Efficient Bit Loading Algorithm for OFDM-NOMA Systems with BER Constraints

T. Assaf, *Member, IEEE*, A. Al-Dweik, *Senior Member, IEEE*, M. S. El Moursi, *Senior Member, IEEE*,
and H. Zeineldin, *Senior Member, IEEE, Member, IEEE*

Abstract

This paper considers applying bit loading to multicarrier non-orthogonal multiple access (NOMA) systems. The objective is to maximize the system total throughput while satisfying the users' individual quality of service (QoS) constraints. Although bit loading is generally an NP-hard problem, even for orthogonal multiple access (OMA), the mutual interference between the users and the dependence of power coefficients and modulation orders are additional challenges that add substantial complexity to the optimization problem. Therefore, we propose in this paper an efficient bit loading algorithm for multicarrier NOMA systems and compares the complexity and throughput with OMA. The obtained results show that NOMA has virtual cognition and hybrid modes of operation, NOMA/OMA, that enables NOMA to outperform OMA by 100% for the two-user scenario. The complexity of the loading process for NOMA is noticeably higher than OMA, which is due to the high computational complexity of bit error rate (BER) computation for NOMA. The obtained results show that the NOMA throughput depends on which user is loaded first, and on the initially loaded bits.

Index Terms

Bit Loading, non-orthogonal multiple access (NOMA), orthogonal multiple access (OMA), bit error rate (BER), spectrum efficiency, computational complexity.

I. INTRODUCTION

The expeditious development of the mobile Internet and Internet of Things (IoT) has introduced various capacity, data rate, and power consumption challenges for the Fifth Generation (5G) wireless communications systems [1]. Consequently, extensive research has been focused in the last few years to develop technologies that can overcome such challenges, such as energy-aware communication and energy harvesting [2], [3], cloud-based radio access

T. Assaf, A. Al-Dweik, and M. S. El Moursi are with the Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi 127788, United Arab Emirates. (E-mail: {tasneem.assaf, arafat.dweik, mohamed.elmoursi}@ku.ac.ae).

A. Al-Dweik is also with the Center for Cyber Physical Systems (C2PS), Khalifa University, Abu Dhabi, UAE, and the Department of Electrical and Computer Engineering, Western University, London, ON, Canada. (E-mail: dweik@fulbrightmail.org).

H. H. Zeineldin is with the Faculty of Engineering, Cairo University, Giza, Egypt and is currently on leave from Khalifa University, Abu Dhabi, UAE. (E-mail: hatem.zeineldin@ku.ac.ae).

networks [4], wireless network virtualization [5], advanced antenna systems [6], [7], and efficient error correction coding [8]. Furthermore, great advancements have been achieved in terms of multiple access technologies such as the non-orthogonal multiple access (NOMA), which may enhance the spectral efficiency with respect to orthogonal multiple access (OMA), by allowing multiple users to share the transmission resources simultaneously. Generally speaking, NOMA schemes can be classified into two main categories, code-domain NOMA [9], [10] and power-domain NOMA [11]–[13], which is the focus of this paper.

Towards further increasing the spectrum efficiency of NOMA, efficient modulation schemes such as orthogonal frequency division multiplexing (OFDM) have been proposed in the literature [14], [15]. Despite the several advantages of OFDM, its bit error rate (BER) performance in fading channels is similar to single-carrier systems. Therefore, adopting adaptive modulation, commonly known as bit loading, can enhance OFDM-NOMA systems in terms of BER and consequently spectrum efficiency. Both BER and throughput in OFDM-NOMA depend on the modulation level and operating environments. In the low signal-to-noise (SNR) regime, increasing the modulation level does not contribute to the total throughput due to elevated detection error. In contrast, in the high SNR regime, increasing the modulation level may significantly improve the total throughput. However, applying bit loading, or adaptive modulation, for OMA [16]–[19] is drastically different from NOMA, because the loading processes for all OMA users are generally independent, i.e., changing the modulation level for any user does not affect the BER of other users. For NOMA, the mutual inter-user interference (IUI) between all users creates BER dependency, i.e., the BER of any user depends on the modulation order of all other users. Moreover, changing the modulation order for a particular subcarrier makes it necessary in certain scenarios to change the power assignment for that subcarrier, which may affect the power assignment for other users as well. Consequently, the bit loading process for NOMA is computationally more demanding than OMA, nevertheless, the throughput gain can be significant.

A. Related Work

Although NOMA has been considered widely in the literature, based on extensive literature search, and to the best of the authors' knowledge, very little work has considered the throughput maximization problem for NOMA systems using bit loading. For example, user pairing and adaptive turbo trellis coded modulation scheme for cognitive radio (CR) in downlink NOMA are proposed in [20]. The adaptive modulation is implemented by switching between multiple modulation orders based on a predefined set of SNR thresholds, and thus, the throughput is not maximized. In [21], a hybrid OMA/NOMA downlink system is considered, where the authors proposed a multi-stages algorithm for user grouping, subcarrier allocation, and bit allocation schemes while considering data-rate and error-rate constraints. However, the number of bits allocated of each OFDM subcarrier is predefined and the modulation order assigned for each user depends on the associated SNR. Therefore, no throughput maximization process is performed. Moreover, the study focuses on power reduction with no effort to maximize the throughput.

Luo and Teh [22] considered maximizing the throughput of a cooperative NOMA system by adaptively switching between OMA and NOMA at the relay node. However, no error rate is guaranteed and the switching is performed by selecting the mode that maximizes the throughput based on channel gain.

B. Motivation and Contribution

Although adaptive modulation, or bit loading, has been widely considered for OMA as reported in [16]–[19], [23] and the references listed therein, it is not the case for NOMA where only very limited research work considered the problem, and with several simplifying assumptions. Therefore, this paper considers the bit loading problem in OFDM-NOMA systems and proposes an efficient bit loading algorithm for the two-user scenario, where the objective is to maximize the total system throughput while satisfying the users' individual BER constraints. The considered bit loading problem is a nonconvex integer non-linear optimization problem, and thus it is NP-hard, which implies that optimal solutions are prohibitively expensive. Moreover, the mutual interference between the users and dependence of the power coefficients and modulation orders are critical challenges that introduce additional complexity to the optimization process. For example, given that the number of possible modulation orders is $\mathcal{M}$, and the number of subcarriers is K , then the total number of combinations to find the optimal modulation orders for a $\mathcal{N}$ OMA users is $\mathcal{N} \times \mathcal{M}^K$ because the subcarriers of each user are loaded independently from other users. In the case of NOMA, the subcarriers of all users have to be loaded simultaneously, and thus, the total number of combinations is $\mathcal{M}^{K\mathcal{N}}$, which is substantially larger than the OMA case.

To resolve the complexity problem, we propose an iterative greedy algorithm that loads the bits of each user separately. The proposed solution is based on the incremental allocation algorithm (IAA) [24], however, different starting points and loading directions, incremental or decremental, are considered. Therefore, the algorithm is greedy within each iteration, but may change directions in different iterations. The impact of the initial loading is also considered to reduce the number of iterations required to provide reliable results. The performance is presented in terms of the throughput per user, total system throughput, and computational complexity, and compared to OMA. The obtained results show that NOMA significantly outperforms OMA in terms of throughput due to the cooperation between the two users. This behavior can be considered as virtual cognitive radio in the sense that the unitized spectrum by one user is utilized by the other user. In such scenarios, the resources of the user with poor channel conditions can be utilized by the other NOMA user, while they are wasted for OMA. Moreover, the bit loading naturally imposes a hybrid OMA/NOMA mode of operation because only one user might be fully allocated to the subcarrier. The complexity of the loading process for NOMA is noticeably higher than OMA, which is due to the high computational complexity required to compute NOMA BER. Moreover, some additional complexity is introduced because of the iterative process. To reduce the complexity, the bit loading process is evaluated using two different initial starting points, namely the fully loaded (FL) and partially loaded (PL). The obtained complexity

results show that the PL approach can reduce the complexity by about 50% for several cases of interest.

C. Paper Organization

The rest of the paper is organized as follows. In Sec. II, the system and channel models are represented. The adaptive bit loading problem formulation is given in III. In IV, the proposed bit loading algorithm is detailed. The complexity of the OMA and NOMA BER computations is derived in Sec. V. Numerical and simulation results are given in Sec. VI. Finally, the work is concluded in Sec. VII.

II. SYSTEM AND CHANNEL MODELS

This work considers a downlink power-domain NOMA serving two users, U_1 and U_2 , where the users' equipment (UE) and the base station (BS) are equipped with a single antenna. The users share K subcarriers of an OFDM symbol through superposition modulation (SM). The multiplexed baseband NOMA symbol can be described as

$$\mathbf{x} = \sqrt{P_T \beta_1} \mathbf{x}^{(1)} + \sqrt{P_T \beta_2} \mathbf{x}^{(2)} \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$, $\mathbf{x}^{(n)} = [x_1^{(n)}, x_2^{(n)}, \dots, x_K^{(n)}]^T$ denotes the modulated data sequence of the n th user, P_T is the BS total transmit power, $\beta_n \in [0, 1]$, is the allocated power coefficient for the n th user. For the rest of the paper, the transmit power P_T is normalized to unity and $\sum_{n=1}^2 \beta_n = 1$. The data symbols $x_k^{(n)} \forall \{k, n\}$ have unit energy, i.e., $\mathbb{E} \left[|x_k^{(n)}|^2 \right] = 1$ where $\mathbb{E}[\cdot]$ denotes the statistical expectation. Each data symbol $x_k^{(n)}$ can be loaded with $b_k^{(n)}$ bits using a particular modulation with an order $M_k^{(n)} = 2^{b_k^{(n)}}$. For a proper detection of NOMA symbols, the power allocation coefficients must satisfy the following constraint [25],

$$\frac{\beta_1}{\beta_2} > \frac{M_k^{(1)} - 1}{M_k^{(2)} - 1} \left(\sqrt{M_k^{(2)}} - 1 \right)^2, \quad \forall k. \quad (2)$$

The NOMA stream $\mathbf{x}$ is converted to an OFDM symbol using the inverse discrete Fourier transformation (IDFT), and a cyclic prefix (CP) is added to prevent inter-symbol interference (ISI). At the receiver, the CP is removed and discrete Fourier transformation (DFT) is employed for demodulation. For the n th user, assuming that the channel remains fixed during one OFDM symbol period and the CP is larger than the maximum delay spread of the channel, the DFT output can be represented as,

$$\mathbf{r}^{(n)} = \mathbf{G}^{(n)} \mathbf{x} + \mathbf{w}^{(n)} \quad (3)$$

where $\{\mathbf{r}^{(n)}, \mathbf{w}^{(n)}\} \in \mathbb{C}^{K \times 1}$, $w^{(n)} \sim \mathcal{CN}(0, \sigma_n^2)$ is the additive white Gaussian noise (AWGN), and $\mathbf{G}^{(n)} \in \mathbb{C}^{K \times K}$ is the n th user channel frequency response matrix, which is given by

$$\mathbf{G}^{(n)} = \text{diag} \left\{ \left[G_1^{(n)}, G_2^{(n)}, \dots, G_K^{(n)} \right] \right\}. \quad (4)$$

and

$$G_k^{(n)} = \sqrt{d_n^{-\alpha}} \sum_{i=0}^{Q^{(n)}} g_i^{(n)} \exp\left(-\frac{j2\pi i k}{K}\right), \quad \forall k \quad (5)$$

where the distance between the n th user and the BS is denoted as d_n , α is the path loss exponent, $g_i^{(n)} \sim \mathcal{CN}(0, \sigma_{g_i}^2)$ denotes the i th multipath component gain and $Q^{(n)} + 1$ represents the number of multipath components. In NOMA systems, it is typically assumed that the first user has lower average channel gain, i.e., $\mathbb{E}\left[\left|\alpha_k^{(1)}\right|^2\right] < \mathbb{E}\left[\left|\alpha_k^{(2)}\right|^2\right]$, where $\alpha_k^{(n)} \triangleq \left|G_k^{(n)}\right|$. Thus, the power coefficients should be allocated in the opposite order of the channel amplitudes, i.e., $\beta_1 > \beta_2$. It should be noted that the coherence time is assumed to be larger than the OFDM symbol duration. After the DFT, the information detection can be performed using the joint multi-user detection (JMuD) [26] or successive interference cancellation detection (SICD) [27].

III. BIT LOADING PROBLEM FORMULATION

The adaptive bit loading problem which aims at maximizing the throughput while satisfying certain user BER constraints can be formulated as

$$\max_{\mathbf{b}^{(n)}, \forall n} \mathcal{T}_T = \frac{1}{K} \sum_{k=1}^K b_k^{(1)} + b_k^{(2)} \quad (6)$$

subject to:

$$C_1 : b_k^{(n)} \in \{0, 2, 4, 6, 8, \dots, b_{\max}^{(n)}\}, \quad \forall \{n, k\} \quad (7a)$$

$$C_2 : \bar{P}^{(n)} = \frac{\sum_{k=1}^K b_k^{(n)} P_k^{(n)}}{\sum_{k=1}^K b_k^{(n)}} \leq \bar{P}_{Th}^{(n)}, \quad \forall n \quad (7b)$$

where $\mathcal{T}_T$ is the total throughput normalized by the total number of subcarriers, $b_k^{(n)}$ is the number of bits assigned to the k th subcarrier of the n th user, $b_k^{(n)} = \log_2 M_k^{(n)}$, $b_{\max}^{(n)}$ is the n th user maximum possible number of bits, $\bar{P}^{(n)}$ is the average conditional BER per OFDM symbol, $P_k^{(n)}$ denotes the instantaneous conditional BER for the k th subcarrier, and $\bar{P}_{Th}^{(n)}$ represents the BER threshold. The objective function in (6) aims at maximizing the total throughput over K subcarriers for the two users. Constraint C_1 (7a) specifies the possible modulation orders $M_k^{(n)}$, which are restricted to square $M_k^{(n)}$ -quadrature amplitude modulation (QAM). Moreover, the k th subcarrier for a particular user can be switched-off by setting $b_k^{(n)} = 0$. Constraint C_2 in (7b) is used to enforce BER fairness among the users by setting an average BER threshold per user, $\bar{P}_{Th}^{(n)}$, which can be selected based on the quality of service (QoS) requirements for each user.

The instantaneous BER $P_k^{(n)}$ in (7b) can be computed as

$$P_k^{(n)} = (1 - \Gamma_k) P_k^{(n)}|_{\Gamma_k=0} + \Gamma_k P_k^{(n)}|_{\Gamma_k=1}$$

where $\Gamma_k \in \{0, 1\}$ is a binary variable that allows computing $P_k^{(n)}$ regardless of the number of bits per user in each subcarrier. More specifically,

$$\Gamma_k = \begin{cases} 0, & \min \{b_k^{(1)}, b_k^{(2)}\} = 0 \\ 1, & \text{otherwise} \end{cases} \quad (8)$$

$P_k^{(n)}|_{\Gamma_k}$ can be expressed as [28], [29],

$$P_k^{(n)}|_{\Gamma_k=0} = \frac{2}{v_k^{(n)} \check{v}_k^{(n)}} \sum_{i=1}^{v_k^{(n)}} \sum_{t=0}^{L_{1,i,k}^{(n)}} D_1^{(n)} Q \left((2t+1) \alpha_k^{(n)} \sqrt{\frac{2\gamma}{E_k^{(n)}} \log_2 \left(M_k^{(n)} \right)} \right), \quad n = 1, 2 \quad (9)$$

$$P_k^{(1)}|_{\Gamma_k=1} = \frac{2}{v_k^{(1)} \sqrt{M_k}} \sum_{i=1}^{v_k^{(1)}} \sum_{t=0}^{L_{1,i,k}^{(1)}} \sum_{l=0}^{\Lambda_k^{(2)}} D_1^{(1)} Q \left(g_1^+ \left(2t+1, 2l - \check{v}_k^{(2)} + 1 \right) \right) \quad (10)$$

and

$$P_k^{(2)}|_{\Gamma_k=1} = \frac{2}{v_k^{(2)} \sqrt{M_k}} \sum_{i=1}^{v_k^{(2)}} \left[\sum_{t=0}^{L_{1,i,k}^{(2)}} \sum_{l=0}^{2\Lambda_k^{(1)}} S_k D_2^{(2)} D_3^{(1)} Q \left(g_2^+ (c_i l, 2t+1) \right) - \sum_{t=0}^{L_{1,i,k}^{(2)}} \sum_{l=1}^{2\Lambda_k^{(1)}} S_k D_2^{(2)} D_3^{(1)} Q \left(g_2^- (c_i l, 2t+1) \right) \right] \quad (11)$$

where $\gamma = E_b/N_0$ is the instantaneous SNR, $Q(\cdot)$ is the Q function and the other symbols are defined in Table I. As can be noted from (6), the bit loading problem is a combinatorial non-linear NP-hard optimization problem. Therefore, in this paper, an efficient bit loading algorithm is proposed to solve the problem.

A. Binary Formulation

In order to reduce the number of decision variables and the implementation complexity of the bit loading problem, an equivalent binary formulation for (6) can be obtained as,

$$\max_{\mathbf{Y}^{(n)}, \forall n} \mathcal{T}_T = \mathbf{u} \left(\mathbf{Y}^{(1)} \boldsymbol{\pi}^{(1)} + \mathbf{Y}^{(2)} \boldsymbol{\pi}^{(2)} \right) \quad (12)$$

subject to:

$$C_1 : \sum_{z^{(n)}=1}^{Z^{(n)}} y_{kz^{(n)}}^{(n)} = 1, \forall \{k, n\} \quad (13a)$$

$$C_2 : \bar{P}_B^{(n)} = \left(\mathbf{P}^{(n)} - \bar{P}_{Th}^{(n)} \mathbf{u} \right) \mathbf{Y}^{(n)} \boldsymbol{\pi}^{(n)} \leq 0, \forall n \quad (13b)$$

where $\mathbf{u} = [u_1, u_2, \dots, u_K]$, $u_k = 1, \forall k$, $\boldsymbol{\pi}^{(n)}$ is a vector that consists of all possible number of bits, $\boldsymbol{\pi}^{(n)} = [0, 2, 4, 8, \dots, b_{\max}^{(n)}]^T$, $Z^{(n)}$ is the length of $\boldsymbol{\pi}^{(n)}$, $\mathbf{P}^{(n)} = [P_1^{(n)}, P_2^{(n)}, \dots, P_K^{(n)}]$, contains the instantaneous BER for the n th user and $\mathbf{Y}^{(n)} \in \mathbb{B}^{K \times Z^{(n)}}$ is a binary matrix that represents the number of bits assigned at every

TABLE I: Symbols definitions.

$v_k^{(n)}$	$\log_2 \sqrt{M_k^{(n)}}$
$\check{v}_k^{(n)}$	$\sqrt{M_k^{(n)}}$
M_k	$M_k^{(1)} \times M_k^{(2)}$
$\Lambda_k^{(n)}$	$\check{v}_k^{(n)} - 1$
$\lambda_{i,t,k}^{(n)}$	$\left\lfloor \frac{t2^{i-1}}{\check{v}_k^{(n)}} \right\rfloor$
S_k	$(-1)^{\left\lfloor \frac{l \times 2^{v_k^{(2)} + i - 1}}{\check{v}_k^{(2)}} \right\rfloor} + \lambda_{i,t,k}^{(2)}$
$L_{1,i,k}^{(n)}$	$(1 - 2^{-i})\check{v}_k^{(n)} - 1$
$g_n^\pm(a, b)$	$\gamma(a\sqrt{\beta_1} \pm b\sqrt{\beta_2})$
$Q(g_1(a, b))$	$Q(g_1^+(a, b)) - Q(g_1^-(a, b))$
$D_1^{(n)}$	$(-1)^{\lambda_{i,t,k}^{(n)}} D_2^{(n)}$
$D_2^{(n)}$	$2^{i-1} - \left\lfloor \frac{t2^{i-1}}{\sqrt{M_k^{(n)}}} + \frac{1}{2} \right\rfloor$
$D_3^{(n)}$	$2^{v_k^{(n)}} - \left\lfloor \frac{l}{2} \times 2^{(i-1)\log_2(\check{v}_k^{(n)} - 1)} + \frac{1}{2} \right\rfloor$
c_i	$2 - \delta_{i,1}$

subcarrier,

$$\mathbf{Y}^{(n)} = \begin{bmatrix} y_{11}^{(n)} & y_{12}^{(n)} & \cdots & y_{1Z^{(n)}}^{(n)} \\ y_{21}^{(n)} & y_{22}^{(n)} & \cdots & y_{2Z^{(n)}}^{(n)} \\ \cdots & \cdots & \cdots & \cdots \\ y_{K1}^{(n)} & \cdots & \cdots & y_{KZ^{(n)}}^{(n)} \end{bmatrix} \quad (14)$$

where $y_{kz^{(n)}}^{(n)}$ is a binary variable to indicate the number of bits assigned for the k th subcarrier. The number of bits allocated to the n th user at the k th subcarrier can be interpreted by knowing the index $z^{(n)}$. For example, assume that $\boldsymbol{\pi}^{(n)} = [0, 2, 4, 8]^T$, $\forall n$, $K = 2$, and the output of (12) is $\mathbf{Y}^{(n)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$. Then, the number of bits at $k = 1$ is 2 because $y_{12}^{(n)} = 1$, and the number of bits at $k = 2$ is 4 because $y_{13}^{(n)} = 1$.

IV. BIT LOADING FOR NOMA

Generally speaking, bit loading for OFDM-NOMA is substantially more complicated than single user's systems, i.e., OMA, which is due to the mutual IUI, complexity of computing the BER, and the relation between the modulation order and power coefficients allocated for each user. Therefore, in this work, we propose a low complexity spectrally efficient algorithm to enable maximizing the total OFDM-NOMA system throughput while satisfying the BER constraints for each user. The proposed algorithm is based on the IAA [16], [17], [30], which is modified to suit the special structure of OFDM-NOMA. In particular, the proposed algorithm maximizes the total system throughput by assigning each user a certain number of bits per subcarrier based on the BER constraints of

both users because the number of bits for a particular user affects the BER of the other user due to the mutual IUI. The algorithm allows allocating a particular subcarrier exclusively for a particular user if the fading is severe for the other user as the BER constraint cannot be maintained. On the other hand, if the fading is not severe, then both users can share that subcarrier.

A. Proposed Bit Loading Algorithm for NOMA

The proposed bit loading algorithm is based on the principle of the IAA for single user with BER constraints [16]. The single user IAA assigns certain number of bits to every subcarrier k while satisfying the BER requirements. This algorithm starts with ordering $\alpha_k^{(n)}, \forall \{k\}$ for the n th user in an ascending manner, then initially allocates the maximum number of bits, $b_{\max}^{(n)}$, to every subcarrier. After that, it decrements the assigned number of bits until the BER constraint is satisfied. It should be noted that for the IAA to be optimal, two conditions have to be met: the $\alpha_k^{(n)}, \forall \{k\}$ need to be ascendingly ordered and at every iteration only the minimum possible number of bits can be reduced. This process is applicable for systems which allocate a single user per subcarrier, however, for NOMA systems, some modifications should be performed to account for the existing mutual interference. The proposed bit loading algorithm starts with user δ and then implements the IAA taken into account the reduced power coefficients, β_δ and the IUI. After that, the algorithm is implemented for the other user. It should be noted that at every iteration, the BER constraint is checked until it is satisfied for n th user. Due to the IUI, the IAA does not guarantee optimality for both users, hence, another algorithm is proposed to increase the number of bits per subcarrier. First, $\alpha_k^{(n)}, \forall \{k\}$ are ordered decrementally to ensure that the subcarrier with good channel condition is loaded with the largest possible modulation order, then the number of bits is increased per subcarrier and both users' BER constraints are checked. This process continues as long as the BER constraints of both users are satisfied.

The proposed bit loading mechanism is described in Algorithm 1 consists of two main sub-algorithms, Algorithm 1.1 and Algorithm 1.2, where $I_S = \begin{cases} 0, & \text{PL} \\ 1, & \text{FL} \end{cases}$ represents the initial solution type, δ denotes the starting user, $s = \begin{cases} 1, & \delta = 1 \\ -1, & \delta = 2 \end{cases}$, $\mathbf{Y}^{(n)}(0)$ is the n th user initial solution and $\hat{k} \in [1, \hat{K}]$ and $\check{k} \in [1, \check{K}]$ are, respectively, the indices for the ascendingly and descendingly ordered subcarriers. Algorithm 1.1 implements the IAA and Algorithm 1.2 which performs the incremental allocation of the bits per subcarrier. Given that the CSI is perfectly estimated, Algorithm 1 starts by inputting several parameters, computes the initial solutions $\mathbf{Y}^{(n)}(0), \forall n$, based on I_S , and computes $\bar{P}_B^{(n)}, \forall n$, as shown in steps 1 – 3. It should be noted I_S is decided based on the individual QoS requirements of the users. Then, if C_2 in (13b) is satisfied for both users, Algorithm 1 terminates and $\mathbf{Y}^{(n)}(0), \forall n$, are considered as the final output, otherwise, Algorithms 1.1 and 1.2 run successively. For the n th user, Algorithm 1.1 starts with the subcarrier that has the lowest $\alpha_k^{(n)}$, decrements its number of bits and validates C_2 . Then, at

the second lowest $\alpha_k^{(n)}$ similar operation is performed and so on until $\bar{P}_B^{(n)} \leq 0$ is true. The ascending ordering of $\alpha_k^{(n)}$ allows the algorithm to start from the worst channel coefficient to reduce IUI and improve the throughput for the subcarriers with better channel conditions. The reduction of the number of bits at each subcarrier $\hat{k}$ is done by moving $y_{\hat{k}z(n)}^{(n)} = 1$ to the left once at each iteration. After every movement, the constraint C_2 is checked against violation and once it is satisfied, the corresponding $\mathbf{Y}^{(n)}$ is considered as the output of Algorithm 1.1.

As stated earlier, Algorithm 1.1 is not optimal, hence, the goal of Algorithm 1.2 is to further increase the individual and total system throughput based on the BER constraints. First, the channel coefficients $\alpha_k^{(n)}, \forall k$, are descendingly ordered to ensure that subcarriers with better $\alpha_k^{(n)}$ are loaded with the largest possible number of bits. The steps 27 – 30 are performed to increase the number of bits by moving $y_{\hat{k}z(n)}^{(n)} = 1$ to the right. After that, at every movement, C_2 is validated for both users and if any of $\bar{P}^{(1)} \geq 0$ and $\bar{P}^{(2)} \geq 0$ is valid, then Algorithm 1.2 ends and the pervious iteration's $\mathbf{Y}^{(n)}$ is considered as the output of Algorithm 1. The algorithm is terminated because there is no possibility to obtain better throughput by further considering subcarriers with weaker channel conditions.

Algorithm 1

```

1. Input:  $K, \gamma, \boldsymbol{\pi}^{(n)}, \bar{P}_{Th}^{(n)}, \beta_n, d_n, \alpha_k^{(n)}, \forall \{n, k\}, I_S, \delta, s, \alpha$ 
2. Compute:  $\mathbf{Y}^{(1)}(0), \mathbf{Y}^{(2)}(0)$ 
3. Compute:  $\bar{P}_B^{(1)}, \bar{P}_B^{(2)}$  from (13b)
4. if  $\bar{P}_B^{(1)} \leq 0 \ \& \ \bar{P}_B^{(2)} \leq 0$  then
5.   Stop & Output  $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}$ 
6. else
7.   for  $\vartheta = 1 : 1 : 2$ 
8.     for  $n = \delta : s : s + \delta$ 
9.       Order:  $\alpha_k^{(n)}, \forall \{n, k\}$ 
10.      Call: Algorithm 1.  $\vartheta$ 
11.    end for
12.  end for
13. end

```

Algorithm 1.1

```

14. for  $\hat{k} = 1 : 1 : \hat{K}$ 
15.   if  $y_{\hat{k}1}^{(n)} \neq 1$  then
16.     Find:  $y_{\hat{k}z(n)}^{(n)} = 1$ 
17.     for  $i = 1 : 1 : Z^{(n)} - 1$ 
18.       Move to Left:  $y_{\hat{k}(z-i)}^{(n)} = 1$ 
19.       Compute:  $\bar{P}_B^{(1)}$  from (13b)
20.       if  $\bar{P}_B^{(n)} \leq 0$  then
21.         Stop & Output  $\mathbf{Y}^{(n)}$ 
22.       end if
23.     end for
24.   end if
25. end for

```

Algorithm 1.2

```

26. for  $\tilde{k} = 1 : 1 : \tilde{K}$ 
27.     if  $y_{\tilde{k}Z^{(n)}}^{(n)} \neq 1$  then
28.         Find:  $y_{\tilde{k}z^{(n)}}^{(n)} = 1$ 
29.         for  $i = 1 : 1 : Z^{(n)} - 1$ 
30.             Move to Right:  $y_{\tilde{k}(z+i)^{(n)}}^{(n)} = 1$ 
31.             Compute:  $\bar{P}_B^{(1)}, \bar{P}_B^{(2)}$  from (13b)
32.             if  $\bar{P}_B^{(1)} \geq 0 \parallel \bar{P}_B^{(2)} \geq 0$  then
33.                 Stop & Output previous  $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}$ 
34.             end if
35.         end for
36.     end if
37. end for

```

B. Initial Solution: $\mathbf{Y}^{(n)}(0), \forall n$

The initial solution $\mathbf{Y}^{(n)}(0), \forall n$ of Algorithm 1 can be categorized into two types, FL and PL.

1) *Fully-Loaded (FL) Initial Solution:* As for the FL initial solution, both U_1 and U_2 are assigned the maximum number of bits $b_{\max}^{(n)}$, i.e., $y_{kZ^{(n)}}^{(n)} = 1, \forall k, \forall n$.

2) *Partially-Loaded (PL) Initial Solution:* In order to reduce the number of iterations required to obtain the final solution of Algorithm 1, $\mathbf{Y}^{(n)}, \forall n$, which satisfies the system requirements, the initial solution $\mathbf{Y}^{(n)}(0), \forall n$, can be computed by optimally solving the bit loading problem for a single user [16]. In other words, to get $\mathbf{Y}^{(1)}(0)$, Algorithm 1 is implemented where β_2 is set to 0. The final output $\mathbf{Y}^{(1)}$ is considered as the initial solution for $U_1, \mathbf{Y}^{(1)}(0)$. It should be noted that the IUI from U_2 does not exist as the allocated power is 0, i.e., $\mathbf{Y}^{(2)} = \mathbf{0}$. The same process is applied to obtain $\mathbf{Y}^{(2)}(0)$ where β_1 is set to 0. It should be noted that the PL initial solutions are optimal and can be obtained using the IAA for single user.

V. COMPUTATIONAL COMPLEXITY

In this section, the computational complexity of the proposed bit loading algorithm is provided. The computational complexity of bit loading algorithms with BER constraints is mostly dominated by the required extensive BER calculations performed during the iterative optimization process [24]. In particular, the complexity is computed in terms of the number of real multiplication ($\times$) and addition ($\pm$) operations, and number of times the $Q(\cdot)$ and $\lfloor \cdot \rfloor$ functions are evaluated in the BER expressions (9), (10) and (11). For more informative comparison, the overall equivalent complexity is also presented [31]. Moreover, the computational complexity for the exponent terms, 2^v where a v is real integer, in the BER expressions is neglected as they can be implemented using simple shift registers. Table II presents the computational complexity for the scenario where a single user is transmitting at subcarrier k while Tables III and IV show the computational complexity of the BER calculations for U_1 and U_2 , respectively,

TABLE II: Computational complexity of $P_k^{(n)}|_{\Gamma_k=0}$

Operation	$P_k^{(n)} _{\Gamma_k=0}$
$\times$	$5v_k^{(n)} + 1$
$\pm$	$3v_k^{(n)} + 4 \sum_{i=1}^{v_k^{(n)}} L_{1,i,k}^{(n)}$
$Q(\cdot)$	$v_k^{(n)} + \sum_{i=1}^{v_k^{(n)}} L_{1,i,k}^{(n)}$
$\lfloor \cdot \rfloor$	$2 \times \left(v_k^{(n)} + \sum_{i=1}^{v_k^{(n)}} L_{1,i,k}^{(n)} \right)$

TABLE III: Computational complexity of $P_k^{(1)}|_{\Gamma_k=1}$.

Operation	$P_k^{(1)} _{\Gamma_k=1}$
$\times$	$1 + 6 \left(\Lambda_k^{(2)} + 1 \right) \left[v_k^{(1)} + \sum_{i=1}^{v_k^{(1)}} L_{1,i,k}^{(1)} \right]$
$\pm$	$v_k^{(1)} \left(6\Lambda_k^{(2)} + 5 \right) + 6 \left(\Lambda_k^{(2)} + 1 \right) \sum_{i=1}^{v_k^{(1)}} L_{1,i,k}^{(1)}$
$Q(\cdot)$	$\left(\Lambda_k^{(2)} + 1 \right) \left[v_k^{(1)} + \sum_{i=1}^{v_k^{(1)}} L_{1,i,k}^{(1)} \right]$
$\lfloor \cdot \rfloor$	$2 \times \left(\Lambda_k^{(2)} + 1 \right) \left[v_k^{(1)} + \sum_{i=1}^{v_k^{(1)}} L_{1,i,k}^{(1)} \right]$

where both are transmitting. It can be noticed that the computational complexity is a function of the modulation order $M_k^{(n)}$ and user's index. Clearly, the computational complexity of $P_k^{(1)}|_{\Gamma_k=1}$ and $P_k^{(2)}|_{\Gamma_k=1}$ is substitutionally larger than $P_k^{(n)}|_{\Gamma_k=0}$ which is due to the IUI introduced in NOMA systems. In order to quantify the number of operations required, Tables V, VI and VII represent numerical computational complexity results for various $M_k^{(n)}$ values. The difference between the number of operations between $P_k^{(n)}|_{\Gamma_k=0}$ and $P_k^{(n)}|_{\Gamma_k=1}, \forall n$, is significant. For example, the overall computational complexity of $P_k^{(1)}|_{\Gamma_k=1}$ and $P_k^{(2)}|_{\Gamma_k=1}$ are, respectively, 2 and 6 times larger than $P_k^{(n)}|_{\Gamma_k=0}$, for $M_k^{(n)} = 4, \forall n$. It can be noticed that the computational complexity increases exponentially as $M_k^{(n)}$ increases when U_1 and U_2 are both transmitting while linearly when one of them is transmitting. Consequently, there is a critical tradeoff between the computational complexity and throughput requirements for OFDM-NOMA systems which should be carefully tackled as a function of $M_k^{(n)}$, BER constraints and user's index.

TABLE IV: Computational complexity of $P_k^{(2)}|_{\Gamma_k=1}$.

Operation	$P_k^{(2)} _{\Gamma_k=1}$
$\times$	$1 + 6 \left(4\Lambda_k^{(1)} + 1 \right) \left(v_k^{(2)} + \sum_{i=1}^{v_k^{(2)}} L_{1,i,k}^{(2)} \right)$
$\pm$	$v_k^{(2)} \left(24\Lambda_k^{(1)} + 5 \right) + \left(24\Lambda_k^{(1)} + 7 \right) \sum_{i=1}^{v_k^{(2)}} L_{1,i,k}^{(2)}$
$Q(\cdot)$	$\left(4\Lambda_k^{(1)} + 1 \right) \left[\sum_{i=1}^{v_k^{(2)}} L_{1,i,k}^{(2)} + v_k^{(2)} \right]$
$\lfloor \cdot \rfloor$	$4 \times \left(4\Lambda_k^{(1)} + 1 \right) \left[\sum_{i=1}^{v_k^{(2)}} L_{1,i,k}^{(2)} + v_k^{(2)} \right]$

TABLE V: Numerical computational complexity of $P_k^{(n)}|_{\Gamma_k=0}$.

$M_k^{(n)}$	Overall Equivalent Complexity	$Q(.)$	$\lfloor . \rfloor$
4	27	1	2
16	62	5	10

TABLE VI: Numerical computational complexity of $P_k^{(1)}|_{\Gamma_k=1}$.

$M_k^{(1)}, M_k^{(2)}$	Overall Equivalent Complexity	$Q(.)$	$\lfloor . \rfloor$
4, 4	63	2	4
16, 16	602	20	40

VI. NUMERICAL RESULTS

This section presents the performance of the proposed algorithm concerning throughput and complexity. The simulation results are obtained using a computing machine that runs Intel Xeon CPU E5-2640 processor, clock frequency of 2.5 GHz, 16 GB RAM, and 64 bit operating system. The multicarrier NOMA system considered is an OFDM system with a total of $K = 64$ for U_1 and U_2 . Also, any subcarrier k can be turned OFF for both of the users, $y_{k1}^{(n)} = 1, \forall n$. In order to obtain a reliable BER performance, the possible modulation orders for each user are set to either QPSK or 16-QAM. Therefore, if both users are transmitting at the k th subcarrier, the superposed signal modulation order is going to be either 16-QAM or 256-QAM. The BER thresholds $\bar{P}_{Th}^{(1)}$ and $\bar{P}_{Th}^{(2)}$ are, respectively, set to 10^{-2} and 10^{-2} , and 10^{-2} and 10^{-3} . For the case where both users are transmitting, β_1 is chosen based on (2), $\beta_1 = [0.92, 0.95, 0.99]$. If a single user is decided to transmit, then $\beta_n = 1$. The two different initial solutions, FL and PL are considered.

The n th channel is modeled as a quasi-static fading channel, where the channel remains fixed during a given OFDM symbol, but changes randomly between adjacent symbols. The Rayleigh frequency selective channel is modelled as a 15-tap multipath channel with normalized delays of $[0, 1, 4, 5, 12, 14]$ and average gains of $[0.4584, 0, 0, 0, 0.147, 0.0928, 0, 0, 0, 0, 0, 0, 0, 0, 0]$ [24]. The distances between the two users and the BS are $d_1 = 1.3m$ and $d_2 = 1m$, and the path loss is assumed $\alpha = 2$. The results are obtained by running and averaging Monte Carlo simulations for 10^4 channel realizations. For each channel realization, Algorithm 1 is run at 26 different average SNR values, $\bar{\gamma}$, ranging from 0 to 50 dB.

Figures 1 and 2 represent the throughputs of U_1 and U_2 , respectively, for different power coefficients, and $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-2}$. Due to the existence of IUI in NOMA systems, the interpretation of the results has to be done while considering the throughput behavior of both users simultaneously. In the presented results,

TABLE VII: Numerical computational complexity of $P_k^{(2)}|_{\Gamma_k=1}$.

$M_k^{(1)}, M_k^{(2)}$	Overall Equivalent Complexity	$Q(.)$	$\lfloor . \rfloor$
4, 4	153	5	20
16, 16	1955	65	260

the starting user, e , and initial solution type are varied and their impacts are investigated. The single user's optimal bit loading solution obtained by running Algorithm 1 where the IUI is totally eliminated is referred to as the "SU" solution [16]. In both figures, the SU solution where $\beta_n = 1$ sets as an upper bound for Algorithm 1 solutions due to the absence of the IUI which results in lower BER and hence higher throughput per user. For fair comparison with the proposed system, another SU solution is obtained for U_1 where $\beta_1 = [0.92, 0.95, 0.99]$ and $\beta_2 = 0$ as shown in Fig. 1. Also, for U_2 , a SU solution is provided such that $\beta_2 = [0.08, 0.05, 0.01]$ and $\beta_1 = 0$ as shown in Fig. 2. As expected, these two solutions do not upper bound the proposed algorithm solutions due to the reduced power coefficients, specifically, over the low SNR region. The main reason is that for low SNR values, it is highly probable that only a single user n will be allowed to transmit to be able to satisfy C_2 who unlike the SU solution will be allocated a unity power, $\beta_n = 1$, thus the throughput becomes higher. As for the high SNR region, the impact of the reduced power allocation becomes less on the BER, therefore, both optimal SU solutions converge to each other and upper bound Algorithm 1 solutions. Moreover, it should be noted that for U_1 , increasing β_1 , increases the throughput for all the algorithm solutions and reduces the gap with the associated SU solutions. This is because as β_1 increases, the IUI impact becomes more negligible, i.e., $M_k^{(2)}$ has less impact on the choice of $M_k^{(1)}$, $\forall k$ in terms of the error performance [29]. Therefore, by excessively increasing β_1 , the throughput of U_1 becomes almost independent from the other user as can be clearly seen in Fig. 1 for $\beta_1 = 0.99$. On contrary, increasing β_2 does not necessarily guarantee enhanced throughput performance as can be noted in Algorithm 1 solutions where $\beta_2 = 0.05$ and 0.08 in Fig. 2. This is due to the error performance of the two stages of SIC and the associated error propagation. Thus, it is important to properly allocate power coefficients which can provide optimal BER performance at both stages to enhance the corresponding throughputs. For general SU bit loading algorithms, there is a proportional relationship between the throughput and SNR [16]–[19]. However, for NOMA systems where users have conflicting error rate requirements, increasing the SNR does not necessarily guarantee increasing the throughput as there are other factors that must be taken into account such as β_n and $M_k^{(n)}$, $\forall k$ as can be seen in both figures. In the proposed algorithm, the channel coefficients $\alpha_k^{(e)}$, $\forall k$ are ordered ascendingly. As a result, we start by lowering $M_k^{(n)}$ for the subcarriers with the lower $\alpha_k^{(e)}$ which cannot afford high $M_k^{(n)}$, hence reducing $\bar{P}^{(n)}$ and allowing other subcarriers with better channel conditions to have higher $M_k^{(n)}$. There are four different solutions that are provided for both users where the initial solution type and starting user, e , are varied as: [FL, 1], [PL, 1], [FL, 2], and [PL, 2]. In Fig. 1, for $\beta_1 = 0.92$, from $\bar{\gamma} = 0$ to 34 dB, both [FL, 2] and [PL, 2] solutions perform better than the other solutions starting from U_1 . Actually, starting the proposed algorithm with U_2 reduces the IUI on U_1 which results in a very similar throughput performance to the optimal SU solution. On the other hand, the [FL, 1] and [PL, 1] solutions allow the existence of IUI coming from U_2 which reduces the chance to satisfy $\bar{P}_{Th}^{(1)}$, hence decreasing the corresponding throughput. In the high SNR range, $\bar{\gamma} = 35$ to 50 dB,

the throughput performance of the four solutions becomes almost identical as the BER reduces greatly and hence the BER threshold can be satisfied, regardless of the initial solution or starting user. There are very small differences between the solutions starting from FL and PL initial solutions which can be neglected. Similar observations can be seen when increasing β_1 . An opposite behavior can be noticed in Fig. 2 where higher throughput performance is achieved when the algorithm starts with U_1 which clearly reduces the IUI on U_2 . Again, both FL and PL solutions result in similar throughput performance. It should be noted that for the low SNR region, when starting with U_1 , the IUI is considered very low and hence the solution is almost identical to the SU solution with $\beta_2 = 1$.

Figures 3 and 4 show, respectively, the throughputs of U_1 and U_2 for different power coefficients, and $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-3}$. Similar observations can be noticed as in Figures 1 and 2. In addition, in the low and moderate SNR regions, U_1 has higher throughput compared to the case where $\bar{P}_{Th}^{(2)} = 10^{-2}$ which is due to the reduced IUI of U_2 . As for U_2 , the achieved throughput is less as it is more challenging to satisfy $\bar{P}_{Th}^{(2)} = 10^{-3}$ constraint.

Figures 5 and 6 present the total throughput of the proposed system for different power coefficients, and $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-2}$, and $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-3}$, respectively. It can be noticed that both the FL and PL initial solutions perform almost identically over the full SNR range. Due to the conflicting BER performance requirements of the individual users, increasing β_1 does not necessarily increase the total throughput over the full SNR range. It can be noted that when $\beta_2 = 0.95$, we can obtain the best throughput performance. Therefore, the need to proper power allocation is crucial in NOMA systems to maximize the total system throughput. Moreover, the best performance can be obtained from the algorithm solution starting from U_1 over the majority of SNR regions as it has more impact on the system throughput performance. Given that K subcarriers are allocated for both SU (with reduced β_n) and NOMA systems, the proposed solution outperform the optimal SU solution in terms of the throughput performance over all possible SNR values and considered scenarios. The difference between the two solutions increases as the SNR increases. For example, in Fig. 3, for $\beta_1 = 0.99$ and $\bar{\gamma} = 10$ dB, the ratio is 48% while for $\bar{\gamma} = 30$ dB, the ratio is 57%.

Figures 7 and 8 depict the modulation distribution for the proposed system at $\bar{\gamma} = 10$ and 30 dB, $\beta_1 = 0.95$, and $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-2}$. Also, Figures 9 and 10 present the modulation distribution at $\bar{\gamma} = 10$ and 30 dB, $\beta_1 = 0.95$, and $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-3}$. The modulation order when single user transmission is decided is denoted as $(M_k^{(n)})^+$. For low SNR values, single user transmission is more likely to occur as it can provide less BER as can be seen in Fig. 7. As stated earlier, a subcarrier k can be turned OFF for both users especially at low SNR region as shown in Figures 7 and 9. This case is referred to as "NULL". Moreover, as U_2 has lower power than U_1 , then it is expected to have lower throughput and lower modulation orders. As for the case where $\bar{\gamma} = 30$ dB, as expected, the NULL percentage is almost zero as C_2 can be satisfied for both users. Also, at such high SNR value, the simultaneous transmission from both users is highly probable with superposed modulation

TABLE VIII: Average computational time (milliseconds) for Algorithm 1, $\beta_1 = 0.92$.

$\bar{\gamma}$	10 dB			30 dB		
$e, \bar{P}_{Th}^{(2)}$	T_{FL}	T_{PL}	Ratio	T_{FL}	T_{PL}	Ratio
1, 10^{-2}	441.43	255.47	58%	142.649051	92.82	65%
1, 10^{-3}	441.46	276.38	63%	331.49	226.81	68%
2, 10^{-2}	350.48	179.93	51%	120.59	63.36	53%
2, 10^{-3}	352.76	201.23	57%	277.52	167.77	60%

orders, 16-QAM or 256-QAM. Due to the power assigned to U_2 , lower modulation levels can be obtained. The results show that as $\bar{P}_{Th}^{(n)}, \forall n$ increases the modulation levels increases as well as the simultaneous transmission percentage.

The computational efficiency is a critical requirement in several delay-sensitive applications, hence, we provide the average computational time required for several cases of interest as shown in Table VIII. The time required to obtain the throughput results varies depending on the initial solution type and starting user. It can be noted that starting with a PL solution reduces the computational time as compared to starting with a FL solution. To quantify the reduction, the ratio percentage is calculated as $\frac{T_{PL}}{T_{FL}}\%$, where T_{PL} and T_{FL} are the time needed to compute Algorithm 1 when starting with PL and FL initial solutions, respectively. The computational time decreases as $\bar{\gamma}$ increases as it becomes easier to satisfy the BER constraints for both users. Moreover, as $\bar{P}_{Th}^{(n)}, \forall n$, decreases, the computational time increases as U_n requires more time to satisfy C_2 . As can be concluded from the table, it is more desirable to start with U_2 and PL initial solution for all considered scenarios.

Furthermore, Figures 11 and 12 show the number of iterations needed to obtain the outputs of Algorithm 1, at $\bar{\gamma} = 20$ dB, $\beta_1 = 0.95$, and $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-2}$, and $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-3}$.

When starting with U_1 , the number of iterations required for U_1 is higher than U_2 which is because of the IUI reduction on U_2 , and vice versa. Moreover, as $\bar{\gamma}$ increases the number of iterations reduces for both users as the average BER threshold can be satisfied faster. Also, $\bar{P}_{Th}^{(n)}, \forall n$, affects the number of iterations, for example, when $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-2}$ are considered, the number of iterations for U_1 is higher than the other case where $\bar{P}_{Th}^{(1)} = 10^{-2}$ and $\bar{P}_{Th}^{(2)} = 10^{-3}$ because of the increased IUI, while U_2 requires higher number of iterations as it takes more iterations to reach $\bar{P}_{Th}^{(2)} = 10^{-3}$.

VII. CONCLUSION

In this work, an adaptive bit loading algorithm for OFDM-NOMA systems is proposed. The objective is to maximize the total system throughput while satisfying the users' BER requirements. The computational complexity due to the mutual interference and the dependence on power coefficients and modulation orders, impact of the

Fig. 1: The throughput for U_1 , where $\bar{P}_{Th}^{(1)} = 10^{-2}$, $\bar{P}_{Th}^{(2)} = 10^{-2}$, $\beta_1 = [0.92, 0.95, 0.99]$.

starting user, and the initial solution type are investigated. The proposed algorithm is based on the IAA, however, different starting points and loading directions, incremental or decremental, are considered. The impact of the initial loading is also considered to reduce the number of iterations and computational time required to provide reliable results. The performance is presented in terms of the throughput per user, total throughput, and computational complexity, and compared to OMA. The obtained results show that NOMA significantly outperforms OMA in terms of throughput due to the cooperation between the two users by 100%. To reduce the complexity, the bit loading process is evaluated using two different initial starting points, FL and PL. The obtained complexity results show that the PL approach can reduce the complexity by about 50% for several cases of interest compared to FL solutions which are commonly used in OMA bit loading algorithms.

REFERENCES

- [1] S. Mattisson, "An overview of 5G requirements and future wireless networks: Accommodating scaling technology," *IEEE Solid-State Circuits Mag.*, vol. 10, no. 3, pp. 54–60, May 2018.
- [2] L. Bariah, S. Muhaidat, and A. Al-Dweik, "Error probability analysis of NOMA-based relay networks with SWIPT," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1223–1226, Jul. 2019.

Fig. 2: The throughput for U_2 , where $\bar{P}_{Th}^{(1)} = 10^{-2}$, $\bar{P}_{Th}^{(2)} = 10^{-2}$, $\beta_2 = [0.08, 0.05, 0.01]$.

- [3] E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 3, pp. 11–21, Jun. 2015.
- [4] M. Kalil, A. Al-Dweik, M. F. Abu Sharkh, A. Shami, and A. Refaey, "A framework for joint wireless network virtualization and cloud radio access networks for next generation wireless networks," *IEEE Access*, vol. 5, pp. 20 814–20 827, 2017.
- [5] M. Kalil, A. Moubayed, A. Shami, and A. Al-Dweik, "Efficient low-complexity scheduler for wireless resource virtualization," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 56–59, Feb. 2016.
- [6] Q. Nadeem, A. Kammoun, M. Debbah, and M. Alouini, "Design of 5G full dimension massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 726–740, Feb. 2018.
- [7] W. Pramudito *et al.*, "Load-aware energy efficient adaptive large scale antenna system," *IEEE Access*, vol. 8, pp. 82 592–82 606, 2020.
- [8] H. Mukhtar, A. Al-Dweik, and A. Shami, "Turbo product codes: Applications, challenges, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 3052–3069, 4th Quart. 2016.
- [9] M. Al-Imari, M. A. Imran, R. Tafazolli, and D. Chen, "Subcarrier and power allocation for LDS-OFDM system," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2011, pp. 1–5.
- [10] F. Kara and H. Kaya, "On the error performance of cooperative-NOMA with statistical CSIT," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 128–131, Jan. 2019.
- [11] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart. 2017.
- [12] J. Ding, J. Cai, and C. Yi, "An improved coalition game approach for MIMO-NOMA clustering integrating beamforming and power allocation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1672–1687, Feb. 2019.

Fig. 3: The throughput for U_1 , where $\bar{P}_{Th}^{(1)} = 10^{-2}$, $\bar{P}_{Th}^{(2)} = 10^{-3}$, $\beta_1 = [0.92, 0.95, 0.99]$.

- [13] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart. 2018.
- [14] A. Tusha, S. Dogan, and H. Arslan, "A hybrid downlink NOMA with OFDM and OFDM-IM for beyond 5G wireless networks," *IEEE Signal Process. Lett.*, vol. 27, pp. 491–495, Mar. 2020.
- [15] A. Adnan, Y. Liu, C. Chow, and C. Yeh, "Demonstration of non-hermitian symmetry (nhs) ifft/fft size efficient ofdm non-orthogonal multiple access (noma) for visible light communication," *IEEE Photon. J.*, vol. 12, no. 3, pp. 1–5, Apr. 2020.
- [16] A. M. Wyglinski, F. Labeau, and P. Kabal, "Bit loading with BER-constraint for multicarrier systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1383–1387, Sept. 2005.
- [17] F. Kalbat, A. Al-Dweik, Y. Iraqi, H. Mukhtar, B. Sharif, and G. K. Karagiannidis, "Direct bit loading with reduced complexity and overhead for precoded OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7169–7173, July 2019.
- [18] Q. Sun, L. Li, A. Tolli, M. Juntti, and J. Mao, "Optimal energy efficient bit and power loading for multicarrier systems," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1194–1197, Apr. 2014.
- [19] D. Wang, Y. Cao, L. Zheng, and Z. Du, "Iterative group-by-group bit-loading algorithms for OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 4131–4135, Oct. 2013.
- [20] Z. Song, Q. Ni, and X. Sun, "Spectrum and energy efficient resource allocation with QoS requirements for hybrid MC-NOMA 5G systems," *IEEE Access*, vol. 6, pp. 37 055–37 069, July 2018.
- [21] J. Tseng, Y. Chen, and C. Wang, "User selection and resource allocation algorithms for multicarrier NOMA systems on downlink beamforming," *IEEE Access*, vol. 8, pp. 59 211–59 224, Mar. 2020.
- [22] S. Luo and K. C. Teh, "Adaptive transmission for cooperative NOMA system with buffer-aided relaying," *IEEE Communications Letters*,

Fig. 4: The throughput for U_2 , where $\bar{P}_{Th}^{(1)} = 10^{-2}$, $\bar{P}_{Th}^{(2)} = 10^{-3}$, $\beta_2 = [0.08, 0.05, 0.01]$.

vol. 21, no. 4, pp. 937–940, 2017.

- [23] D. J. Dechene and A. Shami, “Energy-aware resource allocation strategies for LTE uplink with synchronous HARQ constraints,” *IEEE Transactions on Mobile Computing*, vol. 13, no. 2, pp. 422–433, Dec. 2014.
- [24] Y. Iraqi and A. Al-Dweik, “Adaptive bit loading with reduced computational time and complexity for multicarrier wireless communications,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 3, pp. 2497–2506, June 2020.
- [25] I. Lee and J. Kim, “Average symbol error rate analysis for non-orthogonal multiple access with m -ary QAM signals in Rayleigh fading channels,” *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1328–1331, Aug. 2019.
- [26] T. Assaf, A. Al-Dweik, M. S. E. Moursi, H. Zeineldin, and M. Al-Jarrah, “NOMA receiver design for delay-sensitive systems,” *IEEE Syst. J.*, pp. 1–12, Nov. 2020.
- [27] T. Assaf, A. Al-Dweik, M. E. Moursi, and H. Zeineldin, “Exact BER performance analysis for downlink NOMA systems over Nakagami- m fading channels,” *IEEE Access*, vol. 7, pp. 134 539–134 555, Sep. 2019.
- [28] Kyongkuk Cho and Dongweon Yoon, “On the general BER expression of one- and two-dimensional amplitude modulations,” *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, July 2002.
- [29] T. Assaf, A. J. Al-Dweik, M. S. E. Moursi, H. Zeineldin, and M. Al-Jarrah, “Exact bit error-rate analysis of two-user NOMA using QAM with arbitrary modulation orders,” *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2705–2709, Aug. 2020.
- [30] Y. Iraqi and A. Al-Dweik, “Adaptive bit loading with reduced computational time and complexity for multicarrier wireless communications,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 3, pp. 2497–2506, June 2020.
- [31] M. Tariq, A. Al-Dweik, B. Mohammad, H. Saleh, and T. Stouraitis, “Computational power evaluation for energy-constrained wireless communications systems,” *IEEE Open J. Commun. Soc.*, vol. 1, pp. 308–319, Mar. 2020.

Fig. 5: The total throughput for the two users, where $\bar{P}_{Th}^{(1)} = 10^{-2}$, $\bar{P}_{Th}^{(2)} = 10^{-2}$, $\beta_1 = [0.92, 0.95, 0.99]$.

Fig. 6: The total throughput for the two users, where $\bar{P}_{Th}^{(1)} = 10^{-2}$, $\bar{P}_{Th}^{(2)} = 10^{-3}$, $\beta_1 = [0.92, 0.95, 0.99]$.

Fig. 7: Modulation distributions for U_1 and U_2 , $\bar{\gamma} = 10$ dB, $\beta_1 = 0.95$, $P_{Th}^{(1)} = 10^{-2}$ and $P_{Th}^{(2)} = 10^{-2}$.

Fig. 8: Modulation distribution for U_1 and U_2 , $\bar{\gamma} = 30$ dB, $\beta_1 = 0.95$, $P_{Th}^{(1)} = 10^{-2}$ and $P_{Th}^{(2)} = 10^{-2}$.

Fig. 9: Modulation distribution for U_1 and U_2 , $\bar{\gamma} = 10$ dB, $\beta_1 = 0.95$, $P_{Th}^{(1)} = 10^{-2}$ and $P_{Th}^{(2)} = 10^{-3}$.

Fig. 10: Modulation distribution for U_1 and U_2 , $\bar{\gamma} = 30$ dB, $\beta_1 = 0.95$, $P_{Th}^{(1)} = 10^{-2}$ and $P_{Th}^{(2)} = 10^{-3}$.

Fig. 11: The number of iterations for U_1 and U_2 where $e = 1$, $\bar{\gamma} = 20$ dB and $\beta_1 = 0.95$.

Fig. 12: The number of iterations for U_1 and U_2 where $e = 2$, $\bar{\gamma} = 20$ dB and $\beta_1 = 0.95$.