Fairness-Aware Optimal Power and Discrete Rate Adaptation for K-user NOMA with Imperfect SIC and Adaptive Decoding Order

S. Sruthy, *Student Member, IEEE* and Neelesh B. Mehta, *Fellow, IEEE*

***Abstract*—Practical physical layer limitations such as discrete rate adaptation and imperfect successive interference cancellation (SIC) fundamentally alter the spectral efficiency and fairness of downlink non-orthogonal multiple access (NOMA). For K-user downlink NOMA, we formulate the problem of jointly choosing the modulation and coding schemes and powers of the users to maximize a general weighted α-utility function subject to constraints on the block error rate (BLER) and minimum rate of each user. We account for imperfect SIC and adapt the decoding order. This approach subsumes the weighted sum rate maximization, proportional fairness, and max-min fairness approaches considered separately in the literature. Due to discrete rate adaptation, the optimization problem for any decoding order is NP-hard. We present a novel search algorithm that provably finds the optimal solution with one to five orders of magnitude lower search complexity than exhaustive search. We also propose an alternate approach that employs a smaller set of BLER constraints to reduce complexity further. We derive novel expressions for the outage probability, average sum rate, and weighted α-utility of the optimal solution for two-user NOMA for the farthest-to-nearest decoding order. The analysis is the first to address the general weighted α-utility function. We also extend our approach to account for imperfect channel state information.**

***Index Terms*—Non-orthogonal multiple access, imperfect successive interference cancellation, adaptive decoding, fairness, adaptive modulation and coding, power allocation.**

I. Introduction

Power-domain non-orthogonal multiple access (NOMA) enables a base station (BS) to serve multiple users simultaneously in the same time-frequency resource. In the downlink, the BS superimposes signals of the users with different transmit powers and transmits them simultaneously. The users employ successive interference cancellation (SIC) in their receivers to retrieve their data in the presence of interference from a subset of the other users' signals. With power-domain NOMA, 5G and beyond systems can address demands for higher data rates and fair allocation of resources for cell-center and cell-edge users while meeting their quality-of-service (QoS) requirements. This translates to more connected devices, lower latency, and massive connectivity.

The efficacy of NOMA crucially depends on the following aspects. First, it depends on the system-wide utility function that determines the powers and rates allocated to the users by the NOMA resource allocation algorithm. The algorithm needs to account for the different signal strengths of the users, who are at different distances from the BS and encounter fading, and the interferences their signals cause at the decoders of the other users.

Sum rate, weighted sum rate, sum of the logarithms of the users' rates, which is also referred to as proportional fairness (PF), and the minimum of the users' rates, which is referred to as max-min fairness, are examples of the utility function [1, Ch. 2]. Maximizing the sum rate improves spectral efficiency but results in an unfair allocation. This is because users closer to the BS, who have larger signal-to-noise ratios (SNRs), get assigned more or all resources. On the other hand, proportional fairness and max-min fairness result in a more fair allocation. However, the improvement in fairness comes at the expense of a lower spectral efficiency. This fundamental trade-off between spectral efficiency and fairness is different for NOMA than for the conventional orthogonal multiple access (OMA) scheme, which is used in 5G, as it serves only one user on a time-frequency resource.

Second, the efficacy of NOMA depends on the constraints imposed by the physical layer. Specifically, the BS can transmit packets to the users using only modulation and coding schemes (MCSs) chosen from a pre-specified set, which is also called *discrete rate adaptation*. For example, 5G mandates the use of quadrature phase-shift keying (QPSK), 16-QAM (quadrature amplitude modulation), 64-QAM, and 256-QAM constellations with coding rates that lie in between $78/1024$ and $948/1024$ [2, Table 5.2.2.1-3]. The MCS chosen must satisfy a constraint on the block error rate (BLER), which is the probability that a user's packet is decoded incorrectly. Discrete rate adaptation fundamentally alters the resource allocation problem compared to the widely studied, but idealistic, continuous rate adaptation model for NOMA, which is based on Shannon's capacity formula and where the user rate is any positive real number.

Third, the efficacy depends on the accuracy of SIC, which is employed multiple times by each user. Ideally, SIC should completely remove the interferences from users whose signal have been cancelled. However, in practice, fading, noise, and hardware impairments can lead to a packet decoding error. In such a case, the interference cancellation in all the subsequent stages is adversely affected.

Fourth, the efficacy depends on the order in which the user's signals are successively decoded. Fixed [3]–[16] and adaptive orders [17]–[20], which depend on the users' channel gains, have been considered in the literature. The decoding order

S. Sruthy and N. B. Mehta are with the Electrical Communication Engineering department, Indian Institute of Science, Bengaluru, India (Emails: sruthys@iisc.ac.in, nbmehta@iisc.ac.in).

and imperfect SIC both need to be taken into account in the very design of the resource allocation algorithm to ensure that the BLER constraint of each user is satisfied.

A. Literature Survey

Given the extensive literature on NOMA, we focus on works that address the following four aspects that are most relevant to our work: fairness, rate adaptation, imperfect SIC, and SIC decoding order. We refer the reader to the survey articles in [21], [22] for a broad survey of NOMA.

a) Fairness: A joint user pairing and power allocation algorithm to maximize the PF utility for uplink NOMA is studied in [23]. A PF scheduler and a waterfilling-based power allocation are considered for orthogonal frequency division multiplexing-based downlink NOMA in [24]. A weighted max-min fairness-based power and rate adaptation algorithm is proposed for a vehicular NOMA network in [9]. Users to be scheduled and their powers are numerically determined to maximize the max-min fairness utility of downlink NOMA with multiple transmit and receive antennas in [25]. Joint power allocation and second user assignment to maximize the minimum rate among secondary users are studied for an underlay cognitive downlink network in [26]. An optimal power allocation algorithm to maximize the total α-fairness utility of downlink NOMA is presented in [27].

The optimal power allocation to maximize max-min fairness, weighted sum rate, sum rate, and energy efficiency for downlink NOMA is derived in [6]. A power allocation method to maximize the minimum rate among the users is proposed in [7], [8]. Reference [28] instead minimizes the maximum average video quality degradation probability. Matching theory is applied for power allocation and antenna selection in [29] to balance throughput and fairness. Power allocation is studied in [10]–[12] to maximize the weighted sum rate and the α-fair utility in [30], [31]. The trade-off between sum throughput and the ratio of the minimum to maximum instantaneous rates of the users is studied for downlink NOMA in [32]. However, all the above references assume continuous rate adaptation.

b) Discrete Rate Adaptation: A joint adaptive modulation and power assignment algorithm called dynamic power and rate adaptation (DPRA) for two-user downlink NOMA to achieve a minimum rate for one user and maximize the rate of the other user subject to constraints on the bit error rate (BER) is discussed in [14]. However, the focus is limited to uncoded QAM and the algorithm provides no optimality guarantees. The allocation of powers that maximizes the mutual information for a square-QAM constellation is studied in [33]. Power and uncoded QAM constellation size are adapted in [16] to maximize the throughput. The sum throughput versus fairness trade-off is numerically evaluated for 16-QAM in [32]. However, all the above works limit their attention to two-user NOMA and apply only to uncoded transmissions. While [34] considers the general case of K-user NOMA, it only characterizes the range of powers that can be assigned to each user for uncoded QAM constellations. Sum rate or fairness aspects are not addressed.

c) Imperfect SIC: The outage probability and ergodic rate of a two-user NOMA-based spectrum sharing scheme with imperfect SIC is analyzed in [35]. The outage probability of an overlay cognitive two-user NOMA system with imperfect SIC is analyzed in [36]. The powers allocated to the two users are optimized to maximize the sum throughput. Multi-cell uplink NOMA with imperfect SIC is investigated in [37]. The total power allocated to the users is minimized while satisfying a minimum rate constraint for each user. Clustering, beamforming, and power allocation under imperfect SIC are jointly designed in [38] to maximize the sum rate of K-user NOMA while satisfying a minimum signal-to-interference-plus-noise ratio (SINR) constraint for each user.

However, the above works assume continuous rate adaptation and, with the exception of [38], focus on only two users. Imperfect SIC is modeled by means of an interference leakage term whose power is proportional to the signal power. The applicability of this model is limited in practice for two reasons. First, cyclic redundancy check (CRC) code is widely employed in cellular systems to detect errors. Therefore, the decoding of subsequent stages can be halted in the event of an error. Second, no guidance is available about choosing the specific value of the all-important proportionality constant, which lies between 0 and 1.

d) SIC Decoding Order: The farthest-to-nearest (F2N) decoding order, in which the farthest user's signal is decoded first and the nearest user's signal is decoded last after cancelling the interference from all other users' signals, is considered in several works [3]–[16]. A two-user cooperative NOMA system with untrusted users with adaptive decoding order and imperfect SIC is analyzed in [17]. The goal is to maximize the sum rate and the minimum secrecy rate of two users. A dynamic ordered SIC receiver is introduced in [18], and its outage probability is analyzed. A similar approach under imperfect channel state information (CSI) is further explored in [19]. However, [18], [19] focus on uplink NOMA, while we focus on downlink NOMA. A sub-optimal joint radio resource allocation and SIC ordering algorithm for power-domain sparse code multiple access-based downlink is proposed in [20].

B. Contributions

We see from the above survey that discrete rate adaptation for general K-user NOMA has received limited attention in the literature despite its practical relevance. Furthermore, the practically important aspects that pertain to decoding for discrete rate adaptation, namely imperfect SIC and decoding order, have not been well-studied. Fairness in NOMA too has not been studied with discrete rate adaptation or imperfect SIC; the focus has primarily been on continuous rate adaptation. We address these multiple gaps in the NOMA literature. Table I contrasts the literature and our manuscript.

We study joint optimal power and discrete rate adaptation for downlink K-user NOMA with imperfect SIC and adaptive decoding order. Our goal is to maximize a general fairness-aware weighted α-utility function while satisfying a constraint on the BLER of each user and a QoS constraint on the minimum rate of each user. We make the following contributions.

1) *General Model and Novel Problem Formulation:* For the general case of K-user downlink NOMA, we present a novel formulation of the problem of jointly choosing the MCSs and powers of the users and the decoding order to maximize a general weighted α-utility function under imperfect SIC. For any decoding order, discrete rate adaptation makes this problem a mixed-integer nonlinear programming (MINLP) problem, which is NP-hard. The strength and novelty of this formulation lies in its: (i) generality, as it handles any number of users and any decoding order, and covers a large class of utility functions that have been separately studied in the literature, and (ii) practical relevance, as it accounts for both imperfect SIC and discrete rate adaptation.
2) *Novel Low Complexity Algorithm Based on Structural Insights:* We present a novel algorithm that provably finds the optimal MCS choice and powers (OMCSP) for any decoding order. We then determine the optimal decoding order. The proposed algorithm relies on two novel theoretical results about the structure of the optimal solution. Our first result is a criterion that specifies whether a feasible power allocation that meets the BLER and QoS constraints exists for a given choice of MCSs for the K users. Our second result eliminates potentially many MCS choices once an MCS choice has been determined to be infeasible. Our algorithm has one to five orders of magnitude lower search complexity than exhaustive search depending on K.
3) *Reducing Complexity Further:* We then propose an alternate and novel formulation that reduces the $\mathcal{O}(K^2)$ BLER constraints, which arise due to the incorporation of imperfect SIC in the original problem formulation, to only $\mathcal{O}(K)$ BLER constraints for a given decoding order. This then leads to an even lower complexity algorithm to determine the MCSs and the powers. Our numerical results show that this approach incurs a less than 2% drop in the average weighted α-utility despite its simplicity.
4) *Imperfect CSI:* We then extend the problem formulation and resource allocation algorithm to account for imperfect CSI that occurs in practice due to channel estimation errors and leads to additional interference.
5) *Analysis:* We then derive lower bounds for the average sum rate and average weighted α-utility for the optimal choice of MCSs and powers for two-user NOMA with imperfect SIC and the F2N order. The discrete nature of rate adaptation makes the analysis fundamentally different and novel compared to the analyses in the literature for NOMA with continuous rate adaptation or OMA. Our work is the first to analyze the general weighted α-utility of NOMA with discrete rate adaptation, MCSs that use error correction codes, and imperfect SIC. Our approach leads to bounds whose terms are elegantly expressed in terms of a single function. We also derive a novel upper bound for the outage probability, which is the probability that no feasible allocation of MCSs and powers exists for any decoding order.
6) We present extensive numerical results to quantify the impact of discrete rate adaptation, adaptive decoding, imperfect SIC, and imperfect CSI. Our numerical results indicate that the F2N order is near-optimal for $K \leq 3$ users despite its lower complexity. For $K \geq 4$, adapting the decoding order achieves a higher sum rate, but at the expense of a $K!$ increase in complexity. We observe that the continuous rate adaptation model overestimates performance.

Comments: Our model and approach are comprehensive, novel, and practically more relevant in the following respects:

1) We focus on discrete rate adaptation. While it has been studied in the context of cooperative and cognitive communications (see [39], [40] and the references therein), it has received relatively limited attention in the context of NOMA. In NOMA, the optimal solution and even the existence of a feasible solution are different compared to continuous rate adaptation.
2) The few papers in the literature on discrete rate adaptation for NOMA focus on uncoded QAM, which has limited practical relevance [14], [16], [32]–[34]. Instead, our model applies to the general class of MCSs employed in cellular standards that use various constellations and error correction codes with different rates.
3) We account for the impact of imperfect SIC in each decoding stage on the BLER targets of the K users. Combined with discrete rate adaptation, this leads to an NP-hard MINLP problem, which differs from the leakage interference-based models studied in [35]–[38] that assume continuous rate adaptation. Our approach also addresses imperfect CSI. Such a comprehensive combination is not available in the NOMA literature to the best of our knowledge.
4) We present the optimal solution for the general class of weighted α-utility functions, which subsumes sum rate ($\alpha = 0$) [3], [4], PF ($\alpha = 1$) [23], [24], [29], max-min fairness ($\alpha = \infty$) [6]–[9], [25], [26], [32], and weighted sum rate [10]–[13], [41]–[43]. This enables us to study the trade-off between sum rate and fairness in a systematic and comprehensive manner.
5) Unlike [3]–[16], which assume a fixed SIC decoding order, we determine the optimal decoding order to maximize performance.
6) Our approach applies to the general K-user case, unlike [3], [6], [9], [10], [14], [16], [17], [24], [28], [33]–[35], [37], [43], [44], which focus only on two users. Our approach also incorporates the minimum rate and BLER constraints. These aspects that have not been jointly addressed in the existing literature.
7) No other performance analysis of NOMA covers discrete rate adaptation and imperfect SIC for two-user NOMA. Our analysis is also the first to cover the entire class of weighted α-utility functions.

C. Outline and Notations

Section II describes the system model. Section III formulates the optimization problem and presents our algorithm to find the optimal allocation of MCSs, powers, and decoding order and its simpler variant. It also presents the extension

TABLE I
COMPARISON OF LITERATURE ON FAIRNESS, RATE ADAPTATION, MINIMUM RATE CONSTRAINT, GENERALITY, IMPERFECT SIC, AND ADAPTIVE DECODING IN NOMA

Reference	*Utility function*	*Discrete rate adaptation*	*Minimum rate constraint*	*General K-user case*	*Imperfect SIC*	*Adaptive decoding*
Shen et al. [3],	Sum rate	No	Yes	No (2-user)	No	No
Li et al. [4]	Sum rate	No	No	Yes	No	No
Zhu et al. [6]	Sum rate, max-min, weighted sum rate	No	Yes	No (2-user)	No	No
Xiao et al. [7]	Max-min	No	Yes	Yes	No	No
Liu et al. [25], Xu et al. [26] Timotheou et al. [8]	Max-min	No	No	Yes	No	No
Zheng et al. [9]	Max-min	No	No	No (2-user)	No	No
Van et al. [10]	Weighted sum rate	No	No	No (2-user)	No	No
Wang et al. [12], Salaün et al. [13]	Weighted sum rate	No	No	Yes	No	No
Wang et al. [11], Kim et al. [41], Randrianantenaina et al. [42]	Weighted sum rate	No	Yes	Yes	Yes	No
Chen et al. [23], Youssef et al. [29]	PF	No	No	Yes	No	No
Hojeij et al. [24]	PF	No	No	No (2-user)	No	No
Xu et al. [27], Xu et al. [30]	α-fair	No	No	Yes	No	No
Zeng et al. [31]	α-fair	No	Yes	No (2-user)	No	No
Fang et al. [28]	Min-max	No	No	No (2-user)	No	No
Pastore et al. [32]	Max-min	No, Uncoded 16-QAM	Yes	Yes	No	No
M. Chitra et al. [35], Luo et al. [36]	-	No	Yes	No (2-user)	Yes	No
Zeng et al. [37], Lim et al. [38]	-	No	Yes	Yes	Yes	No
Yu et al. [14] (DPRA)	-	Yes (Uncoded M-QAM)	Yes	No (2-user)	No	No
Iraqi et al. [34], Choi et al. [33], Yahya et al. [16]	-	Yes (Uncoded M-QAM)	No	No (2-user)	No	No
Sruthy et al. [43]	Weighted sum rate	Yes	No	No (2-user)	No	No
Amin et al. [17]	Sum rate, max-min	No	No	No (2-user)	Yes	Yes
Gao et al. [18], Gao et al. [19]	-	No	No	Yes	No	Yes
Zakeri et al. [20]	Sum rate	No	No	Yes	No	Yes
This manuscript	**Weighted α-utility**	**Yes (General MCSs)**	**Yes**	**Yes**	**Yes**	**Yes**

to imperfect CSI. Section IV analyzes the average sum rate and average weighted α-utility of the optimal solution, and the outage probability. Numerical results are presented in Section V. Our conclusions follow in Section VI.

Notations: $\Pr(A)$ denotes the probability of an event A, and $\mathbb{E}[.]$ denotes the expectation. A^C represents the complement of the event A. The cumulative distribution function (CDF) and probability density function (PDF) of a random variable (RV) X are denoted by $F_X(.)$ and $f_X(.)$, respectively. $X \sim \mathcal{CN}(0, \sigma^2)$ means that X is a circularly symmetric complex Gaussian random variable with variance σ^2.

II. SYSTEM MODEL

Consider a downlink NOMA transmission in which the BS transmits simultaneously to K users over a time-frequency resource.[1] Without loss of generality, we index the users in the ascending order of their distances from the BS. Thus, user 1 is the nearest user to the BS and user K is the farthest. The BS transmits the superimposed signals of the K users with a total power of P_{tot}.

Let $\Gamma_k = g_k \ell_k / \sigma^2$ denote the ratio of the channel power gain to the noise power of the k^{th} user. Here, g_k is a unit-power exponential random variable that models fading, ℓ_k is the pathloss, and σ^2 is the noise variance. Furthermore, $\ell_k = \kappa (d_0/d_k)^\eta$, where η is the pathloss exponent, d_0 is the reference distance, κ is a constant, and d_k is the distance between user k and the BS. We first assume perfect CSI at the users and the BS to develop the key ideas. Thus, the BS knows $\Gamma_1, \ldots, \Gamma_K$, as has also been assumed in [4], [7], [8], [11]–[13], [18], [25], [26], [32]. We extend the model to account for imperfect CSI in Section III-C.

Discrete Rate Adaptation and BLER: Let $\Omega = \{0, 1, \ldots, L-1\}$ denote the discrete set of L MCSs that can be used for transmission. The information rate of MCS $m \in \Omega$ is r_m bits/symbol. The MCSs are arranged in the increasing order of their rates: $0 = r_0 < r_1 < r_2 < \cdots < r_{L-1}$. Here, MCS 0 has a rate $r_0 = 0$, i.e., no transmission occurs.

For MCS m, the BLER as a function of the SINR γ, which we denote by $\mathsf{BLER}_m(\gamma)$, is given by [45]

$$\mathsf{BLER}_m(\gamma) = \begin{cases} a_m e^{-b_m \gamma}, & \gamma \geq \lambda_m, \\ 1, & \text{else}, \end{cases} \tag{1}$$

where $a_m > 0$ and $b_m > 0$ are MCS-specific constants and $\lambda_m = \log(a_m)/b_m$. The above formula holds for general constellations with error correction coding. The constants a_m and b_m are tabulated in the literature for different MCSs [45].[2]

Let $T_m(\epsilon)$ be the smallest SINR at which the BLER of MCS m in an additive white Gaussian noise channel is equal to ϵ. We shall refer to it as the decoding threshold of MCS m. It follows from (1) that

$$T_m(\epsilon) = \frac{1}{b_m} \log\left(\frac{a_m}{\epsilon}\right). \tag{2}$$

[1]The design of the scheduler that selects the K users is beyond the scope of this work.

[2]These constants are determined empirically as analytical formulae for them are available only for some uncoded constellations [46].

TABLE II
MCSs FROM 5G NR STANDARD, THEIR RATES (IN BITS/SYMBOL), AND THEIR DECODING THRESHOLDS [2, TABLE 5.2.2.1-3]

m	Modulation	r_m	Decoding threshold (dB) $T_m(0.1)$	$T_m(\frac{0.1}{2})$	$T_m(\frac{0.1}{3})$	$T_m(\frac{0.1}{4})$
0	-	0	-	-	-	-
1	QPSK	0.15	-7.24	-7.02	-6.90	-6.82
2	QPSK	0.23	-6.65	-6.45	-6.33	-6.25
3	QPSK	0.38	-3.93	-3.74	-3.63	-3.56
4	QPSK	0.60	-0.36	-0.16	-0.05	0.02
5	QPSK	0.88	1.41	1.63	1.76	1.85
6	QPSK	1.18	3.19	3.38	3.48	3.56
7	16-QAM	1.48	5.83	6.05	6.18	6.26
8	16-QAM	1.91	7.75	7.97	8.10	8.19
9	16-QAM	2.41	9.55	9.78	9.90	9.99
10	64-QAM	2.73	11.23	11.47	11.61	11.71
11	64-QAM	3.32	13.23	13.49	13.63	13.73
12	64-QAM	3.90	15.63	15.91	16.06	16.16
13	64-QAM	4.52	19.13	19.43	19.59	19.71
14	64-QAM	5.12	22.31	22.73	22.96	23.12
15	64-QAM	5.55	24.55	25.05	25.31	25.49

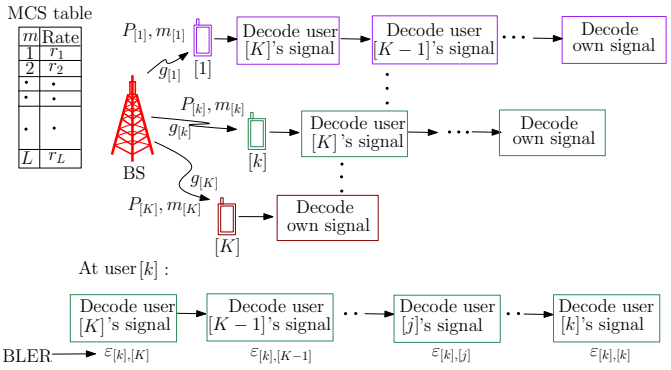


Fig. 1. System model for NOMA with K users that illustrates the powers, MCS indices, and channel gains of all users for a given decoding order $([1],[2],\ldots,[K])$. The MCS table used for discrete rate adaptation is also shown. The multi-stage decoding by user $[k]$ along with the BLER at each stage is shown at the bottom.

Table II shows the MCSs, rates, and decoding thresholds for the MCSs employed in the 5G standard for different BLER targets [2, Table 5.2.2.1-3]. We shall make the following monotonicity assumption about the decoding thresholds:

$$T_m(x) < T_{m+1}(x),\ \forall\ 0 \le x \le 1, \tag{3}$$

i.e., a higher rate MCS has a higher decoding threshold for the same BLER target. It can be numerically verified to hold for the MCSs in Table II.

Weighted α-Utility Function: We consider the following general utility function for a user assigned a rate r [1, Ch. 2]:

$$U_\alpha(r) = \begin{cases} \log(r), & \text{if } \alpha = 1, \\ \dfrac{r^{1-\alpha}}{1-\alpha}, & \text{if } \alpha \neq 1 \text{ and } \alpha \geq 0. \end{cases} \tag{4}$$

The weighted α-utility of the system is equal to $\sum_{k=1}^{K} w_k U_\alpha(r_{m_k})$, where $w_k \geq 1$ is the weight for user k and m_k is the MCS assigned to user k. Intuitively, a larger weight prioritizes the corresponding user. $w_K \geq w_{K-1} \geq \cdots \geq w_1$ ensures a fairer allocation to users further away from the BS.

The strength of our formulation is that it includes as special cases various notions of fairness, which have been separately considered in the literature and have led to limited insights about the sum-rate versus fairness trade-off. It includes sum rate maximization ($\alpha = 0$ and unit weights) [3], [4], proportional fairness (PF) ($\alpha = 1$ and unit weights) [23], [24], [29], α-fair optimization (unit weights) [27], [30], [31], and max-min fairness ($\alpha = \infty$ and unit weights) [6]–[9], [25], [26], [32]. It also includes weighted sum rate fairness ($\alpha = 0$ and different weights) [10]–[13], [41], [42]. The choice of α and the weights enables the system designer to trade-off between fairness and spectral efficiency depending on the specific requirements and constraints of the NOMA system.

A. Decoding Order and SINRs

We first formally define the decoding order. We then specify the SINRs of the K users in terms of the decoding order.

Definition A decoding order π is a permutation of the set $\{1,2,\ldots,K\}$, where $\pi(k)$ denotes the index of the user whose signal is decoded in the $(K-k+1)^{\text{th}}$ decoding stage.

To keep the notation crisp, we shall denote $\pi(k)$ by $[k]$. A decoding order π is fully specified by the K-tuple $([1],[2],\ldots,[K])$; we shall use them interchangeably. For exampe, the F2N order, which is assumed in [3]–[16], sets $[1]=1,[2]=2,\ldots,[K]=K$.[3]

The number of decoding stages at a user depends on the decoding order. At user $[1]$'s receiver, user $[K]$ is decoded first, followed by user $[K-1]$, and so on until user $[1]$. At user $[2]$'s receiver, user $[K]$ is decoded first, followed by user $[K-1]$, and so on until user $[2]$. For example, under the F2N order, user k decodes and cancels the data of the farther users $k+1,\ldots,K$ before decoding its own data. Consequently, user k has $K-k+1$ decoding stages. The system model with the above notation is shown in Figure 1.

Let Ξ denote the set of all decoding orders. It consists of $K!$ elements. For example, for $K=2$, there are two possible decoding orders: 1) $[1]=1$ and $[2]=2$ (F2N order), and 2) $[1]=2$ and $[2]=1$, where the nearest user's data is decoded before that of the farthest user.

Let user k be allocated a power P_k, for $1 \le k \le K$. The SINR $\gamma_{[k],[j]}$ of user $[k]$ for decoding user $[j]$'s data, after canceling the signals of the users $[j+1],\ldots,[K]$ is

$$\gamma_{[k],[j]} = \frac{P_{[j]}\Gamma_{[k]}}{(P_{[1]}+\cdots+P_{[j-1]})\Gamma_{[k]}+1}. \tag{5}$$

For example, $\gamma_{[1],[K]} = P_{[K]}\Gamma_{[1]}/[(P_{[1]}+\cdots+P_{[K-1]})\Gamma_{[1]}+1]$ is the SINR at user $[1]$ for decoding the farther user $[K]$'s data. User $[1]$ first decodes the data of user $[K]$. If the decoding is successful, which can be determined using the CRC check, then user $[1]$ cancels user $[K]$'s data from its received signal. It then decodes user $[K-1]$'s data with SINR $\gamma_{[1],[K-1]} = P_{[K-1]}\Gamma_{[1]}/((P_{[1]}+\cdots+P_{[K-2]})\Gamma_{[1]}+1)$. The user repeats the

[3]The decoding order needs to be conveyed by the BS to the users along with their MCSs and powers.

cancellation and decoding steps until a decoding error occurs or it decodes its own data. When it finally decodes its own data, the SNR is $\gamma_{[1],[1]} = P_{[1]}\Gamma_{[1]}$. In general, a user's receiver stops decoding subsequent stages once the CRC check fails.

B. BLER Constraints with Imperfect SIC

Imperfect SIC affects each stage of the decoding process at the receiver of every user. Consider user $[k]$ in Figure 1. There are $K-k+1$ decoding stages at the user's receiver. It can decode its own data only after successfully decoding the $K-k$ signals of users $[k+1],\ldots,[K]$. A decoding failure at any stage leads to user $[k]$ failing to decode its own data.

Since the SIC error is random, we model it probabilistically. Let $\epsilon_{[k],[j]}$ denote the decoding error probability of user $[k]$ when it decodes user $[j]$'s data, where $k \leq K$ and $j \leq K$. Let the target BLER for user $[k]$ be ε. Therefore, to ensure that the BLER of user $[k]$ does not exceed ε, we must have

$$\epsilon_{[k],[K]}+(1-\epsilon_{[k],[K]})\epsilon_{[k],[K-1]}+\cdots+(1-\epsilon_{[k],[K]})\ldots\epsilon_{[k],[k]} \leq \varepsilon. \quad (6)$$

Since $0 < \epsilon_{[k],[K]} \ll 1$, the above constraint simplifies to $\epsilon_{[k],[K]}+\epsilon_{[k],[K-1]}+\cdots+\epsilon_{[k],[k]} \leq \varepsilon$. This constraint can be satisfied by setting

$$\epsilon_{[k],[K]} = \epsilon_{[k],[K-1]} = \cdots = \epsilon_{[k],[k]} \leq \varepsilon/(K-k+1). \quad (7)$$

For example, no SIC occurs at user $[K]$'s receiver. Therefore, its BLER constraint is given by $\epsilon_{[K],[K]} \leq \varepsilon$. For user $[K-1]$, which first decodes and cancels user $[K]$'s data and then decodes its own data, the BLER constraint is given by $\epsilon_{[K-1],[K]} = \epsilon_{[K-1],[K-1]} \leq \frac{\varepsilon}{2}$. Similarly, at user $[1]$, which has K decoding stages, we have $\epsilon_{[1],[K]} = \epsilon_{[1],[K-1]} = \cdots = \epsilon_{[1],[1]} \leq \frac{\varepsilon}{K}$. Thus, the BLER constraint for a user that has more decoding stages needs to be tighter to counteract the cumulative effect of imperfect SIC. Note the dependence of the BLER constraints on the decoding order.

C. Weighted α-Utility Maximization in NOMA: Formulation

Our goal is to jointly choose the MCSs of the K users, their powers, and the optimal decoding order that maximize the weighted α-utility function while meeting the BLER and minimum rate constraints for each user. Let $m_k \in \Omega$ be the MCS of user k. For a given decoding order $\pi = ([1],[2],\ldots,[K])$, the optimum MCSs and powers are the solution to the following constrained optimization problem:

$$\mathcal{W}_0 : \max_{\substack{m_1\in\Omega,\ldots,m_K\in\Omega,\\ P_1\geq 0,\ldots,P_K\geq 0}} \left\{\sum_{k=1}^{K} w_k U_\alpha(r_{m_k})\right\}, \quad (8)$$

$$\text{s.t. } \mathsf{BLER}_{m_{[k]}}(\gamma_{[1],[k]}) \leq \frac{\varepsilon}{K},\ldots,$$

$$\mathsf{BLER}_{m_{[k]}}(\gamma_{[k],[k]}) \leq \frac{\varepsilon}{K-k+1}, \forall 1\leq k\leq K, \quad (9)$$

$$\sum_{k=1}^{K} P_k \leq P_{\text{tot}}, \quad (10)$$

$$r_{m_k} \geq R_{\min}, \forall 1\leq k\leq K. \quad (11)$$

The constraint in (9) follows from (7) by substituting $\epsilon_{[k],[j]} = \mathsf{BLER}_{m_{[j]}}(\gamma_{[k],[j]})$ and rearranging terms. There are a total of $K(K+1)/2$ BLER constraints in $\mathcal{W}_0$. The total power constraint is captured in (10) and $R_{\min}$ in (11) is the minimum rate required by each user.[4] We note that $\sum_{k=1}^{K} w_{[k]}U_\alpha(r_{m_{[k]}}) = \sum_{k=1}^{K} w_k U_\alpha(r_{m_k})$ and $\sum_{k=1}^{K} P_{[k]} = \sum_{k=1}^{K} P_k$ since the summation is over all the users.

From the definition of the decoding threshold $T_m(.)$ in (2), (9) is equivalent to $\gamma_{[1],[k]} \geq T_{m_{[k]}}(\varepsilon/K),\ldots,\gamma_{[k],[k]} \geq T_{m_{[k]}}(\varepsilon/(K-k+1))$. Let $\Omega_{\min} \subseteq \Omega$ be the set of MCSs whose rate is greater than or equal to $R_{\min}$. Hence, $\mathcal{W}_0$ is equivalent to the following optimization problem:

$$\mathcal{W}_1 : \max_{\substack{m_1,\ldots,m_K,\\ P_1\geq 0,\ldots,P_K\geq 0}} \left\{\sum_{k=1}^{K} w_k U_\alpha(r_{m_k})\right\}, \quad (12)$$

$$\text{s.t. } \gamma_{[1],[k]} \geq T_{m_{[k]}}\left(\frac{\varepsilon}{K}\right),\ldots,$$

$$\gamma_{[k],[k]} \geq T_{m_{[k]}}\left(\frac{\varepsilon}{K-k+1}\right), \forall 1\leq k\leq K, \quad (13)$$

$$\sum_{i=1}^{K} P_k \leq P_{\text{tot}}, \quad (14)$$

$$m_k \in \Omega_{\min}, \forall 1\leq k\leq K. \quad (15)$$

$\mathcal{W}_1$ is a MINLP problem. Even though the non-linearity is confined to the objective function, this class of problems is NP-hard [47, Ch. 15]. Alternatively put, the problem is combinatorial in nature due to discrete rate adaptation. Given the channel gains, the problem requires comparing the weighted α-utility of $\mathcal{O}(L^K)$ feasible MCS choices and choosing the best among them. When we solve $\mathcal{W}_0$ for the $K!$ decoding orders and choose the one that yields the largest weighted α-utility function, the search complexity is $\mathcal{O}(K!L^K)$.

III. Optimal MCS Choice, Powers, and Decoding Order

We shall use the following terminology. Let $\Omega_{\min}^K \triangleq \Omega_{\min}\times\ldots\times\Omega_{\min}$, where $\times$ denotes Cartesian product. For a given MCS choice $\mathbf{m} = (m_1, m_2, \ldots, m_K) \in \Omega_{\min}^K$, where $m_k \in \Omega_{\min}$, $\forall k \in \{1,\ldots,K\}$, we say that a *feasible power allocation* $(P_1, P_2, \ldots, P_K)$ exists for a decoding order π if it satisfies the constraints in (13) and (14). If such a feasible power allocation exists, then $\mathbf{m}$ is feasible for π.

We present the OMCSP algorithm below that provably finds the optimal solution of $\mathcal{W}_1$ (for a given decoding order). It uses two results to reduce the search complexity. The first result in Theorem 1 presents a closed-form criterion to determine whether an MCS choice is feasible or not for a given decoding order. The second result in Corollary 1 eliminates several other MCS choices from the search once an MCS choice is determined to be infeasible for a given decoding order. Then, we search over all decoding orders to determine the optimal one.

When we optimize the decoding order, we call it the OMCSP-Adaptive (*OMCSP-A*) algorithm. When we consider only the F2N order, we call it the *OMCSP-F2N* algorithm.

Theorem *1:* An MCS choice $\mathbf{m} \in \Omega_{\min}^K$, is feasible for a given decoding order $([1],[2],\ldots,[K])$ if and only if

[4]The above formulation can be generalized to ensure that different users have different minimum rate requirements.

the following allocation of powers satisfies the total power constraint $\sum_{i=1}^{K} P^*_{[i]} \leq P_{\text{tot}}$:

$$P^*_{[1]} = \frac{1}{\Gamma_{[1]}} T_{m_{[1]}}\left(\frac{\varepsilon}{K}\right), \tag{16}$$

$$P^*_{[k]} = \max\left\{ T_{m_{[k]}}\left(\frac{\varepsilon}{K}\right)\left[\sum_{q=1}^{k-1} P^*_{[q]} + \frac{1}{\Gamma_{[1]}}\right], \cdots, \right.$$
$$\left. T_{m_{[k]}}\left(\frac{\varepsilon}{K-k+1}\right)\left[\sum_{q=1}^{k-1} P^*_{[q]} + \frac{1}{\Gamma_{[k]}}\right]\right\}, \forall 2 \leq k \leq K. \tag{17}$$

Furthermore, any feasible power allocation must satisfy

$$P_{[1]} \geq P^*_{[1]}, \ldots, P_{[K]} \geq P^*_{[K]}. \tag{18}$$

Proof: The proof is given in Appendix A. ∎

We shall refer to $(P^*_1, \ldots, P^*_K)$ as the *minimum feasible (MF)* power allocation for the decoding order π. Implicit in the above theorem statement is the fact that for an MCS choice, the MF power allocation is not the only feasible solution. In fact, an infinite number of feasible power allocation vectors can exist for a decoding order. The geometry of the feasible power allocation vectors is described in Appendix A. This non-uniqueness is an artifact of discrete rate adaptation.

Corollary *1:* If an MCS choice $\mathbf{m} = (m_1, m_2, \ldots, m_K) \in \Omega^K_{\min}$ is infeasible for a given decoding order, then all MCS choices of the form $(m_1 + i_1, m_2 + i_2, \ldots, m_K + i_K)$, $\forall i_1, \ldots, i_K \geq 0$ are infeasible.

Proof: The proof is given in Appendix B. ∎

We note that nothing can be stated in general about the infeasibility of an MCS choice obtained by other operations such as exchanging the MCSs among the users. Therefore, such operations are not considered.

Infeasibility: It is possible that no feasible solution exists for $\mathcal{W}_1$ for any decoding order. This depends on $\Gamma_1, \ldots, \Gamma_K$, and P_{tot}. In this case, the system is said to be in outage. We analyze the probability of outage in Section IV-B.

A. OMCSP and OMCSP-A Algorithms

The pseudo-code of OMCSP is given in Algorithm 1 for a given decoding order. The algorithm proceeds as follows. The sets $\mathcal{S}$ and $\mathcal{I}$ keep track of the MCS choices that are feasible and infeasible, respectively. The sets $\mathcal{U}$ and $\mathcal{P}$ keep track of the weighted α-utilities and their corresponding MF power allocation of the feasible MCS choices. Initially, $\mathcal{S} = \emptyset$ and $\mathcal{I}$ includes all MCS choices that do not meet the minimum rate constraint of at least one user and are, thus, infeasible.

We start with an MCS choice $\mathbf{m} \in \Omega^K \setminus \mathcal{I}$. Using Theorem 1, we check whether a feasible power allocation exists for it. If $\mathbf{m}$ is feasible, then we include it in $\mathcal{S}$. We find its weighted α-utility and powers and include them in $\mathcal{U}$ and $\mathcal{P}$, respectively. Otherwise, we include all MCS choices of the form $(m_1 + i_1, m_2 + i_2, \ldots, m_K + i_K)$, $\forall i_1, i_2, \ldots, i_K \geq 0$, (which includes $\mathbf{m}$) in $\mathcal{I}$. We then pick another MCS choice from $\Omega^K \setminus (\mathcal{I} \cup \mathcal{S})$ and repeat the above steps. The search stops when no more MCS choices remain, i.e., $\mathcal{I} \cup \mathcal{S} = \Omega^K$. Since Ω is finite, the search will terminate. Finally, we find U_{opt} as the highest weighted α-utility from $\mathcal{U}$. We declare $\mathbf{m}_{\text{opt}}$ and $\mathbf{P}_{\text{MF}}$ as the MCS choice in $\mathcal{S}$ and the MF power allocation in $\mathcal{P}$, respectively, corresponding to U_{opt}. If no feasible solution exists, then $\mathcal{U} = \emptyset$.

Algorithm 1: OMCSP

Input: $\pi, \Gamma_1, \ldots, \Gamma_K, P_{\text{tot}}, \alpha, w_1, w_2, \ldots, w_K, \Omega_{\min}, \varepsilon$
Initialize: $\mathbf{m}_{\text{opt}} = (0, 0, \ldots, 0)$, $\mathbf{P}_{\text{MF}} = (0, 0, \ldots, 0)$, $\mathcal{S} = \emptyset$, $\mathcal{U} = \emptyset$, $\mathcal{P} = \emptyset$, and $[1] = \pi(1), \ldots, [K] = \pi(K)$
Initialize: $\mathcal{I} = \{(m_1, \ldots, m_K) : \exists k \in \{1, \ldots, K\}$ such that $m_k \notin \Omega_{\min}\}$
for $(m_1, m_2, \ldots, m_K) \notin \mathcal{I} \cup \mathcal{S}$ **do**
  if *MF power allocation* $(P^*_1, P^*_2, \ldots, P^*_K)$ *meets total power constraint (use Theorem 1)* **then**
    $\mathcal{S} = \mathcal{S} \cup \{(m_1, m_2, \ldots, m_K)\}$;
    $\mathcal{U} = \mathcal{U} \cup \left\{\sum_{k=1}^{K} w_k U_\alpha(r_{m_k})\right\}$;
    $\mathcal{P} = \mathcal{P} \cup \{(P^*_1, P^*_2, \ldots, P^*_K)\}$;
  else
    $\mathcal{I} = \mathcal{I} \cup \{(m_1 + i_1, m_2 + i_2, \ldots, m_K + i_K),\ \forall i_1, i_2, \ldots, i_K \geq 0\}$;
  end
end
$U_{\text{opt}} \leftarrow$ Highest weighted α-utility in $\mathcal{U}$;
$\mathbf{m}_{\text{opt}} \leftarrow$ MCS choice in $\mathcal{S}$ corresponding to U_{opt};
$\mathbf{P}_{\text{MF}} \leftarrow$ Power allocation in $\mathcal{P}$ corresponding to U_{opt};

Algorithm 2: OMCSP-A

Input: $\Gamma_1, \ldots, \Gamma_K, P_{\text{tot}}, \alpha, w_1, w_2, \ldots, w_K, \Omega_{\min}, \varepsilon$
Initialize: $\mathcal{T} = \emptyset$, $\mathcal{M} = \emptyset$, and $\mathcal{P} = \emptyset$
for $\pi \in \Xi$ **do**
  Find U_{opt}, $\mathbf{m}_{\text{opt}}$, and $\mathbf{P}_{\text{MF}}$ using Algorithm 1
  $\mathcal{T} = \mathcal{T} \cup \{U_{\text{opt}}\}$;
  $\mathcal{M} = \mathcal{M} \cup \{\mathbf{m}_{\text{opt}}\}$;
  $\mathcal{P} = \mathcal{P} \cup \{\mathbf{P}_{\text{MF}}\}$;
end
$\pi^* = \text{argmax}\{\mathcal{T}\}$;
$\mathbf{m}^*_{\text{opt}} \leftarrow$ MCS choice in $\mathcal{M}$ corresponding to π^*;
$\mathbf{P}^*_{\text{MF}} \leftarrow$ Power allocation in $\mathcal{P}$ corresponding to π^*;

Theorem *2:* Given the decoding order, OMCSP finds the optimal MCS choice and MF power allocation for $\mathcal{W}_1$.

Proof: The proof is given in Appendix C. ∎

The pseudo-code for OMCSP-A is presented in Algorithm 2. The sets $\mathcal{T}$, $\mathcal{M}$, and $\mathcal{P}$ store the highest weighted α-utility U_{opt}, the corresponding MCS choice $\mathbf{m}_{\text{opt}}$, and the associated MF power allocation, respectively, for a given decoding order π. Initially, all three sets are empty. The algorithm begins with a decoding order π and determines U_{opt}, $\mathbf{m}_{\text{opt}}$, and $\mathbf{P}_{\text{MF}}$ using Algorithm 1. These values are added to $\mathcal{T}$, $\mathcal{M}$, and $\mathcal{P}$, respectively. This process is repeated for all decoding orders. Finally, the optimal decoding order π^* is identified as the one yielding the highest weighted α-utility in $\mathcal{T}$. The optimal MCS for OMCSP-A, $\mathbf{m}^*_{\text{opt}}$, is the MCS choice corresponding to π^* from $\mathcal{M}$, while the optimal power allocation, $\mathbf{P}^*_{\text{MF}}$, is the associated power assignment from $\mathcal{P}$.

Complexity: To find the optimum MCS choice for a given

decoding order, exhaustive search has to check the feasibility of $\mathcal{O}(L^K)$ MCS choices. Each feasibility check requires $\mathcal{O}(K^2)$ computations in Theorem 1 because there are $K-1$ such maxima to be evaluated. The k^{th} one involves computing the maximum of k numbers, which has $\mathcal{O}(k)$ complexity. When the decoding order is optimized, the complexity increases by a factor $K!$. With the MCS pruning of Algorithm 1, the number of MCS choices whose feasibility needs to be evaluated is lower for OMCSP. The reduction depends on system parameters such as P_{tot} and also the channel realization. We present the numerical complexity results in Section V.

Note: Compared to [15], [16], which also considered discrete rate adaptation, our model and analysis are both different. First, [15], [16] consider only uncoded constellations. Second, they assume per-symbol joint decoding and evaluate the BER. However, we consider all MCSs specified in the standard; these use error correction codes with different code rates. We consider SIC and CRC-based error detection.

B. Lower-Complexity Approach

The optimization problem $\mathcal{W}_0$ has $\mathcal{O}(K^2)$ BLER constraints for a given decoding order. It led to Theorem 1, where $K-1$ maxima each involving up to K terms needed to be computed. Thus, evaluating the feasibility of an MCS choice entails $\mathcal{O}(K^2)$ calculations. We now propose an approach that uses a tighter set of only K constraints and a computationally simpler feasibility check. As we shall see, its average weighted α-utility is within 1.8% of the optimal value.

Lemma *1:* The following constraints imply the constraints in (13):

$$\min\{\gamma_{[1],[k]},\ldots,\gamma_{[k],[k]}\} \geq \max\left\{T_{m_{[k]}}\left(\frac{\varepsilon}{K}\right),\ldots, T_{m_{[k]}}\left(\frac{\varepsilon}{K-k+1}\right)\right\},\ \forall 1\leq k\leq K. \quad (19)$$

Proof: We skip the proof to conserve the space. ∎

From (2), $T_{m_{[k]}}(\varepsilon/K) \geq T_{m_{[k]}}(\varepsilon/(K-k+1))$, $\forall 1 \leq k \leq K$. In the F2N order, since user k is decoded before users $k-1,\ldots,1$, which are closer to the BS, we have $\gamma_{j,k} > \gamma_{k,k}$ with high probability. Hence, (19) simplifies to $\gamma_{k,k} \geq T_{m_k}(\varepsilon/K)$. We extend this to all decoding orders to simplify the above constraint. In this case, Lemma 1 leads to the following optimization problem.

$$\mathcal{W}_2: \max_{\substack{m_1,\ldots,m_K,\\ P_1\geq 0,\ldots,P_K\geq 0}} \left\{\sum_{k=1}^{K} w_k U_\alpha(r_{m_k})\right\}, \quad (20)$$

$$\text{s.t.}\quad \gamma_{[k],[k]} \geq T_{m_{[k]}}\left(\frac{\varepsilon}{K}\right), \forall 1\leq k\leq K, \quad (21)$$

$$\sum_{k=1}^{K} P_k \leq P_{\text{tot}}, \quad (22)$$

$$m_k \in \Omega_{\min}, \forall 1\leq k\leq K. \quad (23)$$

$\mathcal{W}_2$ is a tightening of $\mathcal{W}_1$ for the F2N order. Furthermore, the tightening is exact for $K=2$ under the F2N order because $\gamma_{1,2} > \gamma_{2,2} \geq T_{m_2}(\varepsilon/2) > T_{m_2}(\varepsilon)$.

The equivalent of Theorem 1 for checking whether an MCS choice is feasible in $\mathcal{W}_2$ is then as follows. We skip the proof as it is similar to Appendix A.

Theorem *3:* An MCS choice $\mathbf{m} \in \Omega_{\min}^K$, is feasible for a given decoding order $([1],[2],\ldots,[K])$ if and only if the following power allocation satisfies $\sum_{i=1}^{K} P^*_{[i]} \leq P_{\text{tot}}$:

$$P^*_{[1]} = \frac{T_{m_{[1]}}(\varepsilon/K)}{\Gamma_{[1]}}, \quad (24)$$

$$P^*_{[k]} = \frac{T_{m_{[k]}}(\varepsilon/K)\left[\left(P^*_{[1]}+\cdots+P^*_{[k-1]}\right)\Gamma_{[k]}+1\right]}{\Gamma_{[k]}}, \quad \forall 2\leq k\leq K. \quad (25)$$

In the above theorem, the complexity of computing the feasibility of an MCS choice decreases from $\mathcal{O}(K^2)$ in Theorem 2 to $\mathcal{O}(K)$. Specifically, OMCSP considers 3, 6, and 10 BLER constraints for $K=2$, 3, and 4, respectively. On the other hand, the corresponding number of BLER constraints considered by the lower-complexity approach decreases to 2, 3, and 4. To determine the optimal MCS choice, power allocations, and decoding order for the lower-complexity approach, Algorithms 1 and 2 can be applied after replacing Theorem 1 with Theorem 3.

C. Imperfect CSI

We now update the problem formulation and algorithm to handle imperfect CSI. Let h_k represent the complex baseband fading channel for user k. Let $\hat{h}_k$ be its estimate at the user. Then,

$$h_k = \sqrt{\rho_k}\hat{h}_k + \sqrt{1-\rho_k}v_k, \quad (26)$$

where $\sqrt{\rho_k} \in (0,1]$ is the correlation coefficient between h_k and $\hat{h}_k$, and $v_k \sim \mathcal{CN}(0,1)$ is the estimation error, which is independent of $\hat{h}_k$. For example, in minimum mean square error (MMSE) channel estimation, $\rho_k = P_p\ell_k/(P_p\ell_k+\sigma^2)$, where P_p is the pilot power. As $P_p/\sigma^2 \to \infty$, we get $\rho_k \to 1$ and $|\hat{h}_k|^2 \to |h_k|^2$, which implies perfect CSI.

Using this model, the received signal $y_{[k]}$ at user $[k]$ is

$$y_{[k]} = \sqrt{\rho_{[k]}}\hat{h}_{[k]}\left(\sqrt{P_{[1]}}x_{[1]}+\cdots+\sqrt{P_{[K]}}x_{[K]}\right) + \sqrt{1-\rho_{[k]}}\left(\sqrt{P_{[1]}}x_{[1]}+\cdots+\sqrt{P_{[K]}}x_{[K]}\right)v_{[k]} + n_{[k]}, \quad (27)$$

where $x_{[k]}$ is the data symbol for user $[k]$ and $n_{[k]} \sim \mathcal{CN}(0,\sigma^2)$ is the noise. The second term is additional interference due to imperfect CSI. The SINR corresponding to (5) becomes

$$\gamma_{[k],[j]} = \frac{P_{[j]}\hat{\Gamma}_{[k]}}{\left(\sum_{t=1}^{j-1}P_{[t]}\right)\hat{\Gamma}_{[k]} + \frac{(1-\rho_{[k]})}{\sigma^2}\left(\sum_{t=1}^{K}P_{[t]}\right)+1}, \quad (28)$$

where $\hat{\Gamma}_k = \rho_k|\hat{h}_k|^2/\sigma^2$. Using this updated SINR definition, the optimization problem $\mathcal{W}_0$ can be solved following the approach in Section II-C. Here, the BS knows $\hat{\Gamma}_k$ and $(1-\rho_k)/\sigma^2$, $\forall 1\leq k\leq K$.

IV. Average Sum Rate and Outage Probability Analysis ($K=2$)

We now analyze the average sum rate of the optimal solution and the outage probability of $\mathcal{W}_1$ with perfect CSI.

The analysis is intractable for the general case of K users. This is because for a given decoding order, an MCS choice is optimal only if all MCS choices with a larger weighted α-utility are infeasible. Since the number of such choices scales as $\mathcal{O}(L^{K-1})$, the complexity grows exponentially with K. We, therefore, focus on two-user NOMA and the F2N order in this section. It serves as a bound for the case when the decoding order is adapted. Such an analysis of two-user NOMA with discrete rate adaptation and imperfect SIC is novel compared to the literature. A new challenge that the analysis circumvents is that the optimal MCS choice is not known in closed-form.

A. Average Sum Rate

Let $\Pr(\mathbf{m}_{\text{opt}} = \mathbf{m})$ be the probability that MCS choice $\mathbf{m} = (m_1, m_2)$ is the optimal MCS choice $\mathbf{m}_{\text{opt}}$ that maximizes the weighted α-utility for the F2N order. For $\mathbf{m}$, the weighted α-utility $\zeta_{\mathbf{m}}$ is $\zeta_{\mathbf{m}} \triangleq w_1 U_\alpha(r_{m_1}) + w_2 U_\alpha(r_{m_2})$.

Our approach is as follows. We derive the average sum rate. We first write it in terms of $\Pr(\mathbf{m}_{\text{opt}} = \mathbf{m})$. We then derive expressions for $\Pr(\mathbf{m}_{\text{opt}} = \mathbf{m})$, where the utility function subtly manifests itself.

From the law of total expectation, the sum rate, when averaged over the channel gains of the users, is given by

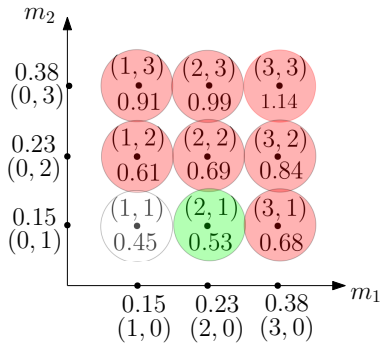
$$\mathbb{E}[r_{m_1} + r_{m_2}] = \sum_{m_1 \in \Omega_{\min}} \sum_{m_2 \in \Omega_{\min}} (r_{m_1} + r_{m_2}) \times \Pr(\mathbf{m}_{\text{opt}} = \mathbf{m}). \quad (29)$$

For an MCS pair $\mathbf{m}$, let $\mathcal{S}_{\mathbf{m}}$ be the set of all MCS pairs whose weighted α-utility exceeds $\zeta_{\mathbf{m}}$. The MCS pair $\mathbf{m}$ is optimum if and only if it is feasible and every MCS pair in $\mathcal{S}_{\mathbf{m}}$ is infeasible. Thus,

$$\Pr(\mathbf{m}_{\text{opt}} = \mathbf{m}) = \Pr(\mathbf{m} \text{ is feasible, all the MCS pairs in } \mathcal{S}_{\mathbf{m}} \text{ are infeasible}). \quad (30)$$

The evaluation of $\Pr(\mathbf{m}_{\text{opt}} = \mathbf{m})$ is challenging for two reasons. First, the events that $\mathbf{m}_{\text{opt}}$ is feasible and an MCS pair in $\mathcal{S}_{\mathbf{m}}$ is infeasible are dependent. This is because the occurrences of these events are determined by the same three SINRs $\gamma_{1,1}, \gamma_{1,2}$, and $\gamma_{2,2}$. Second, $\mathcal{S}_{\mathbf{m}}$ has an involved structure as we explain below. From Corollary 1, it follows that all MCS pairs of the form $(m_1 + i_1, m_2 + i_2)$, $\forall i_1 \geq 0, i_2 \geq 0, i_1 + i_2 > 0$, must be in $\mathcal{S}_{\mathbf{m}}$. *However, due to weighting and the non-linear form of the α-fair utility function, other MCS pairs can also lie in $\mathcal{S}_{\mathbf{m}}$.* This sets our problem apart. We first illustrate this with an example and make observations that lead us to the general analysis.

Example: Figure 2 shows the MCS pairs for $\Omega = \{0, 1, 2, 3\}$ with rates $r_0 = 0, r_1 = 0.15, r_2 = 0.23$, and $r_3 = 0.38$ bits/symbol. These correspond to the first four MCSs in Table II. The weighted α-utility of each MCS pair is shown for $\alpha = 0$, $w_1 = 1$, $w_2 = 2$, and $R_{\min} = 0.1$ bits/symbol. Thus, $\Omega_{\min} = \{1, 2, 3\}$. Consider, for example, the MCS pair $(2, 1)$. Its weighted α-utility is 0.53. For it to be optimal, the MCS pairs $(2, 2)$, $(2, 3)$, $(3, 1)$, $(3, 2)$, and $(3, 3)$ must be infeasible since they clearly have a higher weighted α-utility. However, in addition, the MCS pairs $(1, 2)$ and $(3, 1)$ also have a higher weighted α-utility and must, therefore, be infeasible. Such MCS pairs need to be carefully enumerated and accounted for in the analysis. Another observation we shall exploit in our analysis is that it is sufficient to require that the MCS pairs $(1, 2)$ and $(3, 1)$ are infeasible. Corollary 1 then automatically implies the infeasibility of all the above MCS pairs.



MCS pair	$\mathcal{S}_m$
$(1,1)$	$(1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)$
$(1,2)$	$(1,3), (2,2), (2,3), (3,1), (3,2), (3,3)$
$(1,3)$	$(2,3), (3,3)$
$(2,1)$	$(1,2), (1,3), (2,2), (2,3), (3,1), (3,2), (3,3)$
$(2,2)$	$(1,3), (2,3), (3,3), (3,2), (3,3)$
$(2,3)$	$(3,3)$
$(3,1)$	$(2,2), (3,2), (1,3), (2,3), (3,3)$
$(3,2)$	$(1,3), (2,3), (3,3)$
$(3,3)$	——

Fig. 2. MCS pairs with $\Omega = \{0, 1, 2, 3\}$ and rates $r_0 = 0$, $r_1 = 0.15$, $r_2 = 0.23$, and $r_3 = 0.38$ bits/symbol are shown. The weighted α-utility for each MCS pair is shown for $\alpha = 0$, $w_1 = 1$, $w_2 = 2$, and $R_{\min} = r_1$. Also illustrated is MCS pair $(2, 1)$ and the MCS pairs that need to be infeasible for it to be optimal. At the bottom of the figure, each MCS pair and its $\mathcal{S}_{\mathbf{m}}$ is shown. The infeasibility of the MCS pairs in blue ink does not follow from Corollary 1.

The tight bounding approach that we present below addresses the above two challenges.

General Approach: The utility function in (8) is a monotonically increasing function of the user rate. Therefore, if $\mathbf{m}$ is optimum, then $(m_1 + 1, m_2)$ and $(m_1, m_2 + 1)$, whose weighted α-utilities are greater than $\zeta_{\mathbf{m}}$, must be infeasible. However, since the rates of the users are weighted differently, even MCS pairs of the following forms can have a higher weighted α-utility and must be infeasible[5]:

1) $(m_1 - u, m_2 + 1) \in \mathcal{S}_{\mathbf{m}}$, where $u \geq 0$ is the largest integer such that $\zeta_{m_1 - u, m_2 + 1} > \zeta_{\mathbf{m}}$.
2) $(m_1 + v, m_2 - 1) \in \mathcal{S}_{\mathbf{m}}$, where $v > 0$ is the smallest integer such that $\zeta_{m_1 + v, m_2 - 1} > \zeta_{\mathbf{m}}$.
3) $(m_1 + 1, m_2 - w) \in \mathcal{S}_{\mathbf{m}}$, where $w \geq 0$ is the smallest integer such that $\zeta_{m_1 + 1, m_2 - w} > \zeta_{\mathbf{m}}$.
4) $(m_1 + x, m_2 - 2) \in \mathcal{S}_{\mathbf{m}}$, where $x > 0$ is the smallest integer such that $\zeta_{m_1 + x, m_2 - 2} > \zeta_{\mathbf{m}}$.

[5]Even MCS pairs of the form $(m_1 - 1, m_2 + z)$ and $(m_1 - 2, m_2 + z)$ for some $z > 2$ can have a higher weighted α-utility. However, they do not contribute much to the sum rate while making the analysis more involved.

5) $(m_1-y, m_2+2) \in \mathcal{S}_{\mathbf{m}}$ where $y > 0$ is the largest integer such that $\zeta_{m_1-y,m_2+2} > \zeta_{\mathbf{m}}$.

Let E_0 denote the event that $\mathbf{m}$ is feasible. Let E_u denote the event that $(m_1 - u, m_2 + 1)$ is feasible. Similarly, E_v, E_w, E_x, and E_y denote the events that $(m_1 + v, m_2 - 1), (m_1+1, m_2-w), (m_1+x, m_2-2)$, and (m_1-y, m_2+2), respectively, are feasible. From Corollary 1, the above events imply the infeasibility of many other MCSs.

Using Theorem 3 under the F2N order, E_0 occurs when

$$\Gamma_1 \geq \frac{T_{m_1}(T_{m_2}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2}} \;\&\; \Gamma_2 \geq \frac{T_{m_2}}{P_{\text{tot}}}. \tag{31}$$

E_u occurs when

$$\Gamma_1 \geq \frac{T_{m_1-u}(T_{m_2+1}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2+1}} \;\&\; \Gamma_2 \geq \frac{T_{m_2+1}}{P_{\text{tot}}}. \tag{32}$$

Similarly, E_v, E_w, and E_x respectively occur when

$$\Gamma_1 \geq \frac{T_{m_1+v}(T_{m_2-1}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2-1}} \;\&\; \Gamma_2 \geq \frac{T_{m_2-1}}{P_{\text{tot}}}, \tag{33}$$

$$\Gamma_1 \geq \frac{T_{m_1+1}(T_{m_2-w}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2-w}} \;\&\; \Gamma_2 \geq \frac{T_{m_2-w}}{P_{\text{tot}}}, \tag{34}$$

$$\Gamma_1 \geq \frac{T_{m_1+x}(T_{m_2-2}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2-2}} \;\&\; \Gamma_2 \geq \frac{T_{m_2-2}}{P_{\text{tot}}}. \tag{35}$$

Hence, the probability in (30) can be upper bounded as

$$\Pr(\mathbf{m}_{\text{opt}} = \mathbf{m}) \leq \Pr\left(E_0 \cap E_u^C \cap E_v^C \cap E_w^C \cap E_x^C\right). \tag{36}$$

Using De Morgan's laws, we get

$$\begin{aligned}\Pr\left(E_0 \cap E_u^C \cap E_v^C \cap E_w^C \cap E_x^C\right) &= \Pr(E_0) - \Pr(E_0 \cap E_u) \\ &- \Pr(E_0 \cap E_v) - \Pr(E_0 \cap E_w) \\ &- \Pr(E_0 \cap E_x) + \Pr(E_0 \cap E_u \cap E_v) \\ &+ \Pr(E_0 \cap E_u \cap E_w) + \Pr(E_0 \cap E_u \cap E_x) \\ &+ \Pr(E_0 \cap E_v \cap E_w) + \Pr(E_0 \cap E_v \cap E_x) \\ &+ \Pr(E_0 \cap E_w \cap E_x) - \Pr(E_0 \cap E_u \cap E_v \cap E_w) \\ &- \Pr(E_0 \cap E_u \cap E_v \cap E_x) - \Pr(E_0 \cap E_v \cap E_w \cap E_x) \\ &- \Pr(E_0 \cap E_u \cap E_w \cap E_x) \\ &+ \Pr(E_0 \cap E_u \cap E_v \cap E_w \cap E_x).\end{aligned} \tag{37}$$

We present expressions for each of the probability terms in (37). We show that they can all be written in a unified manner in terms of a common function $\Pi(.)$, which we first define below. Consider the M MCS pairs $\mathbf{n}^{(1)} = \left(n_1^{(1)} n_2^{(1)}\right), \ldots, \mathbf{n}^{(M)} = \left(n_1^{(M)} n_2^{(M)}\right)$. Let $f_2 = \max_{1\leq i\leq M}\left\{n_2^{(i)}\right\}$. Then,

$$\Pi(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}, \ldots, \mathbf{n}^{(M)}) \triangleq \int_{\frac{T_{f_2}}{P_{\text{tot}}}}^{\infty} \frac{\sigma^2}{\ell_2} \exp\left(-\frac{\sigma^2 x}{\ell_2}\right) \times \exp\left(-\frac{\sigma^2}{\ell_1} \max_{1\leq i\leq M}\left\{\frac{T_{n_1^{(i)}}(T_{n_2^{(i)}}+1)x}{xP_{\text{tot}} - T_{n_2^{(i)}}}\right\}\right) dx. \tag{38}$$

The single integral can be simplified using Gauss-Laguerre quadrature to the following easily computable form:

$$\Pi(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}, \ldots, \mathbf{n}^{(M)}) \approx \exp\left(-\frac{T_{f_2}\sigma^2}{P_{\text{tot}}\ell_2}\right) \sum_{i=1}^{p} s_i \times \exp\left(-\frac{\sigma^2}{\ell_1} \max_{1\leq i\leq M}\left\{\frac{\eta(q_i)T_{n_1^{(1)}}(T_{n_2^{(1)}}+1)}{\eta(q_i)P_{\text{tot}} - T_{n_2^{(1)}}}\right\}\right), \tag{39}$$

where $\eta(x) = \frac{x\ell_2}{\sigma^2} + \frac{T_{f_2}}{P_{\text{tot}}}$, p is the number of Gauss-quadrature terms, and q_i and s_i, for $1 \leq i \leq p$, are the Gauss-Laguerre quadrature abscissas and weights, respectively, which are tabulated in [48, Table 25.9].[6]

We organize the terms in (37) on the basis of the number of events they consist of. The probabilities of the one and two event terms in (37) are given by

$$\Pr(E_0) = \Pi(\mathbf{m}), \tag{40}$$
$$\Pr(E_0 \cap E_u) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1)), \tag{41}$$
$$\Pr(E_0 \cap E_v) = \Pi(\mathbf{m}, (m_1 + v, m_2 - 1)), \tag{42}$$
$$\Pr(E_0 \cap E_w) = \Pi(\mathbf{m}, (m_1 + 1, m_2 - w)), \tag{43}$$
$$\Pr(E_0 \cap E_x) = \Pi(\mathbf{m}, (m_1 + x, m_2 - 2)). \tag{44}$$

The probabilities of the three event terms in (37) are given by

$$\Pr(E_0 \cap E_u \cap E_v) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1), (m_1 + v, m_2 - 1)), \tag{45}$$
$$\Pr(E_0 \cap E_u \cap E_w) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1), (m_1 + 1, m_2 - w)), \tag{46}$$
$$\Pr(E_0 \cap E_u \cap E_x) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1), (m_1 + x, m_2 - 2)), \tag{47}$$
$$\Pr(E_0 \cap E_v \cap E_w) = \Pi(\mathbf{m}, (m_1 + v, m_2 - 1), (m_1 + 1, m_2 - w)), \tag{48}$$
$$\Pr(E_0 \cap E_v \cap E_x) = \Pi(\mathbf{m}, (m_1 + v, m_2 - 1), (m_1 + x, m_2 - 2)), \tag{49}$$
$$\Pr(E_0 \cap E_w \cap E_x) = \Pi(\mathbf{m}, (m_1 + 1, m_2 - w), (m_1 + x, m_2 - 2)). \tag{50}$$

The derivation of (40)–(50) is given in Appendix D.

The probabilities of the four event terms in (37) are given by

$$\Pr(E_0 \cap E_u \cap E_v \cap E_w) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1), (m_1 + v, m_2 - 1), (m_1 + 1, m_2 - w)), \tag{51}$$
$$\Pr(E_0 \cap E_u \cap E_w \cap E_x) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1), (m_1 + 1, m_2 - w), (m_1 + x, m_2 - 2)), \tag{52}$$
$$\Pr(E_0 \cap E_v \cap E_w \cap E_x) = \Pi(\mathbf{m}, (m_1 + v, m_2 - 1), (m_1 + 1, m_2 - w), (m_1 + x, m_2 - 2)), \tag{53}$$
$$\Pr(E_0 \cap E_u \cap E_w \cap E_x) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1), (m_1 + 1, m_2 - w), (m_1 + x, m_2 - 2)). \tag{54}$$

[6]The approximation error decreases as p increases. We have found $p = 7$ to be sufficient for the range of parameters of interest.

The probability of the five event term is given by

$$\Pr\left(E_0 \cap E_u \cap E_v \cap E_w \cap E_x\right) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1), (m_1 + v, m_2 - 1), (m_1 + 1, m_2 - w), (m_1 + x, m_2 - 2)). \quad (55)$$

The derivation is similar to Appendix D and is skipped.

Similarly, we can show that

$$\Pr\left(\mathbf{m}_{\text{opt}} = \mathbf{m}\right) \leq \Pr\left(E_0 \cap E_u^C \cap E_v^C \cap E_w^C \cap E_y^C\right). \quad (56)$$

An expression for $\Pr\left(E_0 \cap E_u^C \cap E_v^C \cap E_w^C \cap E_y^C\right)$ can be derived along similar lines as above. Combining (36) and (56) leads to the following tighter bound:

$$\Pr\left(\mathbf{m}_{\text{opt}} = \mathbf{m}\right) \leq \min\left\{\Pr\left(E_0 \cap E_u^C \cap E_v^C \cap E_w^C \cap E_x^C\right), \Pr\left(E_0 \cap E_u^C \cap E_v^C \cap E_w^C \cap E_y^C\right)\right\}. \quad (57)$$

Substituting the above equation in (29) gives an upper bound for the average sum rate.

Asymptotic Insights: We consider the two regimes of small and large P_{tot} to gain more insights.

- $P_{tot} \to \infty$*:* In (38), the lower limit T_{f_2}/P_{tot} in the integral approaches 0. The second exponential term in the integrand approaches 1 and, therefore, so does the integral for any M MCS pairs. It can be shown that the probabilities of all MCSs except MCS pair (L, L) become 0, while that of MCS pair (L, L) becomes 1. Hence, from (29), the average rate approaches r_L.
- $P_{tot} \to 0$*:* In (38), the lower limit T_{f_2}/P_{tot} in the integral approaches ∞ and the integral is zero. Therefore, so does the integral for any M MCS pairs. Hence, $\Pr\left(\mathbf{m}_{\text{opt}} = \mathbf{m}\right) = 0$ for all MCS pairs. Hence, the average rate approaches zero.

B. Average Weighted α-Utility and Outage Probability

Using the law of total expectation, the average weighted α-utility is given by

$$\mathbb{E}\left[\zeta_{\mathbf{m}}\right] = \sum_{m_1 \in \mathcal{Q}_{\min}} \sum_{m_2 \in \mathcal{Q}_{\min}} \zeta_{\mathbf{m}} \Pr\left(\mathbf{m}_{\text{opt}} = \mathbf{m}\right). \quad (58)$$

Substituting the expression for $\Pr(\mathbf{m}_{\text{opt}} = \mathbf{m})$ from Section IV-A in (58) gives the formula for the weighted α-utility.

The outage probability O is given by

$$O = 1 - \sum_{m_1 \in \Omega_{\min}} \sum_{m_2 \in \Omega_{\min}} \Pr\left(\mathbf{m}_{\text{opt}} = \mathbf{m}\right). \quad (59)$$

Substituting the expression for $\Pr(\mathbf{m}_{\text{opt}} = \mathbf{m})$ from Section IV-A in (59) gives the formula for O. Note that O is not a function of α or weights since the existence of a feasible solution for $\mathcal{W}_1$ depends only on its constraints and not its objective function.

The analysis provides an independent validation of the simulation results. It also brings out how the different system parameters affect key performance metrics such as average sum rate, average weighted α-utility, and O.

C. Comparison with OMA with Discrete Rate Adaptation

We now contrast the above analysis with the classical analysis for single-user OMA with discrete rate adaptation [49, Ch. 9]. The average rate of the user k is given by

$$\mathbb{E}[r_m] = \sum_{m=1}^{L} r_m \Pr(T_m \leq \gamma_k < T_{m+1}), \quad (60)$$

$$= \sum_{m=1}^{L} r_m \left(\exp\left(-\frac{\sigma^2 T_m}{P_{\text{tot}} \ell_k}\right) - \exp\left(-\frac{\sigma^2 T_{m+1}}{P_{\text{tot}} \ell_k}\right)\right). \quad (61)$$

We see that (61) is considerably simpler compared to the expressions we derived for $K = 2$.

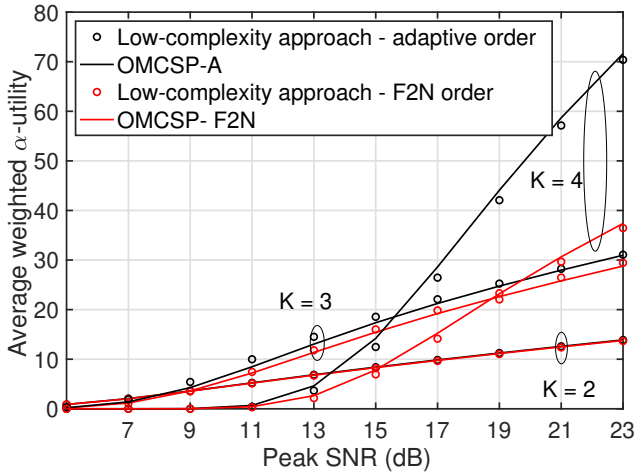


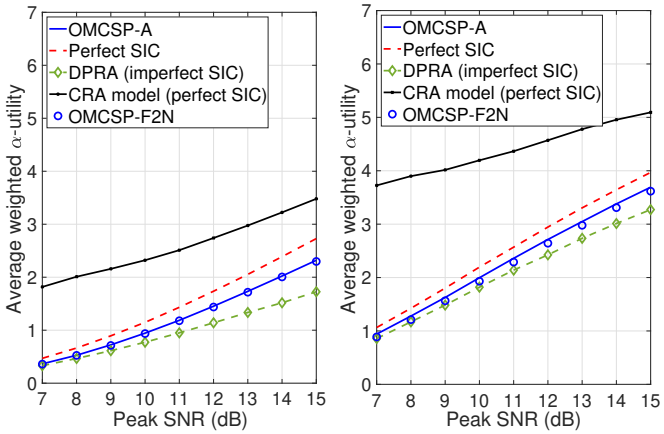
Fig. 3. Average weighted α-utility as a function of the peak SNR $\frac{P_{\text{tot}}\ell_1}{\sigma^2}$ ($\alpha = 0.5$, $w_1 = 1$, $w_2 = 4$, $w_3 = 8$, $w_4 = 12$, $\frac{\ell_1}{\ell_2} = 10$ dB, $\frac{\ell_1}{\ell_3} = 15$ dB, and $\frac{\ell_1}{\ell_4} = 20$ dB).

V. Numerical Results

We now present Monte Carlo simulation results to benchmark the performance and complexity of the proposed method, verify the analysis, and characterize the influence of key system parameters. The MCSs are from Table II [2, Table 5.2.2.1-2]. We set $w_1 = 1$, $\varepsilon = 0.1$, and $R_{\min} = 0.38$ bits/symbol. The results are averaged over $1,000,000$ independent realizations of the channel gains of the users. In the event of an outage, we assign zero power and zero rate to all users.

Figure 3 plots the average weighted α-utilities of OMCSP-A, OMCSP-F2N, and the corresponding lower-complexity approaches as a function of $\frac{P_{\text{tot}}\ell_1}{\sigma^2}$, which we shall refer to as the peak SNR. It shows results for $K = 2$, 3, and 4 users. For $K = 2$, the weighted α-utilities of the OMCSP-F2N and the corresponding lower-complexity approach are identical, as noted in Section III-B. For $K = 3$ and 4, the lower-complexity approach under adaptive order remains within 1.6% and 1.8% of OMCSP-A across all peak SNRs, making it a viable alternative to OMCSP-A.

The weighted α-utility of OMCSP-F2N is within 1.5% of that of OMCSP-A for $K = 2$ and within 7.4% for $K = 3$. Thus, it is sufficient to only consider the F2N order for $K \leq 3$. For $K = 4$, a significant gap emerges between the two approaches. However, this comes at the expense of an increase in complexity, which we study in a subsequent figure. As the



(a) $\alpha = 0$ (b) $\alpha = 0.5$

Fig. 4. Benchmarking: Average weighted α-utility as a function of the peak SNR, $\frac{P_{\text{tot}}\ell_1}{\sigma^2}$, for two values of α ($K = 2$, $\frac{\ell_1}{\ell_2} = 10$ dB, $w_1 = 1$, and $w_2 = 1$).

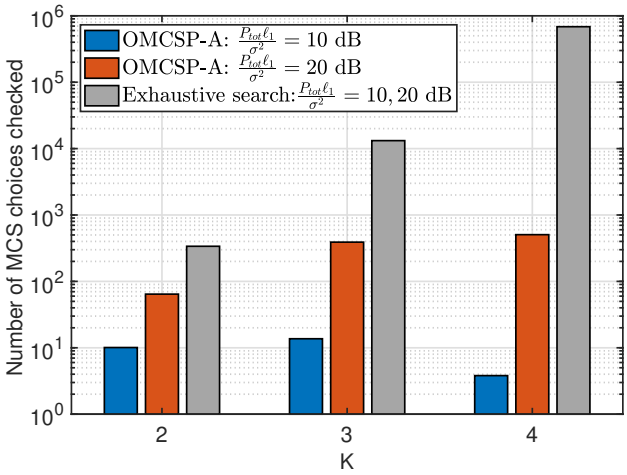


Fig. 5. Complexity: Average number of MCS choices to be checked (in log scale) by OMCSP-A and exhaustive search as a function of number of users K under adaptive order ($\alpha = 1$, $w_1 = 1$, $w_2 = 4$, $w_3 = 8$, $w_4 = 12$, $\frac{\ell_1}{\ell_2} = 10$ dB, $\frac{\ell_1}{\ell_3} = 15$ dB, and $\frac{\ell_1}{\ell_4} = 20$ dB).

peak SNR increases, the weighted α-utility increases for any K as higher rate MCSs can be assigned to the users with higher probabilities. Notice that it is better to serve fewer users at lower SNRs. This is because the minimum rate constraint can be met for fewer users at lower SNRs.

We benchmark the proposed method with the following:

1) *Continuous Rate Adaptation (CRA) Model [3], [4], [6]–[10], [13], [23]–[31], [35]–[38], [41], [42]*: The rate of a user is determined using the Shannon's capacity formula. It is a function of the user's SINR. Each user's optimal power is numerically determined to maximize the weighted α-utility subject to the total power constraint under perfect SIC.[7]
2) *DPRA [14]*: The far user is first assigned its minimum rate and the least power required to support it. Then, the near user's rate is chosen to be as large as possible with the remaining power. Lastly, the far user's rate is increased if any power is still left. However, DPRA is designed for only $K = 2$ and $w_2 = w_1$.
3) *Perfect SIC*: Packet decoding errors are assumed to occur only in the final stage of decoding. Thus, the maximum allowed BLER for all users is the same and is $\varepsilon = 0.1$. All other procedures for determining the optimal MCS choice and power allocations are the same as the proposed method. This idealized scheme helps understand the impact of imperfect SIC in the intermediate stages where other users' signals are decoded.

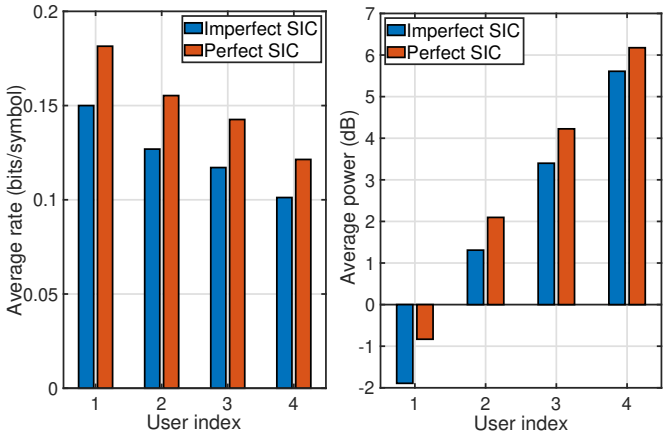
We note that a comparison with [15], [16], [32]–[34], which are the only other papers that consider NOMA with discrete rate adaptation, is not possible. This is because the methods in [16], [32], [33] apply only to uncoded constellations, while [15], [34] focus on the BER of uncoded constellations but not on the rate.

Figures 4(a) and 4(b) compare the average weighted α-utilities of all approaches as a function of $\frac{P_{\text{tot}}\ell_1}{\sigma^2}$. Results are shown for $K = 2$ and two values of α. The weight w_2 is set to 1 as DPRA is designed only for this value. The average weighted α-utility of OMCSP-F2N exceeds that of DPRA by 51.2% for $\alpha = 0$ and by 33.1% for $\alpha = 0.5$ at a peak SNR of 15 dB. We see that the CRA model significantly overestimates the average weighted α-utility. Imperfect SIC causes the average weighted α-utility to decrease by 30-40% compared to perfect SIC. Thus, discrete rate adaptation and imperfect SIC have a significant impact on the performance of NOMA. For $\alpha = 0$ and 0.5, the curves for OMCSP-A and OMCSP-F2N are close to each other for all peak SNRs. As the peak SNR increases, the average weighted α-utilities of all the approaches increase because of the higher odds of assigning higher rate MCSs to the users. The average weighted α-utility is sensitive to α and exhibits a wide variation when α changes from 0 to 0.5.

Figure 5 plots the number of MCS choices to be checked for feasibility by OMCSP-A and exhaustive search as a function of K for two values of the peak SNR. As the peak SNR increases, the number of searches required for each K increases due to the higher probability that a higher rate MCS pair is feasible for a given decoding order. This leads to fewer MCS choices being pruned based on Corollary 1. For $\frac{P_{\text{tot}}\ell_1}{\sigma^2} = 10$ dB, OMCSP-A requires a factor of 33.5, 963.6, and 179949.6 fewer searches than exhaustive search for $K = 2$, 3, and 4, respectively. Thus, a noteworthy one to five orders of magnitude reduction in the search complexity is achieved by OMCSP-A. When only the F2N order is considered, the corresponding reduction factor in the number of searches is 21.4, 194.4, and 4231.2 (figure not shown due to space constraints), which is also noteworthy.

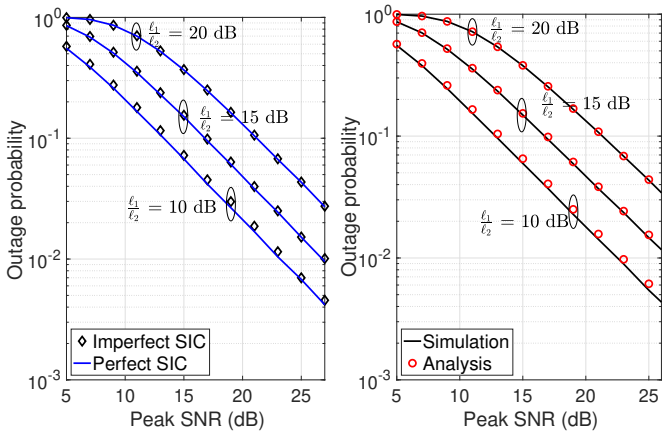
Figure 6(a) shows the average rate of each user with imperfect and perfect SIC under adaptive order for $\alpha = 0.9$ and $K = 4$. Figure 6(b) presents the corresponding average minimum feasible power allocated to each user. The nearest user 1 has the highest average rate and ends up being assigned the lowest power on average. The reverse is true for the farthest user. This holds for both perfect and imperfect SIC.

[7]The BLER is assumed to be zero as it is based on Shannon-theoretic capacity.



(a) Average rate (b) Average power

Fig. 6. Per-user view: Average rate and average power of each user under OMCSP-A ($\alpha = 0.9$, $K = 4$, $\frac{P_{\text{tot}}\ell_1}{\sigma^2} = 15$ dB, $\frac{\ell_1}{\ell_2} = 10$ dB, $\frac{\ell_1}{\ell_3} = 15$ dB, $\frac{\ell_1}{\ell_4} = 20$ dB, $w_1 = 1$, $w_2 = 4$, $w_3 = 8$, and $w_4 = 12$).

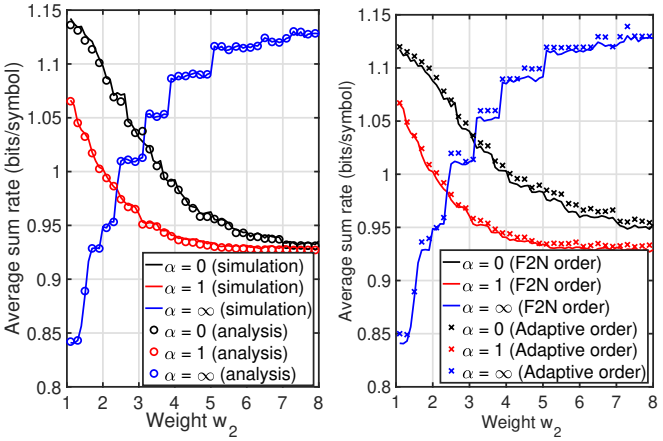


(a) Imperfect SIC vs. perfect SIC (b) Analysis vs. simulations

Fig. 7. Outage probability of OMCSP-A as a function of the peak SNR $\frac{P_{\text{tot}}\ell_1}{\sigma^2}$ with perfect and imperfect SIC and adaptive and F2N order ($K = 2$).

The average rate with imperfect SIC is lower than with perfect SIC for all users. This is because BLER constraints in (9) are tighter than ε due to imperfect SIC. This also increases the odds of an outage, in which case no power is assigned to the users. As a result, the average power assigned to the users with imperfect SIC is lower than with perfect SIC. One observation that is counter-intuitive at first sight is that with imperfect SIC lower powers are assigned to the users. This happens because the lower rates have a smaller decoding threshold, which requires less power for transmission.

Figure 7(a) plots the outage probability O of OMCSP-A with perfect and imperfect SIC as a function of the peak SNR $\frac{P_{\text{tot}}\ell_1}{\sigma^2}$. We show results for three values of $\frac{\ell_1}{\ell_2}$. As $\frac{\ell_1}{\ell_2}$ increases, which corresponds to an increase in the ratio of the distances of the two users from the BS, O increases. This is because the near user's SNR is kept fixed and the far user's SNR decreases. O decreases as the peak SNR increases because the power budget becomes sufficient to ensure that both users can meet their minimum rate constraints. For all



(a) F2N order: Analysis vs. simulations (b) F2N order vs. adaptive order

Fig. 8. Impact of utility function parameters: Average sum rate of OMCSP-F2N as a function of the weight w_2 for different α with imperfect SIC ($K = 2$, $\frac{P_{\text{tot}}\ell_1}{\sigma^2} = 10$ dB, and $\frac{\ell_1}{\ell_2} = 10$ dB).

$\frac{\ell_1}{\ell_2}$ and peak SNRs, the outage probability with imperfect SIC is greater, albeit marginally, than with perfect SIC. Figure 7(b) compares the outage probabilities of OMCSP-F2N from analysis and simulations. The analysis tracks the simulation results well at all SNRs. The outage probability of OMCSP-A is indistinguishable from that of OMCSP-F2N and is not shown to avoid clutter.

Figure 8(a) studies the impact of the utility function parameters. It plots the average sum rate of OMCSP-F2N as a function of w_2 for different values of α. The simulation and analytical results are in good agreement for all values of α and w_2. As w_2 increases, the average sum rate decreases for $\alpha = 0$ and 1 and reaches a floor. The decrease occurs because the far user is allocated higher rate MCSs, and a larger portion of the power must be assigned to it and not the near user 1 to support this. The floor occurs because most of the total power budget is allocated to the far user, leaving no room for further improvement. For the same reasons, the average sum rate decreases when α increases from 0 to 1. However, the trends reverse for max-min fairness ($\alpha = \infty$). As w_2 increases, the average sum rate of OMCSP-F2N increases. This happens because the minimum of r_{m_1} and $w_2 r_{m_2}$ becomes r_{m_1} for larger values of w_2. Thus, as w_2 increases, the algorithm maximizes the near user's rate, which increases the sum rate. We also note the pronounced increase in the average sum rate for $\alpha = \infty$ at $w_2 = 1.3$, 1.9, 2.3, 3.1, 3.7, and 4.9. This is due to the discontinuous switch from one MCS to another in discrete rate adaptation. Such discontinuities also occur for $\alpha = 0$ and 1, but they are not as visible. Figure 8(b) compares the average sum rates of OMCSP-A and OMCSP-F2N for different values of α. The two are close to each other. Thus, the F2N order is near-optimal for all values of α for $K = 2$.

Imperfect CSI: Figure 9 plots the average weighted α-utility of OMCSP-A as a function of the peak SNR for different values of $\rho \triangleq \rho_1 = \cdots = \rho_K$. As ρ decreases, i.e., the channel estimation accuracy decreases, the average weighted α-utility decreases. This is due to an increase in the power required due

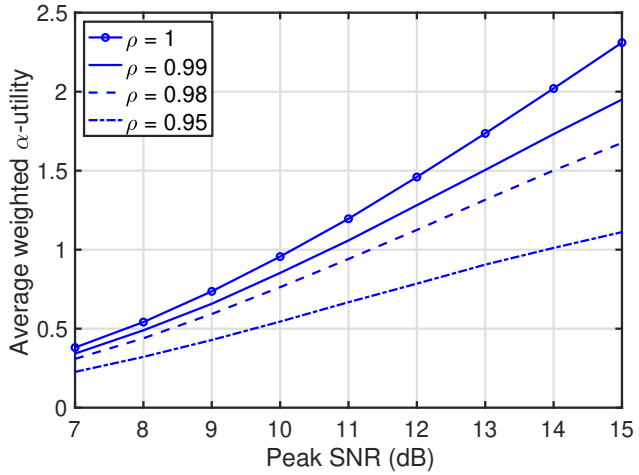


Fig. 9. Imperfect CSI: Average weighted α-utility of OMCSP-A as a function of the peak SNR $\frac{P_{\text{tot}}\ell_1}{\sigma^2}$ for different $\rho_1 = \cdots = \rho_K \triangleq \rho$ ($K = 2$, $\alpha = 0$, $w_2 = 4$, and $\frac{\ell_1}{\ell_2} = 10$ dB).

to the additional estimation error-related interference term in the SINR expression in (28). This also increases the outage probability, which, lowers the average weighted α-utility.

VI. Conclusions

We determined the optimal powers, MCSs, and decoding order that maximized a general weighted α-utility function of K-user downlink NOMA with discrete rate adaptation and imperfect SIC. Our formulation subsumed the sum rate, weighted sum rate maximization, PF, and max-min fairness approaches considered separately in the literature. Imperfect SIC with discrete rate adaptation for a given decoding order led to tighter BLER constraints for users with more decoding stages. To solve the resulting NP-hard optimization problem, we presented a novel algorithm that provably found the optimal MCSs and powers of the K users for a given decoding order. We then determined the optimal decoding order. Our approach had a significantly lower complexity than exhaustive search. The complexity was further reduced by employing a smaller set of BLER constraints. We saw that the F2N order was as good as adapting the decoding order for $K \leq 3$ users. Imperfect CSI led to a lower average weighted α-utility.

We also derived novel bounds for the average weighted α-utility, average sum rate, and the outage probability for two-user NOMA and the F2N order. The expressions substantially differed from for single-user OMA.

An interesting direction for future work is incorporating hardware impairments into the system model. Another avenue is determining the optimal utility parameter values to maximize the sum rate while meeting a specific fairness target. Our methodology can be combined with the design of the scheduler at the BS, which chooses the users to be transmitted to. It can be used in system-level simulation that evaluates the scheduler along with discrete rate adaptation. Our framework can also be extended to coordinated direct and relay transmission (CDRT) [50] and other NOMA-based cooperative communication models.

Appendix

A. Proof of Theorem 1

The SINR constraint of user $[1]$ in (13) becomes $P_{[1]}\Gamma_{[1]} \geq T_{m_{[1]}}(\varepsilon/K)$. In general, for user $[k]$, the BLER constraints become $P_{[k]}\Gamma_{[1]} \geq T_{m_{[k]}}(\varepsilon/K)[(P_{[1]}+P_{[2]}+\cdots+P_{[k-1]})\Gamma_{[1]}+1]$,$\cdots$, $P_{[k]}\Gamma_{[k]} \geq T_{m_{[k]}}(\varepsilon/(K-k+1))[(P_{[1]}+P_{[2]}+\cdots+P_{[k-1]})\Gamma_{[k]}+1]$. Thus, all the inequalities in $\mathcal{W}_1$ can be rewritten as linear functions of $P_{[1]}, P_{[2]}, \ldots, P_{[K]}$.

From $P_{[1]}\Gamma_{[1]} \geq T_{m_{[1]}}(\varepsilon/K)$, it follows that the smallest power $P^*_{[1]}$ required for user $[1]$ is $T_{m_{[1]}}(\varepsilon/K)/\Gamma_{[1]}$, which yields (16). In general, for user $[k]$, its power must satisfy

$$P_{[k]} \geq \max\left\{\frac{T_{m_{[k]}}(\frac{\varepsilon}{K})[(P_{[1]}+\cdots+P_{[k-1]})\Gamma_{[1]}+1]}{\Gamma_{[1]}}, \cdots, \frac{T_{m_{[k]}}(\frac{\varepsilon}{K-k+1})[(P_{[1]}+\cdots+P_{[k-1]})\Gamma_{[k]}+1]}{\Gamma_{[k]}}\right\}. \quad (62)$$

Hence, the smallest power $P^*_{[k]}$ required for user $[k]$ is

$$P^*_{[k]} = \max\left\{\frac{T_{m_{[k]}}(\frac{\varepsilon}{K})[(P^*_{[1]}+\cdots+P^*_{[k-1]})\Gamma_{[1]}+1]}{\Gamma_{[1]}}, \cdots, \frac{T_{m_{[k]}}\left(\frac{\varepsilon}{K-k+1}\right)[(P^*_{[1]}+\cdots+P^*_{[k-1]})\Gamma_{[k]}+1]}{\Gamma_{[k]}}\right\}. \quad (63)$$

The above powers must satisfy the sum power constraint in (14). Thus, a feasible power allocation exists if and only if $\sum_{i=1}^{K} P^*_{[i]} \leq P_{\text{tot}}$.

B. Proof of Corollary 1

If an MCS choice $\mathbf{m} = (m_1, m_2, \ldots, m_K)$ is infeasible for a decoding order $([1], [2], \ldots, [K])$, then it follows from Theorem 1 that $\sum_{j=1}^{K} P^*_{[j]} > P_{\text{tot}}$. Consider the MCS choice $\mathbf{m}' = (m_1, \ldots, m_{k-1}, m_k + i_k, m_{k+1}, \ldots, m_K)$, where $i_k > 0$ for some k. Let $k = [q]$ for the decoding order under consideration. Let $\tilde{P}^*_j$, $\forall j \in \{1, \ldots, K\}$, denote the MF power allocation for $\mathbf{m}'$. Applying the monotonicity assumption of decoding thresholds to the inequalities in Theorem 1, we get $\tilde{P}^*_{[l]} = P^*_{[l]}$, for $1 \leq l < q$, and $\tilde{P}^*_{[l]} \geq P^*_{[l]}$, for $q \leq l \leq K$. Therefore,

$$\sum_{j=1}^{K} \tilde{P}^*_{[j]} \geq \sum_{j=1}^{K} P^*_{[j]} > P_{\text{tot}}. \quad (64)$$

Hence, $\mathbf{m}'$ is also infeasible. Applying this argument successively to different k completes the proof.

C. Proof of Theorem 2

The key observation is that Algorithm 1 moves an MCS choice to $\mathcal{I}$ if and only if the choice is infeasible. The infeasibility is determined using Theorem 1 or Corollary 1. Therefore, all the remaining MCS choices must be feasible and will get included in $\mathcal{S}$ when the search terminates. Their weighted α-utility functions get included in $\mathcal{U}$. Therefore, the algorithm evaluates the weighted α-utility of every feasible MCS choice and identifies the optimal one among them.

D. Single, Two, and Three Event Probability Terms

a) Single Event Term: From (31), it follows that

$$\Pr(E_0) = \Pr\left(\Gamma_1 \geq \frac{T_{m_1}(T_{m_2}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2}}, \Gamma_2 \geq \frac{T_{m_2}}{P_{\text{tot}}}\right). \quad (65)$$

Conditioning on Γ_2, we get

$$\Pr(E_0) = \mathbb{E}\left[\Pr\left(\Gamma_1 \geq \frac{T_{m_1}(T_{m_2}+1)x}{xP_{\text{tot}} - T_{m_2}} \,\middle|\, \Gamma_2 = x\right) \times \mathbb{1}_{\left\{\Gamma_2 \geq \frac{T_{m_2}}{P_{\text{tot}}}\right\}}\right]. \quad (66)$$

Writing the above expectation in terms of the CDF of Γ_1 and the PDF of Γ_2, we get

$$\Pr(E_0) = \int_{\frac{T_{m_2}}{P_{\text{tot}}}}^{\infty} \left[1 - F_{\Gamma_1}\left(\frac{T_{m_1}(T_{m_2}+1)x}{xP_{\text{tot}} - T_{m_2}}\right)\right] f_{\Gamma_2}(x)\, dx. \quad (67)$$

Since Γ_1 and Γ_2 are exponential RVs with means ℓ_1/σ^2 and ℓ_2/σ^2, respectively, we have

$$F_{\Gamma_1}(x) = 1 - \exp\left(-\frac{\sigma^2 x}{\ell_1}\right), \text{ for } x \geq 0, \quad (68)$$

$$f_{\Gamma_2}(x) = \frac{\sigma^2}{\ell_2}\exp\left(-\frac{\sigma^2 x}{\ell_2}\right), \text{ for } x \geq 0. \quad (69)$$

Substituting (68) and (69) in (67), we get

$$\Pr(E_0) = \frac{\sigma^2}{\ell_2}\int_{\frac{T_{m_2}}{P_{\text{tot}}}}^{\infty} \exp\left(-\frac{\sigma^2 x}{\ell_2} - \frac{\sigma^2}{\ell_1}\frac{T_{m_1}(T_{m_2}+1)x}{xP_{\text{tot}} - T_{m_2}}\right) dx, \quad (70)$$

$$= \Pi(\mathbf{m}), \quad (71)$$

where $\Pi(.)$ is defined in (38).

b) Two Event Terms: We first evaluate $\Pr(E_0 \cap E_u)$. From (31) and (32), it follows that

$$\Pr(E_0 \cap E_u) = \Pr\left(\Gamma_1 \geq \max\left\{\frac{T_{m_1}(T_{m_2}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2}}, \frac{T_{m_1-u}(T_{m_2+1}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2+1}}\right\}, \Gamma_2 \geq \frac{T_{m_2+1}}{P_{\text{tot}}}\right). \quad (72)$$

Conditioning on Γ_2, we get

$$\Pr(E_0 \cap E_u) = \mathbb{E}\left[\Pr\left(\Gamma_1 \geq \max\left\{\frac{T_{m_1}(T_{m_2}+1)x}{xP_{\text{tot}} - T_{m_2}}, \frac{T_{m_1-u}(T_{m_2+1}+1)x}{xP_{\text{tot}} - T_{m_2+1}}\right\} \,\middle|\, \Gamma_2 = x\right) \mathbb{1}_{\left\{x \geq \frac{T_{m_2+1}}{P_{\text{tot}}}\right\}}\right]. \quad (73)$$

In terms of the CDF of Γ_1 and the PDF of Γ_2, we get

$$\Pr(E_0 \cap E_u) = \int_{\frac{T_{m_2+1}}{P_{\text{tot}}}}^{\infty} \left[1 - F_{\Gamma_1}\left(\max\left\{\frac{T_{m_1}(T_{m_2}+1)x}{xP_{\text{tot}} - T_{m_2}}, \frac{T_{m_1-u}(T_{m_2+1}+1)x}{xP_{\text{tot}} - T_{m_2+1}}\right\}\right)\right] f_{\Gamma_2}(x)\, dx, \quad (74)$$

$$= \Pi(\mathbf{m}, (m_1 - u, m_2 + 1)). \quad (75)$$

In a similar manner, we can derive the expressions for $\Pr(E_0 \cap E_v)$, $\Pr(E_0 \cap E_w)$, and $\Pr(E_0 \cap E_x)$ in (42), (43), and (44), respectively.

c) Three Event Terms: We first evaluate $\Pr(E_0 \cap E_u \cap E_v)$. From (31), (32), and (33), we get

$$\Pr(E_0 \cap E_u \cap E_v) = \Pr\left(\Gamma_1 > \frac{T_{m_1}(T_{m_2}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2}}, \Gamma_2 \geq \frac{T_{m_2}}{P_{\text{tot}}}, \Gamma_1 > \frac{T_{m_1-u}(T_{m_2+1}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2+1}}, \Gamma_2 \geq \frac{T_{m_2+1}}{P_{\text{tot}}}, \Gamma_1 > \frac{T_{m_1+v}(T_{m_2-1}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2-1}}, \Gamma_2 \geq \frac{T_{m_2-1}}{P_{\text{tot}}}\right). \quad (76)$$

Since $T_{m_2-1} < T_{m_2} < T_{m_2+1}$, (76) simplifies to

$$\Pr(E_0 \cap E_u \cap E_v) = \Pr\left(\Gamma_1 \geq \max\left\{\frac{T_{m_1}(T_{m_2}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2}}, \frac{T_{m_1-u}(T_{m_2+1}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2+1}}, \frac{T_{m_1+v}(T_{m_2-1}+1)\Gamma_2}{\Gamma_2 P_{\text{tot}} - T_{m_2-1}}\right\}, \Gamma_2 \geq \frac{T_{m_2+1}}{P_{\text{tot}}}\right). \quad (77)$$

Conditioning on Γ_2, we get

$$\Pr(E_0 \cap E_u \cap E_v) = \mathbb{E}\left[\Pr\left(\Gamma_1 \geq \max\left\{\frac{T_{m_1}(T_{m_2}+1)x}{xP_{\text{tot}} - T_{m_2}}, \frac{T_{m_1-u}(T_{m_2+1}+1)x}{xP_{\text{tot}} - T_{m_2+1}}, \frac{T_{m_1+v}(T_{m_2-1}+1)x}{xP_{\text{tot}} - T_{m_2-1}}\right\} \,\middle|\, \Gamma_2 = x\right) \mathbb{1}_{\left\{\Gamma_2 \geq \frac{T_{m_2+1}}{P_{\text{tot}}}\right\}}\right]. \quad (78)$$

Writing the above expectation in terms of the CDF of Γ_1 and PDF of Γ_2, and simplifying along lines similar to (67), (70), and (71), we get

$$\Pr(E_0 \cap E_u \cap E_v) = \Pi(\mathbf{m}, (m_1 - u, m_2 + 1), (m_1 + v, m_2 - 1)). \quad (79)$$

Along similar lines, we can derive the expressions for $\Pr(E_0 \cap E_u \cap E_w)$, $\Pr(E_0 \cap E_u \cap E_x)$, $\Pr(E_0 \cap E_v \cap E_w)$, $\Pr(E_0 \cap E_v \cap E_x)$, and $\Pr(E_0 \cap E_w \cap E_x)$ in (46), (47), (48), (49), and (50), respectively.

REFERENCES

[1] R. Srikant and L. Ying, *Communication Networks: An Optimization, Control and Stochastic Networks Perspective*. Cambridge Univ. Press, 2014.

[2] 3GPP, "NR; physical layer procedures for data," 3rd Generation Partnership Project (3GPP), TS 38.214 (v15.2.0), Dec. 2018.

[3] K. Z. Shen, D. K. C. So, J. Tang, and Z. Ding, "Power allocation for NOMA with cache-aided D2D communication," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 529–542, Jan. 2024.

[4] J. Li, X. B. Zhai, H. Qian, R. Zhang, and X. Liu, "Joint trajectory design and power allocation in NOMA-based UAV networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2345–2357, Feb. 2024.

[5] X. Li, C. Li, and Y. Jin, "Dynamic resource allocation for transmit power minimization in OFDM-based NOMA systems," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2558–2561, Dec. 2016.

[6] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.

[7] Z. Xiao, L. Zhu, Z. Gao, D. O. Wu, and X.-G. Xia, "User fairness non-orthogonal multiple access (NOMA) for millimeter-wave communications with analog beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3411–3423, Jul. 2019.

[8] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[9] H. Zheng, H. Li, S. Hou, and Z. Song, "Joint resource allocation with weighted max-min fairness for NOMA-enabled V2X communications," *IEEE Access*, vol. 6, pp. 65 449–65 462, Oct. 2018.

[10] B. Van Nguyen, Q.-D. Vu, and K. Kim, "Analysis and optimization for weighted sum rate in energy harvesting cooperative NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12 379–12 383, Dec. 2018.

[11] J. Wang, Q. Peng, Y. Huang, H.-M. Wang, and X. You, "Convexity of weighted sum rate maximization in NOMA systems," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1323–1327, Sept. 2017.

[12] X. Wang, R. Chen, Y. Xu, and Q. Meng, "Low-complexity power allocation in NOMA systems with imperfect SIC for maximizing weighted sum-rate," *IEEE Access*, vol. 7, pp. 94 238–94 253, Jul. 2019.

[13] L. Salaün, M. Coupechoux, and C. S. Chen, "Joint subcarrier and power allocation in NOMA: Optimal and approximate algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 2215–2230, Mar. 2020.

[14] 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.

[15] 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, Sept. 2021.

[16] ——, "Design and analysis of NOMA with adaptive modulation and power under BLER constraints," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11 228–11 233, Oct. 2022.

[17] I. Amin, D. Mishra, R. Saini, and S. Aïssa, "Power allocation and decoding order selection for secrecy fairness in downlink cooperative NOMA with untrusted receivers under imperfect SIC," *IEEE Trans. on Inf. Forensics and Secur.*, vol. 19, pp. 9406–9418, Sep. 2024.

[18] Y. Gao, B. Xia, K. Xiao, Z. Chen, X. Li, and S. Zhang, "Theoretical analysis of the dynamic decode ordering SIC receiver for uplink NOMA systems," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2246–2249, Oct. 2017.

[19] Y. Gao, B. Xia, Y. Liu, Y. Yao, K. Xiao, and G. Lu, "Analysis of the dynamic ordered decoding for uplink NOMA systems with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6647–6651, Jul. 2018.

[20] A. Zakeri, M. Moltafet, and N. Mokari, "Joint radio resource allocation and SIC ordering in NOMA-based networks using submodularity and matching theory," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9761–9773, Oct. 2019.

[21] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Commun. Mag.*, vol. 1, pp. 179–189, Jan. 2020.

[22] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2192–2235, 4th Qtr. 2020.

[23] L. Chen, L. Ma, and Y. Xu, "Proportional fairness-based user pairing and power allocation algorithm for non-orthogonal multiple access system," *IEEE Access*, vol. 7, pp. 19 602–19 615, Jan. 2019.

[24] M.-R. Hojeij, C. Abdel Nour, J. Farah, and C. Douillard, "Waterfilling-based proportional fairness scheduler for downlink non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 230–233, Feb. 2017.

[25] Y. Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, "Fairness of user clustering in MIMO non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1465–1468, Jul. 2016.

[26] L. Xu, H. Xing, Y. Deng, A. Nallanathan, and C. Zhuansun, "Fairness-aware throughput maximization for underlaying cognitive NOMA networks," *IEEE Syst. J.*, vol. 15, no. 2, pp. 1881–1892, Feb. 2021.

[27] P. Xu and K. Cumanan, "Optimal power allocation scheme for non-orthogonal multiple access with α-fairness," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2357–2369, Oct. 2017.

[28] S. Fang, H. Chen, Z. Khan, and P. Fan, "User fairness aware power allocation for NOMA-assisted video transmission with adaptive quality adjustment," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 1054–1059, Jan. 2022.

[29] M.-J. Youssef, J. Farah, C. A. Nour, and C. Douillard, "Resource allocation in NOMA systems for centralized and distributed antennas with mixed traffic using matching theory," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 414–428, Jan. 2020.

[30] P. Xu, K. Cumanan, and Z. Yang, "Optimal power allocation scheme for NOMA with adaptive rates and α-fairness," in *Proc. IEEE Globecom*, Dec. 2017, pp. 1–6.

[31] J. Zeng, C. Xiao, T. Wu, W. Ni, R. P. Liu, and Y. J. Guo, "Uplink non-orthogonal multiple access with statistical delay requirement: Effective capacity, power allocation, and α fairness," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1298–1313, Feb. 2023.

[32] A. Pastore and M. Navarro, "A fairness–throughput tradeoff perspective on NOMA multiresolution broadcasting," *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 179–186, Jan. 2019.

[33] 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.

[34] 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.

[35] M. Chitra and S. Dhanasekaran, "Performance analysis of NOMA-based spectrum sharing system with imperfect SIC," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8198–8203, Jun. 2023.

[36] L. Luo, Q. Li, and J. Cheng, "Performance analysis of overlay cognitive NOMA systems with imperfect successive interference cancellation," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4709–4722, Aug. 2020.

[37] M. Zeng, W. Hao, O. A. Dobre, Z. Ding, and H. V. Poor, "Power minimization for multi-cell uplink NOMA with imperfect SIC," *IEEE Wireless Commun. Lett.*, vol. 9, no. 12, pp. 2030–2034, Dec. 2020.

[38] B. Lim, W. J. Yun, J. Kim, and Y.-C. Ko, "Joint user clustering, beamforming, and power allocation for mmwave-NOMA with imperfect SIC," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2025–2038, Mar. 2024.

[39] M. Taki and F. Lahouti, "Discrete rate interfering cognitive link adaptation design with primary link spectral efficiency provisioning," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 2929–2939, Sep. 2011.

[40] M. M. Abdallah, A. H. Salem, M.-S. Alouini, and K. A. Qaraqe, "Adaptive discrete rate and power transmission for spectrum sharing systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1283–1289, Apr. 2012.

[41] D.-Y. Kim, H. Jafarkhani, and J.-W. Lee, "Low-complexity dynamic resource scheduling for downlink MC-NOMA over fading channels," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3536–3550, May 2022.

[42] I. Randrianantenaina, M. Kaneko, H. Dahrouj, H. ElSawy, and M.-S. Alouini, "Interference management in NOMA-based fog-radio access networks via scheduling and power allocation," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5056–5071, Aug. 2020.

[43] S. Sruthy and N. B. Mehta, "Power and discrete rate adaptation in wideband NOMA in frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4186–4198, May 2024.

[44] J. Luo, J. Tang, D. K. C. So, G. Chen, K. Cumanan, and J. A. Chambers, "A deep learning-based approach to power minimization in multi-carrier NOMA with SWIPT," *IEEE Access*, vol. 7, pp. 17 450–17 460, Jan. 2019.

[45] G. S. Kesava and N. B. Mehta, "Multi-connectivity for URLLC and coexistence with eMBB in time-varying and frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3599–3611, Jun. 2023.

[46] A. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.

[47] M. Jünger, T. M. Liebling, D. Naddef, G. L. Nemhauser, W. R. Pulleyblank, G. Reinelt, G. Rinaldi, and L. A. Wolsey, Eds., *50 Years of Integer Programming 1958-2008 - From the Early Years to the State-of-the-Art*. Springer, 2010.

[48] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, NY, USA: Dover, 1964.

[49] A. Goldsmith, *Wireless Communications*. Cambridge Univ. Press, 2005.

[50] J.-B. Kim and I.-H. Lee, "Non-orthogonal multiple access in coordinated direct and relay transmission," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 2037–2040, Nov. 2015.