A Framework for Optimizing Multi-cell NOMA: Delivering Demand with Less Resource

Lei You[1], Lei Lei[2], Di Yuan[1], Sumei Sun[3], Symeon Chatzinotas[2], and Björn Ottersten[2]

[1]Department of Information Technology, Uppsala University, Sweden
[2]Interdisciplinary Centre for Security, Reliability and Trust, Luxembourg University, Luxembourg
[3]Institute for Infocomm Research, A*STAR, Singapore
Emails: {lei.you; di.yuan}@it.uu.se, {lei.lei; symeon.chatzinotas; bjorn.ottersten}@uni.lu, sunsm@i2r.a-star.edu.sg

***Abstract*—Non-orthogonal multiple access (NOMA) allows multiple users to simultaneously access the same time-frequency resource by using superposition coding and successive interference cancellation (SIC). Thus far, most papers on NOMA have focused on performance gain for one or sometimes two base stations. In this paper, we study multi-cell NOMA and provide a general framework for user clustering and power allocation, taking into account inter-cell interference, for optimizing resource allocation of NOMA in multi-cell networks of arbitrary topology. We provide a series of theoretical analysis, to algorithmically enable optimization approaches. The resulting algorithmic notion is very general. Namely, we prove that for any performance metric that monotonically increases in the cells' resource consumption, we have convergence guarantee for global optimum. We apply the framework with its algorithmic concept to a multi-cell scenario to demonstrate the gain of NOMA in achieving significantly higher efficiency.**

I. Introduction

To what extent can non-orthogonal multiple access (NOMA) improve network resource efficiency? In two recent surveys [1], [2], the authors pointed out that resource allocation in multi-cell NOMA poses 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 [1], [2] have addressed two-cell scenarios. To the best of our knowledge, enhancing network resource efficiency in multi-cell NOMA with user pairing has not been addressed yet. In [3], the authors proposed two interference alignment based coordinated beamforming methods in two-cell scenarios. Reference [4] uses stochastic geometry to model the inter-cell interference in NOMA. The crucial aspect of multi-cell NOMA consists of capturing mutual influence among cells. In the past decade, a modeling approach that characterizes the inter-cell interference via capturing the mutual influence among the load of cells, i.e., *load coupling*, had been proposed and widely adopted for orthogonal multiple access (OMA) networks [5]–[10]. However, this approach does not apply to NOMA, because time-frequency resource sharing via SIC is not allowed. Whether or not the type of model in OMA can be extended to NOMA has remained open until now.

The main contribution of this paper is that, *we provide a general framework for obtaining optimal user clustering and power allocation in interference-coupled multi-cell NOMA for resource efficiency.* More specifically,

1) We make a significant generalization for the interference models being used in [5]–[10] to multi-cell NOMA. Theoretical analysis in terms of feasibility and computation are provided.
2) Based on 1), a unified optimization framework for jointly optimizing user clustering and power allocation in multi-cell NOMA is derived, to achieve global optimum for *any* performance metric that monotonically increases in the cells' resource consumption.
3) We demonstrate the gain of NOMA in multi-cell scenarios and show NOMA is indeed a promising solution for meeting user demands with less resource than OMA.

II. A Multi-cell Interference Averaging Model

A. Network Model

We consider downlink, and remark that the framework can be straightforwardly extended to uplink. Denote by $\mathcal{I} = \{1, 2, \ldots, n\}$ and $\mathcal{J} = \{n+1, n+2, \ldots, n+m\}$ the sets of cells and UEs, respectively. Denote by $\mathcal{J}_i$ the set of UEs served by cell i, with $i \in \mathcal{I}$. Denote by g_{ij} the path loss factor from cell i to UE j, with $i \in \mathcal{I}$ and $j \in \mathcal{J}$. When using j to refer to one UE in $\mathcal{J}$, i by default indicates j's serving cell, unless stated otherwise.

We use u as a generic notation for UE *cluster* (referred to as "cluster" in the remaining context), i.e. a set consisting of one or multiple users that are allowed to access the same time-frequency resource by SIC. With more users being put in a cluster, the complexity for decoding in SIC grows fast [1], [2]. For the sake of this practical consideration, we follow the assumption used in other references [1]–[3], [11]–[13] that up to two users are clustered together[1]. If there is a need to differentiate between clusters, we put indices on u, e.g., $u_1, u_2, u_3, \ldots$. For UE clustering in cell i, denote by $\mathcal{U}_i$ the set of candidate clusters. We have $\mathcal{U}_i \cap \mathcal{U}_k = \phi$ for any $i \neq k$ with $i, k \in \mathcal{I}$. Similarly, denote by $\mathcal{U}_j$ ($j \in \mathcal{J}$) the set of clusters containing UE j. Let $\mathcal{U} = \bigcup_{i \in \mathcal{I}} \mathcal{U}_i$ (or equivalently $\mathcal{U} = \bigcup_{j \in \mathcal{J}} \mathcal{U}_j$) be the set of all clusters. Note that one UE may belong to multiple user clusters, e.g. $u_1 = \{1, 2\}$, $u_2 = \{1, 3\}$

[1] Reference [11] demonstrated most of the possible performance improvement can be achieved by two-users clustering in NOMA.

with UE 1 belonging to both u_1 and u_2. To keep the generality of our model for extreme case (e.g. there is only one UE in a cell), a cluster may consist of a single UE, i.e. $u_1 = \{1\}$ and $u_2 = \{2\}$.

The time-frequency domain resource that is divided into resource blocks (RBs). Let p_i be the transmission power on any RB in cell i. For any cluster $u = \{j, h\}$ in cell i, RB(s) can be accessed together (i.e., shared) by UEs j and h. On any of the shared RBs, *power splitting* is done on p_i, with p_{ju} and p_{hu} allocated to j and h, respectively, and $p_{ju} + p_{hu} = p_i$. On one RB, for any UE j and any cluster u ($j \in u$), the signal-to-interference and noise ratio (SINR) is computed by:

$$\gamma_{ju} = \frac{p_{ju} g_{ij}}{\underbrace{\sum_{\substack{h \in u: \\ b_u(h) < b_u(j)}} p_{hu} g_{ij}}_{\text{intra-cell}} + \underbrace{\sum_{k \in \mathcal{I} \backslash \{i\}} I_{kj}}_{\text{inter-cell}} + \sigma^2} \tag{1}$$

In (1), σ^2 is the noise power. Parameter I_{kj} refers to the interference from cell k to UE j. In [14] (Chapter 6.2.2, pp. 238) it is shown that, with superposition coding, a user can decode the data of another user with worse channel gain. The user with worse channel condition is subject to intra-cell interference. We use bijection $b_u(j) \rightarrow \{1, 2\}$ ($u \in \mathcal{U}$ and $j \in u$) to represent the decoding order. Based on the bijection, the UE with value 1 decodes the UE with value 2, and the UE with value 2 receives intra-cell interference from the UE with value 1. The decoding order is not constrained by power splitting [14], even though by our numerical results, more power is always allocated to the one with worse channel. The decoding order is fixed. The issue of the influence of inter-cell interference on the decoding order is addressed later in Section II-C.

B. Multi-cell Interference Modeling

Interference modeling based on considering the amount of resource consumption has been widely used for OMA. The method is specified as follows. Denote by ρ_k the proportion of RBs that are allocated for serving UEs in cell k. If cell k is fully loaded, meaning that all RBs are allocated, then $\rho_k = 1$. Another extreme case is that cell k is idle within the time interval in question, and accordingly $\rho_k = 0$. For the two cases, consider any UE j served by cell i. The exact interference j receives from cell k is $I_{kj} = p_k g_{kj}$ and $I_{kj} = 0$, respectively. For the former case, cell k interferes with every RB in cell i. For the latter, no interference is caused by cell k, as none of the RBs in cell k are active when $\rho_k = 0$. For $0 < \rho_k < 1$, a balance is stroked between exactness and simplicity by averaging on the interference within the time-frequency domain, see (2). This interference averaging technique was used in [5]–[10].

$$I_{kj} = p_k g_{kj} \rho_k \tag{2}$$

Intuitively, ρ_k reflects the likelihood that a UE outside cell k receives interference from k. By the definition of ρ_k, it can be interpreted as the *load of cell* k, and *used for measuring the time-frequency resource consumption* of cell k. An explanation of (2) is that, the inter-cell interference incurred by a cell is directly proportional to the cell's load, which has a direct correlation to the number of served UEs and the intensity of the cell's data traffic.

C. Decoding

The modeling complexity increases significantly for NOMA because inter-cell interference influences the decoding order. The modeling task is approached by identifying those clusters of which the decoding orders are decoupled from the inter-cell interference. UEs fulfilling Lemma 1 below are theoretically guaranteed to be independent of the inter-cell interference in respect of decoding in SIC. The proof of Lemma 1 is in the Appendix. Clusters consisting of UEs violating Lemma 1 are excluded from $\mathcal{U}$ and the model complexity is thus significantly reduced.

Lemma 1. *Suppose two users j and h within cluster u are served by cell i ($g_{ij} > g_{ih}$). If $g_{ij}/g_{ih} \geqslant g_{kj}/g_{kh}$ for all $k \in \mathcal{I} \backslash \{i\}$, then $b_u(j) = 1$ and $b_u(h) = 2$.*

In practical consideration, Lemma 1 reduces the user clustering complexity without damaging the performance. As pointed out by [15], [16], the large scale path-loss is a practically reasonable factor for ranking the decoding order. As for user clustering in NOMA, two users with disparate channels from the cell are preferred to be clustered for achieving good performance [11], [13]. Consider a cluster $u = \{j, h\}$ of cell i. If $g_{ij} \gg g_{ih}$, then most likely $g_{ij}/g_{ih} > g_{kj}/g_{kh}$ for $k \in \mathcal{I} \backslash \{i\}$, as the large scale path loss from other cells, tends not to differ as much as from the serving cell i in this case.

D. Cell Load Computation with User Clustering

Denote by d_j the bit demand of UE j with $j \in \mathcal{J}$. Let B be the spectral bandwidth on each RB. Denote by M the total number of RBs. Since the term $B \log(1 + \gamma_{ju})$ represents the capacity of one RB for UE j in cluster u ($j \in u$), the total achievable capacity on all RBs with respect to j and u is computed by

$$c_{ju} = MB \log\left(1 + \gamma_{ju}\right). \tag{3}$$

Denote by x_u the proportion of allocated RB(s) to cluster u. The sum of x_u for $u \in \mathcal{U}_i$ equals the load of cell i, as shown by (4), where $\bar{\rho}$ represents the load limit.

$$\rho_i = \sum_{u \in \mathcal{U}_i} x_u \leqslant \bar{\rho} \tag{4}$$

Note that the term $c_{ju} x_u$ computes the achieved bits for UE j in cluster u with allocated proportion of time-frequency resource x_u. To satisfy the quality-of-service (QoS) requirement, we have for $j \in \mathcal{J}$:

$$\sum_{u \in \mathcal{U}_j} c_{ju} x_u \geqslant d_j. \tag{5}$$

Given $\mathbf{d} = [d_1, d_2, \ldots, d_m]$, the inequalities system (1)–(5) forms a region for $\mathbf{x} = [x_1, x_2, \ldots, x_{|\mathcal{U}|}]$. Within this region, the QoS can be satisfied with the available network resource. Note that the system is non-linear, as ρ_k appears in the logarithm term in (3). The user clustering problem is to select a subset of clusters in $\mathcal{U}$ and respectively allocate resource to each selected cluster. For a cluster u that is not selected, then $x_u = 0$. Note that allocating more resource to one cell's cluster may enhance the QoS of the cell, while causing more interference to others. Besides, selecting sub-optimal clusters may lead to over load of cells or failure of meeting the bit demand. Hence the problem is challenging.

E. Comparison to OMA Modeling

We remark that the models proposed for OMA in [8] are essentially a special case of the NOMA model in this section, i.e. $\mathcal{U} = \{\{1\}, \{2\}, \dots, \{m\}\}$. In this case, the intra-cell interference term disappears from (1). Since any cluster u ($u \in \mathcal{U}$) only contains one UE j ($j \in \mathcal{J}$), the indices "ju" (and the index "u") can be merged (replaced) to (by) j, for (1) and (3)-(5). Parameter x_j and c_j then represent respectively the proportion of allocated RB(s) and the achievable capacity for UE j with $j \in \mathcal{J}$. With all these being done, (3)–(5) can be combined such that $\boldsymbol{x}$ is eliminated, leaving $\boldsymbol{\rho}$ to be the only variable. This system of cell load $\boldsymbol{\rho}$ fulfills the analytical framework of standard interference function (SIF), which enables the computation of the optimal network load settings via fixed-point iterations [7].

Indeed, by viewing the model as a feasibility problem with variables $\boldsymbol{x}$ and $\boldsymbol{\rho}$, the orthogonality in OMA enables decomposition among UEs, in terms of the QoS constraints (5). The resource allocation is thus on UE-level. However, in the general NOMA case, one needs to optimize the split of UE demand across multiple clusters containing the same UE. As a result, the clusters sharing UEs couple with each other. The loss of orthogonality therefore leads to a new dimension of complexity in the analysis.

III. Analytical Results

A. Main Results

In this section, we provide theoretical insights for the proposed model in Section II. The main results are summarized as follows. The model in Section II falls into the framework of SIF with respect to the cell load $\boldsymbol{\rho}$. The proof of this conclusion directly leads to a framework for user clustering and power allocation. Algorithms within this framework are able to solve the problem named *MinF* in (6) ($i \in \mathcal{I}$, $j \in \mathcal{J}$, $u \in \mathcal{U}$) to global optimum, with any real-valued function $F(\boldsymbol{\rho})$ that is monotonically increasing in $\boldsymbol{\rho}$.

$$[\textit{MinF}] \quad \min_{\boldsymbol{\rho}, \boldsymbol{p}, \boldsymbol{x} \geqslant 0} F(\boldsymbol{\rho}) \quad \text{s.t. } (1)\text{–}(5),\ \boldsymbol{p} \in \mathcal{P} \tag{6}$$

In (6), $\mathcal{P}$ is a (finite) set of candidate power allocations for a user cluster. In the main analysis, we temporarily fix the power to one of the candidates in $\mathcal{P}$ until Section III-E. Note that this simplification is made only for the sake of presentation, without any loss of the generality of our conclusions. In Section III-E, we relax this assumption and extend our analytical results to the case with the freedom of power allocation.

B. Single Cell Load Minimization

We start with the much simpler problem that concerns a single cell. Consider any cell i ($i \in \mathcal{I}$). The load minimization problem for cell i is in (7), with indices $u \in \mathcal{U}_i$ and $j \in u$ in (1)–(3) and (5). Variable $\boldsymbol{x}_i$ represents the vector of x_u ($u \in \mathcal{U}_i$).

$$\min_{\rho_i, \boldsymbol{x}_i \geqslant 0} \{\rho_i = \sum_{u \in \mathcal{U}_i} x_u |\ (1)\text{–}(3), (5)\} \tag{7}$$

Note that in (7), the loads of all cells other than i, i.e., ρ_k ($k \in \mathcal{I} \backslash \{i\}$), are treated as parameters instead of variables. With this precondition, the single-cell load minimization is a linear programming (LP) problem and can thus be solved to optimum efficiently. We remark that for any given load of cells $k \in \mathcal{I} \backslash \{i\}$, there is a minimum ρ_i. Thus, one can view the minimum ρ_i as a function of ρ_k, $k \in \mathcal{I} \backslash \{i\}$. For convenience, we use $\boldsymbol{\rho}_{-i}$ to represent the vector of all elements in $\boldsymbol{\rho}$ other than ρ_i. We show in Lemma 2 below the feasibility of (7) for the sake of rigor.

Lemma 2. *The system of inequalities of* (1)–(3),(5) *is always feasible for variables* ρ_i *and* $\boldsymbol{x}_i$.

The proof of Lemma 2 is based on Farkas' lemma [17]. The proof is not shown here due to the limit of space. The problem in (7) can then be defined as a function of $\boldsymbol{\rho}_{-i}$, which gives the minimum load for cell i, as shown in (8):

$$f_i(\boldsymbol{\rho}_{-i}) = \min_{\rho_i, \boldsymbol{x} \geqslant 0} \{\rho_i = \sum_{u \in \mathcal{U}_i} x_u |\ (1)\text{–}(3), (5)\}. \tag{8}$$

Lemma 3. *No infinite discontinuity exists for* $f_i(\boldsymbol{\rho}_{-i})$.

Lemma 3 states that $f_i(\boldsymbol{\rho}_{-i})$ is real-valued in its domain. The lemma is induced by Lemma 2 and that $\rho_i = 0$ is a lower bound of (7), such that the optimal objective in the LP cannot be $-\infty$.

C. Standard Interference Function

Network-wise, we have the function $\boldsymbol{f}(\boldsymbol{\rho})$ defined elementwisely in (9) for $\mathcal{I}$. By Theorem 1, $\boldsymbol{f}(\boldsymbol{\rho})$ is an SIF.

$$\boldsymbol{f}(\boldsymbol{\rho}) = [f_1(\boldsymbol{\rho}_{-1}), f_2(\boldsymbol{\rho}_{-2}), \dots, f_n(\boldsymbol{\rho}_{-n}))] \tag{9}$$

Theorem 1. $\boldsymbol{f}(\boldsymbol{\rho})$ *is an SIF, i.e. the following properties hold:*

1) *(Scalability)* $\alpha \boldsymbol{f}(\boldsymbol{\rho}) > \boldsymbol{f}(\alpha \boldsymbol{\rho})$, $\boldsymbol{\rho} \in \mathbb{R}^n_+$, $\alpha > 1$.
2) *(Monotonicity)* $\boldsymbol{f}(\boldsymbol{\rho}) \geqslant \boldsymbol{f}(\boldsymbol{\rho}')$, $\boldsymbol{\rho} \geqslant \boldsymbol{\rho}'$, $\boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathbb{R}^n_+$.

The proof of Theorem 1 is in the Appendix. Any function satisfying the two properties in Theorem 1 falls into the category of SIF. We explain the main properties of SIF as follows. For the non-linear equation system $\boldsymbol{f}(\boldsymbol{\rho}) = \boldsymbol{\rho}$, if there exists a feasible solution $\boldsymbol{\rho}^* \in \mathbb{R}^n_+$, i.e., equation $\boldsymbol{f}(\boldsymbol{\rho}^*) = \boldsymbol{\rho}^*$ holds, then $\boldsymbol{\rho}^*$ (named as the fixed point of $\boldsymbol{f}(\boldsymbol{\rho})$) uniquely exists. Another property is that, $\boldsymbol{\rho}^*$ can be computed by fixed-point iterations, iteratively by the equation $\boldsymbol{\rho}^{(k)} = \boldsymbol{f}(\boldsymbol{\rho}^{(k-1)})$ with $k \geqslant 1$ and any $\boldsymbol{\rho}^{(0)} \in \mathbb{R}^n_+$. With the existence of $\boldsymbol{\rho}^*$, starting from any $\boldsymbol{\rho}^{(0)} \in \mathbb{R}^n_+$, the iterations eventually converge to $\boldsymbol{\rho}^*$. Denote by $\boldsymbol{f}^k$ ($k > 1$) the function composition of $\boldsymbol{f}(\boldsymbol{f}^{k-1}(\boldsymbol{\rho}))$. We formally state this property in Lemma 4.

Lemma 4. *If $\lim_{k\to\infty} \mathbf{f}^{k}(\boldsymbol{\rho})$ exists for any $\boldsymbol{\rho} \in \mathbb{R}^{n}_{+}$, it exists uniquely for all $\boldsymbol{\rho} \in \mathbb{R}^{n}_{+}$ and is independent of $\boldsymbol{\rho}$.*

D. User Clustering

MinF with fixed power allocation is essentially a user clustering problem. Based on the analysis in Section III-C, we derive sufficient and necessary conditions for *MinF* with fixed power allocation, in terms of its feasibility and optimality, in Theorem 2 and Theorem 3, respectively. The proofs of both theorems are detailed in the Appendix due to their rather technical nature. Note that though the variables in *MinF* (with fixed power) are $\boldsymbol{\rho}$ and $\boldsymbol{x}$, the conditions shown in the two theorems only concern $\boldsymbol{\rho}$. This is because, when evaluating the function $\mathbf{f}(\boldsymbol{\rho})$, $\boldsymbol{x}$ is accordingly computed by solving corresponding LPs in (7). Thus we omit $\boldsymbol{x}$ in our following discussion for the sake of presentation.

Theorem 2. *For fixed-power MinF, $\boldsymbol{\rho}$ ($\boldsymbol{\rho} \leqslant \bar{\rho}\mathbf{1}$) is feasible if and only if the load $\mathbf{f}(\boldsymbol{\rho})$ is feasible and $\boldsymbol{\rho} \geqslant \mathbf{f}(\boldsymbol{\rho})$.*

Besides feasibility, for problem solving, Theorem 2 provides an efficient and effective method for improving any feasible solution to *MinF*. For any feasible solution $\boldsymbol{\rho}$, $\mathbf{f}(\boldsymbol{\rho})$ yields a better one[2]. One can compute $\mathbf{f}(\boldsymbol{\rho})$ and use it to replace $\boldsymbol{\rho}$ as a better solution for *MinF*, by solving n LP problems.

Theorem 3. *Load $\boldsymbol{\rho}^{*}$ is at the optimum of fixed-power MinF if and only if $\boldsymbol{\rho}^{*} = \mathbf{f}(\boldsymbol{\rho}^{*}) \leqslant \bar{\rho}\mathbf{1}$.*

Theorem 3 shows that, the optimal solution of *MinF* is at the fixed point of the function $\mathbf{f}(\boldsymbol{\rho})$. In addition, Theorem 3 reveals the relationship between the feasibility of the proposed model and the function $\mathbf{f}(\boldsymbol{\rho})$. Suppose that we have a feasible solution $\boldsymbol{\rho}$ for *MinF*. Since *MinF* is bounded below by $F(\mathbf{0})$, we conclude the existence of the optimum of *MinF*, which, by Theorem 3, resulting in the existence of the fixed point for $\mathbf{f}(\boldsymbol{\rho})$. Hence the existence of the fixed point for $\mathbf{f}(\boldsymbol{\rho})$ is necessary for the feasibility of *MinF*. Also, if the fixed point of $\mathbf{f}(\boldsymbol{\rho})$ exists (i.e. there is a $\boldsymbol{\rho}^{*}$ such that $\boldsymbol{\rho}^{*} = \mathbf{f}(\boldsymbol{\rho}^{*})$ holds), then $\boldsymbol{\rho}^{*}$ is an optimal solution to *MinF*, showing $\boldsymbol{\rho}^{*}$ feasibility. Therefore the existence of the fixed point for $\mathbf{f}(\boldsymbol{\rho})$ is sufficient for the feasibility of *MinF* as well. The conclusion is summarized in Lemma 5 below.

Lemma 5. *The fixed-power MinF is feasible if and only if the fixed point exists for $\mathbf{f}(\boldsymbol{\rho})$.*

Starting from any $\boldsymbol{\rho}^{(0)} \in \mathbb{R}^{+}_{n}$, we run the fixed point iterations $\boldsymbol{\rho}^{(k)} = \mathbf{f}(\boldsymbol{\rho}^{(k-1)})$ for $k \geqslant 1$. During each iteration, n LPs in (7) for $i \in \mathcal{I}$ are respectively solved. At the convergence, the optimum is reached. One may also strike a balance between the optimality and the computational efficiency. Once we know that $\boldsymbol{\rho}^{(k)}$ is feasible for any $k \geqslant 0$, then all $\boldsymbol{\rho}^{(k+1)}, \boldsymbol{\rho}^{(k+2)}, \ldots$ are feasible as well, and one can terminate the iterations at any step after k, to obtain a sub-optimal solution.

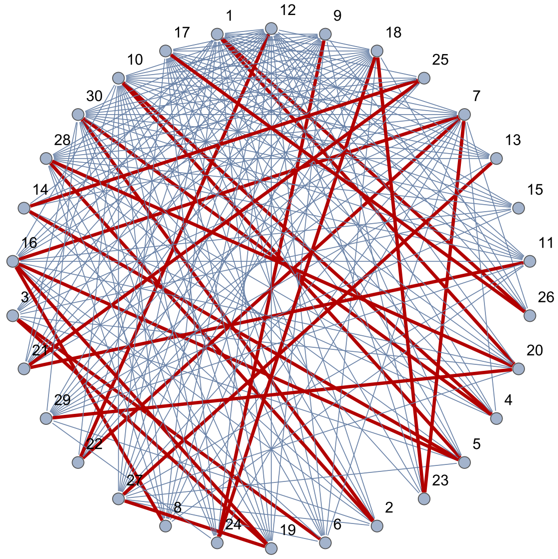

Figure 1. This figure comes from one of our simulations and it is used as an illustration for one cell's user clustering in multi-cell scenarios. There are 30 UEs within this cell. Each vertex represents a UE and each edge a candidate clustering option. Starting from UE 12, all UEs are sorted decreasingly according to its power gain from the cell and arranged clockwise (i.e. UE 12 has the best channel condition and UE 1 has the worst). The highlighted 28 edges are selected among all 181 candidate ones by (10) and (11) via solving LPs. Note that not all the UEs are expected to use NOMA, e.g., UE 15 is not clustered with others. From the visualization, it rarely happens that one UE is clustered with another with similar channel condition (e.g. its near neighbors in the circle).

Mathematically, the corresponding $\boldsymbol{x}^{*}$ for $\boldsymbol{\rho}^{*}$ is formulated in (10) with $i \in \mathcal{I}$.

$$\boldsymbol{x}_{i}^{*} = \arg\min_{\boldsymbol{x}_{i}} f_{i}(\boldsymbol{\rho}^{*}) \tag{10}$$

Accordingly we obtain the optimal user clustering solution, denoted by $\mathcal{U}^{*}$ ($\mathcal{U}^{*} \subseteq \mathcal{U}$), in (11).

$$\mathcal{U}^{*} = \{u \mid x_{u}^{*} > 0, u \in \mathcal{U}\} \tag{11}$$

Figure 1 gives an illustration of user clustering. Note that not all the UEs are expected to use NOMA, and the clustering occurs between UEs with large variation in channel conditions.

For a cell i ($i \in \mathcal{I}$), given the information of other cells' load $\boldsymbol{\rho}_{-i}$, solving $f_{i}(\boldsymbol{\rho})$ is based on local information, making it suitable to be run in a distributed manner. The technique called "asynchronous fixed-point iterations" [18] can be used. It means that for an arbitrary subset $\mathcal{I}^{\text{sub}}$ ($\mathcal{I}^{\text{sub}} \subseteq \mathcal{I}$) one can do fixed-point iterations for $\mathbf{f}(\boldsymbol{\rho})$ by following the rules that 1) $\rho_{i}^{(k)} = f_{i}(\boldsymbol{\rho}^{(k-1)})$ for any $i \in \mathcal{I}^{\text{sub}}$ and 2) $\rho_{i}^{(k)} = \rho_{i}^{(k-1)}$ for any $i \in \mathcal{I} \backslash \mathcal{I}^{\text{sub}}$, and $k \geqslant 1$, without loss of the convergence property. The solution obtained by such an iterative process still possesses feasibility for *MinF*, as can be verified by Theorem 2. The asynchronous fixed-point iterations converge to the fixed point of $\mathbf{f}(\boldsymbol{\rho})$ and optimality holds as well [18][3].

[2] Rigorously speaking, the new solution is only guaranteed to be no worse by Theorem 2. However in fact it is guaranteed to be better unless the old one is already at the optimum. A proof can be easily derived based on Lemma 5.

[3] An intuitive explanation is that, the fixed point is unique, regardless of how we reach it.

Therefore, it is sufficient for a cell to have information of a subset of cells (e.g., the surrounding cells) having major significance in terms of interference. The update for such information is very local and hence easily implemented via the LTE X2 interface.

E. Solving MinF

All the conclusions in Section III still hold with power allocation taken into consideration. An intuitive explanation is provided as follows. First, note that one can decompose the power allocation in terms of cells, as it can be seen in (1) that, any cell i interferes with other cells with p_i, independent of the power splitting scheme being used in cell i. Consider any cell $i \in \mathcal{I}$. Suppose there are in total K_i allocation schemes for cell i. Respectively, denote by $f_i^{[1]}(\boldsymbol{\rho}_{-i}), f_i^{[2]}(\boldsymbol{\rho}_{-i}), \dots, f_i^{[K_i]}(\boldsymbol{\rho}_{-i})$ the function $f_i(\boldsymbol{\rho}_{-i})$ in Section III-C under the K_i candidate power allocations. The load optimization problem in Section III-B evolves to (12).

$$f_i'(\boldsymbol{\rho}_{-i}) = \min\left\{f_i^{[1]}(\boldsymbol{\rho}_{-i}), f_i^{[2]}(\boldsymbol{\rho}_{-i}), \dots, f_i^{[K_i]}(\boldsymbol{\rho}_{-i})\right\} \quad (12)$$

The function in (12) is also SIF, because both the scalability and the monotonicity hold for $f_i'(\boldsymbol{\rho}_{-i})$. We denote by $\mathbf{f}'(\boldsymbol{\rho})$ the vector version of (12) with $i \in \mathcal{I}$. As the SIF properties hold for the new function $\mathbf{f}'(\boldsymbol{\rho})$, all the conclusions in Section III-D naturally remain valid for $\mathbf{f}'(\boldsymbol{\rho})$, accordingly with the notation $\mathbf{f}$ (or f_i) changed to $\mathbf{f}'$ (or f_i') in all the theorems' statements as well as in their corresponding proofs, and the word "fixed-power" in all the theorems' statements can thus be removed. Note that for evaluating the expression of $\mathbf{f}'(\boldsymbol{\rho})$, one needs to solve $\sum_{i=1}^{n} K_i$ instead of n LP problems as for $\mathbf{f}(\boldsymbol{\rho})$.

Denote by $\boldsymbol{\rho}'^{*}$ the fixed point of $\mathbf{f}'(\boldsymbol{\rho})$, i.e. $\boldsymbol{\rho}'^{*} = \mathbf{f}'(\boldsymbol{\rho}'^{*})$. We use $\mathbf{p}_i$ to represent the vector of p_{ju} with $u \in \mathcal{U}_i$ and $j \in u$, and $\mathcal{P}_i$ the candidate set of power allocation schemes for cell i (i.e. the K_i schemes as mentioned above). The optimal power allocation $\mathbf{p}_i^*$ for *MinF* is given by (13), for $i \in \mathcal{I}$. In other words, $\mathbf{p}_i^*$ corresponds to the k_{th} power allocation scheme, which leads to the minimum among all the K_i functions $f_i^{[k]}(\boldsymbol{\rho}_{-i}'^{*})$ ($k \in [1, K_i]$) in (12) at the convergence.

$$\mathbf{p}_i^* = \arg\min_{\mathbf{p}_i \in \mathcal{P}_i} f_i'(\boldsymbol{\rho}'^{*}) \quad (13)$$

The optimal clustering with power allocation $\mathbf{p}^*$ can be obtained by using (10) and (11) with f_i replaced by f_i' and $\boldsymbol{\rho}^*$ replaced by $\boldsymbol{\rho}'^{*}$ respectively.

IV. Numerical Results

A. Simulation Settings

We consider three performance metrics, the total load $\|\boldsymbol{\rho}\|_1$, the maximum load $\|\boldsymbol{\rho}\|_\infty$, and the efficiency of achieved rate on RBs. The rate efficiency is defined as the ratio between the sum of all user demands and the total of consumed RBs. We consider heterogeneous network scenarios in the simulation. Six small cells (SCs) are deployed around one macro cell (MC). The parameter setting is in Table I.

The UE demands are set in correspondence to the value of $\bar{\rho}$, such that the load of at least one cell in OMA reaches the limit $\bar{\rho}$. Two power allocation schemes are used for performance comparison. The "fractional transmit power control" (FTPC) proposed in [19] uses a parameter $\alpha_{\text{FTPC}} \in [0, 1]$ to control the fairness for power splitting among UEs. In [12], power allocation based on a pre-determined power ratio set is suggested, with a proportion $\alpha_{\text{NTT}} \in (0, 0.5)$ for allocating power to the UE with better channel condition. We use "NTT" to represent this power allocation scheme. Two sets of candidate parameter values of α_{FTPC} and α_{NTT} in Table I are used respectively for the two power allocation schemes, in computing (13). Uniform power allocation (i.e. two UEs in NOMA are allocated with the same amount of power), referred to as "Uniform", is used as reference. OMA is used as the baseline for performance benchmarking. The other parameters in Table I are coherent with [20].

Table I
Simulation Parameters.

Parameter	**Value**
Cell radius	500 m
Carrier frequency	2 GHz
Total bandwidth	20 MHz
Cell coverage radius	MC: 500 m; SC: 100 m
Number of users	$\{70, 140, 210, 280, 350\}$
Cell load limit $\bar{\rho}$	$\{0.4, 0.6, 0.8, 1.0\}$
Path loss	COST-231-HATA
Shadowing (Log-normal)	MC: 8 dB standard deviation SC: 4 dB standard deviation
Fading	Rayleigh flat fading
Noise power spectral density	-173 dBm/Hz
Total power on RB	MC: 800 mW SC: 100 mW
α_{FTPC}	$\{0.2, 0.4, 0.6, 0.8\}$
α_{NTT}	$\{0.1, 0.2, 0.3, 0.4\}$

B. Performance Evaluation

In summary, the numerical results show significant improvement by NOMA on resource and rate efficiency. Power allocation plays an important role in enhancing the performance in multi-cell NOMA. NOMA is promising in the scenario with intensive data traffic and high user densities.

In Figure 2, with higher demand, the reduction on total load achieved by NOMA becomes larger, meaning that NOMA is preferred in the scenarios with high traffic density. There is no difference between FTPC and NTT. On the other hand, both are considerably better than Uniform, meaning that power allocation in multi-cell NOMA has significant influence on resource efficiency.

Figure 3 shows the rate efficiency improvement with respect to the network density, with OMA being the baseline. The parameter $\bar{\rho}$ is set to 1.0 such that at least one cell in OMA is in full load. With the increase of the network density, NOMA achieves larger improvement in rate efficiency. Compared to Uniform, both FTPC and NTT lead to better performance, and FTPC has slight advantage over NTT when the user density is large. On the contrary, NTT leads to slightly better performance with low user density. The difference on the rate

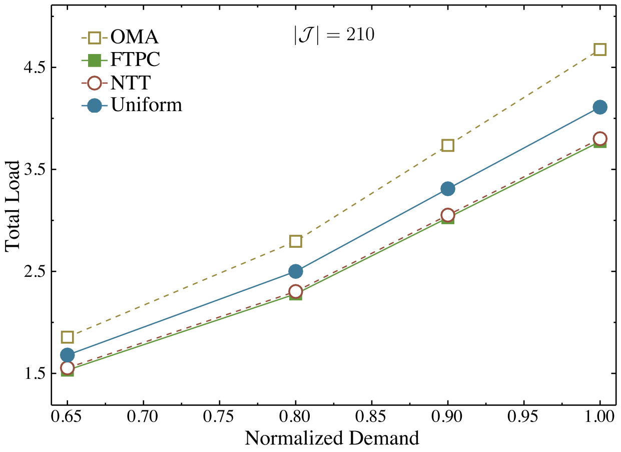

Figure 2. Performance of the total load of the network. The objective is $\mathsf{F}(\boldsymbol{\rho}) = \|\boldsymbol{\rho}\|_1$. The number of UEs is 210.

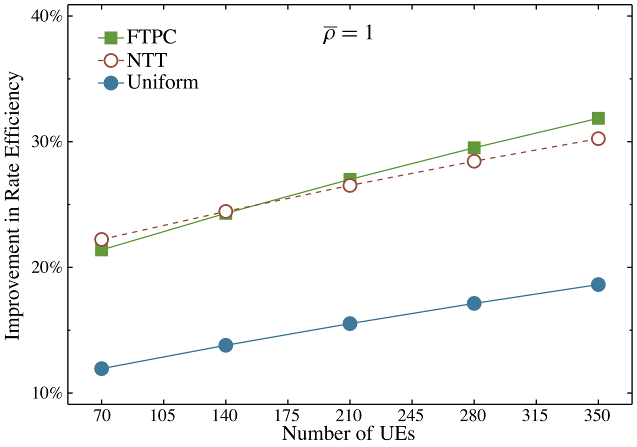

Figure 3. Performance of the improvement in rate efficiency. OMA is the baseline. The load limit $\bar{\rho}$ equals 1.0, meaning that for every data point, at least one cell in OMA is at full load.

performance between Uniform and the other two becomes higher with the increase of the network density.

Table II
PERFORMANCE EVALUATION WITH $\mathsf{F}(\boldsymbol{\rho}) = \|\boldsymbol{\rho}\|_\infty$.

Scheme	**Load Reduction**	**Improvement in Rate Efficiency**
FTPC	19.9%	25.2%
NTT	19.4%	24.2%
Uniform	11.7%	13.0%

For $\mathsf{F}(\boldsymbol{\rho}) = \|\boldsymbol{\rho}\|_\infty$, we minimize the load for the most heavy-loaded cell in the network, and evaluate the performance in terms of its load reduction and rate efficiency improvement. The settings of demands and the number of UEs follow those in Figure 2 and Figure 3. By using OMA as the baseline, the numerical results of improvement are averaged and summarized in Table II, which is coherent with the results in Figure 2 and Figure 3.

V. CONCLUSION

In this paper, multi-cell NOMA has been put into an optimization framework. We conclude that NOMA is a promising technique for raising spectrum efficiency, especially in scenarios with intensive data traffic and high user densities.

VI. ACKNOWLEDGEMENT

This work has been supported in part by the Swedish Research Council and the Luxembourg National Research Fund (FNR) Multi-Annual Thematic Research Programme (CORE) projects SEMIGOD.

REFERENCES

[1] 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 Communications Surveys Tutorials*, in press.

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

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

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

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

[6] I. Viering, M. Dottling, and A. Lobinger, "A mathematical perspective of self-optimizing wireless networks," in *2009 IEEE International Conference on Communications*, 2009, pp. 1–6.

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

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

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

[10] L. You and D. Yuan, "Load optimization with user association in cooperative and load-coupled LTE networks," *IEEE Transactions on Wireless Communications*, in press.

[11] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G non-orthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2016.

[12] "Evaluation methodologies for downlink multiuser superposition transmissions," 3GPP, Tech. Rep. R1-153332, 2014.

[13] 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 *Proceedings of IEEE MILCOM*, 2015, pp. 1170–1175.

[14] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

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

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

[17] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*. Athena Scientific Belmont, MA, 1997.

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

[19] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, “System-level performance evaluation of downlink non-orthogonal multiple access (NOMA),” in *Proceedings of IEEE 24th PIMRC*, 2013, pp. 611–615.

[20] “Requirements for further advancements for evolved universal terrestrial radio access (E-UTRA) (LTE-Advanced),” 3GPP, Tech. Rep. TR 36.913, V13.0.0, 2016.

APPENDIX

PROOF OF LEMMA 1

Proof. The two UEs j and h are in the cluster u and served by cell i. Denote by γ_{hj} and γ_{hh} the SINR at user j and h, in respect of the transmission for user h, in (14) and (15).

$$\gamma_{hj} = \frac{p_{hu}g_{ij}}{p_{ju}g_{ij} + \sum_{k\in\mathcal{J}\backslash\{i\}} p_k g_{kj}\rho_k + \sigma^2} \tag{14}$$

$$\gamma_{hh} = \frac{p_{hu}g_{ih}}{p_{ju}g_{ih} + \sum_{k\in\mathcal{J}\backslash\{i\}} p_k g_{kj}\rho_k + \sigma^2} \tag{15}$$

The condition for UE j to decode UE h is $\gamma_{hj} \geqslant \gamma_{hh}$, i.e.

$$\begin{aligned}\gamma_{hj} \geqslant \gamma_{hh} \Leftrightarrow\ & p_{ju}g_{ij}g_{ih} + g_{ij}\sum_{k\in\mathcal{J}\backslash\{i\}} p_k g_{kh}\rho_k + g_{ij}\sigma^2 \\ & \geqslant p_{ju}g_{ij}g_{ih} + g_{ih}\sum_{k\in\mathcal{J}\backslash\{i\}} p_k g_{kj}\rho_k + g_{ih}\sigma^2 \\ \Leftrightarrow & \sum_{k\in\mathcal{J}\backslash\{i\}} p_k\rho_k(g_{ih}g_{kj} - g_{ij}g_{kh}) \leqslant (g_{ij} - g_{ih})\sigma^2 \end{aligned} \tag{16}$$

Recall that $g_{ij} > g_{ih}$. Thus the right-hand side of (16) is positive. By Lemma 1 that $g_{ij}/g_{ih} \geqslant g_{kj}/g_{kh}$ for $k \in \mathcal{J}\backslash\{i\}$, the left-hand side is negative. Hence (16) holds. □

PROOF OF THEOREM 1

We reformulate the problem in (7) below, for the sake of clarity of the proof.

$$\min_{\rho_i, x_i, r \geqslant 0} \rho_i \tag{17a}$$

$$\text{s.t.} \quad c_{ju} = MB\log\left(1 + \gamma_{ju}(\boldsymbol{\rho}_{-i})\right) \tag{17b}$$

$$\rho_i = \sum_{u\in\mathcal{U}_i} x_u \tag{17c}$$

$$\sum_{u\in\mathcal{U}_j} r_{ju} \geqslant d_j \tag{17d}$$

$$x_u \geqslant \frac{r_{ju}}{c_{ju}(\boldsymbol{\rho}_{-i})} \tag{17e}$$

Proof. (Monotonicity) Suppose $\boldsymbol{\rho}' \leqslant \boldsymbol{\rho}$. We change $c_{ju}(\boldsymbol{\rho})$ to $c_{ju}(\boldsymbol{\rho}')$. According to the monotonicity of the function c_{ju}, we have $c_{ju}(\boldsymbol{\rho}) \leqslant c_{ju}(\boldsymbol{\rho}')$, for any $u \in \mathcal{U}$ and $j \in u$. Note that any feasible solution $(\mathbf{x}, \mathbf{r})$ for the minimization problem with $\boldsymbol{\rho}$ is still feasible for the minimization problem with $\boldsymbol{\rho}'$. Therefore, the minimization problem in (9) is relaxed with $\boldsymbol{\rho}$ being replaced by $\boldsymbol{\rho}'$. Therefore, we have $f_i(\boldsymbol{\rho}') \leqslant f_i(\boldsymbol{\rho})$.

(Scalability) First, note that the equality $\alpha f_i(\boldsymbol{\rho}) = \min_{\mathbf{x},\mathbf{r}\geqslant 0}\{\alpha\rho_i |\ \text{(17b)–(17d)}, r_{ju} \leqslant x_u c_{ju}(\boldsymbol{\rho})\}$, for which we denote the optimal solution by $(\mathbf{x}', \mathbf{r}')$. Consider the problem β, i.e., $\beta : \min_{\mathbf{x},\mathbf{r}\geqslant 0}\{\rho_i |\ \text{(17b)–(17d)}, x_u \geqslant \alpha r_{ju}/c_{ju}(\boldsymbol{\rho})\}$ with $\alpha > 1$. One can verify that $(\alpha\mathbf{x}', \mathbf{r}')$ is a feasible solution to the problem β, with objective value $\alpha f_i(\boldsymbol{\rho})$. Then, the optimum of problem β is no more than $\alpha f_i(\boldsymbol{\rho})$. Suppose we replace $\boldsymbol{\rho}$ with $\alpha\boldsymbol{\rho}$ ($\alpha > 1$) in the minimization problem (9). By the scalability of $1/c_{ju}(\boldsymbol{\rho})$, we have $1/c_{ju}(\alpha\boldsymbol{\rho}) < \alpha/c_{ju}(\boldsymbol{\rho})$. Thus, the minimization problem corresponding to $f_i(\alpha\boldsymbol{\rho})$ is a relaxation of the problem β. For the relaxed minimization problem, the optimal objective value $f_i(\alpha\boldsymbol{\rho})$ is less than that of β. Therefore, we conclude $f_i(\alpha\boldsymbol{\rho}) < \alpha f_i(\boldsymbol{\rho})$. Hence the conclusion. □

PROOF OF THEOREM 2

Lemma 6. *For any* $\boldsymbol{\rho} \geqslant 0$*, if there exists* $i \in \mathcal{J}$ *such that* $\rho_i < f_i(\boldsymbol{\rho}_{-i})$*, then* $\boldsymbol{\rho}$ *is not feasible to* (1)–(5).

Proof. Let $\rho_i' = f_i(\boldsymbol{\rho}_{-i})$. By the definition of f_i, ρ_i' is the minimum value satisfying (1)–(3), and (5) under $\boldsymbol{\rho}_{-i}$. Therefore any ρ_i with $\rho_i < \rho_i'$ causes at least one of the constraints (1)–(3), or (5) being violated with $\boldsymbol{\rho}_{-i}$, meaning that the vector $\boldsymbol{\rho}$ cannot satisfy all constraints (1)–(5). Hence the conclusion. □

Proof. Theorem 2 is proved as follows. By the inverse proposition of Lemma 6, a feasible solution $\boldsymbol{\rho}$ always satisfies $\boldsymbol{\rho} \geqslant \mathbf{f}(\boldsymbol{\rho})$. Now suppose $\boldsymbol{\rho}$ is feasible to *MinF* and consider using $\mathbf{f}(\boldsymbol{\rho})$ as a solution to *MinF*. (Together with the $\mathbf{x}$ obtained when computing $\mathbf{f}(\boldsymbol{\rho})$.) Then $\mathbf{f}(\boldsymbol{\rho})$ satisfies (4). Also $\mathbf{f}(\boldsymbol{\rho})$ together with its $\mathbf{x}$ fulfills (1)–(3) and (5) by the definition of $\mathbf{f}(\boldsymbol{\rho})$. Thus $\mathbf{f}(\boldsymbol{\rho})$ is feasible.

For the sufficiency, note that the feasibility of $\mathbf{f}(\boldsymbol{\rho})$ indicates that $\boldsymbol{\rho}_{-i}$ along with $\mathbf{x}_i$ obtained by solving $f_i(\boldsymbol{\rho}_{-i})$ satisfies (1)–(3), and (5). Combined with the precondition $\rho_i \leqslant \bar{\rho}$ with $i \in \mathcal{J}$, the load $\boldsymbol{\rho}$ is feasible to (1)–(5) (and thus feasible to *MinF*). Hence the conclusion. □

PROOF OF THEOREM 3

Proof. (Necessity) If $\boldsymbol{\rho}^*$ is feasible, then obviously we have $\boldsymbol{\rho}^* \leqslant \bar{\rho}\mathbf{1}$. By Theorem 2, $\mathbf{f}(\boldsymbol{\rho}^*)$ is also feasible and $\mathbf{f}(\boldsymbol{\rho}^*) \leqslant \boldsymbol{\rho}^*$ holds. Also, $\mathbf{f}^k(\boldsymbol{\rho}^*)$ for any $k \geqslant 1$ is a feasible solution. According to Theorem 1, $\mathbf{f}(\boldsymbol{\rho})$ is monotonic in $\boldsymbol{\rho}$, and thus we have $\mathbf{f}^k(\boldsymbol{\rho}^*) \geqslant \mathbf{f}^{k+1}(\boldsymbol{\rho}^*)$ for any $k \geqslant 1$. Based on Lemma 4, we let $\boldsymbol{\rho}' = \lim_{k\to\infty}\mathbf{f}^k(\boldsymbol{\rho}^*)$. Then $\boldsymbol{\rho}' \leqslant \boldsymbol{\rho}^*$ holds, by the above discussion. 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 2, for any feasible $\boldsymbol{\rho}$, $\lim_{k\to\infty}\mathbf{f}^k(\boldsymbol{\rho})$ is feasible and $\lim_{k\to\infty}\mathbf{f}^k(\boldsymbol{\rho}) \leqslant \boldsymbol{\rho}$ holds. By Lemma 4, the limit is unique for any $\boldsymbol{\rho} \geqslant 0$, 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 $\boldsymbol{\rho}^* \leqslant \boldsymbol{\rho}$ for any feasible $\boldsymbol{\rho}$, meaning that $\boldsymbol{\rho}^*$ is optimal for *MinF*. Hence the conclusion. □