Resource Optimization With Load Coupling in Multi-Cell NOMA

Lei You, *Student Member, IEEE*, Di Yuan, *Senior Member, IEEE*, Lei Lei, *Member, IEEE*, Sumei Sun, *Fellow, IEEE*, Symeon Chatzinotas, *Senior Member, IEEE*, and Björn Ottersten, *Fellow, IEEE*

Abstract—Optimizing non-orthogonal multiple access (NOMA) in multi-cell scenarios is much more challenging than the single-cell case because inter-cell interference must be considered. Most papers addressing NOMA consider a single cell. We take a significant step in analyzing NOMA in multi-cell scenarios. We explore the potential of NOMA networks in achieving optimal resource utilization with arbitrary topologies. Towards this goal, we investigate a broad class of problems consisting of optimizing power allocation and user pairing for any cost function that is monotonically increasing in time-frequency resource consumption. We propose an algorithm that achieves global optimality for this problem class. The basic idea is to prove that solving the joint optimization problem of power allocation, user pair selection, and time-frequency resource allocation amounts to solving a so-called iterated function without a closed form. We prove that the algorithm approaches optimality with fast convergence. Numerically, we evaluate and demonstrate the performance of NOMA for multi-cell scenarios in terms of resource efficiency and load balancing.

Index Terms—NOMA, multi-cell, resource allocation.

I. INTRODUCTION

NON-ORTHOGONAL multiple access (NOMA) is considered a promising technique for enhancing resource efficiency [2]–[13]. In two recent surveys [2], [3], the authors pointed out that resource allocation in multi-cell NOMA poses much more research challenges compared to the single-cell case, because optimizing NOMA with multiple cells has to model the interplay between successive interference cancellation (SIC) and inter-cell interference. As one step forward, the investigations in [2], [3] have addressed two-cell

scenarios. In [6], the authors proposed two coordinated NOMA beamforming methods for two-cell scenarios. Reference [8] uses stochastic geometry to model the inter-cell interference in NOMA. Hence the results do not apply for analyzing network with specific given network topology. Reference [9] optimizes energy efficiency in multi-cell NOMA with downlink power control. However, the aspect of determining which users share resource by SIC, i.e., *user pairing*, is not considered. To the best of our knowledge, finding *optimal power allocation and user pairing* simultaneously for enhancing network resource efficiency in multi-cell NOMA without restrictions on network topology has not been addressed yet.

The crucial aspect of multi-cell NOMA consists of capturing the mutual interference among cells; This is a key consideration in SIC of NOMA. Moreover, the cells are coupled with each other because of the interference. Therefore, the cells cannot be optimized independently. For orthogonal multiple access (OMA) networks, a modeling approach had been proposed to characterize the inter-cell interference via capturing the mutual influence among the cells' resource allocations [14]–[36]. The model, named *load-coupling*, refers to the time-frequency resource consumption in each cell as the *cell load*. However, the model does not allow SIC. In our recent work [1], we addressed resource optimization in multi-cell NOMA. However, the system model is constrained by fixed power allocation. How to model joint optimization of power allocation and user pairing and how to solve the resulting problems to optimality have remained open so far.

II. MAIN RESULTS

For multi-cell NOMA, stochastic geometry is adopted to model inter-cell interference [8], which results in difficulties for analysis upon specific network topologies. In this paper, we target optimizing multi-cell NOMA network with any given topology. In the modeling approaches of OMA used by [14]–[36], instead of making micro-level assumptions on the behavior of the resource scheduler or slot-by-slot consideration of inter-cell interference per resource block (RB) in each individual cell, the level of interference generated by a cell is directly related to the amount of allocated time-frequency resource in the cell. This is used to model the coupling relationship of resource allocation among cells, which is shown to be sufficiently accurate for network-level interference characterization [24], [32].

We demonstrate how NOMA can be modeled in multi-cell scenarios by significantly extending the approaches

in [14]–[36], with joint optimization of power allocation and user pairing. One fundamental result under such type of models in OMA is the existence of the equilibrium for resource allocation. However, this modeling approach in NOMA leads to non-closed form formulation of cell load coupling, unlike the case of OMA. The fact poses significant challenges in analysis and problem solving. As one of our main results, we prove that such an equilibrium for resource allocation in NOMA exists as well and propose an efficient algorithm for obtaining the equilibrium. Furthermore, we prove that the equilibrium is the global optimum for resource optimization in multi-cell NOMA and thus a wide class of resource optimization problems can be optimally solved by our algorithm. Because of our analytical results, previous works about OMA with load coupling is a special case of ours, namely, the algorithmic notions and mathematical tools being used in those works of classic multi-cell power control or OMA load coupling thus directly apply to the analysis multi-cell NOMA, suggesting future works on this topic. All our analytical results are based on the extended model.

To the best of our knowledge, this is the first work investigating how to *optimally utilize power and time-frequency resources jointly* in multi-cell NOMA. As a key strength of our modeling approach, it enables to formulate and optimize an entire class of resource optimization problems. Namely, as long as the cost function is monotonically increasing in the cells' time-frequency resource consumption, our proposed framework for multi-cell NOMA applies. Specifically, for solving this class of problems optimally, we derive a *polynomial-time* algorithm S-Cell that gives the optimal power allocation and user pairing, for any given input of inter-cell interference. To address the dynamic coupling of inter-cell interference, we derive a unified algorithmic framework M-Cell that solves the multi-cell resource optimization problems optimally. The algorithm S-Cell serves as a subroutine and is iteratively called by M-Cell. We demonstrate theoretically the *linear convergence* of this process.

The fundamental differences between our investigated problems and single-cell NOMA are summarized as follows. For multiple cells, the resource allocation in one cell affects the interference that the cell generates to other cells. The amounts of required resource to meet the demand for all cells are coupled together, rather than being independent to each other. Optimizing resource allocation within one cell leads to a chain reaction among all other cells. Individual optimization for the cells results in sub-optimality and very inaccurate performance analysis. For multi-cell NOMA, not only the time-frequency resource allocation but also the power split and user pairing in all cells are coupled together for the same reason. Therefore, joint optimization in NOMA leads to a rather complex problem for analysis.

By numerical experiments, optimizing resource utilization by our algorithm enlightens how much we can gain from NOMA in terms of *resource efficiency* and *load balancing*.

III. Cell Load Modeling

Denote by $\mathcal{C} = \{1, 2, \ldots, n\}$ and $\mathcal{J} = \{1, 2, \ldots, m\}$ the sets of cells and user equipments (UEs), respectively. Denote
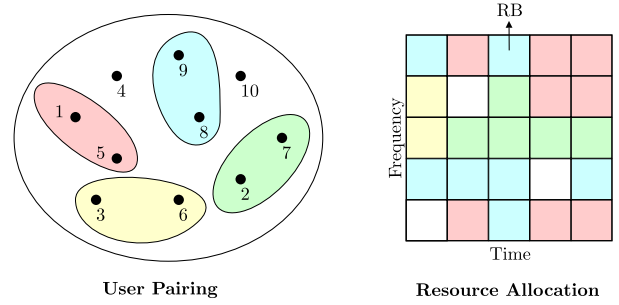

Fig. 1.   This figure illustrates user pairing and time-frequency resource sharing. There are 10 UEs in one cell. Eight form four user pairs $\{1, 5\}$, $\{2, 7\}$, $\{3, 6\}$, $\{8, 9\}$, and the other two UEs 4 and 10 are unpaired. The UEs within one pair share the same time-frequency resource as indicated by the colors.

by $\mathcal{J}_i$ the set of UEs served by cell i ($i \in \mathcal{C}$). When using j to refer to one UE in $\mathcal{J}$, i by default indicates j's serving cell, unless stated otherwise. Downlink is considered in our model.

A. Resource Sharing in NOMA

The resource in time-frequency domain is divided into RBs. In OMA, one RB can be accessed by only one UE. In NOMA, multiple UEs can be clustered together to access the same RB by SIC. Increasing the number of UEs in SIC, however, leads to fast growing decoding complexity [2], [3]. In previous works, it has been demonstrated that most of the possible performance improvement by SIC is reached by pairing as few as two UEs [2]–[7]. Pairing two UEs for resource sharing is illustrated in Figure 1. UEs within one pair share the same RB and the RBs allocated to different pairs do not overlap. We use u as a generic notation for a user pair (referred to as "pair" for simplicity). For cell i ($i \in \mathcal{C}$), denote by $\mathcal{U}_i$ the set of candidate pairs. Suppose there are in total m_i UEs in cell i. Then $|\mathcal{U}_i|$ is up to $\binom{m_i}{2}$. Denote by $\mathcal{V}_j$ ($j \in \mathcal{J}$) the set of pairs containing UE j. Let $\mathcal{U} = \bigcup_{i \in \mathcal{C}} \mathcal{U}_i$ (or equivalently $\mathcal{U} = \bigcup_{j \in \mathcal{J}} \mathcal{V}_j$) be the set of candidate pairs of all cells. Let $s = |\mathcal{U}|$. If there is a need to differentiate between pairs, we put indices on u, i.e., $\mathcal{U} = \{u_1, u_2, u_3, \ldots u_s\}$. Finally, in our model, for UEs we allow for both OMA and NOMA with SIC. For each UE, which mode is used (or both are used) is determined by optimization. In the following, we refer to these two modes as *orthogonal RB allocation* and *non-orthogonal RB allocation*, respectively. In general NOMA, we include both modes.

B. NOMA Downlink

We first consider orthogonal RB allocation in NOMA. Let p_i be the transmission power per RB in cell i ($i \in \mathcal{C}$). Denote by g_{ij} the channel coefficient from cell i to UE j. The signal-to-interference-and-noise ratio (SINR) is:

$$\gamma_j = \frac{p_i g_{ij}}{\sum_{k \in \mathcal{C} \setminus \{i\}} I_{kj} + \sigma^2}. \tag{1}$$

The term I_{kj} denotes the inter-cell interference from cell k to UE j, and is possibly zero. This generic notation is used for the sake of presentation. Later, we use the load-coupling model, where the cell load that reflects the usage of RBs governs the amount of interference. The term σ^2 is the noise power.

We then consider non-orthogonal RB allocation in NOMA. In [37] (Chapter 6.2.2, pp. 238) it is shown that, with superposition coding, one UE of pair u ($u \in \mathcal{U}$) can decode the other by SIC. When there is need to consider the decoding order in u, to be intuitive, we use $\oplus$ to denote the UE that applies interference cancellation, followed by decoding its own signal. And $\ominus$ denotes the UE that only decodes its own signal. Note that both $\oplus$ and $\ominus$ are generic notations and refer to the two different users in any pair u ($u \in \mathcal{U}$) in consideration. For any pair u, p_i is divided to $q_{\oplus u}$ and $q_{\ominus u}$ ($q_{\oplus u} + q_{\ominus u} = p_i$), with $q_{\oplus u}$ and $q_{\ominus u}$ being allocated to $\oplus$ and $\ominus$, respectively. (The generic notation q_{ju} ($j \in u$) denotes the power allocated to UE j.) We remark that $\oplus$ decodes $\ominus$'s signal first and hence $\ominus$'s signal does not compose the interference for $\oplus$. The SINR of $\oplus$ is computed by (2).

$$\gamma_{\oplus u} = \frac{q_{\oplus u} g_{i\oplus}}{\sum_{k \in \mathcal{C} \setminus \{i\}} I_{k\oplus} + \sigma^2}. \tag{2}$$

The UE $\ominus$ is subject to intra-cell interference from $\oplus$, i.e.,

$$\gamma_{\ominus u} = \frac{q_{\ominus u} g_{i\ominus}}{\underbrace{q_{\oplus u} g_{i\ominus}}_{\text{intra-cell}} + \underbrace{\sum_{k \in \mathcal{C} \setminus \{i\}} I_{k\ominus}}_{\text{inter-cell}} + \sigma^2}. \tag{3}$$

Denote by **q** the power allocation of all candidate pairs:

$$\mathbf{q} = \begin{bmatrix} q_{\oplus u_1} & q_{\oplus u_2} & \cdots & q_{\oplus u_s} \\ q_{\ominus u_1} & q_{\ominus u_2} & \cdots & q_{\ominus u_s} \end{bmatrix}.$$

We use $\mathbf{q}_u$ to represent the column of pair u ($u \in \mathcal{U}$) in $\mathbf{q}$, named *power split* for u. We remark that it is not necessary to use all the pairs in $\mathcal{U}$ for resource sharing. Whether or not a pair would be put in use and allocated with RBs is determined by optimization, discussed later in Section III-E. In addition, we remark that the decoding order is not constrained by the power split [37], even though by our numerical results, more power is always allocated to $\ominus$ in optimal solutions. The issue of the influence of inter-cell interference on the decoding order is addressed later in Section III-D.

C. Inter-Cell Interference Modeling

The basic idea is to use the cells' RB consumption levels to characterize respectively the cell's likelihood of interfering to the others. The approach is specified as follows. Denote by ρ_k the proportion of RBs allocated for serving UEs in cell k. The intuition behind the model is partially explained by the two extreme cases $\rho_k = 1$ and $\rho_k = 0$. If cell k is fully loaded, meaning that all RBs are allocated, then $\rho_k = 1$. In the other extreme case, cell k is idle and accordingly $\rho_k = 0$. Consider any UE j served by cell i. The interference j receives from cell k is $I_{kj} = p_k g_{kj}$ or $I_{kj} = 0$ in the two cases, respectively. In general, ρ_k serves as a scaling parameter for interference, see (4). By the interference modeling approach, the cell load directly translates to the scaling effect of interference and therefore the same notation is used for both.

$$I_{kj} = p_k g_{kj} \rho_k. \tag{4}$$

Intuitively, ρ_k reflects the likelihood that a UE outside cell k receives interference from k. Note that ρ_k in fact is the

proportion of time-frequency resource consumption of cell k and hence is referred to as *the load of cell k*.

We remark that this type of interference modeling approach is a suitable approximation for network-level performance analysis, which enables study of inter-cell interference in large-scale multi-cell networks without having to modeling micro-level interference. Detailed system-level simulations (e.g. [24] and [32]) have shown that this type of modeling has sufficient accuracy for cell-level interference characterization. This approach has been widely used and is getting increasingly popular [14]–[36], which however, to our best knowledge, is limited to OMA. We provide analytical results in order to extend the modeling approach to NOMA.

D. Determining Decoding Order

Inter-cell interference affects the decoding order in NOMA, and thus how to model the load-coupling in NOMA is significantly more challenging than OMA. Lemma 1 below resolves this issue by identifying pairs for which the decoding order can be determined independently of interference. As another benefit, it significantly reduces the set of candidate pairs.

Lemma 1: For $u = \{j, h\}$ ($g_{ij} \geq g_{ih}$) in cell i, if $g_{ij}/g_{ih} \geq g_{kj}/g_{kh}$ ($k \in \mathcal{C} \setminus \{i\}$), then SIC at j decodes first the signal for h, followed by decoding its own signal, and, user h does not apply SIC. That is, the decoding order $\oplus = j$ and $\ominus = h$ always hold for the pair independent of interference.

Proof: Denote by γ_{hj} and γ_{hh} respectively in (5) and (6) the SINRs at users j and h for the downlink signal of h.

$$\gamma_{hj} = \frac{q_{hu} g_{ij}}{q_{ju} g_{ij} + \sum_{k \in \mathcal{C} \setminus \{i\}} p_k g_{kj} \rho_k + \sigma^2}. \tag{5}$$

$$\gamma_{hh} = \frac{q_{hu} g_{ih}}{q_{ju} g_{ih} + \sum_{k \in \mathcal{C} \setminus \{i\}} p_k g_{kh} \rho_k + \sigma^2}. \tag{6}$$

With superposition coding, j cancels the interference from h if j can decode any data that h can decode [37], i.e. $\gamma_{hj} \geq \gamma_{hh}$, which reads:

$$q_{ju} g_{ij} g_{ih} + g_{ij} \sum_{k \in \mathcal{C} \setminus \{i\}} p_k g_{kh} \rho_k + g_{ij} \sigma^2$$
$$\geq q_{ju} g_{ij} g_{ih} + g_{ih} \sum_{k \in \mathcal{C} \setminus \{i\}} p_k g_{kj} \rho_k + g_{ih} \sigma^2.$$

Further, $\gamma_{hj} \geq \gamma_{hh}$ if and only if:

$$\sum_{k \in \mathcal{C} \setminus \{i\}} \frac{p_k \rho_k}{\sigma^2} (g_{ih} g_{kj} - g_{ij} g_{kh}) \leq (g_{ij} - g_{ih}). \tag{7}$$

Recall that $g_{ij} \geq g_{ih}$, and therefore the right-hand side of (7) is non-negative. Because of the condition $g_{ij}/g_{ih} \geq g_{kj}/g_{kh}$ for all $k \in \mathcal{C} \setminus \{i\}$ in the statement of the lemma, the left-hand side is non-positive. Hence Lemma 1. $\square$

The result of Lemma 1 is coherent with the previous observations that two UEs with large difference in channel conditions are preferred to be paired [4], [7]. If $g_{ij} \gg g_{ih}$, then most likely the condition in Lemma 1 holds, as the large scale path-loss from other cells, tends not to differ as much as from the serving cell i in this case. Besides, the large scale path-loss is a practically reasonable factor for ranking the decoding order [38], [39]. In Section VII, numerical results further show

that considering the UE pairs as defined by Lemma 1 virtually does not lead to any loss of performance. Lemma 1 is used to filter the candidate pairs set $\mathcal{U}$ (i.e. to drop some candidate pairs from $\mathcal{U}$) so as to reduce computational complexity. From now on, we let $\mathcal{U}_i$ be composed of pairs satisfying Lemma 1.

E. RB Allocation

If UE j ($j \in \mathcal{J}$) is using orthogonal RB allocation, then the achievable capacity[1] of j is (8), with γ_j being (1).

$$c_j = \log(1 + \gamma_j). \tag{8}$$

For non-orthogonal RB allocation, the achievable capacity for j and u ($j \in u$) is computed by $c_{ju} = MB \log(1 + \gamma_{ju})$ with γ_{ju} being (2) or (3). Therefore,

$$c_{ju} = \begin{cases} \log(1 + \gamma_{ju}) & j \in u \\ 0 & j \notin u \end{cases}. \tag{9}$$

For UE j ($j \in \mathcal{J}$), we use x_j to denote the proportion of RBs with orthogonal RB allocation to j. For any pair u ($u \in \mathcal{U}$), denote by x_u the non-orthogonal RB allocation for the two UEs in pair u. We use the vector $\mathbf{x}$ to represent the RB allocation for all the UEs, i.e.,

$$\mathbf{x} = [\ \underbrace{x_1, x_2, \ldots, x_m}_{\text{Orthogonal RB allocation}}\ ,\ \underbrace{x_{u_1}, x_{u_2}, \ldots, x_{u_s}}_{\text{Non-orthogonal RB allocation}}\ ].$$

For any UE j, $x_j = 0$ means that UE j does not use orthogonal RB allocation. Similarly, for any pair u, $x_u = 0$ means that pair u is not put in use. For any UE j, if $x_u = 0$ for all $u \in \mathcal{V}_j$, then it means that UE j only uses orthogonal RB allocation. Resources used by different pairs are orthogonal such that there is no interference among pairs. Denote by $\bar{\rho}$ the cell load limit. By constraining that the sum of them which equals to the load of cell i does not exceed $\bar{\rho}$, the amounts represented by x_j ($j \in \mathcal{J}_i$) and x_u ($u \in \mathcal{U}_i$) do not overlap. Orthogonal RB allocation is considered among the pairs in one cell, meaning that the pairs do not have interference with each other.

$$\underset{\substack{\uparrow \\ \text{Cell} \\ \text{load}}}{\rho_i} = \underbrace{\sum_{j \in \mathcal{J}_i} x_j}_{\substack{\text{Orthogonal} \\ \text{RB proportion}}} + \underbrace{\sum_{u \in \mathcal{U}_i} x_u}_{\substack{\text{Non-orthogonal} \\ \text{RB proportion}}} \leq \underset{\substack{\uparrow \\ \text{Load} \\ \text{limit}}}{\rho_i}. \tag{10}$$

We use $\boldsymbol{\rho}$ to represent the vector of network load, i.e.,

$$\boldsymbol{\rho} = [\rho_1, \rho_2, \ldots, \rho_n].$$

Similarly, we use vector $\bar{\boldsymbol{\rho}}$ to denote the load limits of all cells.

The term $c_j x_j$ computes the bits delivered to UE j with orthogonal RB allocation, because c_j is the achievable capacity of UE j on all RBs and x_j is the proportion of RBs with orthogonal RB allocation. Similarly, the term $c_{ju} x_u$ is the bits delivered to UE j by non-orthogonal RB allocation for pair u. Denote by d_j the bits demand of UE j. The quality-of-service (QoS) requirement is:

$$\underbrace{c_j x_j}_{\substack{\text{Bits delivered} \\ \text{by orthogonal} \\ \text{RB allocation}}} + \underbrace{\sum_{u \in \mathcal{V}_j} c_{ju} x_u}_{\substack{\text{Bits delivered by} \\ \text{non-orthogonal} \\ \text{RB allocation}}} \geq \underset{\substack{\uparrow \\ \text{Bits} \\ \text{demand}}}{d_j}. \tag{11}$$

[1]For the sake of presentation, we use the natural logarithm throughout the paper. We remark that all conclusions hold for the logarithm to base 2.

We remark that d_j is normalized by the RB spectral bandwidth and the total number of RBs, for the sake of presentation. Note that a user can use orthogonal RB allocation individually, or non-orthogonal RB allocation with the other user in the pair, or both, which is subject to optimization. The amount of allocated RBs to a user in OMA or a pair adopting NOMA, is subject to optimization, under the constraint that the overall allocated resource does not exceed limit.

We remark that there is an implicit *pair selection* problem in the above expressions. Note that $|\mathcal{U}|$ increases fast with $|\mathcal{J}|$. It is therefore impractical to simultaneously use all pairs in $\mathcal{U}$. To deal with this issue, each UE is allowed to use up to one pair in $\mathcal{U}$ for optimization, as formulated later in Section IV, though our system model is not limited by this. The problem of pairing and resource allocation is challenging: First, UEs of the same pair are coupled in resource allocation. Second, one can observe that increasing x_u (or x_j) for some pair u (or some UE j) may enhance the throughput of the UEs of u (or UE j). However, since x_u (or x_j) appears in the inter-cell interference term (see (4) and (10)), the increase of x_u (or x_j) results in less available resources for other UEs and leads to more interference. The user pairing selection is not given a priori but is determined by optimization. We remark that whether or not a UE should be allocated with resources with OMA or NOMA, or both, is up to optimization. The overall amount of resource used by NOMA and OMA in the entire network are part of the optimization output.

F. Comparison to OMA Modeling

The models proposed for OMA in [14]–[20] are inherently a special case of our NOMA model. The former is obtained by setting $\mathcal{U} = \phi$. Then, the terms for non-orthogonal RB allocation disappear in (10) and (11) and $\mathbf{x}$ is therefore eliminated in (8)–(11). Also, there is no power split in OMA. Hence (8)–(11) form a non-linear system only in terms of $\boldsymbol{\rho}$. This system falls into the analytical framework of standard interference function (SIF) [40], which enables the computation of the optimal network load settings via fixed-point iterations [17]. However, for the general NOMA case, the resource allocation is not at UE-level. One needs to split a UE's demand between orthogonal and non-orthogonal RB allocations, which results in a new dimension of complexity.

IV. PROBLEM FORMULATION

By successively plugging (1) and (4) into (8), we obtain a function c_j in load $\boldsymbol{\rho}$, i.e., $c_j(\boldsymbol{\rho})$. Similarly, we obtain $c_{ju}(\mathbf{q}, \boldsymbol{\rho})$ from (2), (3), (4), and (9). For pair u ($u \in \mathcal{U}$), we use a binary variable y_u to indicate whether or not the pair u is selected. Define $\mathbf{y}$ as

$$\mathbf{y} = [y_{u_1}, y_{u_2}, \ldots, y_{u_s}].$$

We minimize a generic cost function $F(\boldsymbol{\rho})$ that is *monotonically (but not necessarily strictly monotonically) increasing in each element of $\boldsymbol{\rho}$*. MINF is given below. Constraints (12b) guarantee that the cell load complies to the load limit $\bar{\rho}$. Constraints (12c) state the relationship between RB allocation and cell load. Constraints (12d) and (12e) are for QoS

and power, respectively. Constraints (12f) guarantee that RB allocation occurs only for selected pairs. By constraints (12g), each UE belongs up to one pair such that the selected pairs are mutually exclusive. The variables are cell load $\boldsymbol{\rho}$, power allocation $\mathbf{q}$, RB allocation $\mathbf{x}$, and pair selection $\mathbf{y}$. The variable domains are imposed by (12h) and (12i). Throughout this paper, we use $\mathbf{0}$ to represent zero vector/matrix. For simplicity, the dimension(s) of $\mathbf{0}$ is not explicitly stated.

$$[\text{MINF}] \min_{\boldsymbol{\rho},\mathbf{q},\mathbf{x},\mathbf{y}} \quad F(\boldsymbol{\rho}) \tag{12a}$$

$$\text{s.t. } \rho_i \leq \bar{\rho}, \ i \in \mathcal{C} \tag{12b}$$

$$\rho_i = \sum_{j \in \mathcal{J}_i} x_j + \sum_{u \in \mathcal{U}_i} x_u, \quad i \in \mathcal{C} \tag{12c}$$

$$c_j(\boldsymbol{\rho})x_j + \sum_{u \in \mathcal{V}_j} c_{ju}(\mathbf{q},\boldsymbol{\rho})x_u \geq d_j, \quad j \in \mathcal{J} \tag{12d}$$

$$\sum_{j \in u} q_{ju} = p_i, \quad u \in \mathcal{U}_i, \ i \in \mathcal{C} \tag{12e}$$

$$x_u \leq y_u, \quad u \in \mathcal{U} \tag{12f}$$

$$\sum_{u \in \mathcal{V}_j} y_u \leq 1, \quad j \in \mathcal{J} \tag{12g}$$

$$\boldsymbol{\rho}, \mathbf{q}, \mathbf{x} \geq \mathbf{0} \tag{12h}$$

$$y_u \in \{0,1\}, \quad u \in \mathcal{U} \tag{12i}$$

V. OPTIMIZATION WITHIN A CELL

In multi-cell NOMA, due to the interference among cells, one cell's pair selection may affect the other cells' power splits, and vice versa. Let us consider a simple case in this section. Suppose we optimize the load of one cell i, and the cell load levels of $\mathcal{C} \backslash \{i\}$ are temporarily fixed. This optimization step is a module for solving MINF later in Section VI. We respectively use $\mathbf{q}_i$, $\mathbf{x}_i$, $\mathbf{y}_i$ to denote the corresponding variable elements for power allocation, RB allocation, and pair selection. Vector $\boldsymbol{\rho}_{-i}$ is composed of all elements but ρ_i of $\boldsymbol{\rho}$. We minimize ρ_i under fixed $\boldsymbol{\rho}_{-i}$, as formulated below.

$$\min_{\rho_i,\mathbf{q}_i,\mathbf{x}_i,\mathbf{y}_i} \rho_i \text{ s.t. (12c)–(12i) of cell } i, \text{ with fixed } \boldsymbol{\rho}_{-i}. \tag{13}$$

Since $\boldsymbol{\rho}_{-i}$ is fixed, c_j is a constant and c_{ju} is a function in $\mathbf{q}_i$ only. Different from previous works [41] and [42], this single-cell resource optimization problem is subject to user demand constraints.

The optimization is not straightforward even under fixed inter-cell interference. The optimal power split for one pair is up to how much time-frequency resource is allocated to this pair. In other words, for one pair u, if the amount of RBs allocated to u changes, the optimal power split for u before this change loses its optimality. So the power split $\mathbf{q}$ and the resource allocation $\mathbf{x}$ are coupled together. In addition, the pair selection is a combinatorial problem. Therefore, the power split $\mathbf{q}$, the time-frequency resource allocation $\mathbf{x}$, and the user pair selection $\mathbf{y}$, must be optimized jointly.

Lemma 2: All constraints of (12d) in (13) hold as equalities at any optimum.

Proof: Denote the optimal objective value of (13) by ρ_i' and the optimal orthogonal RB allocation of j ($j \in \mathcal{J}_i$) by x_j'. Suppose strict inequality holds for some j. If $x_j' > 0$, by fixing all other variables except for x_j in (13), one can verify that the solution $x_j' - \epsilon$ ($\epsilon > 0$) is still feasible to (13) as long as ϵ is sufficiently small. In addition, $x_j' - \epsilon$ leads to a lower objective value $\rho_i' - \epsilon$, which contradicts our assumption that ρ_i' is the optimal objective value. If $x_j' = 0$, then j's demand has to be satisfied by non-orthogonal RB allocation and the same argument applies to variable x_u ($j \in u$). $\square$

The first analytical result is that, *the optimal power split is independent of pair selection*, as formulated in Theorem 1. Denote by $\mathcal{Y}_u$ the set of all possible pairing solutions of (13) that include pair u, i.e.,

$$\mathcal{Y}_u = \{\mathbf{y}_i | y_u = 1, \sum_{u \in \mathcal{V}_j} y_u \leq 1, \ j \in \mathcal{J}_i\}.$$

Definition 1: Given a pair selection $\hat{\mathbf{y}}_i$ ($\hat{\mathbf{y}}_i \in \mathcal{Y}_u$), the optimal power split for pair u ($u \in \mathcal{U}_i$), denoted by $\hat{\mathbf{q}}_u$, is the column for pair u in $\hat{\mathbf{q}}_i$, where $\hat{\mathbf{q}}_i$ is obtained by optimally solving (13) for $\hat{\mathbf{y}}_i$.

Theorem 1: Consider u ($u \in \mathcal{U}_i$). The optimal power split for any $\hat{\mathbf{y}}_i$ ($\hat{\mathbf{y}}_i \in \mathcal{Y}_u$) is also optimal for $\hat{\mathbf{y}}_i'$ ($\hat{\mathbf{y}}_i' \in \mathcal{Y}_u$).

Proof: Denote by $\hat{\mathbf{q}}_u$ and $\hat{\mathbf{q}}_u'$ the optimal power splits for $\hat{\mathbf{y}}_i$ and $\hat{\mathbf{y}}_i'$, respectively. Suppose $\hat{\mathbf{q}}_u$ is not optimal for $\hat{\mathbf{y}}_i'$. There are two possibilities: 1) $c_{ju}(\hat{\mathbf{q}}_u) = c_{ju}(\hat{\mathbf{q}}_u')$ ($j \in u$); 2) $c_{ju}(\hat{\mathbf{q}}_u) \neq c_{ju}(\hat{\mathbf{q}}_u')$ for at least one j in u.

For 1), $\hat{\mathbf{q}}_u$ and $\hat{\mathbf{q}}_u'$ result in the same x_j and x_u for satisfying (12d) and are equally good for (13), which conflicts our assumption. Thus $\hat{\mathbf{q}}_u$ is optimal for $\hat{\mathbf{y}}_i'$. We then consider 2) and assume $c_{ju}(\hat{\mathbf{q}}_u) > c_{ju}(\hat{\mathbf{q}}_u')$. By Lemma 2, $\hat{\mathbf{q}}_u'$ makes (12d) become equality under $\hat{\mathbf{y}}_i'$. Replacing $\hat{\mathbf{q}}_u'$ by $\hat{\mathbf{q}}_u$ leads to some slack in (12d) and hence the objective can be improved. This contradicts that $\hat{\mathbf{q}}_u'$ is optimal for $\mathbf{y}_i'$. The same proof applies to $c_{ju}(\hat{\mathbf{q}}_u) < c_{ju}(\hat{\mathbf{q}}_u')$. Hence the conclusion. $\square$

By Theorem 1, the optimal power split is decoupled from pair selection. Next we analytically prove how to find the optimal power split for any pair.

A. Finding Optimal Power Split

Under fixed $\mathbf{y}_i$ ($\mathbf{y}_i \in \mathcal{Y}_u$, $u \in \mathcal{U}_i$), constraints (12g) are removed. Constraints (12f), and (12i) of (13) for all u with $y_u = 0$ in $\mathbf{y}_i$ are removed. Therefore, for each pair $u = \{\oplus, \ominus\}$, we can formulate a problem in (14). Solving this problem yields the optimal power split. In (14), $x_\oplus$ and $x_\ominus$ are the orthogonal RB allocation for $\oplus$ and $\ominus$, respectively. The variable x_u denotes the amount of non-orthogonal RB allocation for u.

$$\min_{\substack{x_\oplus, x_\ominus, x_u \geq 0 \\ \mathbf{q}_u \geq \mathbf{0}}} x_\oplus + x_\ominus + x_u \tag{14a}$$

$$\text{s.t. } c_\oplus(\boldsymbol{\rho}_{-i})x_\oplus + c_{\oplus u}(\mathbf{q}_u, \boldsymbol{\rho}_{-i})x_u \geq d_\oplus \tag{14b}$$

$$c_\ominus(\boldsymbol{\rho}_{-i})x_\ominus + c_{\ominus u}(\mathbf{q}_u, \boldsymbol{\rho}_{-i})x_u \geq d_\ominus \tag{14c}$$

$$q_{\oplus u} + q_{\ominus u} = p_i \tag{14d}$$

For deriving solution method for (14), define function w_j ($j = \oplus$ or $j = \ominus$) of $\boldsymbol{\rho}_{-i}$ as follows.

$$w_j(\boldsymbol{\rho}_{-i}) = \left(\sum_{k \in \mathcal{C} \backslash \{i\}} p_k g_{kj}\rho_k + \sigma^2\right) \Big/ g_{ij}. \tag{15}$$

For $q_{\oplus u}$, one can derive from (2) and (9):

$$q_{\oplus u} = (e^{c_{\oplus u}} - 1)w_{\oplus}(\boldsymbol{\rho}_{-i}). \tag{16}$$

Combining (16) with (14d), $q_{\oplus u}$ and $q_{\ominus u}$ can be eliminated, giving (17) below. Formulation (17) is equivalent to (14). Given $c_{\oplus u}$ and $c_{\ominus u}$, the corresponding $q_{\oplus u}$ and $q_{\ominus u}$ can be obtained from $c_{\oplus u}$ and $c_{\ominus u}$ by (14d) and (16).

$$\min_{\substack{x_{\oplus}, x_{\ominus}, x_u \geq 0 \\ c_{\oplus u}, c_{\ominus u} \geq 0}} x_{\oplus} + x_{\ominus} + x_u \tag{17a}$$

$$\text{s.t. } c_{\oplus}(\boldsymbol{\rho}_{-i})x_{\oplus} + c_{\oplus u}x_u \geq d_{\oplus} \tag{17b}$$

$$c_{\ominus}(\boldsymbol{\rho}_{-i})x_{\ominus} + c_{\ominus u}x_u \geq d_{\ominus} \tag{17c}$$

$$\text{Cv}_u(c_{\oplus u}, c_{\ominus u}, \boldsymbol{\rho}_{-i}) \leq 0 \tag{17d}$$

In (17), the function Cv_u is defined in (32) in the Appendix.[2] One can easily verify that Cv_u is convex in $c_{\oplus u}$ and $c_{\ominus u}$ (with $g_{i\oplus} \geq g_{i\ominus}$). The difficulty of (17) is on the two bi-linear constraints (17b) and (17c). However, they become linear with fixed x_u. To ease the presentation, we define the function below.

$$Z_u(x_u, \boldsymbol{\rho}_{-i}) = x_u + \min_{\substack{x_{\oplus}, x_{\ominus} \geq 0 \\ c_{\oplus u}, c_{\ominus u} \geq 0}} x_{\oplus} + x_{\ominus} \quad \text{s.t. } (17b)–(17d).$$

$$\tag{18}$$

Solving (17) (and equivalently (14)) is to find the minimum of $Z_u(x_u, \boldsymbol{\rho}_{-i})$. The following theorem shows the uniqueness of the minimum of $Z_u(x_u, \boldsymbol{\rho}_{-i})$.

Theorem 2: $Z_u(x_u, \boldsymbol{\rho}_{-i})$ *has unique minimum in* x_u.

Proof: Since the first term x_u in $Z_u(x_u, \boldsymbol{\rho}_{-i})$ is strictly monotonically increasing, to prove that $Z_u(x_u, \boldsymbol{\rho}_{-i})$ has unique minimum, we only need to prove that the remaining part of $Z_u(x_u, \boldsymbol{\rho}_{-i})$ is monotonically (but not necessarily strictly monotonically) decreasing in x_u. For this part, at the optimum (17b) and (17c) hold as equalities because of Lemma 2. Hence, reformulating the problem by replacing the inequalities in (17b) and (17c) with equalities does not lose optimality. With equalities, the variables $x_{\oplus}$ and $x_{\ominus}$ can be represented by $c_{\oplus u}$ and $c_{\ominus u}$:

$$x_{\oplus} = \frac{(d_{\oplus} - c_{\oplus u}x_u)}{c_{\oplus}(\boldsymbol{\rho}_{-i})}, \quad x_{\ominus} = \frac{(d_{\ominus} - c_{\ominus u}x_u)}{c_{\ominus}(\boldsymbol{\rho}_{-i})}. \tag{19}$$

Therefore $x_{\oplus}$ and $x_{\ominus}$ can be eliminated from the objective function. The minimization is thus equivalent to maximizing $c_{\oplus u}/c_{\oplus}(\boldsymbol{\rho}_{-i}) + c_{\ominus u}/c_{\ominus}(\boldsymbol{\rho}_{-i})$. We formulate this maximization problem below.

$$\max_{c_{\oplus u}, c_{\ominus u} \geq 0} \frac{c_{\oplus u}}{c_{\oplus}(\boldsymbol{\rho}_{-i})} + \frac{c_{\ominus u}}{c_{\ominus}(\boldsymbol{\rho}_{-i})} \tag{20a}$$

$$\text{s.t. } c_{\oplus u}x_u \leq d_{\oplus} \tag{20b}$$

$$c_{\ominus u}x_u \leq d_{\ominus} \tag{20c}$$

$$\text{Cv}_u(c_{\oplus u}, c_{\ominus u}, \boldsymbol{\rho}_{-i}) \leq 0 \tag{20d}$$

Constraints (20b) and (20c) originate from the non-negativity requirement of $x_{\oplus}$ and $x_{\ominus}$. Note that (20) is convex. In addition, the feasible region shrinks with the increase of x_u. Then

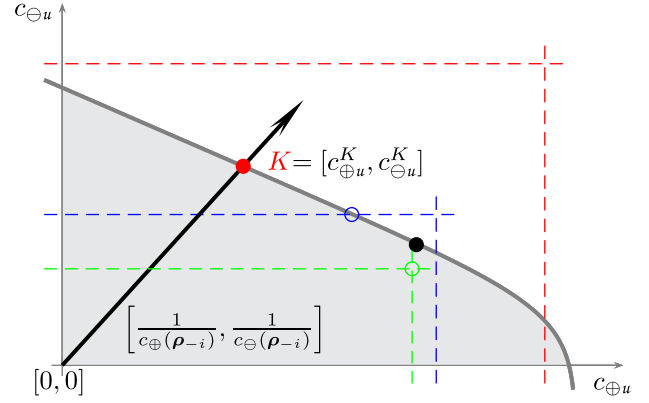[2]We remark that (17d) holds as equality at optimality.

Fig. 2. This figure shows the feasible region of (20) of $c_{\oplus u}$ and $c_{\ominus u}$ for different x_u. The shadowed area below the curve is (20d). The vertical and horizontal dashed lines are the hyperplanes defined by (20b) and (20c), respectively. The hyperplanes are shown by red, blue, and green dashed lines for Case 1, Case 2, and Case 3, respectively. The point K is optimal for Case 1. The blue and green circles represent the optimal solutions for Case 2 and Case 3, respectively. The black dot is the point on the curve where the two hyperplanes intersect with each other (see footnote 4 for the existence of this point.).

the optimum of (20) monotonically decreases with x_u. Hence the theorem. □

Because of Theorem 2, bi-section search of x_u reaches the minimum of $Z_u(x_u, \boldsymbol{\rho}_{-i})$. Note that for any x_u, computing $Z_u(x_u, \boldsymbol{\rho}_{-i})$ needs to solve (20). In the following, we prove how this can be done much more efficiently than employing standard convex optimization.

We remark that (20) is a two-dimensional optimization problem with respect to $c_{\oplus u}$ and $c_{\ominus u}$. Constraints (20b) and (20c) are defined by two hyperplanes $c_{\oplus u} = d_{\oplus}/x_u$ and $c_{\ominus u} = d_{\ominus}/x_u$, respectively. Due to the convexity of the function Cv_u in $c_{\oplus u}$ and $c_{\ominus u}$, the curve $\text{Cv}_u(c_{\oplus u}, c_{\ominus u}, \boldsymbol{\rho}_{-i}) = 0$ in (20d) along with $c_{\oplus u} = 0$ and $c_{\ominus u} = 0$ forms a convex set of $[c_{\oplus u}, c_{\ominus u}]$. The optimum of (20) depends on whether the two hyperplanes intersect with the curve and how they intersect. This leads to three possible cases to be considered, named Case 1, Case 2, and Case 3, respectively. In Case 1, constraints (20b) and (20c) are redundant, and the optimum is determined by the coefficients $1/c_{\oplus}(\boldsymbol{\rho}_{-i})$ and $1/c_{\ominus}(\boldsymbol{\rho}_{-i})$ in the objective function and the curve $\text{Cv}_u(c_{\oplus u}, c_{\ominus u}, \boldsymbol{\rho}_{-i}) = 0$. In Case 2, the optimum is defined by one of the hyperplanes and the curve. In Case 3, constraint (20d) is redundant, and the optimum is determined by the two hyperplanes. With x_u increasing from 0 to ∞, Case 1, Case 2, and Case 3 happen sequentially, and all happen eventually. The three cases are illustrated in Figure 2 with colors. Below we respectively show how to compute the optimum for each case.

In the following, we first compute the optimum in Case 1, represented by $K = [c_{\oplus u}^K, c_{\ominus u}^K]$, which is the intersection of the vector $[1/c_{\oplus}(\boldsymbol{\rho}_{-i}), 1/c_{\ominus}(\boldsymbol{\rho}_{-i})]$ and the curve. The point also leads to a closed-form solution for the optima of all the three cases. Mathematically, point K is solved by applying bi-section search to (21) below.

$$\begin{cases} \text{Cv}_u(c_{\oplus u}, c_{\ominus u}, \boldsymbol{\rho}_{-i}) = 0 \\ c_{\oplus u}c_{\oplus}(\boldsymbol{\rho}_{-i}) = c_{\ominus u}c_{\ominus}(\boldsymbol{\rho}_{-i}). \end{cases} \tag{21}$$

For solving (21), we can first eliminate $c_{\oplus u}$ or $c_{\ominus u}$ in the first equation. This can be done by representing one of $c_{\oplus u}$ and $c_{\ominus u}$ with the other by the second equation. Since there is only one variable in the first equation, one can use bi-section search to find its solution.[3] Then, we compute the value of x_u when at least one hyperplane goes through K, denoted by x_u^K in (22).

$$x_u^K = \min\{d_\oplus / c_{\oplus u}^K, d_\ominus / c_{\ominus u}^K\}. \tag{22}$$

The three cases 1, 2, and 3, indicated by red, blue, and green lines respectively in Figure 2, are as follows.

Case 1: ($x_u \le \min_{j \in u} d_j / c_{ju}^K$): Point K is the optimum of (20), because (20b) and (20c) are redundant. This happens when x_u is sufficiently small (or 0), as shown in Figure 2.

Case 2: ($x_u > \min_{j \in u} d_j / c_{ju}^K$ and $\mathrm{Cv}_u(d_\oplus / x_u, d_\ominus / x_u) > 0$): There exists one point on the curve where both two hyperplanes intersect.[4] We represent this point by the black dot on the curve in Figure 2. In Case 2, one hyperplane intersects with the curve at some point between K and the black dot, and intersects with the other hyperplane on some point above the curve, see Figure 2. The intersection point of the curve and the hyperplane is the optimum of (20). Without loss of generality, we assume K violates (20c), meaning that the hyperplane of (20c) goes through the optimum, as shown by Figure 2. By plugging the equation of the hyperplane into that of the curve, the optimal $c_{\oplus u}$ is a function of x_u. Similarly, if K violates (20c) instead, then $c_{\oplus u} = d_\oplus / x_u$ and the optimal $c_{\ominus u}$ is a function of x_u. To know which hyperplane goes through the optimum, one only needs to check which of $c_{ju}^K x_{ju}^K > d_j$ ($j = \oplus$ or $j = \ominus$) holds. Note that exactly one of the two holds in Case 2. The optimal $c_{\oplus u}$ and $c_{\ominus u}$ are computed respectively by (33) and (34) defined in the Appendix.

Case 3: ($x_u > \min_{j \in u} d_j / c_{ju}^K$ and $\mathrm{Cv}_u(d_\oplus / x_u, d_\ominus / x_u) \le 0$): Constraint (20d) is redundant, as shown in Figure 2. The optimum is the intersection point of the two hyperplanes, computed by $c_{ju} = d_j / x_u$ ($j = \oplus$ or $j = \ominus$).

In summary, the optimal solution of (20) is computed by (23) below ($j = \oplus$ or $j = \ominus$) in closed form, with H_{ju} being (33) or (34) in the Appendix.

$$C_{ju}(x_u, \boldsymbol{\rho}_{-i}) = \begin{cases} c_{ju}^K & \text{Case 1} \\ H_{ju}(x_u, \boldsymbol{\rho}_{-i}) & \text{Case 2} \\ d_j / x_u & \text{Case 3} \end{cases} \tag{23}$$

The function $Z_u(x_u, \boldsymbol{\rho}_{-i})$ computes the amount of resource used for both orthogonal and non-orthogonal RB allocations for the UEs in u. It is optimal to serve the two UEs only by orthogonal RB allocation, if the minimum of $Z_u(x_u, \boldsymbol{\rho}_{-i})$ occurs at $x_u = 0$. In all other cases, $\min_{x_u} Z_u(x_u, \boldsymbol{\rho}_{-i})$ yields the optimal power split for non-orthogonal RB allocations. The algorithm optimally solving (14), named SPLIT, is as follows.

---

[3] The solution is guaranteed to be unique and hence bi-section search applies. This is because, by representing one of $c_{\oplus u}$ and $c_{\ominus u}$ by the other by function Cv_u, one variable is monotonically decreasing in the other, resulting in a unique zero point.

[4] The existence of this point is guaranteed: With the increase of x_u, both hyperplanes will eventually intersect with the curve with two intersection points. By increasing x_u, the distance between the two intersections keeps being smaller. The two intersections will eventually overlap.

> SPLIT($u, \boldsymbol{\rho}_{-i}$)
>
> 1 $x_u^* = \arg\min_{x_u} Z_u(x_u, \boldsymbol{\rho}_{-i})$ // Bi-section search
> 2 Compute $\langle c_{\oplus u}^*, c_{\ominus u}^* \rangle$ by (23) // With $x_u^*, \boldsymbol{\rho}_{-i}$
> 3 Compute $\langle x_\oplus^*, x_\ominus^* \rangle$ by (19) // With $c_{\oplus u}^*, c_{\ominus u}^*, \boldsymbol{\rho}_{-i}$
> 4 Convert $\langle c_{\oplus u}^*, c_{\ominus u}^* \rangle$ to $\langle q_{\oplus u}^*, q_{\ominus u}^* \rangle$ // By (14d), (16)
> 5 **return** $\langle q_{\oplus u}^*, q_{\ominus u}^*, x_\oplus^*, x_\ominus^*, x_u^* \rangle$

B. Optimal Pairing

Denote by $\mathcal{Y}_i$ the set of all candidate pair selections:

$$\mathcal{Y}_i = (\cup_{u \in \mathcal{U}_i} \mathcal{Y}_u) \cup \{\mathbf{0}\}.$$

By obtaining $\min_{x_u} Z_u(x_u, \boldsymbol{\rho}_{-i})$ for all $u \in \mathcal{U}_i$ as shown earlier in Section V-A, enumerating all $\mathbf{y}_i$ in $\mathcal{Y}_i$ gives the optimal solution to (13). This exhaustive search however does not scale, as $|\mathcal{Y}_i|$ is exponential in the number of UEs. By the following derivation, we are able to obtain the optimum of (13) in polynomial time.

Theorem 3: The optimum of (13) *is computed by finding the maximum weighted matching in an undirected graph.*

Proof: To prove the conclusion, an undirected weighted graph $\mathcal{G}_i$ is constructed and explained below.

$$\mathcal{G}_i = \begin{cases} \langle \mathcal{J}_i, \mathcal{U}_i, \mathbf{w} \rangle & |\mathcal{J}_i| \text{ even} \\ \langle \mathcal{J}_i \cup \{\Delta\}, \mathcal{U}_i \cup \{\{j, \Delta\} | j \in \mathcal{J}_i\}, \mathbf{w} \rangle & |\mathcal{J}_i| \text{ odd.} \end{cases} \tag{24}$$

In (24), the graph is represented by a 3-tuple, with the first element being the vertex set, the second element being the edge set, and the third element being the weight vector. Parameter Δ is an auxiliary vertex for odd $|\mathcal{J}_i|$. Without loss of generality, below we focus on odd $|\mathcal{J}_i|$. (All conclusions naturally hold for $|\mathcal{J}_i|$ being even.) By the definition in (24), each UE is corresponding to a vertex. For each pair u in $\mathcal{U}_i$, there is one edge connecting the two UEs in u, associated with weight w_u. We name these as *type-1 edges*. Besides, for each UE j in $\mathcal{J}_i$, there is one extra edge connecting j and the auxiliary vertex Δ, associated with weight w_j. We name these as *type-2 edges*. An illustration is given in Figure 3.

The weight $\mathbf{w}$ is defined as follows, where T is a positive value keeping all weights being positive.

Type-1 edge	$w_u = T - \min_{x_u} Z_u(x_u, \boldsymbol{\rho}_{-i})$ ($u \in \mathcal{U}_i$)
Type-2 edge	$w_j = T - d_j / c_j(\boldsymbol{\rho}_{-i})$ ($j \in \mathcal{J}_i$)

First, we remark that any $\mathbf{y}_i$ is feasible to (13) if and only if all the pairs u with $y_u = 1$ ($u \in \mathcal{U}_i$) form a matching (or an empty edge set) in $\mathcal{G}_i$. Otherwise, there exists j such that $\sum_{u \in \mathcal{V}_j} y_u \ge 2$, and (12g) would be violated. Then, by the definition of weights, minimizing the load ρ_i becomes finding a maximum weighted matching. □

The algorithm S-CELL solving (13) exactly is as follows. Lines 1–8 compute the edge weights of the graph to be constructed. Then we construct the graph in Line 11 and compute
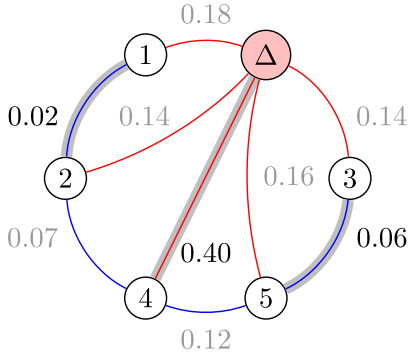
Fig. 3. The figure shows an example of one cell i with five UEs, i.e., $\mathcal{J}_i = \{1,2,3,4,5\}$. Assume the candidate pair set is $\mathcal{U}_i = \{\{1,2\},\{2,4\},\{4,5\},\{3,5\}\}$. The blue edges are type-1. The red edges are type-2. A matching is a set of edges without common vertices (also called independent edge set) and is a pair selection solution. The maximum matching, as highlighted in the figure, is $\{\{1,2\},\{3,5\},\{4,\Delta\}\}$. Note that two paired UEs in the solution of matching does not necessarily imply that the two share resource via NOMA. For any pair u, if x_u happens to be zero in the solution, then there is no RB allocated in non-orthogonal manner to the pair u and hence the two UEs in u are allocated with orthogonal resources.

the maximum matching[5] $\mathcal{U}_i^*$ in Line 12, which by Theorem 3 is the optimal pair selection in cell i. Lines 13–22 assign the obtained solutions to $\langle q_{\oplus u}^*, q_{\ominus u}^*, x_\oplus^*, x_\ominus^*, x_u^* \rangle$ for the pairs in $\mathcal{U}_i^*$. The other pairs are not selected and hence their values in $\mathbf{x}_i^*$, $\mathbf{q}_i^*$, and $\mathbf{y}_i^*$ are zeros.

We remark that in the matching process, if the number of nodes is odd, for the unpaired UE, it is allocated with orthogonal RBs. For two UEs that are paired in the solution of matching, they are not necessarily in non-orthogonal allocation but is up to optimization.

VI. Multi-Cell Load Optimization

This section proposes the algorithmic framework M-CELL for deriving the optimum of MINF, by analyzing sufficient-and-necessary conditions of optimality and feasibility.

A. Revisiting Single-Cell Load Minimization

Recall that for single cell optimization, the optimum of (13) of cell i ($i \in \mathcal{C}$) is a function of the load of other cells $\boldsymbol{\rho}_{-i}$. By Lemma 3 below, this function is well-defined for any non-negative $\boldsymbol{\rho}_{-i}$.

Lemma 3: The problem in (13) *is always feasible.*

Proof: We select some $\mathbf{y}_i$ in $\mathcal{Y}_i$ and fix it in (13) ($\mathcal{Y}_i \neq \phi$ by definition). For each pair u, we fix $\mathbf{q}_u$ to $[p_i/2, p_i/2]^\top$. To prove (13) is feasible, we prove the remaining problem is always feasible. Note that, with $\mathbf{y}_i$ and $\mathbf{q}_u$ being fixed, (13) becomes a linear programming (LP) problem, which is stated below (the equalities are by Lemma 2).

$$\min_{\mathbf{x} \geq \mathbf{0}} \sum_{j \in \mathcal{J}_i} x_j + \sum_{u \in \mathcal{U}_i} x_u, \quad \text{s.t.} \quad c_j x_j + \sum_{u \in \mathcal{V}_j} c_{ju} x_u = d_j. \quad (25)$$

[5]The best known algorithm [43] runs on $\mathcal{G}_i$ in $O((|\mathcal{U}_i| + |\mathcal{J}_i|)\sqrt{|\mathcal{J}_i|})$ (odd $|\mathcal{J}_i|$) or $O(|\mathcal{U}_i|\sqrt{|\mathcal{J}_i|})$ (even $|\mathcal{J}_i|$).

S-CELL$(\boldsymbol{\rho}_{-i})$
  1 **for** $u \in \mathcal{U}_i$
  2      $\langle q_{\oplus u}, q_{\ominus u}, x_\oplus, x_\ominus, x_u \rangle = $ SPLIT$(u, \boldsymbol{\rho}_{-i})$
  3      $w_u = T - (x_\oplus + x_\ominus + x_u)$ **//** $w_u > 0$
  4 **end for**
  5 **if** $|\mathcal{J}_i|$ is odd
  6   **for** $j \in \mathcal{J}_i$
  7          $x_j = d_j/c_j(\boldsymbol{\rho}_{-i})$
  8          $w_j = T - x_j$ **//** $w_j > 0$
  9   **end for**
10 **end if**
11 Construct $\mathcal{G}_i$ by (24)
12 $\mathcal{U}_i^* = $ MAXIMUM-WEIGHTED-MATCHING$(\mathcal{G}_i)$
13 $\mathbf{x}_i^* = \mathbf{0}, \mathbf{q}_i^* = \mathbf{0}, \mathbf{y}_i^* = \mathbf{0}$
14 **for** $u \in \mathcal{U}_i^* \cap \mathcal{U}_i$
15      $x_u^* = x_u, y_u^* = 1$
16   **for** $j \in u$
17           $q_{ju}^* = q_{ju}, x_j^* = x_j$
18   **end for**
19 **end for**
20 **if** $|\mathcal{J}_i|$ is odd
21   Find the $\{j, \Delta\}$ in $\mathcal{U}_i^*$ and let $x_j^* = x_j$
22 **end if**
23 $\rho_i = \sum_{j \in \mathcal{J}_i} x_j^* + \sum_{u \in \mathcal{U}_i} x_u^*$
24 **return** $\langle \rho_i, \mathbf{q}_i^*, \mathbf{x}_i^*, \mathbf{y}_i^* \rangle$

By Farkas' lemma, a group of linear constraints in standard form, i.e. $\mathbf{Ax} = \mathbf{b}$ ($\mathbf{b} \geq \mathbf{0}$), is feasible with $\mathbf{x} \geq \mathbf{0}$ if and only if there does not exist $\mathbf{v}$ such that $\mathbf{v}^\top \mathbf{A} \geq \mathbf{0}^\top$ and $\mathbf{v}^\top \mathbf{b} < 0$. Obviously, there is no $\mathbf{v}$ with $\mathbf{d} \geq \mathbf{0}$ satisfying (26).

$$v_j \geq 0 \ (j \in \mathcal{J}_i) \quad \text{and} \quad \sum_{j \in \mathcal{J}_i} v_j d_j < 0 \qquad (26)$$

Hence (25) is feasible, and the conclusion holds. $\qquad \square$

Let $\lambda_i = |\mathcal{Y}_i|$. For each $\mathbf{y}_i$ in $\mathcal{Y}_i$, we use an integer in $[1, \lambda_i]$ to uniquely index $\mathbf{y}_i$. We refer to all the pair selection solutions in $\mathcal{Y}_i$ as pairing 1, pairing 2, ..., pairing λ_i. Denote by $f_{ik}(\boldsymbol{\rho}_{-i})$ the optimum of (13) under pairing k ($1 \leq k \leq \lambda_i$), i.e.,

$$f_{ik}(\boldsymbol{\rho}_{-i}) = \min_{\rho_i, \mathbf{q}_i, \mathbf{x}_i} \rho_i \quad \text{s.t. (12c)–(12f) and (12h) of cell } i.$$

Let $f_i(\boldsymbol{\rho}_{-i})$ be the optimum[6] of (13). Then we have:

$$f_i(\boldsymbol{\rho}_{-i}) = \min_{k=1,2,\ldots,\lambda_i} f_{ik}(\boldsymbol{\rho}_{-i}). \qquad (27)$$

Network-wisely, we have:

$$\mathbf{f}(\boldsymbol{\rho}) = [f_1(\boldsymbol{\rho}_{-1}), f_2(\boldsymbol{\rho}_{-2}), \ldots, f_n(\boldsymbol{\rho}_{-n})]. \qquad (28)$$

The following theorem reveals a key property of $\mathbf{f}(\boldsymbol{\rho})$.

Theorem 4: $\mathbf{f}(\boldsymbol{\rho})$ *is an SIF, i.e. the following properties hold:*

1) *(Scalability)* $\alpha \mathbf{f}(\boldsymbol{\rho}) > \mathbf{f}(\alpha \boldsymbol{\rho})$, $\boldsymbol{\rho} \geq \mathbf{0}$, $\alpha > 1$.
2) *(Monotonicity)* $\mathbf{f}(\boldsymbol{\rho}) \geq \mathbf{f}(\boldsymbol{\rho}')$, $\boldsymbol{\rho} \geq \boldsymbol{\rho}'$, $\boldsymbol{\rho}, \boldsymbol{\rho}' \geq \mathbf{0}$.

[6]Therefore, $f_i(\boldsymbol{\rho}_{-i})$ equals the ρ_i obtained from S-CELL$(\boldsymbol{\rho}_{-i})$.

Proof: We first prove monotonicity and scalability for $f_{ik}(\boldsymbol{\rho}_{-i})$ ($i \in \mathcal{C}$, $k = 1, 2, \ldots, \lambda_i$). For monotonicity, we prove that $f_{ik}(\boldsymbol{\rho}'_{-i}) \leq f_{ik}(\boldsymbol{\rho}_{-i})$ for $\boldsymbol{\rho}'_{-i} \leq \boldsymbol{\rho}_{-i}$ as follows. Given any non-negative $\boldsymbol{\rho}_{-i}$, we replace $\boldsymbol{\rho}_{-i}$ with $\boldsymbol{\rho}'_{-i}$. Note that $c_{ju}(\boldsymbol{\rho}'_{-i}) \geq c_{ju}(\boldsymbol{\rho}_{-i})$. Thus the replacement makes the solution space of (12d) larger, and the optimum with $\boldsymbol{\rho}'_{-i}$ is no larger than that with $\boldsymbol{\rho}_{-i}$. Therefore $f_{ik}(\boldsymbol{\rho}'_{-i}) \leq f_{ik}(\boldsymbol{\rho}_{-i})$. For scalability, we prove that $f_{ik}(\alpha\boldsymbol{\rho}_{-i}) \leq \alpha f_{ik}(\boldsymbol{\rho}_{-i})$ for $\alpha > 1$ and non-negative $\boldsymbol{\rho}_{-i}$ as follows. Denote the optimal solution of $f_{ik}(\boldsymbol{\rho}_{-i})$ by $\langle \rho''_i, \mathbf{q}''_i, \mathbf{x}''_i \rangle$. We have $f_{ik}(\boldsymbol{\rho}_{-i}) = \rho''_i$. Due to that $1/c_{ju}(\mathbf{q}_i, \boldsymbol{\rho}_{-i})$ and $1/c_j(\boldsymbol{\rho}_{-i})$ are strictly concave in $\boldsymbol{\rho}_{-i}$, the two inequalities

$$\frac{1}{c_j(\alpha\boldsymbol{\rho}_{-i})} < \frac{\alpha}{c_j(\boldsymbol{\rho}_{-i})}, \frac{1}{c_{ju}(\mathbf{q}_i, \alpha\boldsymbol{\rho}_{-i})} < \frac{\alpha}{c_{ju}(\mathbf{q}_i, \boldsymbol{\rho}_{-i})} \quad (29)$$

hold for $\alpha > 1$. Consider the following minimization problem 30, with $\mathbf{y}_i$ being fixed to pairing k.

$$\min_{\rho_i, \mathbf{q}_i, \mathbf{x}_i \geq \mathbf{0}} \rho_i \quad (30a)$$

$$\text{s.t. (12c), (12e) and (12f) of cell } i \quad (30b)$$

$$c_j(\boldsymbol{\rho}_{-i})x_j + \sum_{u \in \mathcal{V}_j} c_{ju}(\mathbf{q}, \boldsymbol{\rho}_{-i})x_u \geq \alpha d_j,$$

$$j \in \mathcal{J}_i \quad (30c)$$

Note that $\langle \alpha\rho''_i, \mathbf{q}''_i, \alpha\mathbf{x}''_i \rangle$ is feasible to (30), with the objective value being $\alpha f_{ik}(\boldsymbol{\rho}_{-i})$. Hence the optimum of (30) is no more than $\alpha f_{ik}(\boldsymbol{\rho}_{-i})$. For $f_{ik}(\alpha\boldsymbol{\rho}_{-i})$, note that the corresponding optimization problem only differs with (30) in (30c). Instead of (30c), in $f_{ik}(\alpha\boldsymbol{\rho}_{-i})$ we have:

$$c_j(\alpha\boldsymbol{\rho}_{-i})x_j + \sum_{u \in \mathcal{V}_j} c_{ju}(\mathbf{q}, \alpha\boldsymbol{\rho}_{-i}) \geq d_j, \ j \in \mathcal{J}_i \quad (31)$$

By Lemma 2, (30c) is equality at the optimum. Then by (29), for any solution of (30), using it for the optimization problem associated with $f_{ik}(\alpha\boldsymbol{\rho}_{-i})$ makes (31) an inequality. (This is because by (29) we obtain $c_j(\boldsymbol{\rho}_{-i})/\alpha < c_j(\alpha\boldsymbol{\rho}_{-i})$ and $c_{ju}(\mathbf{q}_i, \boldsymbol{\rho}_{-i})/\alpha < c_{ju}(\mathbf{q}_i, \alpha\boldsymbol{\rho}_{-i})$). Therefore the problem for $f_{ik}(\alpha\boldsymbol{\rho}_{-i})$ has a lower optimum than (30). Further, the optimum is lower than $\alpha f_{ik}(\boldsymbol{\rho}_{-i})$. Hence $f_{ik}(\alpha\boldsymbol{\rho}_{-i}) < \alpha f_{ik}(\boldsymbol{\rho}_{-i})$.

We then allow k to be variable and consider $f_i(\boldsymbol{\rho}_{-i})$ ($i \in \mathcal{C}$). For $\boldsymbol{\rho}'_{-i} \leq \boldsymbol{\rho}_{-i}$ we have

$$f_i(\boldsymbol{\rho}'_{-i}) = \min_k f_{ik}(\boldsymbol{\rho}'_{-i}) \leq \min_k f_{ik}(\boldsymbol{\rho}_{-i}) = f_i(\boldsymbol{\rho}_{-i})$$

and for $\alpha > 1$ we have

$$f_i(\alpha\boldsymbol{\rho}_{-i}) = \min_k f_{ik}(\alpha\boldsymbol{\rho}_{-i}) < \alpha \min_k f_{ik}(\boldsymbol{\rho}_{-i}) = \alpha f_i(\boldsymbol{\rho}_{-i})$$

Hence the conclusion. $\square$

Given $\boldsymbol{\rho}$, denote by $\mathbf{f}^k$ ($k > 1$) the function composition of $\mathbf{f}(\mathbf{f}^{k-1}(\boldsymbol{\rho}))$ (with $\mathbf{f}^0(\boldsymbol{\rho}) = \boldsymbol{\rho}$). Lemma 4 holds by [40].

Lemma 4: If $\lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho})$ exists, then it exists uniquely for any $\boldsymbol{\rho} \geq \mathbf{0}$.

B. Optimality and Feasibility

Based on Theorem 4, we derive sufficient-and-necessary conditions for MINF in terms of its feasibility and optimality. For any load $\boldsymbol{\rho}$, we say that a load $\boldsymbol{\rho}$ is achievable if and only

if there exist $\mathbf{q}$, $\mathbf{x}$, and $\mathbf{y}$ such that the solution $\langle \boldsymbol{\rho}, \mathbf{q}, \mathbf{x}, \mathbf{y} \rangle$ is feasible to MINF.

Lemma 5: For any $\boldsymbol{\rho} \geq \mathbf{0}$, if there exists $i \in \mathcal{C}$ such that $\rho_i < f_i(\boldsymbol{\rho}_{-i})$, then $\boldsymbol{\rho}$ is not achievable in MINF.

Proof: Let $\rho'_i = f_i(\boldsymbol{\rho}_{-i})$. By the definition of f_i, ρ'_i is the minimum value satisfying (12c)–(12i) under $\boldsymbol{\rho}_{-i}$. Therefore any ρ_i with $\rho_i < \rho'_i$ is not achievable with constraints (12c)–(12i). Hence the conclusion. $\square$

Theorem 5: In MINF, $\boldsymbol{\rho}$ ($\boldsymbol{\rho} \leq \bar{\boldsymbol{\rho}}$) is achievable if and only if $\mathbf{f}(\boldsymbol{\rho})$ is achievable and $\boldsymbol{\rho} \geq \mathbf{f}(\boldsymbol{\rho})$.

Proof: By the inverse proposition of Lemma 5, an achievable $\boldsymbol{\rho}$ always satisfies $\boldsymbol{\rho} \geq \mathbf{f}(\boldsymbol{\rho})$. The necessity is proved as follows. Suppose $\boldsymbol{\rho}$ is achievable for MINF. Consider using $\mathbf{f}(\boldsymbol{\rho})$ as another solution (together with the $\langle \mathbf{q}, \mathbf{x}, \mathbf{y} \rangle$ obtained when computing $\mathbf{f}(\boldsymbol{\rho})$). Then $\mathbf{f}(\boldsymbol{\rho})$ satisfies (12b). Also, $\mathbf{f}(\boldsymbol{\rho})$ together with its $\langle \mathbf{q}, \mathbf{x}, \mathbf{y} \rangle$ fulfills (12c)–(12i) by the definition of $\mathbf{f}(\boldsymbol{\rho})$. Thus, $\mathbf{f}(\boldsymbol{\rho})$ is achievable.

For the sufficiency, note that the achievability of $\mathbf{f}(\boldsymbol{\rho})$ implies that $\boldsymbol{\rho}$ along with $\langle \mathbf{q}, \mathbf{x}, \mathbf{y} \rangle$ obtained by solving $\mathbf{f}(\boldsymbol{\rho})$ satisfies (12c)–(12i). Combined with the precondition $\rho_i \leq \bar{\rho}$ ($i \in \mathcal{C}$), the load $\boldsymbol{\rho}$ is feasible to (12b)–(12i) (and thus achievable in MINF). Hence the conclusion. $\square$

Theorem 5 provides an effective method for improving any sub-optimal solution to MINF. *For any achievable $\boldsymbol{\rho}$, evaluating $\mathbf{f}(\boldsymbol{\rho})$ always yields a better solution.*[7] This conclusion is based on Theorem 5: Suppose $\boldsymbol{\rho}$ ($\boldsymbol{\rho} \geq \mathbf{0}$) is the current cell load, and let $\boldsymbol{\rho}'$ be the function value evaluated at $\boldsymbol{\rho}$, i.e. $\boldsymbol{\rho}' = \mathbf{f}(\boldsymbol{\rho})$. By Theorem 5, we always have $\boldsymbol{\rho}' \leq \boldsymbol{\rho}$.

Recall that $F(\boldsymbol{\rho})$ is the objective function of the problem MINF. Theorem 6 below states that, the fixed point of $\mathbf{f}(\boldsymbol{\rho})$ (along with $\langle \mathbf{q}, \mathbf{x}, \mathbf{y} \rangle$ obtained when computing $\mathbf{f}(\boldsymbol{\rho})$) is optimal to MINF.

Theorem 6: Load $\boldsymbol{\rho}^*$ is the optimum of MINF if (and only if when $F(\boldsymbol{\rho})$ is strictly monotonic) $\boldsymbol{\rho}^* = \mathbf{f}(\boldsymbol{\rho}^*) \leq \bar{\boldsymbol{\rho}}$.

Proof: (Necessity) If $\boldsymbol{\rho}^*$ is optimal (and thus feasible), then obviously we have $\boldsymbol{\rho}^* \leq \bar{\boldsymbol{\rho}}$. By Theorem 5, $\mathbf{f}(\boldsymbol{\rho}^*)$ is also feasible and $\mathbf{f}(\boldsymbol{\rho}^*) \leq \boldsymbol{\rho}^*$. By successively applying Theorem 5, $\mathbf{f}^k(\boldsymbol{\rho}^*)$ for any $k \geq 1$ is a feasible solution and $\mathbf{f}^k(\boldsymbol{\rho}^*) \leq \mathbf{f}^{k-1}(\boldsymbol{\rho}^*)$. Let $\boldsymbol{\rho}' = \lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho}^*)$. Then $\boldsymbol{\rho}' \leq \boldsymbol{\rho}^*$ holds by the above derivation. In addition, note that $\boldsymbol{\rho}'$ is a feasible solution as well. By that $\boldsymbol{\rho}^*$ is optimal for MINF, we have $\boldsymbol{\rho}' = \boldsymbol{\rho}^*$, otherwise $\boldsymbol{\rho}'$ would lead to a better objective value in MINF than $\boldsymbol{\rho}^*$. Hence $\boldsymbol{\rho}^* = \lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho}^*)$, i.e. $\boldsymbol{\rho}^* = \mathbf{f}(\boldsymbol{\rho}^*)$.

(Sufficiency) By Theorem 5, for any feasible $\boldsymbol{\rho}$, $\lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho})$ is feasible and $\lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho}) \leq \boldsymbol{\rho}$ holds. By Lemma 4, the limit remains for any $\boldsymbol{\rho} \geq \mathbf{0}$, and thus $\lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho}) = \lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho}^*)$. Since $\boldsymbol{\rho}^* = \mathbf{f}(\boldsymbol{\rho}^*)$, we have $\boldsymbol{\rho}^* = \lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho})$. Thus $\boldsymbol{\rho}^* \leq \boldsymbol{\rho}$ for any feasible $\boldsymbol{\rho}$, meaning that $\boldsymbol{\rho}^*$ is optimal for MINF.

Hence the conclusion. $\square$

C. The Algorithmic Framework

Starting from any non-negative $\boldsymbol{\rho}^{(0)}$, we compute $\lim_{k \to \infty} \mathbf{f}^k(\boldsymbol{\rho})$ iteratively. During each iteration, n problems

---

[7]Rigorously, Theorem 5 implies that the new solution is not worse. In fact it is guaranteed to be strictly better (with strictly monotonic $F(\boldsymbol{\rho})$) unless the old one is already optimal. A proof can be easily derived based on Theorem 6.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Cell radius	500 m
Carrier frequency	2 GHz
Total bandwidth	20 MHz
Cell load limit $\bar{\rho}$	1.0
Path loss model	COST-231-HATA
Shadowing (Log-normal)	6 dB standard deviation
Fading	Rayleigh flat fading
Noise power spectral density	-173 dBm/Hz
RB power p_i ($i \in \mathcal{C}$)	800 mW
Convergence tolerance (ϵ)	10^{-4}

in (13) for $i \in \mathcal{C}$ are solved. The convergence is guaranteed by Lemma 4. At the convergence, by Theorem 6, the optimum is reached. Note that once $\boldsymbol{\rho}^{(k)}$ is feasible for any $k \geq 0$, then by Theorem 5, all $\boldsymbol{\rho}^{(k+1)}, \boldsymbol{\rho}^{(k+2)}, \ldots$ are feasible as well. One can terminate prematurely to obtain a sub-optimal solution with less computation. M-CELL is outlined below.

M-CELL$(\boldsymbol{\rho}^{(0)}, \epsilon)$
1   $k = 0$
2   **repeat**
3       $k = k + 1$
4   **for** $i \in \mathcal{C}$
5           $\langle \rho_i^{(k)}, \mathbf{q}_i^{(k)}, \mathbf{x}_i^{(k)}, \mathbf{y}_i^{(k)} \rangle = \text{S-CELL}(\boldsymbol{\rho}_{-i}^{(k-1)})$
6   **end for**
7   **until** $\|\boldsymbol{\rho}^{(k)} - \boldsymbol{\rho}^{(k-1)}\|_\infty \leq \epsilon$
8   **if** $\rho_i^{(k)} > \bar{\rho}$ for any i ($i \in \mathcal{C}$)
9   MINF is infeasible
10 **end if**
11 **return** $\langle \boldsymbol{\rho}^{(k)}, \mathbf{q}^{(k)}, \mathbf{x}^{(k)}, \mathbf{y}^{(k)} \rangle$

M-CELL applies fixed point iterations using $\mathbf{f}(\boldsymbol{\rho})$. The convergence of fixed point iterations on $\mathbf{f}(\boldsymbol{\rho})$ is linear [44]. The feasibility check is done by Lines 8 and 9. The infeasibility of MINF implies that at least one cell will be overloaded for meeting user demands. If this happens, we know for sure that the user demands cannot be satisfied. We remark that all the conclusions derived in this section are independent of the implementation of S-CELL in Line 5. As long as the subroutine S-CELL yields the optimal solution to (13), M-CELL achieves the optimum of MINF.[8] Besides, M-CELL possesses the optimality for MINF with any objective function that is monotonically (but not necessarily strictly monotonic) increasing in each element of $\boldsymbol{\rho}$. These two properties make M-CELL an algorithmic framework. To our knowledge, the most efficient S-CELL is what we derived in Section V.

For a cell i ($i \in \mathcal{C}$), given the information of other cells' load $\boldsymbol{\rho}_{-i}$, solving $f_i(\boldsymbol{\rho})$ is based on local information, making M-CELL suitable to run in a distributed manner. A cell can maintain the information of a subset of cells (e.g.,

---

[8]With filtered $\mathcal{U}$, the proposed M-CELL is proved to converge to the global optimum of MINF. Without filtered $\mathcal{U}$ (or for an arbitrarily selected set of candidate pairs $\mathcal{U}$), M-CELL is still applicable to MINF though there is no theoretical guarantee of convergence or optimality, as the decoding order for each pair may change in the iteration process.
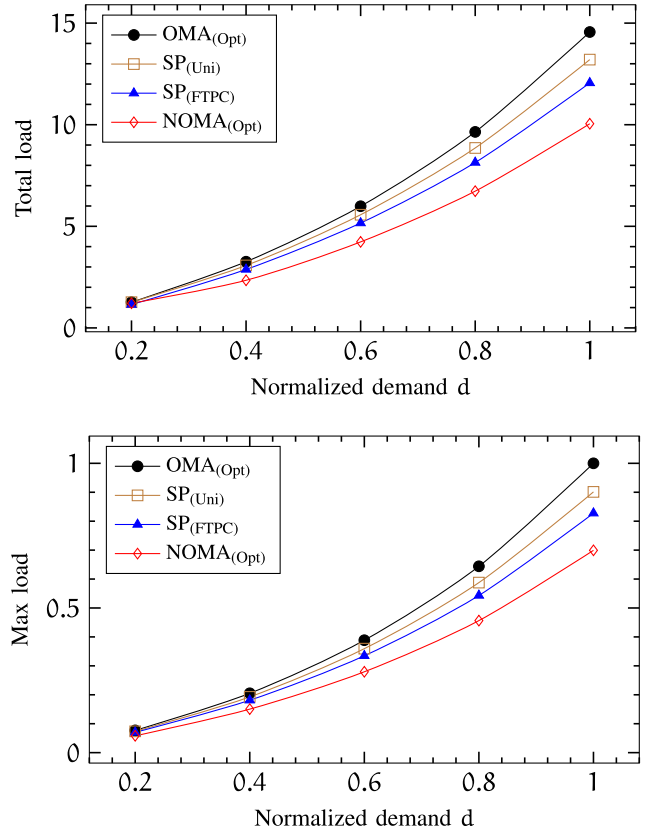

Fig. 4. This figure illustrates the total and maximum load in function of normalized demand. At $d = 1.0$, the network reaches its resource limit such that any larger demand cannot be satisfied by OMA$_{(\text{Opt})}$. SP$_{(\text{Uni})}$ and SP$_{(\text{FTPC})}$ are two sub-optimal NOMA power allocation schemes, for which pair selections are optimally computed.

the surrounding cells) having major significance in terms of interference, and exchange the information with other cells periodically, which can be implemented via the LTE X2 interface. The technique called "asynchronous fixed-point iterations" [40] can be used. The asynchronous fixed-point iterations converge to the fixed point that is the same as obtained by its synchronized version. Intuitively, the fixed point is unique, regardless of how we reach it.

VII. PERFORMANCE EVALUATION

We use a cellular network of 19 cells. To eliminate edge effects, wrap-around technique [45] is applied. Inside each cell, 30 UEs are randomly and uniformly distributed. In each cell, there are in total $\binom{30}{2} = 435$ possible choices for user pairing in NOMA. User demands are set to be a uniform value d. In the simulations, d is normalized by $M \times B$ in (8) and (9), and belongs to $(0, 1]$. With OMA, the network reaches the resource limit at $d = 1.0$, i.e., any $d > 1.0$ leads to at least one cell being overload in OMA. Other parameters are given in Table I.

We consider two objectives for performance evaluation: resource efficiency and load balancing. For resource efficiency, the objective function is $F(\boldsymbol{\rho}) = \sum_{i \in \mathcal{C}} \rho_i$, i.e., to minimize the total network time-frequency resource consumption (or cells' average resource consumption if divided by n). For load balancing, we adopt min-max fairness and the objective function is $F(\boldsymbol{\rho}) = \max_{i \in \mathcal{C}} \rho_i$. Section VII-A and
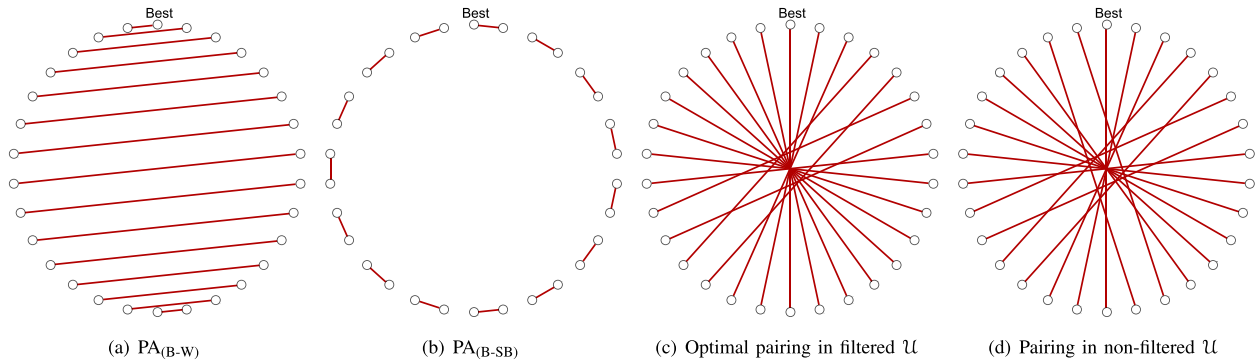
Fig. 5. This figure illustrates pair selection in a typical cell of 30 UEs. The UEs are represented by the vertices on a circle. The UE marked "Best" at the top position has the best channel condition. The UEs are arranged clock-wisely in the descending order of channel conditions. The edges are selected pairs. Each subfigure represents one pairing method. Figure 5(a) and Figure 5(b) show PA$_{(B-W)}$ and PA$_{(B-SB)}$, respectively. Figure 5(c) shows optimal pairing after filtered $\mathcal{U}$. Figure 5(d) shows the pairing solution obtained with non-filtered $\mathcal{U}$. Note that in all cases the candidate pairs are not related to inter-cell interference. The interference affects the subroutine S-CELL in M-CELL.
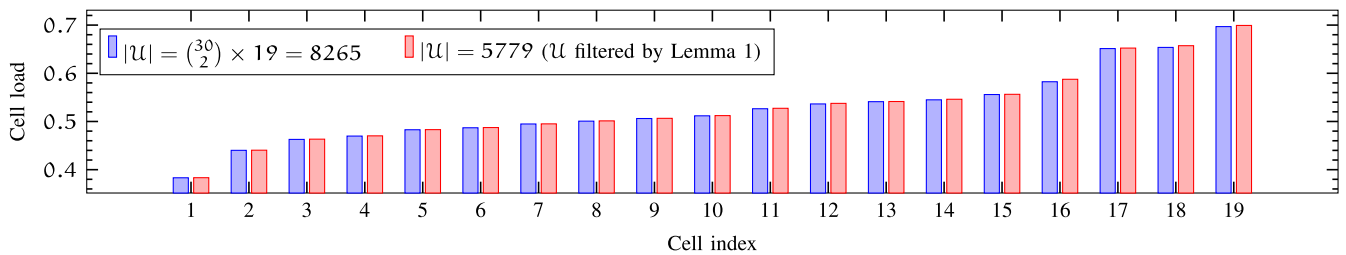
Fig. 6. This figure shows the optimized load levels of all 19 cells. The cells are numbered in an ascending order of load. The blue bars show the computed cell load under $\mathcal{U}$ composed of all $\binom{30}{2} \times 19$ pairs. The red bars show the minimum cell load with pairs satisfying Lemma 1.

Section VII-B provide results for power allocation and user pairing, respectively. The optimal OMA, named OMA$_{(Opt)}$, is obtained by fixing **y** to **0** in MINF and solving the remaining problem to optimality.[9] The proposed optimal NOMA solution is named NOMA$_{(Opt)}$ subsequentially.

A. Power Allocation

We use OMA$_{(Opt)}$ as baseline. As for NOMA, the pairing candidate set $\mathcal{U}$ initially covers all pairs of UEs in each cell. Then, those pairs not fulfilling Lemma 1 are dropped from $\mathcal{U}$. We then use M-CELL to compute NOMA$_{(Opt)}$. Besides the optimal NOMA, we implement two other sub-optimal NOMA power split schemes for comparison. One is named "SP$_{(Uni)}$", in which the power p_i splits equally between $q_{\oplus u}$ and $q_{\ominus u}$ for any pair $u = \{\oplus, \ominus\}$ ($u \in \mathcal{U}$). The other is "fractional transmit power control" (FTPC), named SP$_{(FTPC)}$, using a parameter to control the fairness for power split. We set this parameter to be 0.4 as recommended in [46]. Under both SP$_{(Uni)}$ and SP$_{(FTPC)}$, we use the method in Section V-B to compute the optimal pair selection. Both two power split schemes are easily accommodated by M-CELL.

Figure 4 shows the total load and the maximum load in function of normalized demand. As expected, the cell load levels monotonically increase with user demand. At high user demand, NOMA$_{(Opt)}$ dramatically improves the load performance. For $d = 1.0$, it achieves 31% better performance

than OMA$_{(Opt)}$ for both total load and maximum load. The two sub-optimal solutions SP$_{(Uni)}$ and SP$_{(FTPC)}$ also result in load improvement than OMA$_{(Opt)}$. Compared to the two sub-optimal solutions, the improvement achieved by NOMA$_{(Opt)}$ over OMA$_{(Opt)}$ is doubled or more. On average, by using the same amount of time-frequency resource, NOMA$_{(Opt)}$ delivers 33% more bits demand than OMA$_{(Opt)}$. Besides, SP$_{(FTPC)}$ achieves better performance than SP$_{(Uni)}$, as the former takes into account the channel conditions in power split. Generally, in SP$_{(FTPC)}$, UE with worse channel is allocated with more power.

In summary, power allocation has considerably large influence on NOMA. Even if the UE pairs are optimally selected, sub-optimal power allocations in NOMA have significant deviation from optimal NOMA.

B. User Pairing

We study the influence of user pairing by considering two sub-optimal ones [4], named "PA$_{(B-W)}$" and "PA$_{(B-SB)}$", respectively. Suppose we sort the UEs in descending order of their channel conditions. In PA$_{(B-W)}$, the UE with the best channel condition is paired with the UE with the worst, and the UE with the second best is paired with one with the second worst, and so on. In PA$_{(B-SB)}$, the UE with the best channel condition is paired with the one with the second best, and so on. See Figures 5(a) and 5(b) for an illustration. In addition, we examine to what extend pair filtering (by Lemma 1) affects performance. For filtered $\mathcal{U}$, optimal pair selection is done by Section V-B. For non-filtered $\mathcal{U}$, we apply M-CELL even though there is no theoretical guarantee on optimality. Convergence, however, is observed for all the
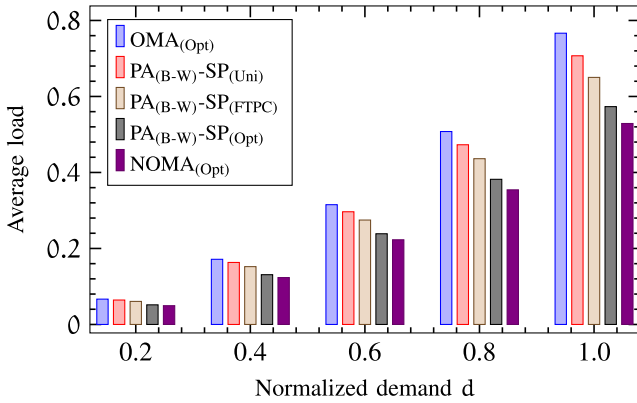
---

[9]With **y** being fixed to **0** in MINF, the variables **q** and **x** disappear. Then we modify Line 5 of M-CELL to be "$\rho_i^{(k)} = \sum_{j \in \mathcal{I}_i} d_j / c_j (\boldsymbol{\rho}_{-i}^{(k-1)})$" and Line 11 to be "**return** $\boldsymbol{\rho}^{(k)}$". The modified M-CELL gives the optimal load for OMA (see [17] for further details).

Fig. 7. This figure evaluates PA$_{(B-W)}$, combined with three power split schemes, SP$_{(Uni)}$, SP$_{(FTPC)}$, and SP$_{(Opt)}$. In SP$_{(Opt)}$, the power split is optimal for each pair. OMA$_{(Opt)}$ and NOMA$_{(Opt)}$ are baselines.
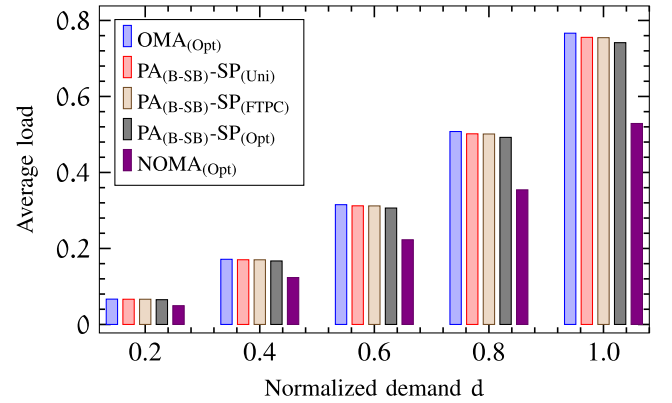

Fig. 8. This figure evaluates PA$_{(B-SB)}$, combined with three power split schemes, SP$_{(Uni)}$, SP$_{(FTPC)}$, and SP$_{(Opt)}$. In SP$_{(Opt)}$, the power split is optimal for each pair. OMA$_{(Opt)}$ and NOMA$_{(Opt)}$ are baselines.

instances we considered. Figures 5(c) and 5(d) illustrated the resulted selection patterns.

In Figure 6, we show the load levels of all 19 cells with $d = 1.0$, under both filtered and non-filtered $\mathcal{U}$. In this specific scenario, $|\mathcal{U}|$ is reduced from $\binom{30}{2} \times 19 = 8265$ to 5779 after being filtered by Lemma 1. We choose $d = 1.0$ because the performance difference among the solutions is the largest. There is very slight difference in cell load levels between the two cases. Numerically, the differences between them are only 0.1% and 0.5% for average and maximum cell load, respectively. This result is coherent with Figure 5(c) and Figure 5(d). One can see that the patterns of the two pair selection solutions are almost identical. Thus, pair filtering by Lemma 1 is effective in reducing the number of candidate pairs, with virtually no impact on performance.

In Figure 7 and Figure 8, we respectively evaluate PA$_{(B-W)}$ and PA$_{(B-SB)}$, combined with three power split schemes SP$_{(Uni)}$, SP$_{(FTPC)}$, and SP$_{(Opt)}$. In SP$_{(Opt)}$, we use the algorithm SPLIT to compute the optimal power split for each pair. All of SP$_{(Uni)}$, SP$_{(FTPC)}$, and SP$_{(Opt)}$ are put into the framework of M-CELL but with fixed pair selection PA$_{(B-W)}$ or PA$_{(B-SB)}$. In addition, OMA$_{(Opt)}$ and NOMA$_{(Opt)}$ are also included for comparison as baselines.

One can see that all the NOMA schemes outperform OMA$_{(Opt)}$. In Figure 7, with PA$_{(B-W)}$, SP$_{(FTPC)}$ outperforms SP$_{(Uni)}$. SP$_{(Opt)}$ beats the other two. On one hand, there is non-negligible gap in load performance between SP$_{(Opt)}$ and NOMA$_{(Opt)}$, even though in SP$_{(Opt)}$, the power split is optimal for the PA$_{(B-W)}$ pairing. Hence pair selection plays an important role for NOMA performance. On the other hand, SP$_{(Opt)}$ yields significantly load improvement compared to OMA$_{(Opt)}$, and we conclude that PA$_{(B-W)}$ is a good sub-optimal pair selection for NOMA. Indeed, PA$_{(B-W)}$ pairs the UEs in a greedy way, aiming at maximizing the diversity of channel conditions of paired UEs. As shown in Figure 5(c), the optimal pair selection has a similar trend. The difference is that optimal pairing has a more "global view" than PA$_{(B-W)}$. In Figure 8, under PA$_{(B-SB)}$, SP$_{(Uni)}$, SP$_{(FTPC)}$, and SP$_{(Opt)}$ improve the load very slightly. All of the three are far from the global optimum and the gap is large under high user demands. We conclude
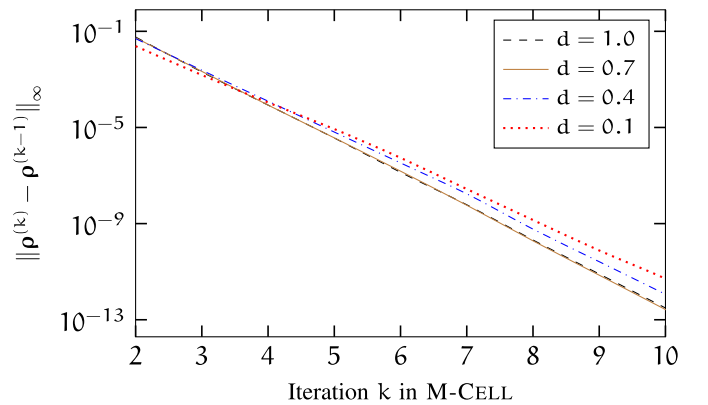

Fig. 9. This figure shows the norm $\| \cdot \|_\infty$ in function of iteration k in M-CELL, under the uniform demands 0.1, 0.4, 0.7, and 1.0, respectively.

that PA$_{(B-SB)}$ is not as effective as PA$_{(B-W)}$ in terms of network load optimization.

As the overall conclusion, jointly optimizing power allocation and user pairing is important for the performance of NOMA.

C. Convergence Analysis

We show the convergence performance of M-CELL in Figure 9, for demands 0.3, 0.5, 0.7, and 1.0, respectively. Initially, $\rho_i^{(0)} = 1$ ($i \in \mathcal{C}$). We observe that M-CELL converges very fast. With higher demand, the convergence becomes slightly faster. High accuracy is reached after a very few iterations. For all the demands consider in the figure, even if we terminate M-CELL after a very few iterations, the obtained solution is close to the optimum.

VIII. CONCLUSIONS

This paper has investigated optimal resource management in multi-cell NOMA, with power allocation and user pairing being considered simultaneously. Joint optimization of both is shown to be very important for NOMA performance. The proposed system model admits a mixed use of OMA and NOMA for the users. Therefore, network architectures that support various multiple access techniques can be analyzed under this model. Finally, as for future work, the paper suggests that

$$\mathrm{Cv}_u(c_{\oplus u}, c_{\ominus u}, \boldsymbol{\rho}_{-i}) = \log\left[\frac{w_\oplus(\boldsymbol{\rho}_{-i})e^{(c_{\oplus u}+c_{\ominus u})} + (w_\ominus(\boldsymbol{\rho}_{-i}) - w_\oplus(\boldsymbol{\rho}_{-i}))e^{c_{\ominus u}}}{p_i + w_\ominus(\boldsymbol{\rho}_{-i})}\right]. \tag{32}$$

$$H_{\oplus u}\left(x_u, \boldsymbol{\rho}_{-i}\right) = \begin{cases} d_\oplus/x_u & c_{\oplus u}^K x_u^K > d_\oplus \\ \log\left[\dfrac{(p_i + w_\ominus(\boldsymbol{\rho}_{-i}))/e^{d_\oplus/x_u} + w_\oplus(\boldsymbol{\rho}_{-i}) - w_\ominus(\boldsymbol{\rho}_{-i})}{w_\oplus(\boldsymbol{\rho}_{-i})}\right] & \text{Otherwise} \end{cases} \tag{33}$$

$$H_{\ominus u}\left(x_u, \boldsymbol{\rho}_{-i}\right) = \begin{cases} d_\ominus/x_u & c_{\ominus u}^K x_u^K > d_\ominus \\ \log\left[\dfrac{p_i + w_\ominus(\boldsymbol{\rho}_{-i})}{w_\oplus(\boldsymbol{\rho}_{-i})e^{d_\ominus/x_u} + w_\ominus(\boldsymbol{\rho}_{-i}) - w_\oplus(\boldsymbol{\rho}_{-i})}\right] & \text{Otherwise} \end{cases} \tag{34}$$

mathematical tools in SIF are useful for analyzing multi-cell NOMA. In summary, NOMA is a promising technique for spectrum efficiency enhancement and cell load balancing.

Appendix

We remark that the function (32), as shown at the top of this page, is first referred to in Section V-A and is used in the formulation (17). The two functions (33) and (34), as shown at the top of this page, are first referred to in Section V-A and are used in the algorithm SPLIT.

References

[1] L. You, L. Lei, D. Yuan, S. Sun, S. Chatzinotas, and B. Ottersten, "A framework for optimizing multi-cell NOMA: Delivering demand with less resource," in *Proc. IEEE GLOBECOM*, Dec. 2017, pp. 1–7.

[2] 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, Sep. 2017.

[3] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor. (Nov. 2016). "Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges." [Online]. Available: https://arxiv.org/abs/1611.01607

[4] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[5] *Evaluation Methodologies for Downlink Multiuser Superposition Transmissions*, document R1-153332, 3GPP, 2014.

[6] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 84–87, Jan. 2017.

[7] J. Kim, J. Koh, J. Kang, K. Lee, and J. Kang, "Design of user clustering and precoding for downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE MILCOM*, Oct. 2015, pp. 1170–1175.

[8] H. Tabassum, E. Hossain, and M. J. Hossain. (Oct. 2016). "Modeling and analysis of uplink non-orthogonal multiple access (NOMA) in large-scale cellular networks using Poisson cluster processes." [Online]. Available: https://arxiv.org/abs/1610.06995

[9] 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.

[10] L. P. Qian, Y. Wu, H. Zhou, and X. Shen, "Joint uplink base station association and power control for small-cell networks with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5567–5582, Sep. 2017.

[11] B. Di, L. Song, Y. Li, and G. Y. Li, "Non-orthogonal multiple access for high-reliable and low-latency V2X communications in 5G systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2383–2397, Oct. 2017.

[12] L. P. Qian, Y. Wu, H. Zhou, and X. Shen, "Dynamic cell association for non-orthogonal multiple-access V2S networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2342–2356, Oct. 2017.

[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] L. You and D. Yuan, "Joint CoMP-cell selection and resource allocation in fronthaul-constrained C-RAN," in *Proc. WiOpt Workshop*, May 2017, pp. 1–6.

[15] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Optimal cell clustering and activation for energy saving in load-coupled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6150–6163, Nov. 2015.

[16] I. Viering, M. Dottling, and A. Lobinger, "A mathematical perspective of self-optimizing wireless networks," in *Proc. IEEE ICC*, Jun. 2009, pp. 1–6.

[17] I. Siomina and D. Yuan, "Analysis of cell load coupling for LTE network planning and optimization," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2287–2297, Jun. 2012.

[18] A. J. Fehske, I. Viering, J. Voigt, C. Sartori, S. Redana, and G. P. Fettweis, "Small-cell self-organizing wireless networks," *Proc. IEEE*, vol. 102, no. 3, pp. 334–350, Mar. 2014.

[19] L. You, D. Yuan, N. Pappas, and P. Värbrand, "Energy-aware wireless relay selection in load-coupled OFDMA cellular networks," *IEEE Commun. Letters*, vol. 21, no. 1, pp. 144–147, Jan. 2017.

[20] L. You and D. Yuan, "Load optimization with user association in cooperative and load-coupled LTE networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3218–3231, Mar. 2017.

[21] I. Siomina, A. Furuskär, and G. Fodor, "A mathematical framework for statistical QoS and capacity studies in OFDM networks," in *Proc. IEEE PIMRC*, Sep. 2009, pp. 2772–2776.

[22] K. Majewski and M. Koonert, "Conservative cell load approximation for radio networks with Shannon channels and its application to LTE network planning," in *Proc. 6th Adv. Int. Conf. Telecommun.*, May 2010, pp. 219–225.

[23] E. Pollakis, R. L. G. Cavalcante, and S. Stańczak, "Base station selection for energy efficient network operation with the majorization-minimization algorithm," in *Proc. IEEE SPAWC*, Jun. 2012, pp. 219–223.

[24] A. J. Fehske and G. P. Fettweis, "Aggregation of variables in load models for interference-coupled cellular data networks," in *Proc. IEEE ICC*, Jun. 2012, pp. 5102–5107.

[25] A. J. Fehske, H. Klessig, J. Voigt, and G. P. Fettweis, "Concurrent load-aware adjustment of user association and antenna tilts in self-organizing radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1974–1988, Jun. 2013.

[26] C. K. Ho, D. Yuan, and S. Sun, "Data offloading in load coupled networks: A utility maximization framework," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1921–1931, Apr. 2014.

[27] R. L. G. Cavalcante, S. Stanczak, M. Schubert, A. Eisenblaetter, and U. Tuerke, "Toward energy-efficient 5G wireless communications technologies: Tools for decoupling the scaling of networks from the growth of operating power," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 24–34, Nov. 2014.

[28] S. Tombaz, S. W. Han, K. W. Sung, and J. Zander, "Energy efficient network deployment with cell DTX," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 977–980, Jun. 2014.

[29] B. Błaszczyszyn, M. Jovanovic, and M. K. Karray, "Performance laws of large heterogeneous cellular networks," in *Proc. WiOpt*, May 2015, pp. 597–604.

[30] C. K. Ho, D. Yuan, L. Lei, and S. Sun, "Power and load coupling in cellular networks for energy optimization," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 509–519, Jan. 2015.

[31] R. L. G. Cavalcante, S. Stańczak, J. Zhang, and H. Zhuang, "Low complexity iterative algorithms for power estimation in ultra-dense load coupled networks," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 6058–6070, Nov. 2016.

[32] H. Klessig, D. Öhmann, A. J. Fehske, and G. P. Fettweis, "A performance evaluation framework for interference-coupled cellular data networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 938–950, Feb. 2016.

[33] R. L. G. Cavalcante, Y. Shen, and S. Stańczak, "Elementary properties of positive concave mappings with applications to network planning and optimization," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1774–1783, Apr. 2016.

[34] Q. Liao, "Dynamic uplink/downlink resource management in flexible duplex-enabled wireless networks," in *Proc. ICC Workshops*, May 2017, pp. 625–631.

[35] R. L. G. Cavalcante, M. Kasparick, and S. Stańczak, "Max-min utility optimization in load coupled interference networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 705–716, Feb. 2017.

[36] D. A. Awan, R. L. G. Cavalcante, and S. Stanczak. (Oct. 2017). "A robust machine learning method for cell-load approximation in wireless networks." [Online]. Available: https://arxiv.org/abs/1710.09318

[37] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[38] G. Geraci, M. Wildemeersch, and T. Q. S. Quek, "Energy efficiency of distributed signal processing in wireless networks: A cross-layer analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 1034–1047, Feb. 2016.

[39] M. Wildemeersch, T. Q. S. Quek, M. Kountouris, A. Rabbachin, and C. H. Slump, "Successive interference cancellation in heterogeneous networks," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4440–4453, Dec. 2014.

[40] 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.

[41] B. Di, L. Song, and Y. Li, "Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7686–7698, Nov. 2016.

[42] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8580–8594, Dec. 2016.

[43] S. Micali and V. V. Vazirani, "An $O(v|v|\ c\ |E|)$ algorithm for finding maximum matching in general graphs," in *Proc. 21st Annu. Symp. Found. Comput. Sci.*, Oct. 1980, pp. 17–27.

[44] H. R. Feyzmahdavian, M. Johansson, and T. Charalambous, "Contractive interference functions and rates of convergence of distributed power control laws," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4494–4502, Dec. 2012.

[45] D. Huo, "Clarification on the wrap-around hexagon network structure," Lucent Technol., Whippany, NJ, USA, Tech. Rep. C802.20-05/15, 2005.

[46] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE PIMRC*, Sep. 2013, pp. 611–615.

Di Yuan (M'03–SM'15) received the M.Sc. degree in computer science and engineering and the Ph.D. degree in optimization from the Linköping Institute of Technology in 1996 and 2001, respectively. After his Ph.D., he was an Associate Professor and then a Full Professor with the Department of Science and Technology, Linköping University, Sweden. He was a Guest Professor with the Technical University of Milan (Politecnico di Milano), Italy, in 2008, and a Senior Visiting Scientist with Ranplan Wireless Network Design Ltd, U.K., in 2009 and 2012. In 2011 and 2013 he was part time with Ericsson Research, Sweden. In 2014 and 2015, he was a Visiting Professor with the University of Maryland, College Park, MD, USA. In 2016, he joined Uppsala University, Sweden, as a Chair Professor. His current research mainly addresses network optimization of 4G and 5G systems and capacity optimization of wireless networks. He has served on the management committees of four European Cooperations in field of Scientific and Technical Research (COST) actions, an invited lecturer of the European Network of Excellence EuroNF, and a principal investigator of several European FP7 and Horizon 2020 projects. He was a co-recipient of the IEEE ICC 2012 Best Paper Award, and a supervisor of the Best Student Journal Paper Award by the IEEE Sweden Joint VT-COM-IT Chapter in 2014. He is an Area Editor of *Computer Networks*.

Lei Lei (S'12–M'17) received the B.Eng. and M.Eng. degrees from Northwestern Polytechnic University, Xi'an, China, in 2008 and 2011, respectively, and the Ph.D. degree from the Department of Science and Technology, Linköping University, Sweden, in 2016. He was a Research Assistant with the Institute for Infocomm Research, Agency for Science, Technology, and Research, Singapore, in 2013. Since 2016, he has been a Research Associate with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. His current research interests include resource allocation and optimization in 5G-satellite networks, wireless caching, energy-efficient communications, and machine learning in wireless communications. He received the IEEE Sweden Vehicular Technology-Communications-Information Theory Joint Chapter Best Student Journal Paper Award in 2014.

Lei You (S'15) received the B.Eng. and M.Eng. degrees from Qingdao University, China, in 2012 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Department of Information Technology, Uppsala University, Sweden. He was a Visiting Researcher with Ranplan Wireless Network Design Ltd., U.K., from 2015 to 2016, and with Converge ICT, Athens, in 2016. He served as a Project Manager for the EU Horizon 2020 Marie-Sklodowska Curie Project DECADE from 2015 to 2016. His current research interests include mathematical optimization and machine learning in the domain of mobile communications and networking.

Sumei Sun (F'16) is currently the Head of the Communications and Networks Cluster, and the Lead Principal Investigator of the Industrial Internet of Things Research Program with the Institute for Infocomm Research, Agency for Science, Technology, and Research, Singapore. Her current research interests are in smart communications and networks, next-generation machine-type communications, and Industrial Internet of Things. She has been serving as a member of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS Steering Committee since 2018. She has been the Vice Director of the IEEE Communications Society Asia Pacific Board since 2016. She has contributed to the IEEE conferences in different roles, e.g., as a member of the IEEE Communications Society Globecom/ICC Management and Strategy Standing Committee, the Executive Vice Chair of Globecom 2017, and the Track Co-Chair of the IEEE VTC 2017 Spring. She has been serving as an Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS since 2015 and an Area Editor of the IEEE TVT since 2017. She is a Distinguished Lecturer of the IEEE Vehicular Technology Society from 2014 to 2018.

Symeon Chatzinotas (S'06–M'09–SM'13) received the M.Eng. degree in telecommunications from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003, and the M.Sc. and Ph.D. degrees in electronic engineering from the University of Surrey, Surrey, U.K., in 2006 and 2009, respectively. He is currently the Deputy Head of the SIGCOM Research Group, Interdisciplinary Centre for Security, Reliability, and Trust, University of Luxembourg, Luxembourg, and a Visiting Professor with the University of Parma, Italy. He was involved in numerous research and development projects for the National Center for Scientific Research Demokritos, Institute of Informatics Telecommunications, the Center of Research and Technology Hellas, Institute of Telematics and Informatics, and the Mobile Communications Research Group, Center of Communication Systems Research, University of Surrey. He has over 300 publications, 2500 citations, and an H-Index of 27 according to Google Scholar. His research interests include multiuser information theory, co-operative/cognitive communications, and wireless networks optimization. He was a co-recipient of the 2014 Distinguished Contributions to Satellite Communications Award, and the Satellite and Space Communications Technical Committee, the IEEE Communications Society, and the CROWNCOM 2015 Best Paper Award.

Björn Ottersten (S'87–M'89–SM'99-F'04) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. He has held research positions with the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg. In 1991, he was appointed as a Professor of Signal Processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. From 1992 to 2004, he was the Head of the Department for Signals, Sensors, and Systems, KTH. From 1996 to 1997, he was the Director of Research with ArrayComm, Inc., a start-up in San Jose, CA, USA, based on his patented technology. From 2004 to 2008, he was the Dean of the School of Electrical Engineering, KTH. He is currently the Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. As Digital Champion of Luxembourg, he acts as an adviser to the European Commission.

He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, and 2013, and he has received seven other IEEE conference papers best paper awards. He is a fellow of EURASIP. He was a recipient of the IEEE Signal Processing Society Technical Achievement Award in 2011 and the European Research Council Advanced Research Grant twice, from 2009 to 2013 and again from 2017 to 2022. He has served as the Editor-in-Chief for *EURASIP Signal Processing Journal*, an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and as a member of the Editorial Board of the *IEEE Signal Processing Magazine*. He is currently a member of the Editorial Boards of the *EURASIP Journal on Advances in Signal Processing* and *Foundations and Trends® in Signal Processing*.