Resource Allocation for Mixed Numerology NOMA

Stephen McWade, Mark F. Flanagan, Juquan Mao, Lei Zhang and Arman Farhang

***Abstract*—6G wireless networks will require the flexibility to accommodate an extremely diverse set of service types. Accommodating different quality of service (QoS) requirements for these service types necessitates the use of mixed numerologies, where services using different subcarrier spacings or symbol durations coexist in the same frequency band. Non-orthogonal multiple access (NOMA) techniques can potentially be used to accommodate users with different numerologies while also gaining the performance benefits associated with NOMA. To achieve the full performance benefits of a mixed numerology NOMA (MN-NOMA) system, resource allocation is paramount. However, the coexistence of mixed numerologies changes the nature of the interference that each user experiences. In this letter, we approach the problem of optimizing subcarrier and power allocation for maximizing the spectral efficiency of MN-NOMA. In particular, we propose a two-stage sub-optimal approach to solve this problem. Numerical results show that the proposed approach provides performance gains over existing benchmark schemes of up to 14% and 12% in spectral efficiency and fairness, respectively.**

***Index Terms*—Mixed numerologies, NOMA, multi-service, resource allocation.**

I. Introduction

Future mobile networks require a high degree of flexibility as well as an ability to simultaneously provide service to multiple users with different service types and quality of service (QoS) requirements (e.g. ultra-reliable and low-latency communications (URLLC), massive machine-type communications (mMTC), vehicle-to-everything (V2X) communications, etc.). The sixth generation of communication networks (6G) is envisioned to have an even wider variety of service types [1] and will require an even higher degree of flexibility to achieve this. Service types with different QoS requirements will need to use different numerologies, which here refers to waveform parameters such as subcarrier spacing (SCS), symbol duration and cyclic prefix (CP) length. For example, V2X communication requires robustness against Doppler spread and thus a smaller symbol duration (which implies a wider SCS for the OFDM waveform). On the other hand, mMTC services require robustness against delay spread and thus a smaller SCS [2]. A one-size-fits-all numerology is obviously very difficult to design. This then leads to the problem of how best to accommodate services with mixed numerologies in order to achieve the required flexibility.

One of the standard approaches to this problem is to separate the system bandwidth into smaller adjacent bandwidth parts (BWPs) with each BWP having a different numerology for its service type. Throughout this letter, we refer to this approach as mixed numerology orthogonal multiple access (MN-OMA). The OFDM subcarriers of mixed numerologies are not orthogonal to each other and this causes inter-numerology interference (INI) which in turn degrades system performance [3]. An alternative to this approach is to use mixed numerology non-orthogonal multiple access (MN-NOMA). In this scenario, users with different numerologies share time and frequency resources and are multiplexed in another domain such as the power or code domain [4]. Superposition coding and successive interference cancellation (SIC) are used to decode the users' signals. Our previous work [5] analyzed interference in a two-user MN-NOMA scenario under a simple power allocation in order to isolate the effect of mixed numerologies on interference. Additionally, we showed that MN-NOMA can accommodate users with different numerologies while providing improved spectral efficiency (SE) over MN-OMA.

The topic of resource allocation in NOMA has received significant attention in recent years [6]. For example, the authors of [7] use an iterative water filling (IWF) based method to maximize the SE of an uplink multi-carrier NOMA system and show that this achieves a global optimum. The authors of [8] use a greedy algorithm for subcarrier and power allocation to maximize the SE of an uplink multi-carrier NOMA system while imposing a limit on the number of users that can use the same subcarrier. Resource allocation is equally important for MN-NOMA in order to achieve its full potential. However, the presence of mixed numerologies changes the nature of the interference experienced by the users, which in turn makes the problem of optimal resource allocation more challenging.

While some research work exists on resource allocation for MN-OMA [9], to the best of our knowledge, there has been little research on the topic of optimizing resource allocation for MN-NOMA. The authors of [10] tackle the problem of power allocation of a multinumerology NOMA system with a constraint on system fairness. However, [10] uses an exhaustive search method to solve the optimization problem, which is impractical for deployment in future wireless networks as it scales very poorly with the number of users. The authors of [11] consider the allocation of time-frequency resource blocks for MN-NOMA but this work does not consider power allocation which is an important aspect of MN-NOMA systems. Another gap in the existing works on both MN-OMA and MN-NOMA is that they almost exclusively consider a small number of users (e.g., 2 or 3 users only [5], [9], [10]).

This letter addresses these gaps in the literature with the fol-

S. McWade and M. Flanagan are with the School of Electrical and Electronic Engineering, University College Dublin, Ireland (email: stephen.mcwade@ucdconnect.ie, mark.flanagan@ieee.org). J. Mao is with the Institute for Communications Systems, University of Surrey, UK (email: juquan.mao@surrey.ac.uk). L. Zhang is with the School of Engineering, University of Glasgow, UK (email: Lei.Zhang@glasgow.ac.uk). A. Farhang is with CONNECT, The Telecommunications Research Centre, Trinity College Dublin, Ireland. (email: arman.farhang@tcd.ie).

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant number 17/RC-PhD/3479 and supported in part by a grant from SFI under Grant number 19/FFP/7005. It was also supported in part by the U.K. Engineering and Physical Sciences Research Council under Grant number EP/S02476X/1.

lowing contributions: (i) We outline a generic analytical model for mixed numerology uplink NOMA where any number of users with different numerologies overlap in both time and frequency domains and are decoded via SIC at the receiver. (ii) We formulate an optimization problem for maximizing the SE of this MN-NOMA system subject to a minimum rate requirement for each user. Additionally, in order to reduce SIC complexity and error propagation effects, we consider a scenario where a limitation is placed on the number of users that can occupy the same subcarrier. (iii) Since the optimization problem is intractable, we utilize a two-stage sub-optimal approach to solve it. Stage 1 uses an iterative greedy algorithm to allocate subcarriers to users. Stage 2 uses a successive convex approximation (SCA) based approach to optimize the power allocation for MN-NOMA users.

Notations: Superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transpose and Hermitian transpose, respectively. Bold lower-case characters are used to denote vectors and bold upper-case characters are used to denote matrices. $\mathbf{X} = \mathrm{diag}(\mathbf{x})$ is a diagonal matrix with the elements of the vector $\mathbf{x}$ on its main diagonal. $\mathbf{x} = \mathrm{diag}(\mathbf{X})$ is a column vector whose elements consist of the main diagonal of the matrix $\mathbf{X}$, and $\otimes$ represents the Kronecker product. The $p \times p$ identity matrix and $p \times q$ all-zero matrix are denoted by $\mathbf{I}_p$ and $\mathbf{0}_{p\times q}$, respectively.

II. System Model

We consider a mixed numerology multi-carrier uplink NOMA system with K users. Each user uses CP orthogonal frequency division multiplexing (CP-OFDM) modulation with its own specified SCS Δf_i, $i \in \{0, \ldots, K-1\}$, i.e., its own numerology. The SCSs of users i and j are related by $\Delta_{i,j} = \frac{\Delta f_i}{\Delta f_j} = \frac{q_i}{q_j}$, where $q_i = 2^{\mu_i}$ is the scaling factor of user i's numerology as per 5G NR [12] and $\mu_i \in \{0, 1, 2, 3...\}$. User i has $N_i = \frac{B}{\Delta f_i}$ subcarriers available, where B is the system bandwidth. Each user has a corresponding CP length $N_{\mathrm{cp},i}$, which is scaled depending on the user's numerology to maintain alignment of the time domain symbols. The total symbol length for user i is therefore $N_{\mathrm{T},i} = N_i + N_{\mathrm{cp},i}$.

We define the power allocation vector of user i as $\mathbf{p}_i = \left[\sqrt{p_{i,0}}\ \sqrt{p_{i,1}}\ \cdots\ \sqrt{p_{i,N_i-1}}\right]^{\mathrm{T}}$ where $p_{i,n}$ is the power allocated to subcarrier $n \in \{0, 1, \ldots, N_i - 1\}$ of user i. We also define the system-level power allocation vector as $\mathbf{p} = \left[\mathbf{p}_0^{\mathrm{T}}\ \mathbf{p}_2^{\mathrm{T}}\ \cdots\ \mathbf{p}_{K-1}^{\mathrm{T}}\right]^{\mathrm{T}}$. In order to accommodate a limit on the number of users that share a subcarrier, we denote the $N_i \times 1$ subcarrier allocation vector of user i as $\mathbf{x}_i = [x_{i,0}\ x_{i,0}\ \ldots\ x_{i,N_i-1}]^{\mathrm{T}}$. Here, $x_{i,n} = 1$ if subcarrier n is allocated to user i and $x_{i,n} = 0$ if it is not. We define the system-level subcarrier allocation vector as $\mathbf{X} = \left[\mathbf{x}_0^{\mathrm{T}}\ \mathbf{x}_2^{\mathrm{T}}\ \ldots\ \mathbf{x}_{K-1}^{\mathrm{T}}\right]^{\mathrm{T}}$.

The transmitted symbol for user i is given by

$$\mathbf{s}_i = \mathbf{A}_{\mathrm{cp},i}\mathbf{F}_i^{\mathrm{H}}\mathrm{diag}(\mathbf{x}_i)\mathrm{diag}(\mathbf{p}_i)\mathbf{d}_i, \tag{1}$$

where $\mathbf{F}_i$ is the N_i-point unitary discrete Fourier transform (DFT) matrix of user i in which the (l,k) element is $\frac{1}{\sqrt{N_i}}e^{-j\frac{2\pi}{N_i}lk}$. The matrix $\mathbf{A}_{\mathrm{cp},i} = [\mathbf{I}_{\mathrm{cp,i}}, \mathbf{I}_{N_i}]^{\mathrm{T}}$ is the CP addition matrix of user i, where $\mathbf{I}_{\mathrm{cp,i}}$ is composed of the final $N_{\mathrm{cp},i}$ columns of $\mathbf{I}_{N_i}$. The vector $\mathbf{d}_i$ is the vector of data-bearing symbols for user i. For each user's channel, we consider a linear time invariant channel model with channel impulse response $\bar{\mathbf{h}}_i = [\bar{h}_{i,0}, \ldots, \bar{h}_{i,L_i-1}]^{\mathrm{T}}$ where L_i is the channel length. Each user's signal passes through its respective channel and these signals are superimposed at the base station. These signals are then decoded at the base station using SIC. We assume that users are decoded in the order of their indices, i.e., user 0 is decoded first, then user 1, and so on, with user $K-1$ being decoded last. When a user is decoded, it only experiences interference from the users yet to be decoded.

We consider a 2-user case as an example, where user i experiences interference from user j. If $\Delta f_i < \Delta f_j$, then the time domain symbol of user i overlaps with $\Delta_{j,i}$ time domain symbols of user j. The received signal is given by

$$\mathbf{r} = \mathbf{H}_i\mathbf{s}_i + \mathbf{H}_j\tilde{\mathbf{s}}_j + \mathbf{w} \tag{2}$$

where $\mathbf{H}_i$ is a Toeplitz matrix with first column $\left[(\bar{\mathbf{h}}_i)^{\mathrm{T}}, \mathbf{0}_{1\times(N_{T,i}-L^{(i)}-1)}\right]^{\mathrm{T}}$ and first row $\left[\bar{h}_{i,0}, \mathbf{0}_{(1)\times(N_{T,i}-1)}\right]$ and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ represents additive white Gaussian noise. The vector $\tilde{\mathbf{s}}_j$ is the concatenation of the $\Delta_{j,i}$ overlapping user j symbols. This concatenated symbol is given by

$$\tilde{\mathbf{s}}_j = \left[\mathbf{I}_{\Delta_{j,i}} \otimes (\mathbf{A}_{\mathrm{cp},j}\mathbf{F}_j^{\mathrm{H}}\mathrm{diag}(\mathbf{x}_j)\mathrm{diag}(\mathbf{p}_j))\right]\tilde{\mathbf{d}}_j, \tag{3}$$

where $\tilde{\mathbf{d}}_j$ is a vector of concatenated data-bearing symbols of user j. The decoded user i signal is given by

$$\mathbf{y}_i = \mathbf{F}_i\mathbf{R}_{\mathrm{cp},i}\mathbf{r} \tag{4}$$

where $\mathbf{R}_{\mathrm{cp},i} = \left[\mathbf{0}_{N_i\times N_{\mathrm{cp},i}}, \mathbf{I}_{N_i}\right]$ is the CP removal matrix of user i. Using (1) and (2), (4) can be expanded as

$$\mathbf{y}_i = \mathbf{\Psi}_i\mathrm{diag}(\mathbf{x}_i)\mathrm{diag}(\mathbf{p}_i)\mathbf{d}_i + \mathbf{F}_i\mathbf{R}_{\mathrm{cp},i}\mathbf{H}_i\tilde{\mathbf{s}}_j + \boldsymbol{\omega}_i, \tag{5}$$

where $\boldsymbol{\omega}_i = \mathbf{F}_i\mathbf{R}_{\mathrm{cp},i}\mathbf{w}$ and $\mathbf{\Psi}_i$ is a square diagonal matrix with diagonal elements equal to the frequency response of the channel for user i, i.e., $\mathbf{F}_i\mathbf{R}_{\mathrm{cp},i}\mathbf{H}_i\mathbf{A}_{\mathrm{cp},i}\mathbf{F}_i^{\mathrm{H}} = \mathrm{diag}(\mathbf{h}_i)$ where $\mathbf{h}_i = \mathbf{F}_i[\bar{\mathbf{h}}_i, \mathbf{0}_{(N_i-L_i)\times 1}]^{\mathrm{T}}$. An interference matrix can be calculated from the second term of (5) via

$$\mathbf{\Gamma}^{(i\leftarrow j)} = \mathbf{F}_i\mathbf{R}_i\mathbf{H}_j\left[\mathbf{I}_{\Delta_{j,i}} \otimes (\mathbf{A}_j\mathbf{F}_j^{\mathrm{H}}\mathrm{diag}(\mathbf{x}_j)\mathrm{diag}(\mathbf{p}_j))\right]. \tag{6}$$

$\mathbf{\Gamma}^{(i\leftarrow j)}$ is an $N_i \times \Delta_{j,i}N_j$ matrix where the (n, o_m)-th element $\Gamma_{n,o_m}^{(1\leftarrow 2)}$ contains the INI weight on subcarrier n of user i from the o-th subcarrier of the m-th overlapping symbol of user j where $o_m = m(N_j) + o$ for $m = 0, \ldots, \Delta_{i,j} - 1$ and $o = 0, \ldots, N_j - 1$.

Alternatively, if $\Delta f_i > \Delta f_j$, then there are $\Delta_{i,j}$ overlapping user i symbols in the duration of a single user j symbol. The received signal is given by

$$\mathbf{r} = \mathbf{H}_i\tilde{\mathbf{s}}_i + \mathbf{H}_i\mathbf{s}_j + \mathbf{w}, \tag{7}$$

and the m-th decoded user i symbol, where $1 \le m \le \Delta_{i,j}$, is given by

$$\mathbf{y}_{i,m} = \mathbf{F}_i\mathbf{R}_{\mathrm{cp},i}\mathbf{C}_m^{(i\leftarrow j)}\mathbf{r}, \tag{8}$$

where $\mathbf{C}^{(i\leftarrow j)} = \left[\mathbf{0}_{N_{T,i}\times(m-1)N_{T,i}}, \mathbf{I}_{N_{T,i}}, \mathbf{0}_{N_{T,i}\times(\Delta_{i,j}-m)N_{T,i}}\right]$ isolates the overlapping part of the user j symbol. Using (1) and (7), (8) can be expanded as

$$\mathbf{y}_{i,m} = \mathbf{\Psi}_i\mathrm{diag}(\mathbf{x}_i)\mathrm{diag}(\mathbf{p}_i)\mathbf{d}_{i,m} + \mathbf{F}_i\mathbf{R}_{\mathrm{cp},i}\mathbf{C}_m^{(i\leftarrow j)}\mathbf{H}_i\mathbf{s}_j + \boldsymbol{\omega}_i. \tag{9}$$

The interference matrix for symbol m of user i is therefore given by

$$\mathbf{\Gamma}^{(i\leftarrow j)} = \mathbf{F}_i \mathbf{R}_{\mathrm{cp},i} \mathbf{C}_m^{(i\leftarrow j)} \mathbf{H}_j \mathbf{A}_j \mathbf{F}_j^{\mathrm{H}} \mathrm{diag}(\mathbf{x}_j) \mathrm{diag}(\mathbf{p}_j), \quad (10)$$

which is an $N_i \times N_j$ matrix where the (n, o)-th element $\Gamma_{n,o}^{(i\leftarrow j)}$ defines the INI coefficient on subcarrier n of user i from the o-th subcarrier of user j.

For both cases, the interference matrix can be used to calculate the mean-squared error (MSE) interference on the victim user,

$$\boldsymbol{\gamma}^{(i\leftarrow j)}(\mathbf{x}_j, \mathbf{p}_j) = \mathrm{diag}(\mathbf{\Gamma}^{(i\leftarrow j)} (\mathbf{\Gamma}^{(i\leftarrow j)})^{\mathrm{H}}), \quad (11)$$

which is a vector of length N_i whose j-th element is equal to the MSE of the corresponding subcarrier of user i due to interference from user j. The signal-to-interference-plus-noise ratio (SINR) on subcarrier n of user i, expressed as a function of $\mathbf{x}$ and $\mathbf{p}$, is given by

$$\Lambda_{i,n}(\mathbf{x}, \mathbf{p}) = \left(\frac{x_{i,n} p_{i,n} |h_{i,n}|^2}{\sum_{j=i+1}^{K} \gamma_n^{(i\leftarrow j)}(\mathbf{x}_j, \mathbf{p}_j) + \sigma^2} \right), \quad (12)$$

where $|h_{i,n}|^2$ is the channel gain on subcarrier n of user i and σ^2 denotes the noise power. The achievable rate of user i on subcarrier n is given by

$$R_{i,n}(\mathbf{x}, \mathbf{p}) = \frac{B}{N_i} \log_2 \left(1 + \Lambda_{i,n}(\mathbf{x}, \mathbf{p})\right), \quad (13)$$

and the achievable sum-rate for the entire system can be expressed as

$$R(\mathbf{x}, \mathbf{p}) = \sum_{i=0}^{K-1} \sum_{n=0}^{N_i-1} R_{i,n}(\mathbf{x}, \mathbf{p}). \quad (14)$$

III. Resource Allocation

We define $\mathcal{N} = \{0, \ldots, N_b - 1\}$ as the set of subcarriers for the base numerology, i.e., the numerology with the smallest SCS. The set of subcarriers for all other numerologies are subsets of $\mathcal{N}$ and are related by the relevant scaling factors. We then define the set of users that can simultaneously use subcarrier $n \in \mathcal{N}$ as $\mathcal{K}_n = \{i : n \bmod q_i = 0\}$. Our optimization problem aims to maximize the SE of the system and can be formulated as

$$\max_{\mathbf{x},\mathbf{p}} \quad \sum_{i=0}^{K-1} \sum_{n=0}^{N_i-1} R_{i,n}(\mathbf{x}, \mathbf{p}) \quad (15a)$$

$$\text{s.t.} \quad \sum_{n=0}^{N_i-1} p_{i,n} \le P_i, \quad (15b)$$

$$p_{i,n} \ge 0, \quad (15c)$$

$$\sum_{n=0}^{N_i-1} R_{i,n} \ge R_{\min}, \forall i \in \{0, \ldots, K-1\}, \quad (15d)$$

$$x_{i,n} \in \{0, 1\}, \quad (15e)$$

$$\sum_{i \in \mathcal{K}_n} x_{i, \frac{n}{q_i}} \le U, \forall n \in \mathcal{N}, \quad (15f)$$

where (15b) is the power constraint for each user, (15c) ensures that the power on a subcarrier cannot be negative, and (15d) ensures that each user achieves a rate of at least $R_{\min}$. Constraints (15e) and (15f) ensure that the number of users allocated to any subcarrier does not exceed U.

Algorithm 1 Subcarrier Allocation and Power Initialization

1: Initialize $\mathbf{X} = \mathbf{1}$
2: **repeat**
3: Use IWF for each user with SC allocation $\mathbf{X}$ to obtain power allocation $\mathbf{p}$
4: Calculate $R(\mathbf{x}, \mathbf{p})$ using (13) and (14)
5: For each $n \in \mathcal{N}$, $\mathcal{S}_n = \{i : p_{i, \frac{n}{q_i}} \ge 0\}$
6: $U_{\max} = \max |\mathcal{S}_n|$ and $\mathcal{N}_{\max} = \{n : |\mathcal{S}_n| = U_{\max}\}$
7: **if** $U_{\max} \ge U$ **then**
8: $(i^*, n^*) = \arg \min_{n \in \mathcal{N}, i \in \mathcal{K}_n} (R_{i,n}^-)$
9: Set $x_{i^*, \frac{n^*}{q_{i^*}}} = 0$
10: **end if**
11: **until** $U_{\max} = U$
12: **return** $\mathbf{X}^* = \mathbf{X}$ and $\mathbf{p}_0 = \mathbf{p}$

It is clear that the objective function is non-convex due to the binary constraint (15e) and due to the nature of the interference term in (12). This makes finding an optimal solution difficult. Instead, we propose a two-stage solution. In Stage 1, we use an iterative greedy algorithm to allocate subcarriers to users and initialize the power allocation using IWF. However, due to the presence of mixed numerologies, IWF does not provide an optimal power allocation. Therefore, in Stage 2, we use SCA to optimize the power allocation given the subcarrier allocation from Stage 1.

Stage 1: Subcarrier Allocation and Power Initialization

The proposed greedy algorithm starts by relaxing constraint (15f) and setting $x_{i,n} = 1$ for all subcarriers of all users. IWF is then used to allocate power for each user. We define the set of users with positive power on each subcarrier as $\mathcal{S}_n = \{i : p_{i, \frac{n}{q_i}} \ge 0\}$. This power allocation may lead to some subcarriers having too many active users, i.e., $|\mathcal{S}_n| > U$. One user is then removed from the overloaded subcarriers in each iteration of the algorithm.

It is important to note that in MN-NOMA, users deploying one subcarrier can affect the rate of users that are allocated to other subcarriers due to INI. The user to be removed from the overloaded subcarrier should therefore be the user which reduces the overall sum rate the least. This can be seen in Line 8 of Algorithm 1, where the user chosen is the one that minimizes $R_{i,n}^- \triangleq R(\mathbf{x}, \mathbf{p}) - R(\mathbf{x}_{\setminus i,n}, \mathbf{p})$ where $\mathbf{x}_{\setminus i,n}$ is simply $\mathbf{x}$ with $x_{i,n} = 0$. The user which minimizes the reduction in overall rate is then de-allocated from subcarrier n by setting $x_{i,n} = 0$. This process is repeated until there are no more overloaded subcarriers remaining. This provides the subcarrier allocation and initial power allocation for Stage 2. A detailed description of the procedure can be found in Algorithm 1. This algorithm is an iterative greedy algorithm which performs the low-complexity IWF procedure at each step until no overloaded subcarriers remain. Since a user is always removed from an overloaded subcarrier at each step, the algorithm is guaranteed to terminate after a finite number of iterations. It is worth noting that the case where all the users are initially active on all the subcarriers results in an upper

Algorithm 2 Power Allocation

Initialize $t = 0$; Initialize $\mathcal{M}_i$, $\boldsymbol{\rho}(t) = \boldsymbol{\rho}_0$ using Algorithm 1, $\mathbf{q}(t) = \log_2(\boldsymbol{\rho}_0)$, $\epsilon = 10^{-6}$
repeat
  Update $\boldsymbol{\gamma}^{(i\leftarrow k)}$ using (11) for all users.
  Update $\alpha_{i,n}$ and $\beta_{i,n}$ using (17) for each subcarrier of each user
  Solve (19a) to obtain $\mathbf{q}(t)$
  $\boldsymbol{\rho}(t) \leftarrow 2^{\mathbf{q}(t)}$
  Use (13) and (14) to obtain $R(t)$
  $t \leftarrow t + 1$
until $R(t) - R(t-1) < \epsilon$
return $\boldsymbol{\rho}^* = \boldsymbol{\rho}(t)$

limit on the number of iterations of $N_b(K-U)$. However, our simulations showed that this bound is loose in practice, as the average number of iterations was significantly lower than $N_b(K-U)$.

Stage 2: Power Allocation Optimization

Given the subcarrier allocation $\mathbf{X}^*$ from Stage 1, we define the set of active subcarriers for each user as $\mathcal{M}_i = \{n : x_{i,n} = 1\}$ with elements $\{\mathcal{M}_{i,0}\ \mathcal{M}_{i,1}\ \dots\ \mathcal{M}_{i,N_{c,i}}\}$. We then define the active power allocation vectors $\boldsymbol{\rho}_i = \left[\sqrt{p_{i,\mathcal{M}_{i,0}}}\ \sqrt{p_{i,\mathcal{M}_{i,1}}}\ \cdots\ \sqrt{p_{i,\mathcal{M}_{i,N_{c,i}}}}\right]^{\mathrm{T}}$ where $N_{c,i}$ is the number of active subcarriers for user i. The active power vector for the whole system is defined as $\boldsymbol{\rho} = \left[\boldsymbol{\rho}_0^{\mathrm{T}}\ \boldsymbol{\rho}_1^{\mathrm{T}}\ \cdots\ \boldsymbol{\rho}_{K-1}^{\mathrm{T}}\right]^{\mathrm{T}}$.

We can now express the optimization problem as:

$$\max_{\boldsymbol{\rho}} \quad \sum_{i=0}^{K-1}\sum_{n\in\mathcal{M}_i} R_{i,n}(\boldsymbol{\rho}) \tag{16a}$$

$$\text{s.t.} \quad \sum_{n\in\mathcal{M}_i} p_{i,n} \le P_i, \tag{16b}$$

$$p_{i,n} \ge 0, \tag{16c}$$

$$\sum_{n\in\mathcal{M}_i} R_{i,n} \ge R_{\min}, \forall i \in \{0,\dots,K-1\}. \tag{16d}$$

It is clear that the objective function is still non-convex due to (13), and in fact has a difference-of-convex (DC) structure. Additionally, the interference term is given by $\sum_{j=i+1}^{K}\gamma_n^{(i\leftarrow j)}(\mathbf{p}_j)$ where $\gamma_n^{(i\leftarrow j)}(\mathbf{p}_j)$ is given in (11). This is in contrast to single numerology NOMA (SN-NOMA) where the interference term is simply given by $\sum_{j=i+1}^{K} p_{j,n}|h_{j,n}|^2$. This change in the interference terms caused by the different numerologies means that the method used in [7] for converting the SN-NOMA problem into a convex problem is not applicable to MN-NOMA. Consequently, an IWF approach no longer guarantees a global optimum. Instead, we utilize the approach from [13] to relax the non-convex problem (16a) and deal with the DC structure. We note that (16a) contains a term of the form $\log(1+\Lambda)$, due to (13). This allows us to utilize the lower bound

$$\alpha\log(\Lambda) + \beta \le \log(1+\Lambda), \tag{17}$$

which is tight for a given Λ at the values of $\alpha = \frac{\Lambda}{1+\Lambda}$ and $\beta = \log(1+\Lambda) - \frac{\Lambda}{1+\Lambda}\log\Lambda$. We apply this bound to approximate (16a) as $\sum_{i=0}^{K-1}\sum_{n\in\mathcal{M}_i}\frac{B}{N_i}\left(\alpha_{i,n}\log_2(\Lambda_{i,n}(\boldsymbol{\rho})) + \beta_{i,n}\right)$. However, this relaxed expression is still non-convex in $\boldsymbol{\rho}$. Therefore, we define vector $\mathbf{q}$ where $q_{i,j} = \log_2(\rho_{i,j})$ and use variable substitution to rewrite the objective function as

$$\sum_{i=0}^{K-1}\sum_{n\in\mathcal{M}_i}\bar{R}_{i,n}(\mathbf{q}) = \sum_{i=0}^{K-1}\sum_{n\in\mathcal{M}_i}\frac{B}{N^{(i)}}\left(\alpha_{i,n}\log_2(\Lambda_{i,n}(2^{\mathbf{q}})) + \beta_{i,n}\right), \tag{18}$$

where $2^{\mathbf{q}}$ is an element-wise operation on the vector $\mathbf{q}$. The relaxed optimization problem can now be formulated as

$$\max_{\mathbf{q}} \quad \sum_{i=0}^{K-1}\sum_{n\in\mathcal{M}_i}\bar{R}_{i,n}(\mathbf{q}) \tag{19a}$$

$$\text{s.t.} \quad \sum_{n\in\mathcal{M}_i} 2^{q_{i,n}} \le P_i, \tag{19b}$$

$$\sum_{n\in\mathcal{M}_i}\bar{R}_{i,n} \ge R_{\min}, \forall i \in \{0,\dots,K-1\}. \tag{19c}$$

Expanding the term $\log_2(\Lambda_{i,n}(2^{\mathbf{q}}))$ in (18) reveals it to be comprised of a linear term and a convex log-sum-exp term. Therefore, it is convex in $\mathbf{q}$. Additionally, the constraints are both sums of convex terms. The relaxed optimization problem is therefore convex and can be solved using standard convex optimization tools such as CVX [14].

The resulting SCA power allocation procedure is summarized in Algorithm 2. Algorithm 2 is an iterative process for maximizing the lower bound on the sum rate. In each iteration, the bound is improved by updating $\alpha_{i,n}$ and $\beta_{i,n}$ for each user's subcarriers and then solving (19a) to obtain the updated power allocation. This process is repeated until convergence. Algorithm 2 results in a monotonically increasing objective function which is guaranteed to converge in polynomial time. It should be noted that since Algorithm 2 maximizes the lower bound on the sum rate, the solution it provides is sub-optimal.

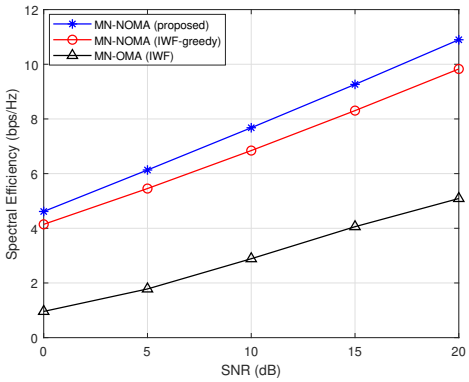

Fig. 1. SE for MN-NOMA using our proposed algorithm, MN-NOMA using the IWF-Greedy algorithm from [8], and MN-OMA using IWF, at different SNR levels.

IV. Numerical Results

This section presents numerical results to showcase the effectiveness of our proposed algorithm. As benchmarks, we consider the application of the IWF based greedy algorithm from [8] to the MN-NOMA system, and we also consider an MN-OMA system using IWF for power allocation. For MN-OMA, we assume that no guardbands exist between the users,

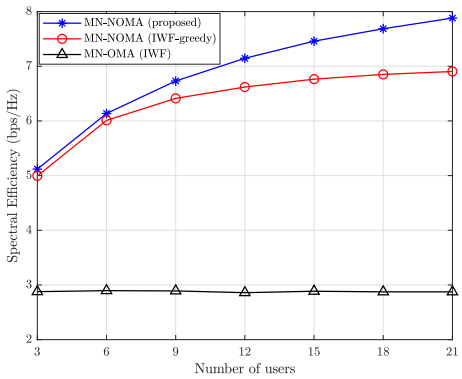

Fig. 2. SE for MN-NOMA using our proposed algorithm, MN-NOMA using the IWF-Greedy algorithm from [8], and MN-OMA using IWF, at different different numbers of users.

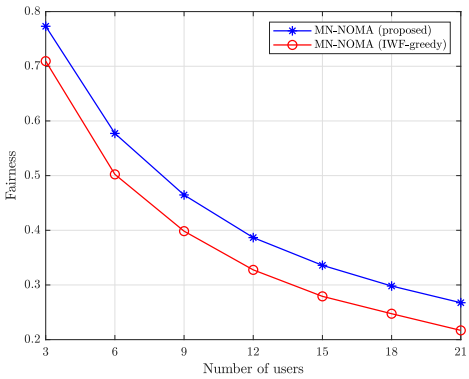

Fig. 3. Fairness for MN-NOMA using our proposed algorithm and the IWF-Greedy algorithm from [8], for different numbers of users.

which represents the best-case scenario for SE. Each user has a power constraint of 1 mW per subcarrier to ensure that they use the same average power over a given time period. For the minimum rate constraint of the proposed algorithm, we set $R_{\min} = 0.5$ bps/Hz. The Extended Vehicular A model is used to model the small-scale fading of each user, and Monte Carlo simulation is used to average the results over 1000 random channel instances. Three numerologies are used with DFT sizes of 512, 256 and 128, as per 5G NR specifications [12]. The CP length is 7% of the time domain symbol. For each instance in the Monte Carlo simulation, the number of users using each numerology is the same. A limit of $U = 2$ users per subcarrier is adopted.

Fig. 1 shows the SE of MN-NOMA using the proposed algorithm compared to the benchmark schemes for different SNR conditions for an 18-user case. It can be seen that the proposed algorithm outperforms the benchmark schemes in both cases, especially at high SNR. It can also be seen that both MN-NOMA schemes significantly outperform the MN-OMA scheme, confirming the benefits of MN-NOMA over MN-OMA. Fig. 2 shows the SE of the proposed algorithm compared to the benchmark schemes with increasing number of users and a fixed SNR of 10 dB. The proposed algorithm outperforms the benchmark schemes by up to 14% at a higher number of users.

Fig. 3 compares the fairness of the proposed algorithm with that of the benchmark IWF scheme. Here, fairness is measured using Jain's fairness index [15], $F = (\sum_{i=0}^{K-1} R_i)^2 / K \sum_{i=1}^{K} R_i^2$. We consider a varying number of users for a fixed SNR of 10 dB. It can be observed from the figure that the proposed algorithm provides a consistent improvement in fairness of approximately 12% compared to the benchmark scheme. This is due to the minimum rate constraint (19c) which ensures that each user is getting a fairer share of the sum rate.

V. Conclusion

In this paper, we have studied the topic of resource allocation optimization for maximizing SE in MN-NOMA. We have presented a generic system model for K-user uplink MN-NOMA and we have proposed a two-stage sub-optimal algorithm for maximizing the SE subject to a minimum rate constraint on each user. Stage 1 initializes the power allocation and allocates users to subcarriers using a greedy algorithm, and Stage 2 then uses SCA to optimize the power allocation. The superiority of the proposed method over benchmark schemes, in both SE and fairness, has been demonstrated numerically.

References

[1] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[2] L. Zhang, A. Farhang, G. Feng, and O. Onireti, *Radio Access Network Slicing and Virtualization for 5G Vertical Industries*. Wiley, 2020.

[3] X. Zhang, L. Zhang, P. Xiao, D. Ma, J. Wei, and Y. Xin, "Mixed Numerologies Interference Analysis and Inter-Numerology Interference Cancellation for Windowed OFDM Systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7047–7061, 2018.

[4] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A Survey of Non-Orthogonal Multiple Access for 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[5] S. McWade, M. F. Flanagan, L. Zhang, and A. Farhang, "Interference and Rate Analysis of Multinumerology NOMA," *IEEE International Conference on Communications (ICC)*, 2020.

[6] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal Multiple Access for 5G and Beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.

[7] M. Zeng, N. Nguyen, O. A. Dobre, Z. Ding, and H. V. Poor, "Spectral and Energy-Efficient Resource Allocation for Multi-Carrier Uplink NOMA Systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 9293–9296, 2019.

[8] M. Al-Imari, P. Xiao, and M. A. Imran, "Receiver and resource allocation optimization for uplink NOMA in 5G wireless networks," *2015 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 151–155, 2015.

[9] J. Mao, L. Zhang, P. Xiao, and K. Nikitopoulos, "Interference Analysis and Power Allocation in the Presence of Mixed Numerologies," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2020.

[10] A. T. Abusabah and H. Arslan, "NOMA for Multinumerology OFDM Systems," *Wireless Communications and Mobile Computing*, vol. 2018, Jan. 2018.

[11] R. J. Wang, C. H. Wang, G. S. Lee, D. N. Yang, W. T. Chen, and J. P. Sheu, "Resource Allocation in 5G with NOMA-Based Mixed Numerology Systems," *GLOBECOM - IEEE Global Communications Conference*, pp. 1–6, 2020.

[12] *3GPP TS 38.211*, 3rd Generation Partnership Project (3GPP), 2016, v15.2.0.

[13] J. Papandriopoulos and J. S. Evans, "Low-Complexity Distributed Algorithms for Spectrum Balancing in Multi-User DSL Networks," *IEEE International Conference on Communications (ICC)*, vol. 7, pp. 3270–3275, 2006.

[14] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[15] R. K. Jain, D.-M. W. Chiu, W. R. Hawe *et al.*, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, 1984.