda_{b,i,k} = 1, \quad \forall b \in B, \quad i, k \in U_b, \quad k \neq i \) is equivalent to \( k \rightarrow i \) in cell \( b \). Similar to the related works, we assume that the perfect channel state information (CSI) of all the users is available at the scheduler. The channel gain from BS \( j \in B \) to user \( i \in U_b \) is denoted by \( g_{j,b,i} \). The allocated power from BS \( b \) to user \( i \in U_b \) is denoted by \( p_{b,i} \). After performing (perfect) SIC at each user \( l \in U_b \setminus \Phi_{b,i} \), the received signal of user \( i \in U_b \) at user \( k \in \{ i \} \cup \Phi_{b,i} \) is given by

\[
y_{b,i,k} = \sqrt{p_{b,i} g_{b,i,b} s_{b,i}} + \sqrt{\sum_{j \in \Phi_{b,i}} p_{b,j} g_{b,j,b} s_{b,j}} + \sum_{j \in U_b \setminus \{ i \}} \sqrt{p_{b,j} g_{b,j,b} s_{j,b}} \tag{1}
\]

where the first and second terms are the received desired signal and INI of user \( i \in U_b \) at user \( k \in U_b \), respectively. The third term represents the ICICI at user \( k \in U_b \). Moreover, \( N_{b,k} \) is the AWGN at user \( k \in U_b \) with zero mean and variance \( \sigma^2_{b,k} \). Without loss of generality, assume that \( |n_{b,i}| = 1, \quad \forall b \in B, \quad i \in U_b, \) and \( h_{b,b,k} = |g_{b,b,k}|^2 \) [14], [16], [25]. According to (1), the SINR of user \( k \in \Phi_{b,i} \) for decoding and canceling the signal of user \( i \in U_b \) is \( \gamma_{b,i,k} = \frac{p_{b,i} h_{b,b,k}}{\sum_{j \in \Phi_{b,i}} p_{b,j} h_{b,b,k} + (h_{b,k} + \sigma^2_{b,k})} \),

where \( \sum_{j \in \Phi_{b,i}} h_{b,b,k} \) is the INI power of user \( i \in U_b \) received at user \( k \in U_b \), and \( h_{b,k} = \sum_{j \in U_b \setminus \{ i \}} p_{b,j} h_{b,j,b} \) is the received ICICI power at user \( k \in U_b \). For the case that \( k = i \), \( \gamma_{b,i,i} \) denotes the SINR of user \( i \in U_b \) for decoding its desired signal \( s_{b,i} \) after SIC. For convenience, let \( h_{b,b,i} \equiv h_{b,i,i} \) and \( \gamma_{b,i,i} \equiv \gamma_{b,i} \). Furthermore, let us denote the matrix of all the decoding indicators by \( \lambda = [\lambda_{b,i,k}], \forall b \in B, i, k \in U_b \), in which \( \lambda_{b} = [\lambda_{b,i,k}], \forall i, k \in U_b \), represents the decoding indicator matrix of users in \( U_b \). Moreover, \( \mathbf{p} = [p_{b,i}], \forall b \in B, i \in U_b \), is the power allocation matrix of all the users, in which \( p_{b,i} \) is the \( i \)-th row of this matrix indicating the power allocation vector of users in cell \( b \). According to the Shannon’s capacity formula, the achievable spectral efficiency of user \( i \in U_b \) after successful SIC is obtained by [1], [2]

\[
R_{b,i}(\mathbf{p}, \lambda_b) = \min_{k \in \{ i \} \cup \Phi_{b,i}} \left\{ \log_2 \left( 1 + \frac{p_{b,i} h_{b,b,k}}{\sum_{j \in \Phi_{b,i}} p_{b,j} h_{b,b,k} + (h_{b,k}(\mathbf{p}_{-b}) + \sigma^2_{b,k})} \right) \right\} \tag{2}
\]

Note that the set \( \Phi_{b,i} \) depends on \( \lambda_b \), although it is not explicitly shown in (2). The centralized total spectral efficiency maximization problem is formulated by

\[
\text{JSPA :} \quad \max_{\mathbf{p}, \lambda_b \geq 0, \lambda_b \in \{ 0, 1 \}} \sum_{b \in B} \sum_{i \in U_b} R_{b,i}(\mathbf{p}, \lambda_b) \tag{3a}
\]

s.t. \( \sum_{b \in B} \sum_{i \in U_b} p_{b,i} \leq P_{b}^{\text{max}}, \quad \forall b \in B, \tag{3b} \)

\[
R_{b,i}(\mathbf{p}, \lambda_b) \geq R_{b,i}^{\min}, \quad \forall b \in B, \quad i \in U_b, \tag{3c}
\]

\[
\lambda_{b,i,k} = 1, \quad \forall b \in B, i, k \in U_b, \tag{3d}
\]

\[
\lambda_{b,i,h,\frac{b}{i}} = 1, \quad \forall b, i, h \in U_b, \tag{3e}
\]

where (3b) and (3c) are the per-BS maximum power and per-user minimum rate constraints, respectively. \( P_{b}^{\text{max}} \) denotes the maximum power of BS \( b \), and \( P_{b}^{\text{min}} \) is the minimum spectral efficiency demand of user \( i \in U_b \). The rest of the constraints are described above.

### III. SOLUTION ALGORITHMS FOR THE SUM-RATE MAXIMIZATION PROBLEM

In this section, we propose globally optimal and suboptimal solutions for the main problem (3) (under the centralized/decentralized resource management frameworks). Finally, we compare the computational complexity of the proposed resource allocation algorithms.

#### A. Centralized Resource Management Framework

In this subsection, we first propose a globally optimal JSPA algorithm for problem (3). Then, by considering a fixed SIC decoding order in (3), we propose two suboptimal rate adoption and power allocation algorithms.

1) **Globally Optimal JSPA Algorithm:** Problem (3) can be classified as a mixed-integer non-linear programming (MINLP) problem.

\[ \text{Corollary 1: Problem (3) is strongly NP-hard for} \ |U_b| = 1, \forall b \in B. \]

\[ \text{Proof:} \quad \text{Consider a single user within each cell, i.e.,} \ |U_b| = 1, \forall b \in B. \text{In this way, (2) can be rewritten as} \]

\[ R_{b,1}(\mathbf{p}) = \log_2 \left( 1 + \frac{p_{b,1} h_{b,b,1}}{\sum_{j \in \Phi_{b,1}} p_{b,j} h_{b,b,j} + (h_{b,1}(\mathbf{p}_{-b}) + \sigma^2_{b,1})} \right), \quad \forall b \in B. \text{Moreover, problem (3) is rewritten as the following power allocation problem} \]

\[ \max_{\mathbf{p} \geq 0} \sum_{b \in B} R_{b,1}(\mathbf{p}) \quad \text{s.t.} \ (3b), (3c). \]

The latter problem is equivalent to problem \( (P_1)' \) in [26]. In Theorem 1 in [26], it is proved that the sum-rate maximization problem \( (P_1)' \) is strongly NP-hard, due to the
nonconcavity of sum-rate function in \( p \). In other words, for \(|U_b| = 1\), \( \forall b \in B \), multi-cell NOMA with \( B \) cells is identical to the point-to-point interference-limited network with \( B \) transmitter-receiver pairs.

Corollary 1 can be generalized to \(|U_b| > 1\) for some \( b \in B \). However, the complete proof is more difficult, and we consider it as future work. Let us define the power consumption coefficient of BS \( b \) as \( \alpha_b \in [0, 1] \) such that \( \sum_{i \in U_b} p_{b,i} = \alpha_b P_b \max \).

The received ICI power at user \( i \in U_b \) can be reformulated by \( I_{b,i} (\alpha_b) = \sum_{j \in \Phi_{b,i}} \alpha_j P_j \max \ h_{b,j,i} \).

Let \( \alpha = [\alpha_b]_{1 \times B} \) be the BSs power consumption coefficient vector. Although the joint optimization of \( \alpha, p, \) and \( \lambda \) is very challenging, for any given \( \alpha \), the optimal \( (\lambda_b^*, p_b^*) \) can be obtained in closed form as follows:

**Corollary 2:** In multi-cell NOMA, the optimal decoding order for the user pair \( i, k \in U_b \) is \( k \rightarrow i \) if and only if \( \tilde{h}_{b,i} (\alpha_b) \leq \tilde{h}_{b,k} (\alpha_b) \), where \( \tilde{h}_{b,i} (\alpha_b) = \frac{h_{b,i} (\alpha_b)}{h_{b,i} (\alpha_b) + \sigma_{b,i}^2} \). Let \( \alpha = [\alpha_b]_{1 \times B} \).

The additional notes on the optimal decoding order can be found in Appendix A of the supplementary material in [24].

**Corollary 3:** For any given \( \alpha_b \) and subsequently \( \tilde{h}_{b,i} (\alpha_b) = \tilde{h}_{b,k} (\alpha_b) \), at the optimal \( \lambda_b^* \), we have

\[
\log_2 \left( 1 + \frac{p_{b,i} \tilde{h}_{b,i}}{\sum_{j \in \Phi_{b,i}} p_{b,j} \tilde{h}_{b,j} + (\tilde{h}_{b,i} + \sigma_{b,i}^2)} \right) \leq \log_2 \left( 1 + \frac{p_{b,k} \tilde{h}_{b,k}}{\sum_{j \in \Phi_{b,k}} p_{b,j} \tilde{h}_{b,j} + (\tilde{h}_{b,k} + \sigma_{b,k}^2)} \right). \tag{4}
\]

According to (2) and (4), at the optimal \( \lambda_b^* \), we have

\[
R_{b,i} (p, \lambda_b^*) = \log_2 \left( 1 + \frac{p_{b,i} \tilde{h}_{b,i}}{\sum_{j \in \Phi_{b,i}} p_{b,j} \tilde{h}_{b,j} + (\tilde{h}_{b,i} + \sigma_{b,i}^2)} \right).
\]

According to Corollary 2, the optimal SIC ordering \( (\lambda_b^*) \) and power control of interfering cells \( (\alpha_b^*) \) cannot be decoupled in general. However, for any given \( \alpha \), and subsequently \( \lambda_b^* \), the optimal power \( p_b^* \) can be obtained in closed form as follows:

**Proposition 1:** Assume that \( \alpha \) is fixed. For convenience, let \(|U_b| = M\), and \( k > i \) if \( \tilde{h}_{b,k} (\alpha_b) > \tilde{h}_{b,i} (\alpha_b) \). According to Corollary 2, the decoding order \( M \rightarrow M - 1 \rightarrow \cdots \rightarrow 1 \) is optimal. The optimal powers in \( p_b^* \) can be obtained in closed form as follows:

\[
p_{b,i}^* = \left[ \sum_{j=1}^{i-1} (1 - \beta_{b,j}) \beta_{b,j} + \frac{1}{\tilde{h}_{b,i}} \right]^+ \leq \left[ \sum_{j=1}^{i-1} (1 - \beta_{b,j}) \beta_{b,j} \right]^+, \quad \forall i = 1, \ldots, M - 1, \tag{5}
\]

and

\[
p_{b,M}^* = \left[ \alpha_b P_b \max - \sum_{i=1}^{M-1} \beta_{b,i} \left( \prod_{j=1}^{i-1} (1 - \beta_{b,j}) \alpha_b P_b \max \right) + \frac{1}{\tilde{h}_{b,i}} \right]^+ \leq \left[ \sum_{j=1}^{i-1} (1 - \beta_{b,j}) \beta_{b,j} \right]^+.
\]

where \( \beta_{b,i} = \frac{\sum_{k=j}^{i-1} \alpha_k P_k \max}{\sum_{k=j+1}^{M} \tilde{h}_{b,k}} \) if \( \beta_{b,i} \neq 0 \). According to the above:

- Remark 1 shows that in the high SINR regions, the optimal powers are approximately insensitive to the channel gains. Hence, (7) and (8) are valid for the fast fading and/or imperfect CSI scenarios, where the CSI variations are small such that the optimal decoding order remains constant.
- Remark 1 shows that the weaker user may get less power than the stronger user, depending on its minimum rate demand. For \( M = 2 \), Remark 2 shows that the weaker user gets less power when its minimum rate demand is less than 1 bps/Hz.

The performance of these approximations is numerically evaluated in Subsection IV-F of the supplementary file [24]. According to Proposition 1, for any given feasible \( \alpha \), the optimal powers can be obtained in closed form. Therefore, the globally optimal \( \alpha \) should satisfy the necessary and sufficient condition \( \sum_{b \in B} R_b^\alpha (\alpha^*) \geq \sum_{b \in B} R_b^\alpha (\alpha) \), for any feasible \( \alpha \). Based on Proposition 1, for any given feasible \( \alpha \), the optimal value in cell \( b \) can be calculated in closed form as \( R_b^\alpha (\alpha) = \sum_{i=1}^{M-1} R_b^{\min} (\alpha^*) \log_2 \left( 1 + p_{b,M}^* (\alpha) \tilde{h}_{b,M} \right) \), where \( p_{b,M}^* (\alpha) \) is obtained by (6). Hence, the globally optimal...
Algorithm 1 Optimal JSPA for Sum-Rate Maximization Problem

1: Initialize the step size $\epsilon_0 \ll 1$, and $R_{\text{tot}} = 0$.
2: for each sample $\tilde{\alpha}$ do
3: Update $\tilde{h}_{b,i} = \frac{h_{b,i}}{\tilde{h}_{b,i}}$, $\forall b \in B$, $i \in U_b$, where $\tilde{h}_{b,i} = \sum_{j \in B} \tilde{\alpha}_j p_{\max} h_{j,b,i}$.
4: Update $\lambda$ according to $\lambda_{b,i,k} = 1$ if $\tilde{h}_{b,k} > \tilde{h}_{b,i}$, or equivalently update users index according to $k > i$ if $\tilde{h}_{b,k} > \tilde{h}_{b,i}$.
5: Find $p$ according to (5) and (6).
6: if $\left( \sum_{b \in B} \sum_{i \in U_b} R_{b,i}(p, \lambda) > R_{\text{tot}} \right)$ then
   Update $R_{\text{tot}}^* = \left( \sum_{b \in B} \sum_{i \in U_b} R_{b,i}(p, \lambda) \right)$, $p^* = p$, and $\lambda^* = \lambda$.
end if
7: end for
8: The outputs $\lambda^*$ and $p^*$ are the optimal solutions.

The pseudo code of the proposed JSPA method is presented in Alg. 1. This algorithm needs to explore all the possible values of $\alpha$. Since the number of samples $S_{\alpha}$ for each $\alpha \in \alpha_{1 \times B}$ is finite, the greedy search on $\alpha$ falls into an $\epsilon$-suboptimal solution such that when $S_{\alpha} \rightarrow \infty$, then $\epsilon \rightarrow 0$. Our proposed globally optimal, or more precisely, $\epsilon$-suboptimal JSPA algorithm utilizes both the exploration of different values of $\alpha_{b} \in \alpha_{1 \times B}$ and the distributed power allocation optimization (calculating sum-rate $\sum_{b \in B} R_{b}^m(\alpha)$).

The complexity of exhaustive search by a factor of $\alpha$ is independent from the number of users, since for any given $\alpha$, the optimal value $\sum_{b \in B} R_{b}^m(\alpha)$ is formulated in closed form; 2) The complexity of finding optimal SIC ordering is negligible since for a fixed $\alpha$, the optimal decoding order is obtained based on Corollary 2.

Since Alg. 1 is still exponential in the number of BSs, it is important to check the feasibility of problem (3) with a low-complexity algorithm before performing Alg. 1. The feasibility problem of (3) can be formulated by

$$\min_{p \geq 0, \lambda \in \{0,1\}} f(p, \lambda) \text{ s.t. (3b)-(3f)},$$

where $f(p, \lambda)$ can be any objective function such that the intersection of the feasible domain of (3b)-(3f) is a subset of the feasible domain of $f(p, \lambda)$. Finding a feasible solution for (10) is challenging, due to the binary variables in $\lambda$.

For the case that $f(p) = \sum_{b \in B} \sum_{i \in U_b} p_{b,i}$, the well-known iterative distributed power minimization framework can globally solve (10) with a fast convergence speed [14]. The structure of the iterative distributed framework as well as the closed-form of optimal powers in each iteration for solving (10) are presented in Subsection III-A1 of the supplementary file [24].

In Subsection IV-D in [24], we discussed about the impact of feasible/infeasible initial points on the convergence of this algorithm.

It is proved that the optimal $\alpha^*$ in the total power minimization problem is indeed a component-wise minimum [14]. It means that for any feasible $\alpha = [\tilde{\alpha}_b], \forall b \in B$ in (10), it can be guaranteed that $\alpha_{b}^* \leq \tilde{\alpha}_b$. As a result, the exploration area of each $\alpha_{b}$ in Alg. 1 can be reduced from $[0,1]$ to $[\alpha_{b}^{\min}, 1]$, where $\alpha_{b}^{\min}$ denotes the optimal power consumption coefficient of BS $b$ in the total power minimization problem. The lower-bounds $\alpha_{b}^{\min}, \forall b \in B$ can significantly reduce the complexity of Alg. 1 for the case that $\alpha_{b}^{\min}$ grows, e.g., when the minimum rate demand of users increases.

2) Suboptimal SIC Ordering: Joint Rate and Power Allocation: According to Corollary 2, since $\lambda^*$ in (3) depends on $\alpha^*_{b}, \forall b$, any fixed decoding order before power allocation is potentially suboptimal. For a fixed decoding order, Corollary 3 may not hold in the complete feasible region of $\alpha$. In other words, a user with higher decoding order may have lower CINR for some power consumption level of interfering cells. The main problem (3) for the given $\lambda$, or equivalently fixed $\Phi_{b,i}$, can be rewritten as

$$\max_{p \geq 0} \sum_{b \in B} \sum_{i \in U_b} R_{b,i}(p) \text{ s.t. (3b), (3c)}. \quad (11)$$

Constraint (3c) can be equivalently transformed to a linear form as follows:

$$2p_{b,i}^{\min} \left( \sum_{j \in \Phi_{b,i}} p_{b,j} h_{b,k} + I_{b,k} + \sigma_{b,k}^2 \right) \leq \left( p_{b,i} + \sum_{j \in \Phi_{b,i}} p_{b,j} \right) h_{b,k} + I_{b,k} + \sigma_{b,k}^2, \forall b \in B, i, k \in U_b, k \in \{i\} \cup \Phi_{b,i}. \quad (12)$$

Hence, the feasible region of (11) is affine, so is convex in $p$. However, (11) is still strongly NP-hard, due to the nonconcavity of the sum-rate function with respect to $p$ (see Corollary 1). The optimal powers in (5) and (6) are derived based on the optimal decoding order. Therefore, Alg. 1 is not applicable for solving (11). Besides, the complexity of exhaustive search for solving (11) is $S_{P}^{BM} \times (M!)^{B}$, where $S_{P}$ is the total number of samples for each $p_{b,i} \in p$. Hence, Alg. 1 reduces the complexity of exhaustive search by a factor of $S_{P}^{BM} \times (M!)^{B}$ when $S_{\alpha} = S_{P}$. In fact, Alg. 1 has two main advantages: 1) The complexity is independent from the number of users, for any given $\alpha$, the optimal value $\sum_{b \in B} R_{b}^m(\alpha)$ is formulated in closed form; 2) The complexity of finding optimal SIC ordering is negligible since for a fixed $\alpha$, the optimal decoding order is obtained based on Corollary 2.
Algorithm 2 Suboptimal JRPA Algorithm Based on the Sequential Programming

1: Initialize \( r(0) \), iteration index \( t = 0 \), and tolerance \( \epsilon_s \ll 1 \).
2: while \( \| r(t) - r(t-1) \| > \epsilon_s \) do
3: \( \text{Set } t := t + 1 \).
4: \( \text{Update the approximation parameter } \hat{g}(r_{b,i}(t)) \) based on \( r^{(t-1)} \).
5: \( \text{Find } (r^{*,t}, \hat{p}^{(t)}) \) by solving the convex approximated problem (22).
6: end while
7: \( \text{Set } \hat{p}_{b,i}^* = e^{\hat{p}_{b,i}} \), \( \forall b \in B, \ i \in U_b \). The output \( p^{(t)} \) is adopted for the network.

following problem as

\[
\text{JRPA: } \max_{p \geq 0, \ r \geq 0} \sum_{b \in B} \sum_{i \in U_b} r_{b,i} \text{ s.t. (3b)},
\]

\[
r_{b,i} \geq R_{b,i}^{\text{min}}, \ \forall b \in B, \ i \in U_b,
\]

\[
r_{b,i} \leq \log_2 \left( 1 + \frac{p_{b,i} h_{b,k}}{\sum_{j \in \Phi_{b,i}} p_{b,j} h_{b,k} + I_{b,k} + \sigma_b^2} \right),
\]

\forall b \in B, \ i, k \in U_b, \ k \in \{i\} \cup \Phi_{b,i},
\]

where \( r = [r_{b,i}], \forall b \in B, \ i \in U_b \). Although the objective function (13a) and constraints (3b) and (13b) are affine, problem (13) is still nonconvex due to the nonconvexity of (13c). The derivations of the proposed sequential programming method for solving (13) is presented in Appendix B. The pseudo code of our proposed JRPA algorithm is shown in Alg. 2.

The decoding order for some user pairs is independent from the ICI, so it can be determined prior to power allocation optimization. For a specific channel gain condition, we prove that the CNR-based decoding order is optimal for a user pair independent from power allocation.

Theorem 1 (SIC Sufficient Condition): For each user pair \( i, k \in U_b \) with \( h_{b,i} \) and \( h_{b,k} \) satisfy the SIC sufficient condition by \( \Phi_{b,i}^{\text{CNR}} = \{ k \in U_b \setminus \{i\} \mid \alpha_{b,i,k} = 1 \} \), \( \frac{h_{b,k}}{\sigma_{b,k}} \geq \frac{h_{b,i}}{\sigma_{b,i}} \), \( \forall b \in B, \ i \in U_b \). The CNR-based decoding order is optimal in cell \( b \), meaning that the number of constraints in (13c) will be reduced to one for each user pair within cell \( b \).

3) Theorem 1 can be considered as a necessary condition in user grouping of the multicarrier NOMA systems. However, it is still unknown how much the system performance decreases if we limit user grouping based on Theorem 1 which can be considered as future work.

Similar to problem (10), it can be shown that the feasible region of (11) is the same as the feasible region of the following total power minimization problem as

\[
\min_{p \geq 0, \ r \geq 0} \sum_{b \in B} \sum_{i \in U_b} p_{b,i} \text{ s.t. (3b), (12)}. \tag{15}
\]

Since the objective function and all the constraints are affine, problem (15) is a linear program, so it is convex. Hence, (15) can be solved by using the Dantzig’s simplex method or interior-point methods (IPMs) [27]. The solution of (15) can be utilized as an initial feasible point for Alg. 2.

3) Suboptimal SIC Ordering: Power Allocation for a Fixed Rate Region: In Subsection III-A2, we showed that for any fixed, thus potentially suboptimal, decoding order, Corollary 3 may not hold at any feasible power allocation, so the joint power allocation and rate adoption is necessary to achieve the maximum sum-rate which is still lower than the users sum-capacity. There are a number of research studies in multi-cell NOMA with suboptimal CNR-based decoding order while enforcing users to achieve their channel capacity, i.e., \( R_{b,i} = \log_2 \left( 1 + \frac{p_{b,i} h_{b,k}}{\sum_{j \in \Phi_{b,i}} p_{b,j} h_{b,k} + I_{b,k} + \sigma_b^2} \right) \), \( \forall b \in B, \ i \in U_b \) by imposing the SIC necessary condition on power allocation. To guarantee that user \( i \in U_b \) achieves its Shannon’s capacity for decoding its desired signal after SIC, the condition in (4) should be satisfied for each user \( k \in \Phi_{b,i} \). For any fixed \( \lambda \), so fixed \( \Phi_{b,i} \), (4) can be rewritten as

\[
\frac{h_{b,i}}{I_{b,k}(\alpha_{b,k}) + \sigma_{b,k}^2} \leq \frac{h_{b,k}}{I_{b,k}(\alpha_{b,k}) + \sigma_{b,k}^2} \text{ for } \forall b \in B, \ i, k \in U_b.
\]

Actually, constraint (16), known as SIC necessary condition [16], [19], [28], states that in the FRPA scheme, the feasible region of \( \alpha_{b,k} \) is limited such that \( h_{b,i}(\alpha_{b,k}) \leq h_{b,k}(\alpha_{b,k}) \) always holds, in which \( h_{b,i}(\alpha_{b,k}) = \frac{h_{b,i}}{I_{b,k}(\alpha_{b,k}) + \sigma_{b,k}^2}, \ l = i, k \). According to Corollary 2, the feasible region of \( \alpha_{b,k} \) is limited
such that the decoding order \( \lambda \) remains optimal. The FRPA problem is formulated by

\[
\text{FRPA} : \max_{p \succeq 0} \sum_{b \in B} \sum_{i \in U_b} R_{b,i}(p) \quad (17a)
\]

s.t. (3b), (16),

\[
\hat{R}_{b,i}(p) \geq R_{b,i}^{\min}, \quad \forall b \in B, \ i \in U_b. \quad (17b)
\]

**Remark 3:** For any fixed \( \lambda \), problem (17) is identical to the main problem (3) with SIC necessary condition (16).

**Proof:** As is mentioned above, (16) is to guarantee that the given decoding order \( \lambda \) remains optimal by limiting the feasible region of \( \alpha_{-,b} \). Therefore, when (16) is imposed to the main problem (3) with the given \( \lambda \), Corollary 2 holds. It means that, \( R_{b,i}(p) = \hat{R}_{b,i}(p) \), \( \forall b \in B, \ i \in U_b \).

Subsequently, the objective function (3a) can be transformed to \( \sum_{b \in B} \sum_{i \in U_b} \hat{R}_{b,i}(p) \) formulated in (17a). For any fixed \( \lambda \), the constraints (3d)-(3f) will be removed from (3). Therefore, the resulting problem of (3) has the same objective function and also the same constraints as in (17), so these problems are identical. \( \square \)

**Remark 4:** Assume that \( \lambda \) is fixed. For any given feasible \( \alpha \) satisfying (16), the optimal powers of (17) can be obtained by using Proposition 1.

**Proof:** According to Remark 3, we proved that (17) is identical to (3) with SIC necessary condition (16). Moreover, it is stated that \( \lambda \) is optimal in the feasible region of \( \alpha_{-,b} \) satisfying (16). It means that with (16), we guarantee that \( \hat{h}_{b,i}(\alpha_{-,b}) \leq \hat{h}_{b,k}(\alpha_{-,b}) \) independent from \( p \).

According to Appendix A, for any given \( \alpha \) satisfying (16), problem (17) can be equivalently divided into \( B \) single-cell NOMA subproblems. In each cell \( b \), we find \( \alpha_b \) by solving (19) and, the proof is completed. \( \square \)

The problem (17) is strongly NP-hard, since for \( |U_b| = 1 \), \( \forall b \in B \), the multi-cell NOMA network is identical to the point-to-point interference-limited network with \( B \) transmitter-receiver pairs (see Corollary 1). Besides, the optimal \( \alpha^* \) in (17) should satisfy the following necessary and sufficient conditions: 1) It results maximum sum-rate:

\[
\sum_{b \in B} R_{b}^{\text{tot}}(\alpha^*) \geq \sum_{b \in B} R_{b}^{\text{tot}}(\alpha), \quad \text{for any feasible } \alpha, \quad \text{where}
\]

\[
R_{b}^{\text{tot}}(\alpha) = \sum_{i=1}^{M-1} R_{b,i}^{\min} + \log_2 \left( 1 + p_{b,M}(\alpha) \hat{h}_{b,M} \right);
\]

2) SIC necessary conditions:

\[
\frac{h_{b,i}(\alpha_{-,b}) + \sigma_{i,i}^b}{h_{b,k}(\alpha_{-,b}) + \sigma_{k,k}^b} \leq \frac{h_{b,k}(\alpha_{-,b}) + \sigma_{i,i}^b}{h_{b,k}(\alpha_{-,b}) + \sigma_{k,k}^b}, \quad \forall b \in B, \ i, k \in U_b, \ k \in \Phi_{b,i}.
\]

Therefore, similar to Alg. 1, we perform a greedy search on the possible values of \( \alpha \) to find optimal, or more precisely, \( \epsilon \)-suboptimal \( \alpha^* \) such that the latter two conditions are satisfied. The pseudo code of the proposed algorithm for solving the FRPA problem is presented in Alg. 3. Although our proposed FRPA algorithm scales well with any number of multiplexed users, it is still exponential in the number of BSs (see Subsection III-A1). One solution is to apply the well-known suboptimal power allocation based on the sequential programming method proposed in [16, 19, 21].

**Fact 1:** In multi-cell NOMA, the SIC necessary condition (16) for each user pair \( i, k \in U_b, \ k \in \Phi_{b,i} \) implies additional maximum power constraints on the BSs in \( B \setminus \{b\} \).

### Algorithm 3 Optimal Power Allocation for Fixed SIC Ordering and Rate Region

1. Initialize the step size \( \epsilon_\alpha = \frac{1}{b_{\max}} \ll 1 \), where \( S_\alpha \) is the number of samples for each \( \alpha_b \), and \( R_{\text{tot}} = 0 \).

2. for each sample \( \hat{\alpha} \) do

3. Update \( \hat{h}_{b,i} = \frac{h_{b,i}}{h_{b,k} + \sigma_{i,i}^b}, \quad \forall b \in B, \ i \in U_b, \) where \( \hat{I}_{b,i} = \sum_{j \in B} \alpha_j p_{j,max} h_{j,b,i} \).

4. if \( \frac{h_{b,i}}{h_{b,k} + \sigma_{i,i}^b} \leq \frac{h_{b,k}}{h_{b,k} + \sigma_{k,k}^b}, \quad \forall b \in B, \ i, k \in U_b, \ k \in \Phi_{b,i} \) then

5. Find \( p \) according to (5) and (6).

6. if \( \sum_{b \in B} \sum_{i \in U_b} R_{b,i}(p) > R_{\text{tot}} \) then

Update \( R_{\text{tot}}^* = \left( \sum_{b \in B} \sum_{i \in U_b} R_{b,i}(p) \right), \) and \( p^* = p \).

end if

7: end if

8: end for

9: The output \( p^* \) is the optimal solution of (17).

**Proof:** Please see Appendix D.

**Corollary 4:** Restricting the rate region of users under the suboptimal decoding order results in additional limitations on the BSs power consumption. For the user pairs satisfying the SIC sufficient condition, the decoding order is optimal, and subsequently the SIC constraint (16) will be completely independent from the power allocation (see Corollary 3). As a result, the negative side impact of the SIC necessary condition on power allocation will be eliminated.

The feasible solution for (17) can be obtained by solving the following total power minimization problem as

\[
\min_{p \succeq 0} \sum_{b \in B} \sum_{i \in U_b} p_{b,i} \quad \text{s.t.} \quad (3b), (16), (17b). \quad (18)
\]

Problem (18) is a linear program which can be solved by using the Dantzig’s simplex method.

### B. Decentralized Resource Management Frameworks

1) Fully Distributed JSPA Framework: Although the globally optimal solution in Subsection III-A1 achieves the channel capacity of users, the complexity of Alg. 1 is still exponential in the number of BSs. Moreover, the centralized framework would cause a large signaling overhead. Here, we propose a fully distributed resource allocation framework in which each BS independently allocates power to its associated users. Actually, we divide the main problem (3) into \( B \) single-cell NOMA problems.

**Fact 2:** At the optimal point of sum-rate maximization problem of single-cell NOMA, the BS operates in its maximum available power. It means that the power constraint (3b) is active for each BS \( b \), i.e., \( \sum_{i \in U_b} p_{b,i} = P_{b,max} \), \( \forall b \in B \) in the fully distributed framework.

**Proof:** Please see Appendix G in [24].

According to Fact 2, we set \( \alpha^*_b = 1 \), \( \forall b \in B \). Based on the given \( \alpha^* \), the optimal decoding order of users under the fully
distributed framework can be easily obtained by Corollary 2. According to Subsection III-A1, the optimal power \( p^* \) under the fully distributed framework can be obtained by using Proposition 1.

2) Semi-Centralized JSPA Framework: The fully distributed framework works well for the case that the ICI levels are significantly low, so \( \alpha^*_b \to 1, \forall b \in B \) in problem (3). For the case that the ICI levels are high, e.g., at femto-cells underlying a single MBS [20], [22], [29], the fully distributed framework may seriously degrade the spectral efficiency of femto-cell users. In the following, we propose a semi-centralized JSPA framework in which we assume that the low-power FBSs operate in their maximum power, while the MBS power consumption is obtained by the joint power allocation and SIC ordering of all the users. Let \( b = 1 \) be the MBS’s index, and \( b = 2, \ldots, B \) be the index of FBSs. In this framework, we assume that \( \alpha_b = 1, \forall b = 2, \ldots, B \). Then, we utilize Alg. 1 to find the globally optimal \( p^* \) and \( \lambda^* \) for problem (3). This algorithm performs a greedy search on \( 0 \leq \alpha_1 \leq 1 \). Hence, the complexity of finding optimal JSPA is on the order of total number of samples for \( \alpha_1 \), i.e., \( S_\alpha \). Actually, the computational complexity of this algorithm is independent from the number of FBSs and users which is a good solution for the large-scale systems. The performance of the semi-centralized framework depends on the FBSs optimal power consumption in (3). For the case that \( \alpha^*_b \to 1, \forall b = 2, \ldots, B \) in (3), the performance gap between the semi-centralized and centralized frameworks tends to zero.

C. Computational Complexity Comparison Between Resource Allocation Algorithms

In this section, we compare the computational complexity of the proposed different resource allocation algorithms for solving the sum-rate maximization problem (3). To simplify the analysis, we assume that each cell has \( M \) users. In this comparison, we apply the barrier method with inner Newton’s method to achieve an \( \epsilon \)-suboptimal solution for a convex problem. The number of barrier iterations required to achieve \( m = \epsilon \)-suboptimal solution is exactly \( T = \lceil \log_{\log \mu} (\epsilon t(0)) \rceil \), where \( m \) is the total number of inequality constraints, \( t(0) \) is the initial accuracy parameter for approximating the functions in inequality constraints in standard form, and \( \mu \) is the step size for updating the accuracy parameter \( t \) [27]. The number of inner Newton’s iterations at each barrier iteration \( i \) is denoted by \( N_i \). In general, \( N_i \) depends on \( \mu \) and how good is the initial points at the barrier iteration \( i \). The computational complexity of solving a convex problem is thus on the order of total number of Newton’s iterations obtained by

\[
C_{\text{cnvx}} = \sum_{i=1}^{Y} N_i. 
\]

For the case that the sequential programming converges in \( Q \) iterations, the complexity of this method is on the order of \( C_{\text{SP}} = Q \sum_{q=1}^{Q} \sum_{i=1}^{Y} N_{q,i} \), where \( N_{q,i} \) denotes the number of inner Newton’s iterations at the \( i \)-th barrier iteration of the \( q \)-th sequential iteration. For convenience, assume that \( N_{q,i} = N, \forall q = 1, \ldots, Q, i = 1, \ldots, Y \). Hence, we have \( C_{\text{cnvx}} = N Y = N \lceil \log_{\log \mu} (\epsilon t(0)) \rceil \), and

\[
C_{\text{SP}} = Q N Y = Q N \lceil \log_{\log \mu} (\epsilon t(0)) \rceil. 
\]

The complexity order of different solution algorithms for (3) is presented in Table I.

In this table, \( S \) denotes the number of samples for each \( p_{b,i} \) or \( \alpha_b \). Moreover, the optimality status is for only the simplified problem. For example, FRPA finds the globally optimal powers in \( p \) and subsequently \( \alpha \), when \( \lambda \) and \( r \) are fixed. It does not mean that the output is globally optimal solution for the main problem (3). Actually, only the first and second rows in Table I can find the globally optimal solution of (3), and the rest of the algorithms are indeed suboptimal. It is noteworthy that in Table I, we considered the highest possible computational complexity (worst case) of each algorithm.

IV. Numerical Results

In this section, we evaluate the performance of our proposed resource allocation algorithms via MATLAB Monte Carlo simulations over 10000 channel realizations. This comparison is divided into two subsections: 1) performance comparison among our proposed JSPA, JRPA, and FRPA algorithms to demonstrate the benefits of optimal SIC ordering, and rate adoption for any suboptimal decoding order (see Table I); 2) performance comparison among the centralized and decentralized resource allocation frameworks to demonstrate the effect of optimal \( \alpha^* \) (ICI management) in the main problem (3). In our simulations, we adopt the commonly-used (suboptimal) CNR-based decoding order [16], [19] for the JRPA and FRPA algorithms. Due to the space limitations, the convergence of the iterative algorithms for solving (10) and (13), and the performance of approximated optimal powers in Remark 1 are presented in Subsections IV-D, IV-E, and IV-F of the supplementary file [24], respectively. The source code is available on GitLab [23].

A. Simulation Settings

Here, we consider a two-tier HetNet consisting of one FBS underlying a MBS.4 Within each cell, there is one BS at the center of a circular area and \( U_n \) users inside it [14]. The system parameters and their corresponding notations are shown in Table II. The network topology and exemplary users placement are shown in Fig. 1.

4Although in practice there are larger number of FBSs, in this experiment, we aim to fundamentally investigate the impact of ICI from MBS to FBS and vice versa in optimal decoding order of users.
B. Performance of Centralized Resource Allocation Algorithms

In this subsection, we compare the performance in terms of outage probability, and users total spectral efficiency of our proposed JSPA, JRPA, and FRPA algorithms. Note that FRPA finds the globally optimal solution of the sum-rate maximization problem in [16] and the single-carrier-based downlink power allocation problem in [19] with significantly reduced computational complexity (see Table I).

1) Outage Probability Performance: The outage probability of the JSPA, JRPA, and FRPA schemes is obtained based on optimally solving their corresponding total power minimization problems. For each scheme, the outage probability is calculated by dividing the number of infeasible problem instances by total number of samples [14]. Fig. 2 shows the impact of order of NOMA clusters and minimum rate demands on the outage probability of the JSPA, JRPA, and FRPA problems. According to Theorem 1, the ICI may affect the optimal decoding order of user pairs which cannot satisfy the SIC sufficient condition. We call these user pairs as the pairs depending on ICI when the CNR-based decoding order is applied. In Fig. 2(a), we calculate the average number of user pairs depending on ICI (denoted by $\Psi$) for different distances among BSs, and different order of NOMA clusters in the CNR-based decoding order. The parameter $\Psi$ is increased by 1) increasing $|U_b|$, which inherently decreases the LHS of (14); 2) increasing the inter-cell channel gain $h_{j,b,i}$, due to increasing the RHS of (14). The second case for the femto-cell user pairs is inversely proportional to the BSs distance, due to the existing path loss. As shown, the impact of $|U_b|$ is higher than the impact of BSs distance. The wide coverage of macro-cell results in large differences between the users channel gains, however the CNR of macro-cell users is typically low reducing the LHS of (14). As a result, $\Psi$ is also affected by $|U_m|$. It is noteworthy that the position of FBS (BSs distance) in the coverage area of MBS does not have significant impact on $\Psi$ for the macro-cell users. For the case that at least one user pair within a cell does not satisfy the SIC sufficient condition, the CNR-based decoding order in that cell may not be optimal, resulting in reduced spectral efficiency and increased outage.

In Figs. 2(b)-2(d), we observe that there exist significant performance gaps between the JSPA and JRPA schemes, which shows the superiority of finding the optimal decoding order in multi-cell NOMA. Besides, JRPA significantly reduces the outage probability of the FRPA scheme which shows the importance of rate adoption when a suboptimal decoding order is applied. The large performance gap between JSPA and FRPA shows that the SIC necessary condition in (16) severely restricts the feasible region of the FRPA problem (see Fact 1) resulting in high outage, specifically for the larger order of NOMA clusters (see Fig. 2(b)). Last but not least, in Fig. 2(b), we observe that the larger order of NOMA clusters results in higher outage probability even for the optimal JSPA algorithm.

2) Average Total Spectral Efficiency Performance: Fig. 3 investigates the impact of order of NOMA clusters and minimum rate demands on the average total spectral efficiency of users. Here, we set the sum-rate to zero when the problem is infeasible. As shown, JSPA always outperforms the JRPA and FRPA algorithms which shows the superiority of finding the optimal decoding order in multi-cell NOMA. Besides, JRPA significantly reduces the outage probability of the FRPA scheme which shows the importance of rate adoption when a suboptimal decoding order is applied. The large performance gap between JSPA and FRPA shows that the SIC necessary condition in (16) seriously restricts the feasible region of the FRPA problem (see Fact 1) resulting in high outage, specifically for the larger order of NOMA clusters (see Fig. 2(b)). Last but not least, in Fig. 2(b), we observe that the larger order of NOMA clusters results in higher outage probability even for the optimal JSPA algorithm.

### Table I

| Algorithm | Complexity | Framework | Optimal | $\alpha$ | $\rho$ |
|-----------|------------|-----------|---------|----------|--------|
| JSPA (Alg. 1) | $Q N \left[ \log \left( \frac{\log (B + B M + B \left( M - 1 \right))}{\log (M \mu)} \right) \right]$ | Centralized | ✓ | ✓ | ✓ |
| Exhaustive Search | | Centralized | ✓ | ✓ | ✓ |
| JRPA (Alg. 2) | $Q N \left[ \log \left( \frac{\log \left( B + B M + B \left( M - 1 \right) \mu \right)}{\log \left( M \mu \right)} \right) \right]$ | Centralized | ✓ | ✓ | ✓ |
| FRPA Monotonic Optimization [19, 21] | $Q N \left[ \log \left( \frac{\log (B + B M + B \left( M - 1 \right))}{\log (M \mu)} \right) \right]$ | Centralized | ✓ | ✓ | ✓ |
| Sequential Program [19, 21] | $Q N \left[ \log \left( \frac{\log (B + B M + B \left( M - 1 \right))}{\log (M \mu)} \right) \right]$ | Centralized | ✓ | ✓ | ✓ |

### Table II

**System Parameters**

| Parameter | Notation | Value | Parameter | Notation | Value |
|-----------|----------|-------|-----------|----------|-------|
| Coverage of MBS | × | 500 m | Lognormal shadowing standard deviation | × | 8 dB |
| Coverage of FBS | × | 40 m | Small-scale fading | × | Rayleigh fading |
| Distance between MBS and FBS | × | 200 m | AWGN power density | $N_0$ dBm | $-174$ dBm |
| Number of macro-cell users | $|U_m|$ | 3 | | | |
| Number of femto-cell users | $|U_f|$ | 3 | | | |
| Minimum distance of users to MBS | × | 20 m | MBS transmit power | $P_{\text{max}}$ dBm | 46 dBm |
| Minimum distance of users to BS | × | 20 m | FBS transmit power | $P_{\text{max}}$ dBm | 30 dBm |
| User distribution model | × | Uniform | Minimum rate of macro-cell users | $R_{\Omega}$ bps/Hz | 0.5 bps/Hz |
| | | | Minimum rate of femto-cell users | $R_{\Omega}$ bps/Hz | 0.5 bps/Hz |
| | | | Minimum rate of cell users | $R_{\Omega}$ bps/Hz | 0.5 bps/Hz |
| | | | Minimum rate of users | $R_{\Omega}$ bps/Hz | 0.5 bps/Hz |
| | | | Step size of each $\alpha$ | $\epsilon_{\Omega}$ | $10^{-3}$ |
| | | | Tolerance of Alg. 2 | $\epsilon_{\Omega}$ | $10^{-3}$ |
| | | | Tolerance of Alg. 2 | $\epsilon_{\Omega}$ | $10^{-3}$ |

### Table III

**Subsection III-B2**

| Subsection III-B2 | $S$ | Semi-Centralized | ✓ | ✓ | ✓ | α |
| Subsection III-B1 | 1 | Fully Distributed | ✓ | ✓ | ✓ | |
optimal value of (13) (see Table I). In Subsection IV-E of [24], we show that JRPA is a near to optimal algorithm, so this lower-bound is significantly tighten. Subsequently, the performance gap between JRPA and FRPA is indeed the lower-bound of the exact performance gain of rate adoption for the CNR-based decoding order. As can be seen, FRPA has a lower performance compared to JRPA. In Fig. 3(a), we observe that for $R_{m}^{\min} = R_{f}^{\min} = 1$ bps/Hz, the negative side impact of increasing the order of NOMA cluster is higher than the multi-user diversity gain. As a result, increasing the order of NOMA clusters results in lower total spectral efficiency. Another reason is increasing the outage probability (shown in Fig. 2(b)) which can significantly affect the average total spectral efficiency. For $|U_{m}| = |U_{f}| = 2$, we observed that the performance gap is low between JSPA and JRPA, due to decreasing $\Psi$ in Fig. 2(a). This performance gap grows when $|U_{b}|$ increases. As is expected, Figs. 3(b) and 3(c) show that the total spectral efficiency of users is decreasing in minimum rate demands.
More importantly, there exist huge performance gaps between JSPA and FRPA for any number of users and minimum rate demands.

C. Performance of Centralized and Decentralized Frameworks

In this subsection, we compare the performance of the centralized, semi-centralized, and fully distributed JSPA frameworks (shown in Table I).

1) Outage Probability Performance: Fig. 4 evaluates the outage probability of different resource allocation frameworks. As can be seen, the fully distributed framework results in huge outage probability. However, the outage probability gap between the semi-centralized and centralized frameworks is decreasing with larger $|U_m|$ (Fig. 4(a)), and/or higher $R_{min}$ (Figs. 4(b) and 4(c)). The large performance gap between the semi-centralized and fully distributed frameworks shows the importance of ICI management from MBS to femto-cell users. It is noteworthy that the feasible point for the semi-centralized framework is obtained based on the greedy search on $\alpha$ with stepsize $\alpha = 10^{-2}$. Although it can be shown that further reducing $\alpha$ would not cause significant higher total spectral efficiency, $\alpha = 10^{-2}$ is not good enough for calculating.
outage probability, $e_\alpha$, can be easily reduced to $10^{-4}$ in the semi-centralized framework, however we set $e_\alpha = 10^{-2}$ for both the centralized and semi-centralized frameworks to have a fair comparison. For the same stepsize $e_\alpha$, the feasible region of the semi-centralized problem is a subset of the feasible region of its corresponding centralized problem.

2) Average Total Spectral Efficiency Performance: The performance of the decentralized frameworks depends on how good is the approximation of the power consumption of the BSs compared to the centralized framework. Figs. 5(a)-5(c) show the impact of order of NOMA clusters and minimum rate demands on the FBS power consumption coefficient $\alpha_f$ at the optimal point of the centralized framework. As is expected, larger $|\mathcal{U}_f|$ and/or $R_{min}^f$ results in larger FBS power consumption. Moreover, we observe that increasing $|\mathcal{U}_m|$ and/or $R_{min}^m$ decreases $\alpha_f$. However, the impact of $|\mathcal{U}_m|$ and/or $R_{min}^m$ on $\alpha_m$ is quite low and negligible, due to the low ICI level from low-power FBS to macro-cell users in average. More importantly, we observe that in most of the cases, the FBS operates in up to 90% of its available power. It is noteworthy that in both the decentralized frameworks, we assume that the FBS operates in 100% of its available power. Figs. 5(d)-5(f) evaluate the impact of order of NOMA clusters and minimum rate demands on the MBS power consumption coefficient $\alpha_m$ at the optimal point of the centralized and semi-centralized frameworks. As is expected, $\alpha_m$ is directly proportional to $|\mathcal{U}_m|$ and/or $R_{min}^m$, while is inversely proportional to $|\mathcal{U}_f|$ and/or $R_{min}^f$. More importantly, we observe that

1) $\alpha_m$ in the semi-centralized framework is always upper-bounded by $\alpha_m$ in the centralized framework. This is due to larger $\alpha_f$ in the semi-centralized framework compared to the centralized framework.

2) The MBS typically operates in less than 60% of its available power. Hence, the fully distributed framework with $\alpha_m = 1$ results in significantly degraded spectral efficiency at femto-cell users, due to the high ICI from the MBS to femto-cell users.

Last but not least, we observe that the MBS power consumption gap between the centralized and semi-centralized frameworks is quite low (see Figs. 5(d)-5(f)).

Fig. 6 investigates the total spectral efficiency of users in the centralized and decentralized frameworks. According to the discussions for Fig. 5, we observed that in most of the cases, the performance gap between the centralized (optimal) and semi-centralized frameworks is quite low, specifically for the lower order of the femto-cell NOMA cluster. Hence, the semi-centralized framework with its low computational complexity (see Table I) is a good candidate solution for the larger-scale systems. Besides, the fully distributed framework results in quite low performance, due to the discussions for Fig. 5.

V. CONCLUDING REMARKS AND FUTURE RESEARCH DIRECTIONS

In this paper, we addressed the problem of optimal joint SIC ordering and power allocation in multi-cell NOMA systems to achieve the maximum users sum-rate. For the given total power consumption of BSs, we obtained the optimal powers in closed form. Then, we proposed a globally optimal JSPA algorithm with a significantly reduced computational complexity. For any fixed decoding order, we addressed the problem of joint rate and power allocation to maximize users sum-rate. We showed that for specific channel conditions, the CNR-based decoding order is optimal for a user pair independent from ICI. We also devised two decentralized resource allocation frameworks. Extensive numerical results are provided to evaluate the outage probability/sum-rate gains achieved by JSPA compared to FRPA and JRPA with CNR-based decoding order. The numerical results show that FRPA leads to a very high outage when the number of multiplexed users increases. JRPA has a near-optimal performance, however, there is still a certain performance gap due to suboptimal CNR-based decoding order. Moreover, the fully distributed framework has a poor performance, because of high ICI from MBS to FBS-users. The semi-centralized framework has a near-optimal performance with significant reduced complexity, such that it scales very well with larger number of FBSs and users. The fast semi-centralized algorithm is a good solution for practical implementations.

A list of possible extensions of this work is provided in the following.

1) Multi-Cluster NOMA: The analysis of this work can be directly applied to NOMA with multiple clusters, where each user occupies one subchannel, and each subchannel has specific power budget, e.g., the system
models in [11], [12], [16]. For the case that the power budget of clusters is not predefined [13], an additional inter-cluster power allocation is required.

2) Hybrid NOMA: In Hybrid NOMA, also known as multicarrier NOMA, each user can occupy more than one subcarrier. For the case that different symbols are sent over different subcarriers, the analysis of this paper can be applied, however more efforts are required to tackle the nonconvex per-user minimum rate constraints over all the assigned subcarriers. Another challenge is subcarrier allocation. The optimal joint power and subcarrier allocation is proved to be strongly NP-hard [30]–[32].

3) Decentralized Recourse Allocation Frameworks: Our semi-centralized framework is specified for a two-tier HetNet including a single MBS and arbitrary number of FBSs and users. A more complicated scheme is when we have more than one MBS each having arbitrary number of FBSs and users. One future work would be how to generalize the analysis in this paper to a $N$-tier wireless network with efficient power control between the tiers, as well as power control among BSs within each tier.

4) Energy Efficiency Maximization Problem: Another important direction of the closed-form of optimal powers obtained in this paper is energy efficiency maximization problem in both the single-cluster and multi-cluster NOMA.

5) NOMA with Imperfect CSI: We showed that the approximated closed-form of optimal powers which are completely independent from channel gains are nearly optimal. However, the decoding order among users depends on the channel gains. One interesting topic would be how to find a robust decoding order for imperfect CSI while applying the approximated closed-form of optimal powers derived in this paper.

6) Applications of Theorem 1. SIC Sufficient Condition in Multi-Cell NOMA: Since dynamically changing decoding orders based on CINR of users through power control among the cells is still challenging for sum-rate maximization problem, there are three schemes which can be analyzed and numerically compared in multi-cell multicarrier NOMA:

a) Considering CNR-based decoding order with FRPA scheme.

b) Considering CNR-based decoding order and imposing Theorem 1 in user grouping (similar to applying Lemma 1 in user grouping in [28]).

c) Considering CNR-based decoding order and applying JRPA algorithm.

In case (c), we consider the complete rate region of users instead of limiting user grouping by imposing Theorem 1 in case (b).

It is worth noting that the analysis in this work as well as all the existing works on single-antenna NOMA cannot be generalized to the multi-antenna systems which are nondegraded in general [3]. The optimal successive decoding order criteria in nondegraded multi-antenna BCs depends on beamforming.

Moreover, when SC-SIC is applied to the multi-antenna BCs, the users may operate in lower rate than their channel capacity for decoding their own signals [3]. The analysis shows that NOMA looses significant multiplexing gain in multi-antenna systems [33].

APPENDIX A

PROOF OF PROPOSITION 1

According to Corollary 3, the achievable spectral efficiency of each user $i \in \mathcal{U}_b$ for the fixed $\alpha_{-b}$ and optimal decoding order $M \rightarrow M - 1 \rightarrow \ldots \rightarrow 1$ can be formulated by

$$
\tilde{R}_{b,i}(p_b) = \log_2 \left( 1 + \frac{p_{b,i} \hat{h}_{b,i}(\alpha_{-b})}{\sum_{j=i+1}^{M} p_{b,j} \hat{h}_{b,i}(\alpha_{-b}) + 1} \right).
$$

Note that $\hat{h}_{b,k}(\alpha_{-b}) > \hat{h}_{b,i}(\alpha_{-b})$ for each $i, k \in \mathcal{U}_b$, $k > i$. Moreover, $\tilde{R}_{b,i}$ is independent from $p_{b,j}$ for the given $\alpha$, meaning that (3) can be equivalently divided into $B$ single-cell NOMA sub-problems. In cell $b$, we find $p_{b}^*$ by solving the following sub-problem as

$$
\max_{p_{b} \geq 0} \sum_{i=1}^{M} \tilde{R}_{b,i}(p_b) \quad (19a)
$$

s.t. $\sum_{i \in \mathcal{U}_b} p_{b,i} = \alpha_b P_b^\text{max}$, $\tilde{R}_{b,i}(p_b) \geq R_{b,i}^\text{min}, \forall i \in \mathcal{U}_b. \quad (19b)$

Similar to single-cell NOMA, it can be easily shown that the Hessian of (19a) is negative definite in $p_{b}^*$. The Slater’s condition holds in (19) since it is convex and there exists $p_{b}^* \geq 0$ satisfying (19c) with strict inequalities. Therefore, the strong duality holds in (19). As a result, the KKT conditions are satisfied and the optimal solution $p^*$ can be obtained by using the Lagrange dual method [27]. In Appendix B of the supplementary file [24], we prove that at the optimal $p_{b}^*$, the allocated power to all the users with lower decoding order, i.e., users $i = 1, \ldots, M - 1$, is to only maintain their minimum spectral efficiency demand. Moreover, only the NOMA cluster-head user $M$ deserves additional power. Then, we obtain the optimal powers in closed form.

APPENDIX B

JOINT POWER ALLOCATION AND RATE ADOPTION ALGORITHM

By taking $\ln$ from the both sides of (13c), we have

$$
\ln (2^{r_{b,i}} - 1) + \ln \left( \sum_{j \in \mathcal{C}_{b,i}} p_{b,j} \hat{h}_{b,k} + I_{b,k} + \sigma_{b,k}^2 \right) \leq \ln (p_{b,i} \hat{h}_{b,k}), \forall b \in \mathcal{B}, i, k \in \mathcal{U}_b, k \in \{i\} \cup \Phi_{b,i}.
$$

Now, let $p_{b,i} = e^{\bar{p}_{b,i}}$, and subsequently $I_{b,i}(\bar{p}_{-b}) =$
The objective function (20) is affine, so is concave on $r$. Constraint (13b) is affine, so is convex. Constraint (20b) is nonconvex since log-sum-exp is convex [27]. However, (20c) is nonconvex. In the left hand side of (20c), it can be easily shown that the first term $\ln(2^{r_{b,i}^{(t-1)}} - 1)$ is strictly concave on $r_{b,i}$ which makes (20) nonconvex and strongly NP-hard. Now, we apply the iterative sequential programming method. At each iteration $t$, we approximate the term $g(r_{b,i}^{(t)}) = \ln(2^{r_{b,i}^{(t-1)}} - 1)$ to its first-order Taylor series around $r_{b,i}^{(t-1)}$ obtained from prior iteration ($t - 1$) as follows:

$$\hat{g}(r_{b,i}^{(t)}) = g(r_{b,i}^{(t-1)}) + g'(r_{b,i}^{(t-1)}) (r_{b,i}^{(t)} - r_{b,i}^{(t-1)}), (21)$$

where $g'(r) = \frac{\partial}{\partial r} g(r)$. By substituting $g(r_{b,i}^{(t)})$ with its affine approximated form $\hat{g}(r_{b,i}^{(t)})$ in (21), problem (20) at iteration $t$ will be approximated to the following convex form as

$$\max_{\hat{r}_{b,i}^{(t)}, r_{b,i}^{(t)} \geq 0} \sum_{b \in B} \sum_{i \in U_b} r_{b,i}^{(t)} \quad \text{s.t.} \quad (13b), (20b)$$

$$\hat{g}(r_{b,i}^{(t)}) + \ln\left(\sum_{j \in F_b} e^{\hat{g}(r_{b,i}^{(t)})}h_{b,k} + \sum_{j \in F_b \setminus k} \left(\sum_{l \in U_d} e^{\hat{g}(r_{b,i}^{(t)})}h_{l,b,i}\right)\right) + \sigma_{b,k}^2 \geq \hat{p}_{b,i} \ln(h_{b,k}), \forall b \in B, i, k \in U_b,$n

where $i \in \{i\} \cup \Phi_{b,i}$. (22b)

It can be shown that (22) satisfies the KKT conditions [21], so it can be solved by using the Lagrange dual method, or IPMs [27]. In the sequential programming, we first initialize $r^{(0)}$. At each iteration $t$, we solve (22) and find $(r^{*(t)}, \hat{p}^{*(t)})$ according to the updated $\hat{g}(r_{b,i}^{(t)})$ based on $r^{*(t-1)}$. We continue the iterations until the convergence is achieved.

The solution of (22) remains in the feasible region of the main problem (20). This is due to the fact that at each iteration $t$, we have $\hat{g}(r_{b,i}^{(t)}) \geq g(r_{b,i}^{(t)})$. Let $(\hat{r}^{*(t)}, \hat{p}^{*(t)})$ be the feasible solution of (22). It implies that (22b) is satisfied.

Thus, we have

$$\hat{p}_{b,i} \ln(h_{b,k}) - \ln\left(\sum_{j \in F_b} e^{\hat{g}(r_{b,i}^{(t)})}h_{b,k} + \sum_{j \in F_b \setminus k} \left(\sum_{l \in U_d} e^{\hat{g}(r_{b,i}^{(t)})}h_{l,b,i}\right)\right) + \sigma_{b,k}^2 \geq \hat{g}(r_{b,i}^{(t)}), \forall b \in B, i, k \in U_b, k \in \{i\} \cup \Phi_{b,i}.$$

Since $\hat{g}(r_{b,i}^{(t)}) \geq g(r_{b,i}^{(t)})$, we can guarantee that (20c) is satisfied, meaning that (22) remains in the feasible region of (20). It can be easily shown that the sequential programming generates a sequence of improved feasible solutions, such that it converges to a stationary point which is a local maxima of (20) [21]. The performance and convergence of this algorithm is numerically evaluated in Subsection IV-E in [24].

APPENDIX C

PROOF OF THEOREM 1

For each user pair $i, k \in U_b$, if (4) holds at any power level $p_{b,i}$, the decoding order $k \rightarrow i$ is optimal. Note that (4) for the user pairs in cell $b$ is completely independent from $p_b$ (see Corollary 2). Let us rewrite (4) for the user pair $i, k \in U_b$ with $\frac{h_{k,b}}{\sigma_{b,k}} > \frac{h_{i,k}}{\sigma_{b,i}}$

$$\frac{h_{b,k} - h_{b,i}}{\sigma_{b,k} - \sigma_{b,i}} \geq \sum_{j \in U_b \setminus k, \sigma_{b,j}} \frac{p_{j,i}}{\sigma_{b,i} \sigma_{b,b}} \left(\frac{h_{j,b,k} h_{b,b} - h_{j,b,i} h_{b,k}}{H_{j,b,i,k}}\right). (23)$$

Assume that the left-hand side (LHS) is non-negative (which corresponds to applying the CNR-based decoding order). If each term $H_{j,b,i,k}$ in the right-hand side (RHS) is non-positive, (23) holds for any $p_b$ [28]. The condition $H_{j,b,i,k} \leq 0$ is equivalent to $\frac{h_{b,j,k}}{\sigma_{b,k}} \geq \frac{h_{j,b,i}}{\sigma_{b,i}}$. The latter sufficient condition implies that for $\frac{h_{j,b,i}}{\sigma_{b,i}} \geq \frac{h_{b,j,k}}{\sigma_{b,k}}$, if $H_{j,b,i,k} \leq 0, \forall j \in B \setminus \{b\}$, the decoding order $k \rightarrow i$ is optimal. This condition is first proposed in Lemma 1 in [28] to guarantee that (23) holds independent from $p$. In this paper, we state that the latter independency guarantees that the decoding order $k \rightarrow i$ is optimal independent from $p$.

Now, we consider a more general case, where $H_{j,b,i,k} > 0$ for at least one neighboring BS $j$. In this case, Lemma 1 in [28] cannot guarantee any optimality for the decoding order $k \rightarrow i$. However, we show that there exists a sufficient condition to guarantee the optimality of the decoding order $k \rightarrow i$. In this line, we find an upper-bound constant for the term in the RHS of (23). If the LHS of (23) is greater than that of the upper-bound term for the RHS of (23), it is guaranteed that (23) holds independent from $p$, so the decoding order $k \rightarrow i$ is optimal. We first define $Q_{b,i,k} = \{j \in B \setminus \{b\} | H_{j,b,i,k} > 0\}$ for the user pair $i, k \in U_b$ with $\frac{h_{b,i}}{\sigma_{b,i}} > \frac{h_{b,k}}{\sigma_{b,k}}$. Obviously, $H_{j,b,i,k} \leq 0$ for each neighboring BS $j \in B \setminus \{Q_{b,i,k} \cup \{b\}\}$.
In this way, the following inequality is always guaranteed

$$\sum_{j \in Q_{b,i,k}} \frac{p_{j}}{\sigma_{b,i,k}} H_{j,b,i,k} \geq \sum_{j \in D_{b}} \left( \sum_{i \in U_{j}} p_{j,i} \right) \sigma_{b,i,k} \sum_{j \in D_{b}} \sigma_{b,i,k} H_{j,b,i,k}. \quad (24)$$

In other words, the upper-bound of the RHS of (23) can be achieved if we assume that each BS $j \in Q_{b,i,k}$ (or equivalently each BS $j$ with $H_{j,b,i,k} > 0$) operates at its maximum power budget $P_{j}^{\text{max}}$, and each BS $j \notin Q_{b,i,k}$ (or equivalently each BS $j$ with $H_{j,b,i,k} \leq 0$) is deactivated. According to (23) and (24), the following inequality always holds

$$h_{b,k} - h_{b,i} \geq \sum_{j \in Q_{b,i,k}} \left( \sum_{i \in D_{j}} p_{j,i} \right) H_{j,b,i,k}. \quad (25)$$

Thus, if $h_{b,k} - h_{b,i} \geq \sum_{j \in Q_{b,i,k}} \left( \sum_{i \in D_{j}} p_{j,i} \right) H_{j,b,i,k}$, we guarantee that (23) holds independent from $p$, so the decoding order $k \rightarrow i$ is optimal, and the proof is completed.

**APPENDIX D

PROOF OF FACT 1**

For convenience, let all the channel gains be normalized by noise, and the CNR-based decoding order is applied. According to (23), for each user pair $i, k \in U_{b}$, $k \in \Phi_{b,i}$ or equivalently $h_{b,k} \geq h_{b,i}$, the constraint (16) can be rewritten as:

$$h_{b,k} - h_{b,i} \geq \sum_{j \notin Q_{b,i,k}} \left( \sum_{i \in D_{j}} p_{j,i} \right) H_{j,b,i,k}. \quad (26)$$

where $H_{j,b,i,k} = (h_{b,k} - h_{b,i} - h_{j,b,i,k})$. Constraint (26) can be equivalently transformed to $(B - 1)$ maximum power constraints including all the BSs in $B \setminus \{b\}$. The power consumption of each BS $j \in B \setminus \{b\}$ is upper-bounded by

$$\sum_{i \in D_{b}} p_{j,i} \leq \left( \frac{h_{b,k} - h_{b,i}}{\Psi_{j,b,i,k}(p_{-(j,b)})} \right) \sum_{i \in D_{b} \setminus \{j\}} \sum_{i \in D_{j}} p_{j,i} H_{j,b,i,k}.$$

In general, the maximum power constraint (3b) and SIC constraint (16) can be equivalently combined as

$$\sum_{i \in D_{b}} p_{b,i} \leq \min \left\{ \frac{P_{b}^{\text{max}}}{\Psi_{b,i,k}(p_{-(b,j)})}, \min_{i \in D_{j}, k \in \Phi_{b,i}} \psi_{b,i,k}(p_{-(j,b)}) \right\}, \forall b \in B. \quad (27)$$

The negative side impact of the SIC necessary constraint (16) can be observed in (27). For the case that

$$\min_{i \in D_{j}, k \in \Phi_{b,i}} \psi_{b,i,k}(p_{-(j,b)}) < P_{b}^{\text{max}}, \quad (16)$$

imposes additional limitations on the total power consumption of BS $b$ which restricts the feasible region of (3b).

**REFERENCES**

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.

[2] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[3] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), “The capacity region of the Gaussian multiple-input–multiple-output broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[4] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (NOMA) for cellular future radio access,” in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1–5.

[5] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, “A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[6] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, “Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges,” *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[7] W. Shin, M. Vaesi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, “Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges,” *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 176–183, Oct. 2017.

[8] P. Xu, Z. Ding, X. Dai, and H. V. Poor, “A new evaluation criterion for non-orthogonal multiple access in 5G software defined networks,” *IEEE Access*, vol. 3, pp. 1633–1639, 2015.

[9] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadhi, H. Yanikomeroglu, and S. M. Sait, “A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks,” *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2192–2235, 4th Quart., 2020.

[10] M. Vaesi, R. Schober, Z. Ding, and H. V. Poor, “Non-orthogonal multiple access: Common myths and critical questions,” *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 174–180, Oct. 2019.

[11] M. S. Ali, H. Tabassum, and E. Hossain, “Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems,” *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[12] M. S. Ali, E. Hossain, A. Al-Dweik, and D. I. Kim, “Downlink power allocation for CoMP-NOMA in multi-cell networks,” *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3982–3998, Sep. 2018.

[13] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, “On optimal power allocation for downlink non-orthogonal multiple access systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.

[14] Y. Fu, Y. Chen, and C. W. Sung, “Distributed power control for the downlink of multi-cell NOMA systems,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6207–6220, Sep. 2017.

[15] L. Lei, L. You, Y. Yang, D. Yuan, S. Chazinotas, and B. Ottersten, “Load coupling and energy optimization in multi-cell and multi-carrier NOMA networks,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11323–11337, Nov. 2019.

[16] Z. Yang, C. Pan, W. Xu, Y. Pan, M. Chen, and M. Elkashlan, “Power control for multi-cell networks with non-orthogonal multiple access,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 927–942, Feb. 2018.

[17] D. Ni, L. Hao, Q.-T. Tran, and X. Qian, “Transmit power minimization for downlink multi-cell multi-carrier NOMA networks,” *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2459–2462, Dec. 2018.

[18] R. D. Yates, “A framework for uplink power control in cellular radio systems,” *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[19] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, “Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems,” *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[20] J. Zhao, Y. Liu, K. K. Chai, A. Nallanathan, Y. Chen, and Z. Han, “Spectrum allocation and power control for non-orthogonal multiple access in HetNets,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5825–5837, Sep. 2017.

[21] A. Zappone, E. Björnson, L. Sanguinetti, and E. Jorswieck, “Globally optimal energy-efficient power control and receiver design in wireless networks,” *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2844–2859, Jun. 2017.

[22] K. Wang, Y. Liu, Z. Ding, A. Nallanathan, and M. Peng, “User association and power allocation for multi-cell non-orthogonal multiple access networks,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5284–5298, Nov. 2019.
Sepehr Rezvani (Student Member, IEEE) received the M.Sc. degree in electrical and computer engineering from Tarbiat Modares University, Tehran, Iran. He is currently pursuing the Ph.D. degree with the Department of Information Theory and Communication Systems, Technische Universität Braunschweig, Braunschweig, Germany. His research interests include signal processing, analysis and applications of multiple access techniques, and optimization theory in wireless communication systems.

Eduard Axel Jorswieck (Fellow, IEEE) was born in Berlin, Germany, in 1975. He received the Dipl.-Ing. degree in computer engineering and the Ph.D. degree in electrical engineering from TU Berlin in 2000 and 2004, respectively.

From 2006 to 2008, he worked as a Post-Doctoral Research Fellow and an Assistant Professor with the Signal Processing Department, KTH Stockholm, Sweden. From 2008 to 2019, he was the Head of the Chair of Communications Theory and a Full Professor at TU Dresden, Germany. Since 2020, he has been the Managing Director of the Institute of Communications Technology, the Head of the Chair for Communications Systems, and a Full Professor at Technische Universität Braunschweig (TU Braunschweig), Brunswick, Germany. His main research interests are in the broad area of communications, signal processing, and applied information theory. He has published more than 135 journal articles, 15 book chapters, three monographs, and some 285 conference papers on these topics. In 2006, he received the IEEE Signal Processing Society Best Paper Award. Together with colleagues, he has received the Best Paper and the Best Student Paper Awards at IEEE CAMSAP 2011, IEEE WCSP 2021, IEEE SPAWC 2021, IEEE ICUFN 2018, PETS 2019, and ISWCS 2019. Since 2017, he has been serving as the Editor-in-Chief for the EURASIP Journal on Wireless Communications and Networking. He has served on the editorial boards for IEEE SIGNAL PROCESSING LETTERS, IEEE TRANSACTION ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

Nader Mokari Yamchi (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Tarbiat Modares University, Tehran, Iran, in 2014. He joined the Department of Electrical and Computer Engineering, Tarbiat Modares University, as an Assistant Professor, in October 2015, where he is currently an Associate Professor. His research interests cover many aspects of wireless technologies with a special emphasis on wireless networks. In recent years, his research has been funded by Iranian Mobile Telecommunication Companies and Iranian National Science Foundation (INSF). He was also involved in a number of large scale network design and consulting projects in the telecom industry. He received the Best Paper Award at ITU K-2020. His thesis received the IEEE Outstanding Ph.D. Thesis Award. He has been elected as an IEEE Exemplary Reviewer in 2016 by IEEE Communications Society.

Mohammad Reza Javan (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Shahid Beheshti University, Tehran, Iran, in 2003, the M.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, in 2006, and the Ph.D. degree in electrical engineering from Tarbiat Modares University, Tehran, in 2013. He is currently a Faculty Member with the Department of Electrical and Robotics Engineering, Shahrood University of Technology, Shahrood, Iran. His research interests include design and analysis of communication networks with emphasis on the application of optimization theory and artificial intelligence.