Design and Analysis of NOMA With Adaptive Modulation and Power Under BLER Constraints

DOI:
10.1109/TVT.2022.3184974

Document Version
Accepted author manuscript

Design and Analysis of NOMA with Adaptive Modulation and Power Under BLER Constraints

Hamad Yahya, *Graduate Student Member, IEEE,* Emad Alsusa, *Senior Member, IEEE,* and Arafat Al-Dweik, *Senior Member, IEEE*

***Abstract*—In this work, we derive closed-form expressions for the throughput of non-orthogonal multiple access (NOMA) and use the expressions to maximize the throughput. The design considers a packet-based transmission where the base station optimizes the modulation orders and power coefficients while satisfying the block error rate requirement for each user. The optimization problem is solved using integer and mixed-integer programming, where the computational complexity is reduced using an efficient stopping criterion, segment-line search, and quantized signal to noise ratios (SNRs). The analytical and simulation results show that selecting the appropriate power and modulation orders can improve the throughput by about 3 dB at high SNRs. The impact of the SNR quantization is evaluated, and the obtained results show that the proposed system can tolerate large quantization steps, which enables reducing the complexity of the optimization and adaptation processes.**

***Index Terms*—Non-orthogonal multiple access (NOMA), block error rate (BLER), throughput, adaptive modulation and power.**

I. Introduction

RECENTLY non-orthogonal multiple access (NOMA) has attracted much attention as it can achieve improved spectral efficiency, support massive connectivity and serve users with diverse quality of service (QoS) requirements [1]. To augment NOMA advantages, adaptive modulation and coding (AMC) have been widely considered [2]–[8]. For example, Wang *et al.* [2] propose an asymmetric adaptive modulation algorithm based on block error rate (BLER) requirements for uplink (UL) two-user NOMA using only binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK). The authors of [3] propose constructing an AMC and power coefficients map for the downlink (DL) via an offline link-level NOMA simulator. Similar to the link-level simulator approach, [4] designs an algorithm aiming at maximizing the minimum user rate while achieving user fairness. The authors of [5] consider a two-user DL multi-carrier NOMA and propose an efficient adaptive modulation algorithm to maximize the throughput while satisfying individual users' bit error rate (BER) constraints. However, the throughput is evaluated only using simulation and the system adopts fixed power assignment (FPA). Similarly, Cejudo *et al.* [6] design an efficient resource allocation algorithm to maximize the sum rate while satisfying the BER constraints.

Hamad Yahya and Emad Alsusa are with the Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, U.K. (email: hamad.mohamadaliyahya@manchester.ac.uk, e.alsusa@manchester.ac.uk)

Arafat Al-Dweik is with the Center for Cyber Physical Systems (C2PS), Khalifa University, Abu Dhabi 127788, UAE, (e-mail: arafat.dweik@ku.ac.ae, dweik@fulbrightmail.org). The work of A. Al-Dweik was supported by Khalifa University competitive internal research Award, grant no. CIRA-2020-056.

Jia *et al.* [7] derived the average spectral efficiency (SE) for the UL two-user NOMA while assuming FPA and adaptive square quadrature amplitude modulation (QAM). Yu *et al.* [8] proposed a joint adaptive modulation and power assignment (PA) algorithm for the DL to achieve a minimum SE target for one of the users and maximize the SE of the other user without sacrificing the BER performance. It is worthy to mention that [7], [8] used the single-user square QAM BER expression rather than the NOMA exact BER expressions. In [9], the necessary and sufficient conditions are derived for the power ranges of NOMA with arbitrary QAM modulation orders and number of users to prevent error floors when successive interference cancellation (SIC) is used. However, the derived power ranges do not specify the optimal power required to achieve a certain BER. Choi [10] studied the PA problem for two-user DL NOMA considering square QAM to maximize the mutual information. Dong *et al.* [11] found the optimal power and phase for two-user UL NOMA with square QAM to maximize the minimum Euclidian distance of the received sum-constellation at the receiver side. Han *et al.* [12] derived BER and symbol error rate expressions for DL NOMA with joint-Gray labeling considering QPSK and arbitrary number of users. Moreover, the BER of DL NOMA system is derived in [1] for arbitrary number of users and QAM modulation orders, where the optimal PA is found considering statistical channel state information (CSI).

In [13], adaptive Bose–Chaudhuri–Hocquenghem coding with constellation rotation is proposed to improve the NOMA BER. Qiu *et al.* [14] propose an algebraic constellation rotation design for a two-user DL NOMA in block fading channels. However, BLER is evaluated via computer simulation. Approximate BLER expressions for short-packet communication NOMA are derived in [15]–[20]. In [15], a two-user DL NOMA is considered with fountain-coding, [16] considered multiple-input-multiple-output-NOMA system, [17], [18] studied Internet of Things networks, [19] studied intelligent reflecting surface-assisted NOMA, and [20] studied residual hardware impairments and channel estimation errors.

Despite the advantages of NOMA, its BER remains generally the main limitation due to the inter-user interference (IUI). Therefore, the power and modulation order for each user have to be carefully selected to ensure that all users can jointly achieve their QoS requirements. Consequently, the power and modulation order should be selected based on the channel conditions and QoS for each user. Therefore, this work considers:

- The design of a NOMA system with adaptive power and QAM modulation orders.
- An efficient low-complexity algorithm is proposed to real-

ize the modulation and power adaptation process.
- To enable efficient design, we derive exact closed-form expressions for the instantaneous throughput (iTp) and closed-form approximation for the average throughput (aTp).
- A novel approach is developed to correlate the selected modulation orders and power, and thus, no side information about the power is required at the receiver side.

It is worth highlighting that [6] considered resource allocation for multicarrier NOMA under BER constraints. Also, the throughput is evaluated through Monte-Carlo simulation. While we consider the DL NOMA, [7] considers the UL NOMA, and derive the information-theoretic SE under the single-user uncoded BER constraint, FPA, and adaptive square QAM.

II. System And Channel Models

This work considers a packet-switched single cell that consists of a base station (BS) that serves two-user NOMA using a single resource block. The BS adapts the modulation orders and power coefficients to maximize the iTp while satisfying the users' instantaneous BLER. The adaptation is applied on a packet-by-packet basis while assuming a block fading channel that is static for L symbols, which is the packet duration [14], [15]. Without loss of generality, the near and far users are denoted as U_1 and U_2, and the allocated power coefficients are α_1 and α_2, where $\alpha_1 + \alpha_2 = 1$. The BS informs the users about the transmission parameters and the users perfectly report their CSI and QoS requirements to the BS through an error-free feedback channel. The received signal for $U_n \, \forall n \in \{1, 2\}$ is

$$\mathbf{y}_n = \mathbf{H}_n(\sqrt{\alpha_1}\mathbf{x}_1 + \sqrt{\alpha_2}\mathbf{x}_2) + \mathbf{w}_n \tag{1}$$

where $\mathbf{y}_n \in \mathbb{C}^{L\times 1}$, the data packet $\mathbf{x}_n = [x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(L)}]^T$, the symbols are drawn uniformly from a BPSK constellation or a rectangular QAM. For each data symbol, $\mathbb{E}[|x_n^{(l)}|^2] = 1 \; \forall l$, and $\mathcal{M}_n = \log_2 M_n$, where M_n is the modulation order, $\mathbb{E}[\cdot]$ denotes the statistical expectation, and $(\cdot)^T$ denotes the transpose operation. The additive white Gaussian noise vector $\mathbf{w}_n \sim \mathcal{CN}(0, N_0\mathbf{I}_L)$ where $\sigma^2_{w_n} = N_0/2$ and $\mathbf{I}_L$ is the identity matrix with a size of $L \times L$. The complex channel gain for single-input-single-output channel is $\mathbf{H}_n \in \mathbb{C}^{L\times L}$, which corresponds to block Rayleigh fading over a one packet interval, that is $\mathbf{H}_n = h_n\mathbf{I}_L$, $h_n = \check{h}_n \times d_n^{-\lambda/2}$, $\check{h}_n \sim \mathcal{CN}(0, 1)$, d_n is the distance between the BS and the nth user, and λ is the pathloss exponent. The small-scale channel gains are assumed to be independent and identically distributed. In this work, we consider using a symbol-by-symbol joint-multiuser maximum likelihood detector with perfect CSI [1, Eq. (5)], which is generally equivalent to SIC where both account for the imperfections caused by the IUI [9], [21],

$$\{\widehat{x}_1, \widehat{x}_2\} = \arg\min_{x_i\in\chi_i} \left| y_n - h_n \sum\nolimits_{i=1}^{2} \sqrt{\alpha_i} x_i \right|^2 \tag{2}$$

where x_i are the trail symbols taken from the user's reference constellation χ_i.

III. Adaptation Process at the Base Station

Inspired by the AMC systems in [3], a tabulated adaptation is proposed based on quantizing the instantaneous signal to noise ratio (SNR) of both users, which is given as $\gamma_n = \frac{2|h_n|^2}{N_0}$. The quantization levels for both users are identical, and the number of quantization levels is given by $k = \frac{\gamma_{\max}-\gamma_{\min}}{\Delta\gamma} + 1$, where $\gamma_{\min}$ is the cut-off outage level that satisfies both users with minimum modulation orders, $\gamma_{\max}$ is the maximum SNR to be considered, and $\Delta\gamma$ determines the resolution of the quantization. Therefore, $k \times k$ transmission modes (TMs) are formed in a grid such that each $\mathrm{TM}_{i,j}, \forall i, j \in \{1, \ldots, k\}$, is characterized by the $(\alpha_1, \mathcal{M}_1, \mathcal{M}_2)$ tuple. The conditions to select $\mathrm{TM}_{i,j}$ are $\rho^1_{i,j} \le \gamma_1 < \rho^1_{i+1,j}$ and $\rho^2_{i,j} \le \gamma_2 < \rho^2_{i,j+1}$, where $\rho^1_{i,j} = \gamma_{\min} + (i-1)\Delta\gamma$ and $\rho^2_{i,j} = \gamma_{\min} + (j-1)\Delta\gamma$. It is worth noting that the boundary conditions are $\rho^1_{k+1,j} = \rho^2_{i,k+1} = \infty$. A look-up table (LUT) is used to store the TMs conditions and the corresponding tuple $(\alpha_1, \mathcal{M}_1, \mathcal{M}_2)$. It is noted that as $\Delta\gamma$ gets smaller, a higher resolution grid with larger number of TMs will be created, which provides a better approximation of the continuous adaptation scenario.

Considering FPA and adaptive power assignment (APA), the instantaneous BLER can be expressed as [22, Eq. 6.21],

$$P_{e_n}|\gamma_n = 1 - (1 - P_{B_n}|\gamma_n)^{\mathcal{M}_n L} \tag{3}$$

where $P_{B_n}|\gamma_n$ is the instantaneous BER which depends on $\alpha_1, \mathcal{M}_1$ and $\mathcal{M}_2$. In practice, IUI causes imperfections in the detection and affects the BER. Therefore, the instantaneous BER expressions accounting for these imperfections can be written in general as [1, Eq. (23)],

$$P_{B_n}|\gamma_n = \frac{1}{2^{\mathcal{M}-2}\mathcal{M}_n} \sum\nolimits_q c_q Q\left(\sqrt{|\Delta_q|^2 \gamma_n}\right) \tag{4}$$

where $Q(\cdot)$ is the Gaussian complementary cumulative distribution function (CDF), $\mathcal{M} = \sum_n \mathcal{M}_n$, c_q is a weighting coefficient and Δ_q is the distance between the constellation points and the decision regions' boundaries [1]. It is worthy to mention that including the definitions of c_q and Δ_q for the arbitrary $[\mathcal{M}_1, \mathcal{M}_2]$ settings are omitted from the work due to space limitations. However, the Matlab code to generate (4) is available in [23]. Note that the factor $2^{\mathcal{M}-2}$ becomes $2^{\mathcal{M}-1}$ if $\mathcal{M}_n = 1, \forall n$. In addition, joint-Gray labeling can be used to improve the BER performance, however, exact closed-from BER expressions for the joint-Gray labeling considering arbitrary QAM orders is still an open research problem [12]. Hence, joint-Gray labeling is considered as a future work. Furthermore, the iTp for each user is defined as the probability of receiving a certain block successfully multiplied by the number of bits per symbol [24],

$$R_n|\gamma_n = (1 - P_{e_n}|\gamma_n)\,\mathcal{M}_n. \tag{5}$$

Therefore, the iTp is defined as the sum of the instantaneous users' throughputs, i.e. $\sum_{n=1}^{2} R_n|\gamma_n \triangleq R$. The LUT generation has an objective to maximize R by optimizing the TM parameters. For the FPA scheme, the LUT generation is an integer optimization problem,

$$(\mathcal{M}_1^*, \mathcal{M}_2^*) = \arg\max_{\mathcal{M}_1,\mathcal{M}_2} R \tag{6}$$

while the LUT generation for the APA scheme is a mixed-integer optimization problem,

$$(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*) = \arg\max_{\alpha_1,\mathcal{M}_1,\mathcal{M}_2} R. \tag{7}$$

Both optimization problems are subject to satisfaction of a predefined BLER threshold, τ_n, such that

$$P_{e_n}|\gamma_n \leq \tau_n, \forall n. \tag{8}$$

A. Adaptive Modulation with FPA

A possible FPA scheme is the equally spaced constellation points PA which depends on the selected modulation orders. Hence, informing the users about the modulation orders is sufficient to identify the selected power coefficients at the BS because such PA results in a constellation diagram that is similar to the single-user case, has equally spaced constellation points, with a modulation order of $M_1 \times M_2 \times \cdots \times M_N$. Moreover, the bit mapping of the constellation points still follows the NOMA conventional mapping. The equally spaced constellation points PA condition for $N = 2$ case can be found considering the real part of the nearest constellation point to the origin, $s_0 = -\Lambda_1\sqrt{\alpha_1\kappa_1^{-1}} + \sqrt{\alpha_2\kappa_2^{-1}}$, and the point next to it, $s_1 = -(\Lambda_1 - 2)\sqrt{\alpha_1\kappa_1^{-1}} + \sqrt{\alpha_2\kappa_2^{-1}}$, where κ_n and Λ_n are respectively the normalization factor and width for U_n constellation [1]. Hence, $2s_0 = s_1 - s_0$ ensures equally spaced constellation points. After some mathematical manipulation and generalization, the condition can be expressed as follows,

$$\frac{\alpha_n}{\alpha_{n+1}} = \frac{\kappa_n}{(\Lambda_n + 1)^2 \kappa_{n+1}}, \forall n \in \{1, 2, \ldots, N-1\} \tag{9}$$

which results in a system of equations that need to be solved to find $\alpha_1, \alpha_2, \ldots, \alpha_{N-1}$. By noting that $\alpha_2 = 1 - \alpha_1$ for $N = 2$, the closed-form PA can be expressed as

$$\alpha_1 = \frac{\kappa_1}{\Lambda_1\kappa_2(\Lambda_1 + 2) + \kappa_1 + \kappa_2} \tag{10}$$

which is more general than [6, Eq. (10)] because the latter is applicable for square QAM only, while (10) is applicable for the general case of rectangular QAM. The objective function in (6) has a low complexity as the solver only needs to compute $P_{e_n}|\gamma_n$ and $R_n|\gamma_n$ $\forall n$ using (3)–(5) at the FPA found in (10) for a limited number of cases that correspond to all possible combinations of $\mathcal{M}_1$ and $\mathcal{M}_2$, and select the pair that maximizes R while satisfying (8). While noting that the computational complexity to compute (4) increases as $\mathcal{M}_n$ becomes larger [1], using brute-force search will have a time complexity of $\mathcal{O}\left(\mathcal{M}_{\max}^2\right)$, where $\mathcal{M}_{\max}$ is the maximum $\mathcal{M}_n$ allocated $\forall n$. Considering the fact that increasing the modulation order would increase the BLER for a fixed SNR can be useful to define search heuristics. Hence, reducing the optimization complexity. Therefore, the search for $(\mathcal{M}_1^*, \mathcal{M}_2^*)$ can be performed in ascending order of $\mathcal{M}$, and the search would stop once the BLER constraints are violated. The function for solving (6) can be described as $f : \left(\rho_{i,j}^1, \rho_{i,j}^2, \tau_1, \tau_2\right) \longmapsto (\mathcal{M}_1^*, \mathcal{M}_2^*)$, where f finds $(\mathcal{M}_1^*, \mathcal{M}_2^*)$ that maximizes R under $(\rho_{i,j}^1, \rho_{i,j}^2, \tau_1, \tau_2)$ conditions. Note that $(\mathcal{M}_1^*, \mathcal{M}_2^*) \longmapsto \alpha_1^*$ using (10).

B. Adaptive Modulation with APA

Solving the optimization problem in (7) is generally not straightforward because the objective function exhibits multiple local maximum points for R, as depicted in Fig. 1. Such behavior is due to the fact that certain α_1 values cause constellation points overlap, and hence detection ambiguity. These points can be found by considering the following condition,

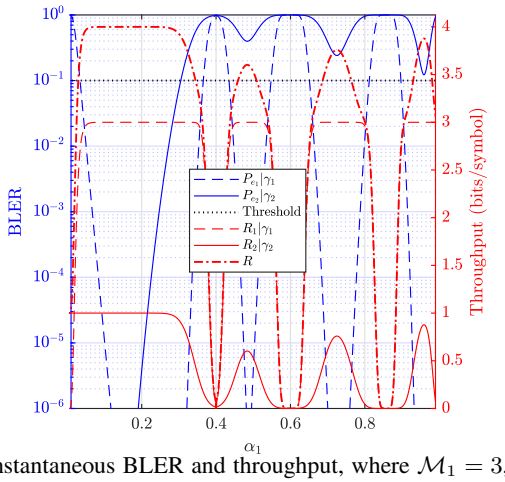

Fig. 1: Instantaneous BLER and throughput, where $\mathcal{M}_1 = 3$, $\mathcal{M}_2 = 1$, $\gamma_1 = 32$ dB, $\gamma_2 = 29$ dB.

$$(a_\mathrm{R} - a_\mathrm{L})\sqrt{\alpha_1\kappa_1^{-1}} + (b_\mathrm{R} + b_\mathrm{L})\sqrt{(1-\alpha_1)\kappa_2^{-1}} = 0 \tag{11}$$

where $a_\mathrm{R}, a_\mathrm{L} \in \{\pm 1, 3, \ldots, \Lambda_1\}$, $b_\mathrm{R}, b_\mathrm{L} \in \{\pm 1, 3, \ldots, \Lambda_2\}$. Solving (11) for all possible combinations of $a_\mathrm{R}, a_\mathrm{L}, b_\mathrm{R}, b_\mathrm{L}$ results in a set of $\alpha_1 > 0$ and $\alpha_1 < 1$ values that impose the constellation points overlap, which can be expressed as

$$\mathbb{A} = \frac{\kappa_1 (b_\mathrm{R} + b_\mathrm{L})^2}{\kappa_1 (b_\mathrm{R}^2 + b_\mathrm{L}^2 + 2b_\mathrm{R}b_\mathrm{L}) + \kappa_2 (a_\mathrm{R}^2 + a_\mathrm{L}^2 - 2a_\mathrm{R}a_\mathrm{L})}. \tag{12}$$

Therefore, to avoid the local maximum problem, it is important to find all local maximum points by segmenting the search range and eventually finding the global maximum. It is worth noting that after the first overlap with respect to α_1, the constellation points become very close to each other, which degrades the performance severely. Nonetheless, the Euclidean distance between the constellation points starts increasing after the last overlap with respect to α_1. Therefore, investigating the two extreme ranges with respect to α_1, i.e., $[0, \min(\mathbb{A})]$ and $[\max(\mathbb{A}), 1]$, is sufficient to find the global maximum. Hence, the optimization complexity is reduced compared to the full-line search.

Furthermore, since brute-force search for $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$ is computationally expensive, a stopping criterion similar to the FPA case is considered to reduce complexity. The pseudocode for the mixed-integer optimization problem is given in Algorithm 1. The function in line 12, $\mathrm{fconopt}(\cdot)$, is based on line-search and used to solve constrained maximization problems where the arguments are as follows, (7) is the objective function, x_0 is the initial guess, lb and ub are the lower and upper bounds of the optimization variable, (8) are the non-linear constraints. The algorithm solving (7) can be described as $g : \left(\rho_{i,j}^1, \rho_{i,j}^2, \tau_1, \tau_2\right) \longmapsto (\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$, where g finds $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$ that maximizes R under $(\rho_{i,j}^1, \rho_{i,j}^2, \tau_1, \tau_2)$.

C. Practical Considerations

Since the BS informs the users about $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$, the adaptation signaling overhead should be considered for $\Delta\gamma \to 0$ cases as the LUT size becomes massive. Assuming that the maximum and minimum modulation orders are 256 and 2, respectively, i.e. $1 \leq \mathcal{M}_n \leq 8$. This represents 8 possibilities, which requires 3 bits to represent $\mathcal{M}_n, n \in (1, 2)$, or $3N$ bits for N users. Moreover, since $0 < \alpha_1 < 1$ then using 8 to 16 bits provides negligible quantization error. Therefore, the total required bits to

Algorithm 1: Adaptive modulation and APA

Input: $\left(\rho_{i,j}^{1}, \rho_{i,j}^{2}, \tau_1, \tau_2\right)$
Output: $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$
$\gamma_1 = \rho_{i,j}^{1}$, $\gamma_2 = \rho_{i,j}^{2}$, $R^* = \mathcal{M}_1^* = \mathcal{M}_2^* = 0$
for $\mathcal{M}_1 = 1 : \mathcal{M}_{\max}$ **do**
  for $\mathcal{M}_2 = 1 : \mathcal{M}_{\max}$ **do**
    if $\mathcal{M}_1 + \mathcal{M}_2 < \mathcal{M}_1^* + \mathcal{M}_2^*$ **then**
      continue
    Compute $\mathbb{A}$ using (12)
    $\alpha_{1,\text{seg}}^{\min} = \min(\mathbb{A})$, $\alpha_{1,\text{seg}}^{\max} = \max(\mathbb{A})$
    $\textbf{key}_l = [0,\ \alpha_{1,\text{seg}}^{\max}]$, $\textbf{key}_u = [\alpha_{1,\text{seg}}^{\min},\ 1]$
    for $k = 1 : 2$ **do**
      $\text{lb} = \text{key}_l^{(k)}$, $\text{ub} = \text{key}_u^{(k)}$
      $x_0 = 0.5\,(\text{lb} + \text{ub})$
      $\left[\check{f}_{\text{val}}^{(k)}, x_{\text{seg}}^{(k)}\right] = \text{fconopt}\left((7), x_0, \text{lb}, \text{ub}, (8)\right)$
      if $P_{e_n}|\gamma_n > \tau_n,\ \forall n$ **then**
        $\check{f}_{\text{val}}^{(k)} = \text{NaN}$
    $f_{\text{val}}^{(\mathcal{M}_1,\mathcal{M}_2)} = \max\left(\check{\mathbf{f}}_{\text{val}}\right)$
    $\alpha_1^{(\mathcal{M}_1,\mathcal{M}_2)} = x_{\text{seg}}^{\left(\arg\max\left(\check{\mathbf{f}}_{\text{val}}\right)\right)}$
    if $f_{\text{val}}^{(\mathcal{M}_1,\mathcal{M}_2)} > R^*$ **then**
      $R^* = f_{\text{val}}^{(\mathcal{M}_1,\mathcal{M}_2)}$
      $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*) = \left(\alpha_1^{(\mathcal{M}_1,\mathcal{M}_2)}, \mathcal{M}_1, \mathcal{M}_2\right)$
    if $\mathcal{M}_2 > 1$, $f_{\text{val}}^{(\mathcal{M}_1,\mathcal{M}_2)} = f_{\text{val}}^{(\mathcal{M}_1,\mathcal{M}_2-1)} = \text{NaN}$ **then**
      break
  if $\mathcal{M}_1, > 1$, $f_{\text{val}}^{(\mathcal{M}_1,\mathcal{M}_2)} = f_{\text{val}}^{(\mathcal{M}_1-1,\mathcal{M}_2)} = \text{NaN}$ **then**
    break

convey the power information can be $16(N-1)$, which is a significant signaling overhead when compared to the modulation orders. Therefore, to limit the signaling overhead, $\Delta\gamma$ should be optimized to limit the LUT size, or FPA should be considered.

IV. Throughput Analysis

This section starts by deriving the probability of each TM as well as the average BLER, and then the aTp.

A. Probability of Transmission Modes

Each TM has lower and upper SNR thresholds for the instantaneous SNRs in order to be selected. The lower and upper γ_n thresholds for $\text{TM}_{i,j}$ will be denoted as $\rho_{i,j}^{l,1} = \rho_{i,j}^{1}$, $\rho_{i,j}^{l,2} = \rho_{i,j}^{2}$, $\rho_{i,j}^{\text{u},1} = \rho_{i+1,j}^{1}$ and $\rho_{i,j}^{\text{u},2} = \rho_{i,j+1}^{2}$. Hence, the probability of $\text{TM}_{i,j}$ can be calculated as follows,

$$\Pr\left(\text{TM}_{i,j}\right) = \prod_{n=1}^{2} \Pr\left(\rho_{i,j}^{l,n} \le \gamma_n < \rho_{i,j}^{\text{u},n}\right) = \prod_{n=1}^{2} \varpi_n \quad (13)$$

where $\varpi_n = F_{\gamma_n}\left(\rho_{i,j}^{\text{u},n}\right) - F_{\gamma_n}\left(\rho_{i,j}^{l,n}\right)$ and $F_{\gamma_n}(\cdot)$ is the CDF of γ_n. Furthermore, the probability of transmission, i.e., $\Pr(\mathcal{M}_1, \mathcal{M}_2) > 0$, is $P_X = \sum_{i,j} \Pr\left(\text{TM}_{i,j}\right)$. Moreover, the average number of transmitted bits for U_n is $S_n = \sum_{i,j} \Pr\left(\text{TM}_{i,j}\right) \mathcal{M}_n^{*(i,j)}$.

On the other hand, the aTp for U_n depends on the average BLER and average number of transmitted bits. Hence, R_n is given as [24, Eq. (18)] $R_n = (1 - P_{e_n})\, S_n$, while the aTp of the system is $\overline{R} = R_1 + R_2$.

B. Average BLER Analysis

The average BLER of U_n can be found by considering the average BLER for all TMs. Hence,

$$P_{e_n} = \frac{1}{P_X} \sum\nolimits_{i,j} \Pr\left(\text{TM}_{i,j}\right) P_{e_n}^{(i,j)} \quad (14)$$

where

$$P_{e_n}^{(i,j)} = \frac{1}{\varpi_n} \int_{\rho_{i,j}^{l,n}}^{\rho_{i,j}^{\text{u},n}} \left(1 - (1 - P_{B_n}|\gamma_n)^{b_n}\right) f_{\gamma_n}(\gamma_n)\ \mathrm{d}\gamma_n \quad (15)$$

$$= 1 - \sum_{r=0}^{b_n} \binom{b_n}{r} \frac{(-1)^r}{\varpi_n} \underbrace{\int_{\rho_{i,j}^{l,n}}^{\rho_{i,j}^{\text{u},n}} (P_{B_n}|\gamma_n)^r f_{\gamma_n}(\gamma_n)\ \mathrm{d}\gamma_n}_{I_n} \quad (16)$$

and $b_n = \mathcal{M}_n^{*(i,j)} L$, $\binom{\cdot}{\cdot}$ is the binomial coefficient, $f_{\gamma_n}(\gamma_n)$ is the probability density function of γ_n. It is worth noting that the Binomial expansion theorem is used to simplify (15) such that $(1-z)^{b_n} = \sum_{r=0}^{b_n} \binom{b_n}{r} (-z)^r$. In addition, the integral, I_n, involves a multinomial expression since $P_{B_n}|\gamma_n$, (4), is a sum of Q-function terms, i.e. $Q(\cdot) + \cdots + Q(\cdot)$. Consequently, evaluating I_n in its current form leads to intractable analysis. Therefore, $P_{B_n}|\gamma_n$ is relaxed by considering a tight approximation that accounts for the most dominant $Q(\cdot)$ term only [25],

$$P_{B_n}|\gamma_n \approx \beta Q\left(\sqrt{\xi \gamma_n}\right) \quad (17)$$

where $\xi = \min_q |\Delta_q|^2$ and $\beta = \frac{1}{2^{\mathcal{M}-2}\mathcal{M}_n} \sum_q c_q$, $q = \arg\min_q |\Delta_q|^2$. Moreover, since the integral involves $Q^r(\cdot)$, its evaluation is still an open research problem for any integer. Hence, $Q(z) \approx \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi(z^2+1)}}$ is used which is a tight approximation for a wide range of argument values [26]. Therefore,

$$I_n \approx \frac{\exp(\Omega)\, \Omega^{\frac{r-2}{2}} \beta^r}{\xi \overline{\gamma}_n (2\pi)^{r/2}} \left[\Gamma\left(\frac{2-r}{2}, \Omega(\xi \rho_{i,j}^{l,n} + 1)\right) - \Gamma\left(\frac{2-r}{2}, \Omega(\xi \rho_{i,j}^{\text{u},n} + 1)\right)\right] \quad (18)$$

and $\Omega = \frac{r}{2} + \frac{1}{\xi \overline{\gamma}_n}$, the incomplete upper Gamma function is denoted by $\Gamma(a, x) = \int_x^\infty \exp(-t) t^{a-1}\, \mathrm{d}t$, while $\overline{\gamma}_n = \frac{2\mathbb{E}\left[|h_n|^2\right]}{N_0}$.

V. Numerical Results and Discussions

This section presents the aTp analytical and numerical results for the FPA and APA, as well as the optimal TM parameters. It is assumed that U_1 and U_2 are at distances d_1 and $1.67 d_1$ from the BS, and the pathloss exponent is $\lambda = 2.7$. Unless stated otherwise, the packet consists of $L = 32$ symbols and the BLER requirement is $\tau_n = 0.01, \forall n$, while $\text{SNR} \triangleq 1/2\sigma_w^2$.

Figs. 2 and 3 show the selected TMs parameters, $(\alpha_1, \mathcal{M}_1, \mathcal{M}_2)$ using (6)–(8). The x and y-axis represent γ_1 and γ_2, respectively. Color coding is used to represent $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_1 + \mathcal{M}_2$, and α_1. The figures, are generated using $\Delta\gamma = 3$ dB, $\gamma_{\min} = 18$ dB, $\gamma_{\max} = 42$ dB, and $\mathcal{M}_{\max} = 8$. As can be noted from the figures, while $\mathcal{M}_n$ of an individual user using FPA may outperform an individual user with the APA in certain scenarios, $\sum_n \mathcal{M}_n$ for the APA consistently outperforms the FPA for $\gamma_1 > \gamma_2$ or $\gamma_1 < \gamma_2$. However, when $\gamma_1 = \gamma_2$ both schemes offer similar $\sum_n \mathcal{M}_n$ which, indicates that the FPA is optimal only for $\gamma_1 = \gamma_2$. Moreover, $\mathcal{M}_1$ and

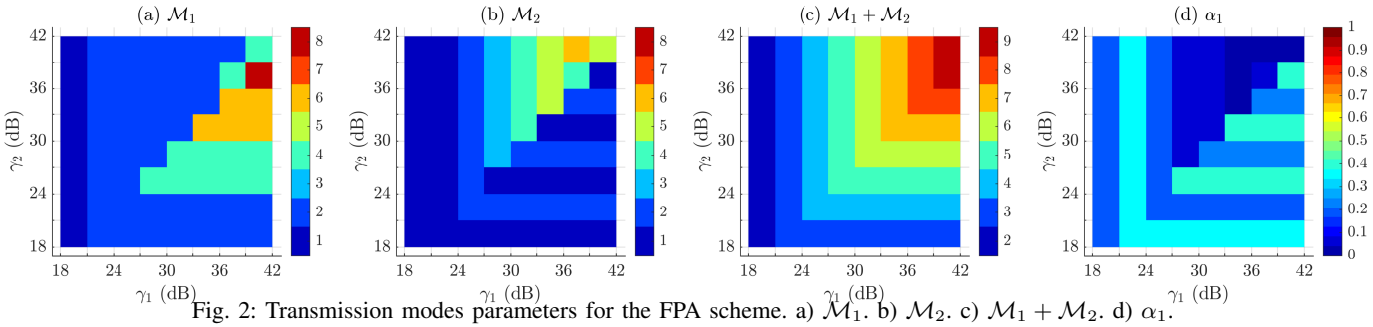

Fig. 2: Transmission modes parameters for the FPA scheme. a) $\mathcal{M}_1$. b) $\mathcal{M}_2$. c) $\mathcal{M}_1 + \mathcal{M}_2$. d) α_1.

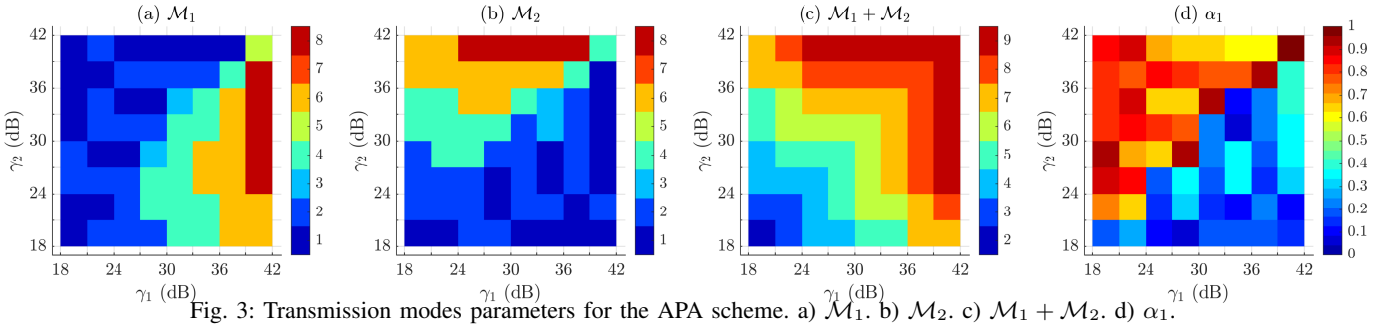

Fig. 3: Transmission modes parameters for the APA scheme. a) $\mathcal{M}_1$. b) $\mathcal{M}_2$. c) $\mathcal{M}_1 + \mathcal{M}_2$. d) α_1.

$\mathcal{M}_2$ for the APA in Figs. 3a and 3b are symmetrical about the right diagonal, i.e., $\mathcal{M}_1|(\gamma_1, \gamma_2) = \mathcal{M}_2|(\gamma_2, \gamma_1)$. For example, $\mathcal{M}_1|(21\text{ dB}, 24\text{ dB}) = \mathcal{M}_2|(24\text{ dB}, 21\text{ dB}) = 2$.

Fig. 4 shows that the analytical and simulation aTp results match closely for various parameters, which confirms the accuracy of the derived approximation. Fig 4a considers $\Delta\gamma = 1$ dB and $L = 32$, and it shows that the APA performance gain is proportional to SNR. For example, the APA outperforms the FPA by about 3 dB at 6 bits/symbol. Furthermore, R_1 is higher than R_2 because U_1 has better channel conditions. Fig. 4b evaluates the effect of increasing $\Delta\gamma$ at a narrower SNR span, where the throughput degrades slightly when $\Delta\gamma$ increases from 1 to 3 dB. However, the degradation becomes significant for $\Delta\gamma = 5$ dB. Based on the observed trend, using $\Delta\gamma < 1$ will not have a tangible impact on the throughput, while the complexity will increase significantly.

Furthermore, Fig. 4c analyzes the effect of changing L for $\Delta\gamma = 3$ dB. The figure shows that increasing L degrades aTp, which is due to the need for lower BER to achieve the same BLER. The proposed algorithm can be extended for $N > 2$, however, at the expense of additional complexity. Moreover, increasing the number of users beyond 2 degrades the BER, and thus, the BLER QoS requirements will be mostly violated causing a throughput degradation at low and moderate SNRs. For example, increasing the number of users from 2 to 3 causes a BER degradation of 10 dB at BER of 10^{-3} [1, Fig. 14]. It is also worth noting that increasing L caused throughput degradation because the BLER is proportional to L in uncoded systems, and in coded systems when the packet length is increased while the codeword length is fixed, i.e., when multiple codewords are grouped to increase the packet length. If the packet length is equal to the codeword length, then increasing the packet length improves the BLER because the error correction capability in this case improves.

Fig. 5 shows that the exact and approximated average BLER results exhibit a close match for wide range of SNRs,and the average BLER does not exceed 0.01, which is the threshold.

Fig. 6 verifies the time complexity of the proposed stopping

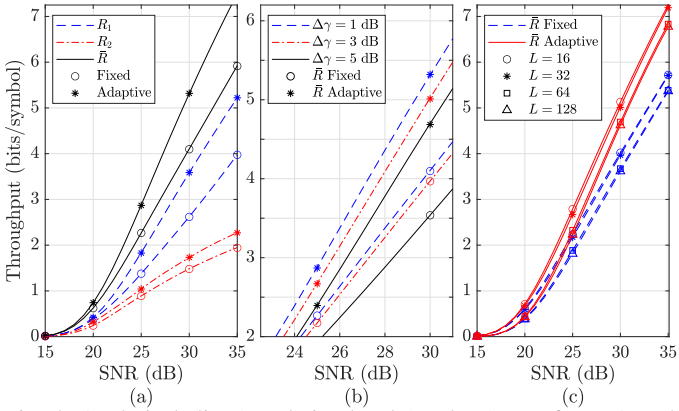

Fig. 4: Analytical (lines) and simulated (markers) aTp for FPA and APA schemes. a) $L = 32$, $\Delta\gamma = 1$ dB. b) $L = 32$, $\Delta\gamma = \{1, 3, 5\}$ dB. c) $L = \{16, 32, 64, 128\}$, $\Delta\gamma = 3$ dB.

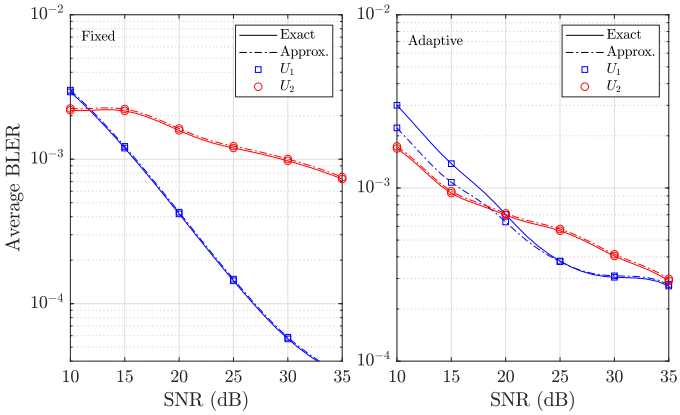

Fig. 5: Exact and approximated average BLER for fixed and adaptive PA schemes, where $L = 32$ and $\Delta\gamma = 3$ dB.

criteria at different pairs of γ_1 and γ_2. For the FPA, the stopping criteria reduce the complexity by 10 to 1400 fold, as compared to brute-force search. When the stopping criteria was adopted for the APA, the proposed segment-line search reduces the time complexity by about 30 to 116 fold when compared to full

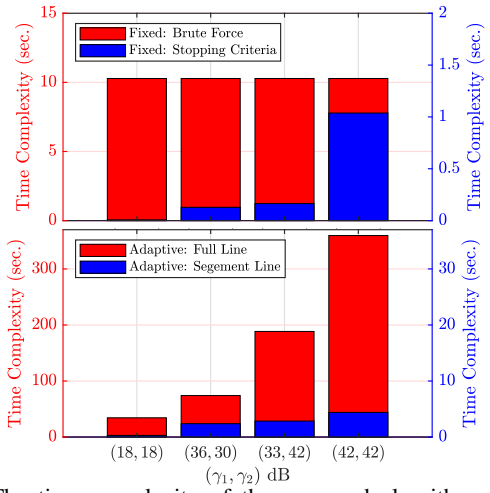

Fig. 6: The time complexity of the proposed algorithm compared with different benchmarks at various instantaneous SNRs.

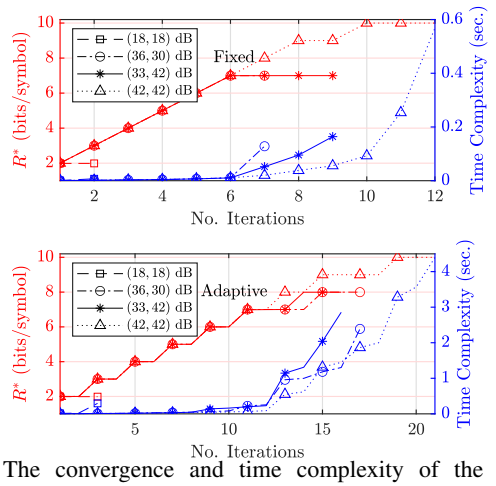

Fig. 7: The convergence and time complexity of the proposed algorithm at various instantaneous SNRs.

line search. Moreover, Fig. 7 presents the convergence and time complexity for various pairs of γ_1 and γ_2. It is observed that APA requires more iterations to converge as compared to FPA. The reason is that the former searches in two power ranges, while the latter only checks the feasibility at a fixed power. In addition, it is noted that the time complexity increases exponentially with the number of iterations, which is due to the increase in the BLER expressions size at larger iteration numbers and higher modulation orders.

VI. Conclusions

This work derived closed-form expressions for the throughput over Rayleigh block fading channels for two-user NOMA systems using LUT-based adaptive modulation with FPA and APA. Reduced complexity integer and mixed-integer optimization problems were formulated and solved. The results show that the FPA scheme provides similar performance to the APA scheme when both users experience identical channel conditions, and that APA outperforms FPA in other channel conditions. Moreover, the system throughput demonstrated high tolerance to the granularity of the quantization levels. Improving the performance at low SNRs is crucial, hence, future work will consider short-packet theory, channel coding and joint Gray-labeling to improve the performance in such scenarios.

References

[1] H. Yahya, E. Alsusa, and A. Al-Dweik, "Exact BER analysis of NOMA with arbitrary number of users and modulation orders," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6330–6344, Sep. 2021.
[2] K. Wang *et al.*, "Asymmetric adaptive modulation for uplink NOMA systems," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7222–7235, Nov. 2021.
[3] H. Y. Hsieh, M. J. Yang, and C. H. Wang, "Fair resource allocation using the MCS map for multi-user superposition transmission (MUST)," in *IEEE PIMRC*, Spain, Dec. 2016, pp. 1–7.
[4] S. Kim, H. Kim, and D. Hong, "Joint power allocation and MCS selection in downlink NOMA system," in *IEEE PIMRC*, Italy, Dec. 2018, pp. 1–4.
[5] T. Assaf, A. J. Al-Dweik, M. Elmoursi, and H. Zeineldin, "Efficient bit loading algorithm for OFDM-NOMA systems with BER constraints," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 423–436, Jan. 2022.
[6] E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource allocation in multicarrier NOMA systems based on optimal channel gain ratios," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 635–650, Jan. 2022.
[7] H. Jia and L. Musavian, "Performance analysis for NOMA with M-QAM modulation," in *IEEE VTC*, Belgium, Jun. 2020, pp. 1–5.
[8] W. Yu, H. Jia, and L. Musavian, "Joint adaptive M-QAM modulation and power adaptation for a downlink NOMA network," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 783–796, Feb. 2022.
[9] Y. Iraqi and A. Al-Dweik, "Power allocation for reliable SIC detection of rectangular QAM-based NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8355–8360, Aug. 2021.
[10] J. Choi, "On the power allocation for a practical multiuser superposition scheme in NOMA systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 438–441, Mar. 2016.
[11] Z. Dong *et al.*, "Uplink non-orthogonal multiple access with finite-alphabet inputs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5743–5758, Sep. 2018.
[12] W. Han, X. Ma, D. Tang, and N. Zhao, "Study of SER and BER in NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3325–3340, Apr. 2021.
[13] M. AbdelMoniem *et al.*, "Enhanced NOMA system using adaptive coding and modulation based on LSTM neural network channel estimation," *Applied Sciences*, vol. 9, no. 15, July 2019.
[14] M. Qiu, Y. Huang, and J. Yuan, "Downlink non-orthogonal multiple access without SIC for block fading channels: An algebraic rotation approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 3903–3918, Aug. 2019.
[15] Z. Zheng, L. Yuan, and F. Fang, "Performance analysis of fountain coded non-orthogonal multiple access with finite blocklength," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1752–1756, Aug. 2021.
[16] D.-D. Tran *et al.*, "Short-packet communications for MIMO NOMA systems over Nakagami-m fading: BLER and minimum blocklength analysis," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3583–3598, Apr. 2021.
[17] T.-H. Vu *et al.*, "Short-packet communications in NOMA-CDRT IoT networks with cochannel interference and imperfect SIC," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2022.
[18] ——, "Performance analysis and deep learning design of wireless powered cognitive NOMA IoT short-packet communications with imperfect CSI and SIC," *IEEE Internet Things J.*, pp. 1–1, 2021.
[19] ——, "Intelligent reflecting surface-aided short-packet non-orthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2022.
[20] N.-P. Le and K. Le, "Uplink NOMA short-packet communications with residual hardware impairments and channel estimation errors," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2022.
[21] T. Assaf *et al.*, "NOMA receiver design for delay-sensitive systems," *IEEE Systems J.*, vol. 15, no. 4, pp. 5606–5617, Dec. 2021.
[22] B. Sklar, *Digital Communications: Fundamentals and Applications*, ser. Novell-Bibliothek. Prentice-Hall PTR, 2001.
[23] H. Yahya, E. Alsusa, and A. Al-Dweik, "Exact BER analysis of NOMA with arbitrary number of users and modulation orders," 2021. [Online]. Available: https://dx.doi.org/10.21227/mxh7-pk27
[24] B. Zhang *et al.*, "IRS-assisted short packet wireless energy transfer and communications," *IEEE Wireless Commun. Lett.*, pp. 1–1, Nov. 2021.
[25] H. Yahya, E. Alsusa, and A. Al-Dweik, "Hybrid cognitive-radio NOMA with blind transmission mode identification and BER constraints," *TechRxiv*, Aug. 2021. [Online]. Available: https://doi.org/10.36227/techrxiv.15157944.v1
[26] G. K. Karagiannidis and A. S. Lioumpas, "An improved approximation for the Gaussian Q-function," *IEEE Commun. Lett.*, vol. 11, no. 8, pp. 644–646, Aug. 2007.