Distributed Rate Control in Downlink NOMA Networks with Reliability Constraints

Onel L. A. López, *Student Member, IEEE*, Hirley Alves, *Member, IEEE*, Matti Latva-aho, *Senior Member, IEEE*

Abstract—Non-orthogonal multiple access (NOMA) has been identified as a promising technology for future wireless systems due to its performance gains in spectral efficiency when compared to conventional orthogonal schemes (OMA). This gain can be easily translated to an increasing number of served users, but imposes a challenge in the system reliability which is of vital importance for new services and applications of coming cellular systems. To cope with these issues we propose a NOMA rate control strategy that makes use only of topological characteristics of the scenario and the reliability constraint. We attain the necessary conditions so that NOMA overcomes the OMA alternative, while we discuss the optimum allocation strategies for the 2-user NOMA setup when operating with equal rate or maximum sum-rate goals. In such scenario we show that the user with the largest target error probability times the ratio between the average receive signal power and the average interference power, should be scheduled to be decoded first for optimum performance. We compare numerically the performance of our allocation scheme with its ideal counterpart requiring full CSI at the BSs and infinitely long blocklength, and show how the gap increases as the reliability constraint becomes more stringent. Results also evidence the benefits of NOMA when the co-interference can be efficiently canceled, specially when the goal is to maximize the sum-rate.

Index Terms—NOMA, rate control, power allocation, reliability constraints, optimal user ordering.

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) has been widely recognized as a promising technology for future wireless systems due to its superior spectral efficiency compared to conventional orthogonal multiple access (OMA) [1], [2]. Compared to OMA, NOMA can exploit the channel diversity more efficiently via smart interference management techniques such as successive interference cancellation (SIC) [3], thus, achieving the performance enhancements. A unified framework for NOMA is provided in [4] where the authors review the principles of various NOMA schemes in different domains. As one of the dominant NOMA schemes, power-domain NOMA, where different users are allocated different power levels according to their channel conditions, has become a strong candidate. These setups are based on the combination of superposition coding at the transmitter side, by allocating the same frequency/time/spatial resource to multiple receivers and multiplexing them on the power domain, and then extracting the intended signals from the composite data using SIC at these receivers [1]. Power domain NOMA has received extensive research interests and readers can refer to [5] for a comprehensive survey on its recent progress towards 5G systems and beyond. Particularly, the work in [6] studies the tradeoff between data rate performance and energy consumption by examining the problem of energy-efficient user scheduling and power optimization in 5G NOMA heterogeneous networks. The scenario of NOMA-based cooperative relay network is explored in [7], while coordinated multipoint transmissions in downlink NOMA cellular systems are analyzed in [8].

A. Related Works

Many of the existing studies focus on the system performance when considering a single NOMA cell/cluster [9]–[15]. The performance of NOMA in a cellular downlink scenario with randomly deployed users is investigated in [9]. Authors show that NOMA can achieve superior performance in terms of ergodic sum rates; however, that depends critically on the choices of the users' targeted data rates and allocated power. In [10], the impact of user pairing on the performance of two NOMA systems, NOMA with fixed power allocation and cognitive-radio-inspired NOMA, is characterized. An uplink power control scheme is proposed in [11] in order to achieve diverse arrived power in an uplink NOMA system, while authors analyze its outage performance and the achievable sum data rate. Interestingly, authors in [12] investigate the theoretical performance comparison between NOMA and conventional OMA systems from an optimization point of view, while in the context of machine type communications (MTC) authors in [13] propose a game theory power control algorithm that prioritizes first the communication reliability and once the reliability is satisfied it focuses on the power consumption issues. Also, energy-efficient NOMA design problem for two downlink receivers that have strict reliability and finite blocklength (latency) constraints have been considered in [14], while in [15] the physical-layer transmission latency reduction enabled by NOMA in short-packet communications has been highlighted. Finally, the combination of NOMA with the multiple-input multiple-output (MIMO) technology is explored in [16] along with the limitations and future research directions in the area.

Above works do not deal with inter-cell interference, which is a pervasive problem in most of the existing wireless networks and could significantly limit the performance of NOMA deployments. In that sense some recent research works have considered the performance characterization of large-scale NOMA systems [17]–[23]. Using stochastic geometry,

Authors are with the Centre for Wireless Communications (CWC), University of Oulu, Finland. {onel.alcarazlopez,hirley.alves,matti.latva-aho}@oulu.fi

the performance of uplink NOMA in terms of the rate coverage and average achievable rate is characterized in [17] using Poisson cluster process and considering both perfect and imperfect SIC. Additionally, authors in [18] proposes and evaluate NOMA as a massive-MTC enabler technology in data aggregation networks, while identifying the power constraints on the devices sharing the same channel in order to attain a fair coexistence with purely OMA setups. As a continuation, the relaying phase, where the aggregatted data is forwarded to the base station, is considered in [19], and the system performance is investigated in terms of average number of users that are simultaneously served. In the context of vehicle-to-everything (V2X) communications, NOMA has been explored in [20] where the Base Station (BS) performs semi-persistent scheduling and allocates time-frequency resources in a non-orthogonal manner, while the vehicles autonomously perform distributed power control when broadcasting their safety information to the neighborhood. System coverage and average achievable rate of the m-th rank user is derived in [21] assuming that the BSs locations follow a homogeneous Poisson Point Process (PPP) in downlink NOMA, and authors show that NOMA can bring considerable performance gain compared to OMA when SIC error is low. Also, the performance of two-user downlink NOMA is investigated [22] in a multi-tier cellular network where the macro cell BSs use the massive MIMO technology and each small cell adopts user pairing to implement two-user NOMA transmission. Finally, authors in [23] developed an analytical framework to derive the meta distribution of the Signal-to-Interference Ratio (SIR) in large-scale co-channel uplink and downlink NOMA networks with one NOMA cluster per cell. Notice that such meta distribution tool provides a more precise characterization of the performance of a typical transmission link than the standard success probability, since it is able of characterizing the fraction of users that perform with a given reliability, whereas the standard success probability just characterizes the fraction of users that are in coverage.

On the other hand, NOMA works have usually been aimed at improving the spectral [9]–[12] and energy [14], [22] efficiency of the system, the coverage probability [17], [21], [22], and also recently have focused in providing massive connectivity [18], [19] for future Internet of Things (IoT) scenarios. Besides the massive access problem, many IoT use cases with stringent delay and reliability constraints, e.g., ultra reliable MTC (uMTC), have been identified for the coming years and constitute important challenges for emerging massive wireless networks. In that regard, the works [13]–[15] analyze reliability and delay metrics but in a single cell setup, while authors in [20] do evaluate a multi-cell setup but in a V2X network. By using stochastic geometry tools, authors in [23] maximize the success probability with and without latency constraints in uplink and downlink NOMA setups with perfect SIC. Additionally, optimizing the performance of these networks has usually relied on power control, sub-channel allocation and user selection and/or user pairing mechanisms, while the transmission rate is another degree of freedom that could be also exploited. In fact, the latter has been already explored efficiently in ad-hoc [24] and OMA [25], [26] networks with reliability constraints, where the proposed rate allocation schemes are based on easy-to-obtain topological characteristics and on the (Rayleigh) fading statistics[1]. Notice that rate allocation mechanisms are promising since their impact is merely local, in the sense that a node varying its transmission rate does not influence in the performance of any other node[2], thus, when properly designed they are suitable for distributed implementations.

B. Contributions and Organization

To the best of our knowledge, a multi-cell downlink NOMA scenario where User Equipments (UEs) operate with reliability constraints and imperfect SIC has not been explored yet, thus, in this work we aim at filling that gap. Our main contributions are

- We propose a NOMA rate allocation scheme to be performed by BSs in order to meet the reliability constraints of their associated UEs. The advantage of this scheme is that no instantaneous Channel State Information (CSI) is required at the transmission side, while it only relies on easy-to-get information such as the average receive power from the desired signal and from the interfering BSs at the target UE, and the reliability constraint. Notice that this scheme is closely related to the one we first proposed in [25] and extended in [26], but there we just analyze an OMA network with multiple antennas at the receiver, while operating in a non-orthogonal fashion raises many other practical concerns;
- We characterize analytically the distribution of the allocated SIR threshold, thus, the distribution of the allocated rate, in Poisson Networks;
- We attain the necessary conditions so that NOMA overcomes the OMA alternative, while we discuss the optimum allocation strategies for the 2-UEs NOMA setup when operating with equal rate or maximum sum-rate goals. Specifically, we find the optimum rate and power allocation profile and show that the UE with the largest value of target error probability times the ratio between the average receive power from the desired signal and the average interference power, should be scheduled to be decoded first for optimum performance;
- We compare numerically the performance of our allocation scheme with its ideal counterpart requiring full CSI at the BSs and infinitely long blocklength, and show how the gap increases as the reliability constraint becomes more stringent. Results also evince the benefits of NOMA when the co-interference can be efficiently canceled, specially when the goal is to maximize the sum-rate.

Next, Section II establishes the system model and assumptions. In Section III we present the rate allocation scheme, and characterize the distribution of the adopted SIR threshold in large-scale networks. Section IV discusses the optimum power allocation strategy in the 2-UEs NOMA setup, while

---

[1]Specifically, [24] focuses on an interference-limited Poisson bipolar network, while in [25], [26] we did similar but for cellular networks with multiple antennas at the receiver side.

[2]In case of NOMA, it would affect the nodes associated to the same NOMA cluster, but this is still local and easy to handle.
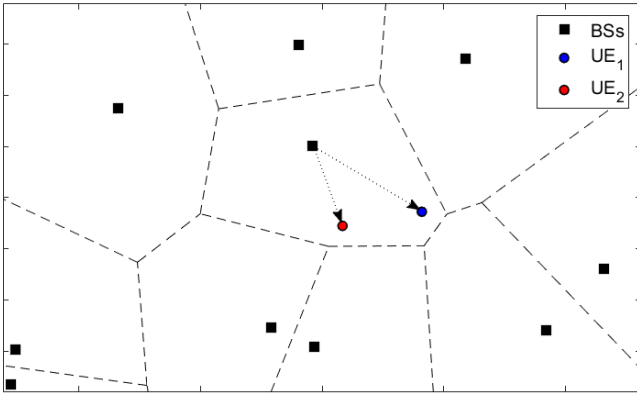
Fig. 1. Illustration of the system model for $M = 2$.

Section VI presents the numerical results. Finally, Section VII concludes the paper.

Notation: $\mathbb{E}[\,\cdot\,]$ and $\mathbb{E}[\,\cdot\,|A]$ denote expectation and expectation conditioned on event A, respectively, while $\Pr(B)$ and $\Pr(B|A)$ are the probability of event B, and $\Pr(B)$ conditioned on A, respectively. $f_X(x)$ and $F_X(x)$ are the Probability Density Function (PDF) and Cumulative Distribution Function (CDF) of random variable (RV) X, respectively. $X \sim \mathrm{Exp}(1)$ is an exponential distributed RV with unit mean, e.g., $f_X(x) = \exp(-x)$ and $F_X(x) = 1 - \exp(-x)$; while $f_{r_i}(x) = 2\pi\lambda x \exp(-\pi\lambda x^2)$ denotes the PDF of the nearest neighbor distance in a two-dimensional PPP with density λ, which is given in [27]. $d^m f(x)/dx^m$ is the m-th derivative of function $f(x)$ with respect to x, while $\min(x, y)$ is the minimum between the values of x and y. Finally, $\mathcal{L}_X(s)$ denotes the Laplace transform of RV X, while $\mathcal{L}^{-1}\{\cdot\}(x)$ is the inverse Laplace transform operation.

II. System Model

Consider a multi-cell downlink cellular network where BSs are spatially distributed according to a 2-D homogeneous PPP Φ with density λ. We assume each UE is associated with the closest BS, namely the UEs in the Voronoi cell of a BS are associated with it, resulting in coverage areas as shown in Fig. 1. UEs are placed such that the distance between UE_i and its associated BS, is r_i. We also consider an interference-limited wireless system given a dense deployment of small cells and hence the impact of noise is neglected throughout the paper[3]. All BSs transmit at the power P_T. We adopt a channel model that comprises standard path-loss with exponent α and Rayleigh fading.

A NOMA group, $\mathcal{M}$, consists of M UEs, UE_i, $i = 1, 2, ..., M$, and notice that an illustration of the system model for $M = 2$ is shown in Fig. 1. Each UE_i tries to decode the interfering NOMA signals destined to UE_m, $m \leq i$, and then remove the interfering NOMA signals from the received signal, in a successive manner. Assume the transmission power of signal for UE_i is P_i with $\sum_{i \in \mathcal{M}} P_i = P_T$. Our system model is similar to [21], although, we do not establish the

decoding order based on the channel gains normalized by the interference since such approach requires full knowledge not only of the channel coefficients and path-loss but also of the interference experienced each time at each UE, which is very difficult to obtain. Instead, the BS allocates the transmit power such that $P_1 \geq ... \geq P_M$ for guaranteeing the desired decoding order[4] and some criteria for feasibility and optimum performance are provided in Section IV. Finally, CSI is only required at the UEs and not at transmitter side.

A. Signal Model

We denote the signal intended to UE_i as x_i where $\mathbb{E}[|x_i|^2] = 1$. According to NOMA principle [2], the transmitted signal at BS is coded as the composite signal from all the UEs sharing the orthogonal resource,

$$x = \sum_{i=1}^{M} \sqrt{P_i} x_i. \tag{1}$$

Thus, in an interference limited system the received signal at UE_i can be represented as

$$y_i = \sqrt{h_i r_i^{-\alpha}} x + I_i, \tag{2}$$

where $h_i \sim \mathrm{Exp}(1)$ is the power channel gain coefficient, and

$$I_i = \sum_{j \in \Phi \setminus \{b_0\}} g_j r_{j,i}^{-\alpha} P_T \tag{3}$$

is the inter-cell interference, which is the sum of the power received from all BSs excluding the serving BS denoted as b_0. In (3), $g_j \sim \mathrm{Exp}(1)$ and $r_{j,i}$ are the Rayleigh fading power coefficient of interfering channel, and transmission distance from interfering BS j to UE_i, respectively.

B. SIR after SIC

Instead of the real SIR at the receiving antenna of a UE, we are more interested in the SIR after SIC which is directly related to the performance of each particular UE. Notice that UE_1 does not need to perform interference cancellation and directly treats $x_j, j \geq 2$, as interference since it comes the first in the decoding order. On the other hand, UE_2 first decodes x_1 and removes it from the received composite signal y_2, based on which UE_2 can further decode x_2. Thus, UE_m first decodes signals $x_j, j < m$, and remove them from the received composite signal y_m, based on which UE_m can further decode x_m. Assuming successful decoding but with error propagation [28] given by the parameter $\mu \in [0, 1]$ measuring the imperfection of SIC, the NOMA SIR of the $i-$th signal at UE_i can be expressed as

$$\mathrm{SIR}_i = \frac{h_i r_i^{-\alpha} P_i}{h_i r_i^{-\alpha}\left[\mu \sum_{j=1}^{i-1} P_j + \sum_{j=i+1}^{M} P_j\right] + I_i}. \tag{4}$$

Notice that $\sum_{j=1}^{i-1} P_j = 0$ and $\sum_{j=i+1}^{M} P_j = 0$ for $i = 1$ and $i = M$, respectively.

---

[3] However, the impact of the noise could easily be incorporated without substantial changes.

[4] For instance, if two UEs, named UE_A and UE_B, are served such that the BS allocates $P_B > P_A$, then UE_B and UE_A signals are decoded in that order at the UEs since for each of them the signal associated with UE_B arrives always with greater power. Thus, $\mathrm{UE}_B = \mathrm{UE}_1$ and $\mathrm{UE}_A = \mathrm{UE}_2$.

III. Distributed rate control under reliability constraint

We are interested in finding the SIR threshold for each link $i \in \mathcal{M}$, γ_i, such that the conditional link success probability, $\mathbb{P}(\mathrm{SIR}_i > \gamma_i | \Phi)$ [24], is equal to $1 - \epsilon_i$, then, the transmitter decides on its rate as $R_i = \log_2(1 + \gamma_i)$. Allocating the rate in such way guarantees that all the UEs meet their reliability requirements since $\mathbb{P}(\mathrm{SIR}_i > \gamma_i | \Phi)$ is defined for each network realization. Notice that for practical feasibility the allocation should rely on information easy to obtain at the BS; for instance, allocating the rate based on the path-loss experienced by each of the interfering signals is not suitable.

Lemma 1. *For a given realization of Φ, the maximum SIR threshold γ_i that guarantees the required reliability for the i−th link is*

$$\gamma_i = \frac{\varphi_i^* P_i}{P_T + \varphi_i^* \left(\mu \sum_{j=1}^{i-1} P_j + \sum_{j=i+1}^{M} P_j \right)}, \quad (5)$$

where φ_i^* is the unique real positive solution of

$$\prod_{j \in \Phi \setminus \{b_0\}} (1 + \varphi_i r_i^\alpha r_{j,i}^{-\alpha}) = \frac{1}{1 - \epsilon_i}. \quad (6)$$

Proof. See Appendix A.

Remark 1. *If $M = 1$ we have that $\gamma_1 = \varphi_1^*$, thus, φ_i^* represents the SIR threshold that would be required for UE_i if it was operating alone in the channel, hence $\varphi_i^* \geq \gamma_i$ holds always. Additionally,*

$$\frac{d\gamma_i}{d\varphi_i^*} = \frac{P_i P_T}{\left(P_T + \varphi_i^* \left(\mu \sum_{j=1}^{i-1} P_j + \sum_{j=i+1}^{M} P_j \right) \right)^2} > 0, \quad (7)$$

therefore, γ_i is an increasing function of φ_i^*.

For practical values of ϵ_i we next propose a very tight and simple approximation for φ_i^* that is suitable when computing (5) for rate allocation.

Theorem 1. *φ_i^* approximates accurately to*

$$\varphi_i^* \approx \frac{r_i^{-\alpha}}{\sum_{j \in \Phi \setminus \{b_0\}} r_{j,i}^{-\alpha}} \epsilon_i \quad (8)$$

when $\epsilon_i \leq 10^{-1}$.

Proof. See Appendix B.

Notice that our results are not restrictive to the adopted path-loss model. In fact, $r_i^{-\alpha} / \sum_{j \in \Phi \setminus \{b_0\}} r_{j,i}^{-\alpha}$ can be expressed in a generalized way as the quotient between the average receive power from the desired signal and from the interfering BSs at the given location, and in that aspect lies the main advantage of using (8). Fig. 2a corroborates the accuracy of (8), which is independent of the specific deployment and path-loss exponent. As expected, the less strict the reliability constraint, the larger φ_i^*, hence the larger γ_i.
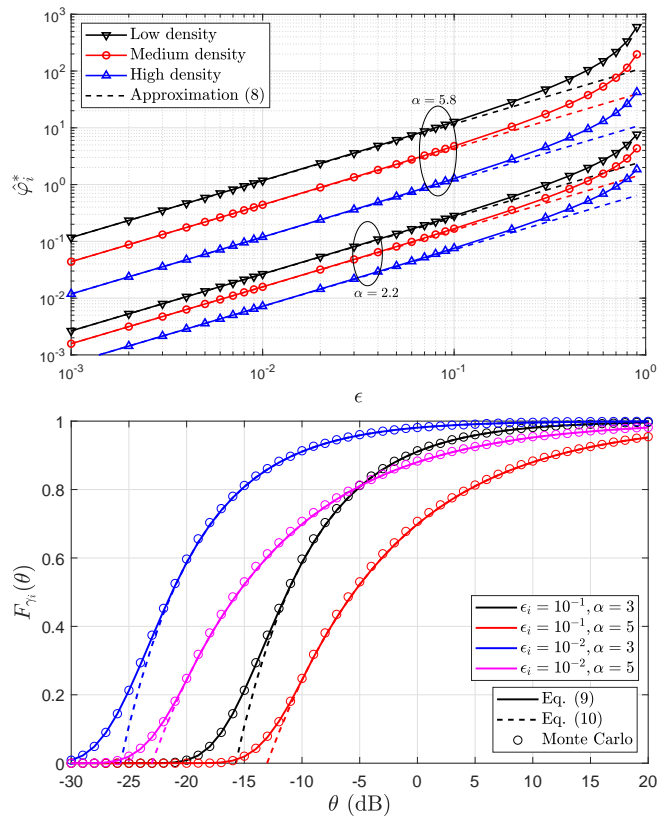

Fig. 2. (a) φ_i^* as a function of ϵ_i, for high density, $r_{j,i} = 40 + 10j$, medium density, $r_{j,i} = 40 + 20j$ and low density, $r_{j,i} = 40 + 30j$, example deployments (top). We set $r_i = 30$m. (b) $F_{\gamma_i}(\theta)$ for OMA scheme ($P_i = P_T$ and $\gamma_i = \varphi_i^*$) (bottom). Since $F_{\gamma_i}(\theta)$ does not depend on λ we choose an arbitrary value of $\lambda = 10^{-4}/\mathrm{m}^2$ ($100/\mathrm{km}^2$) for obtaining the Monte Carlo results.

A. Distribution of the SIR threshold

Herein we find the distribution of the SIR threshold allocated to each UE for satisfying its reliability constraint. The distribution is taken with respect to the PPP spatial randomness and the results are useful for characterizing the links' rate performance in general large-scale networks.

Theorem 2. *The distribution of the SIR threshold allocated to every UE_i for satisfying its reliability constraint ϵ_i according to (5) and (8), is given by*

$$F_{\gamma_i}(\theta) = 1 - \mathcal{L}^{-1} \left\{ \frac{1}{s \, {}_1F_1(-\delta, 1-\delta, -s)} \right\} (z_i(\theta)), \quad (9)$$

$$\underset{z_i(\theta) \leq 1}{\overset{z_i(\theta) > 1}{\lesseqqgtr}} 1 - \mathrm{sinc}(\delta) z_i(\theta)^\delta, \quad (10)$$

in downlink NOMA networks with PPP distributed BSs. In (9) and (10) we use $\delta = 2/\alpha$ and

$$z_i(\theta) = \left(\frac{P_i}{\theta} - \mu \sum_{j=1}^{i-1} P_j - \sum_{j=i+1}^{M} P_j \right) \frac{\epsilon_i}{P_T}. \quad (11)$$

Proof. See Appendix C.

Notice that for $z_i(\theta) > 1$ there is no closed-form analytical expression for $F_{\gamma_i}(\theta)$ since it is required computing

the Laplace inversion of a nasty function which includes an hypergeometric term, thus, numerical methods are required (see [29] for algorithmic implementations of numeric Laplace inversion). Still, (9) is of enormous importance since Monte Carlo simulation of these scenarios are particularly time consuming. Another important observation is that $F_{\gamma_i}(\theta)$ does not depend on the network density λ which matches previous results as in [30, Section III] for Poisson cellular networks without rate control and reliability restrictions.

Fig. 2b corroborates (9) and (10) since the numerical evaluation of (9) and the Monte Carlo results match, while (10) matches when $\theta \geq \epsilon_i$ and works as a lower bound in the complementary region. For a fixed reliability constraint it is shown that as the path-loss exponent increases, the curves shift to the right and greater data rates are feasible. This is because we are considering an interference-limited setup and notice that the interference is affected more than the receive power of the desired signal itself since interfering signals endure larger distances. While for fixed path-loss exponent the allocated data rates decrease as the reliability constraints become more strict and in such case the gap between curves with different values of ϵ_i matches their quotient according to (8). For instance, the gap between the curves with $\epsilon_i = 10^{-1}$ and $\epsilon_i = 10^{-2}$ in Fig. 2b is 10dB which matches $10^{-1}/10^{-2}$.

IV. POWER ALLOCATION AND FEASIBILITY

Herein, we derive and analyze the main conditions that are necessary for the model previously described to be feasible, while we discuss the optimum power allocation strategy for two transmission rate allocation problems: equal-rate (Subsection IV-A) and maximum sum-rate (Subsection IV-B). Notice that the transmission rate, $\mathcal{R}$, differs conceptually from the achievable rate, which is a more relevant performance metric since accounts also for the unsuccessful attempts; however, in practical scenarios where $\epsilon_i \leq 10^{-1}$ both metrics approximate well. This and the fact that dealing with the transmission rate is analytically easier are the main reasons why we selected it as the performance metric.

For analytical tractability we focus on the $M = 2$ setup, but notice that even when some existing results show that NOMA with more devices may provide a better performance gain [21], this may not be practical. The reason is that considering processing complexity for SIC receivers, especially when SIC error propagation is considered, 2-users NOMA is actually more practical in reality [31], [32]. Finally, let's set $P_1 = \beta P_T$, $P_2 = (1 - \beta)P_T$, where $\beta \in [1/2, \ 1]$ since by NOMA definition and SIC operation: $P_1 \geq P_2$. Therefore, parameter β completely determines the power allocation profile.

A. Equal-rate allocation

Herein we consider the scenario in which both UEs are scheduled with the same transmission rate $\mathcal{R} = \log_2(1+\gamma_1) = \log_2(1+\gamma_2)$. Therefore, maximum fairness is attained which is practically advantageous when UEs require transmissions with very similar QoS constraints, e.g., similar ϵ_i and data rates, although our results hold independently of this. For this

case and according to Remark 1: $\gamma_1 = \gamma_2 < \min(\varphi_1^*, \varphi_2^*)$. We now present the following result.

Theorem 3. *The power allocation profile*

$$\beta = \frac{\varphi_1^* + \varphi_2^* + 2\varphi_1^*\varphi_2^*}{2(1-\mu)\varphi_1^*\varphi_2^*} +$$
$$- \frac{\sqrt{(\varphi_1^* + \varphi_2^*)^2 + 4\varphi_1^*\varphi_2^*(\varphi_1^* + \mu\varphi_2^* + \mu\varphi_1^*\varphi_2^*)}}{2(1-\mu)\varphi_1^*\varphi_2^*} \quad (12)$$

guarantees the same performance in terms of transmission rate for both UEs.

Proof. See Appendix D.

The reachable SIR threshold can be easily calculated by substituting (12) into any of γ_1 or γ_2 expressions, which yields

$$\gamma_1 = \gamma_2 = -\frac{\varphi_1^* + \varphi_2^*}{2(\varphi_1^* + \mu\varphi_2^*(1 + \varphi_1^*))} +$$
$$+ \frac{\sqrt{(\varphi_1^* + \varphi_2^*)^2 + 4\varphi_1^*\varphi_2^*(\varphi_1^* + \mu(1 + \varphi_1^*)\varphi_2^*)}}{2(\varphi_1^* + \mu\varphi_2^*(1 + \varphi_1^*))}. \quad (13)$$

Corollary 1. *For equal-rate allocation, the optimum decoding order for the UEs is $\varphi_2^* \geq \varphi_1^*$.*

Proof. See Appendix E.

Let's say UE_A and UE_B are about to be served, then, Corollary 1's result states how they should be ordered at the BS for optimum performance. This is, one should calculate φ^* for each UE according to (8) and then if $\varphi_A^* > \varphi_B^* \rightarrow UE_1 = UE_B$, $UE_2 = UE_A$, otherwise if $\varphi_A^* < \varphi_B^* \rightarrow UE_1 = UE_A$, $UE_2 = UE_B$.

Fig. 3a shows the ratio φ_1^*/φ_2^* for maximum transmission rate as a function of $\varphi_1^* + \varphi_2^*$. Since $\varphi_2^* \geq \varphi_1^*$, the ratio is always smaller or equal than 1. Notice that as $\varphi_1^* + \varphi_2^*$ increases and/or μ decreases, it is more desirable scheduling together for sharing the same spectrum resources those UEs having contrasting values of φ_i^*. According to Theorem 1, the way to control this lies on checking the UEs reliability constraints and the ratio between average signal and interference power. The more heterogeneous (homogeneous) the product of these parameters, the more (less) contrasting the values of φ_i^*. Additionally, according to the figure when SIC always fails the best performance comes when $\varphi_1^* = \varphi_2^*$.

A comparison with the OMA counterpart is required, especially because under OMA setup, both UEs require orthogonal spectrum resources, either in time or frequency. Although optimum OMA design usually conduces to different partition of resources, this is not practically viable since requires cumbersome multi-user synchronization and scheduling tasks for partitioning the time, and could be even infeasible when bandwidth partition is required. Therefore, herein we assume equal time/frequency partition. Since φ_i^* represents the SIR threshold required for UE_i for satisfying its reliability constraint if it is operating alone in the channel, its maximum allocable transmit rate is $\frac{1}{2}\log_2(1 + \varphi_i^*)$, hence it is not the same for both UEs. The goal in this subsection lies in achieving equal-rate performance, thus, the rate under the OMA setup is determined by the worst UE performance as $\frac{1}{2}\log_2(1 + \min(\varphi_1^*, \varphi_2^*)) = \frac{1}{2}\log_2(1 + \varphi_1^*)$. Therefore, both UEs may be allocated with the same transmission rate, but
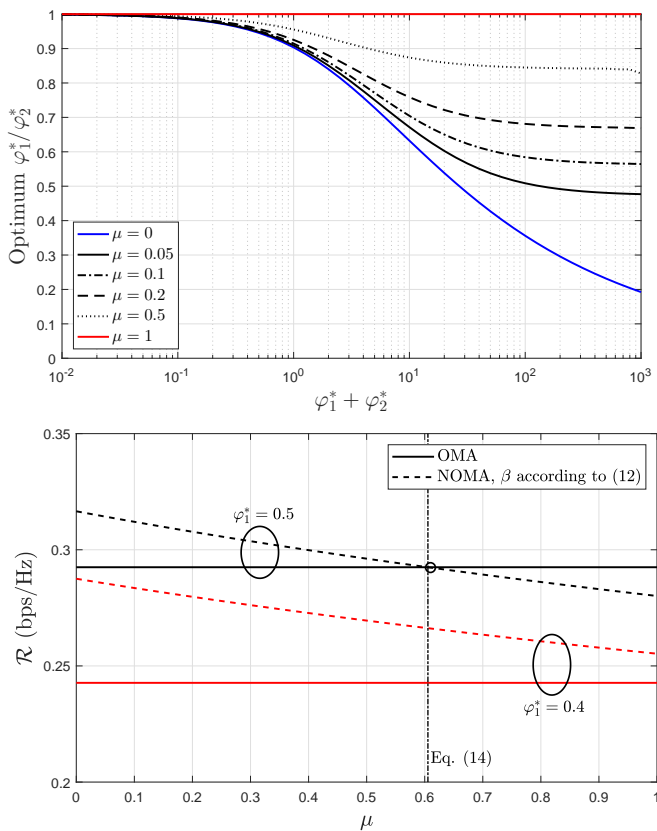
Fig. 3. (a) Optimum ratio φ_1^*/φ_2^* as a function of $\varphi_1^* + \varphi_2^*$ for $\mu \in \{0, 0.05, 0.1, 0.2, 0.5, 1\}$ (top); (b) Transmission rate in bps/Hz as a function of μ, for NOMA and OMA, $\varphi_2^* = 0.6$ and $\varphi_1^* \in \{0.4, 0.5\}$ (bottom).

while UE_1 performs with ϵ_1, UE_2 performs with an outage performance smaller than ϵ_2. Hence, we present the following result.

Corollary 2. *For equal-rate allocation and $M = 2$, NOMA outperforms OMA when*

$$\mu < \frac{(\varphi_2^* - \varphi_1^*)(1 + \sqrt{1 + \varphi_1^*})}{\varphi_1^* \varphi_2^* \sqrt{1 + \varphi_1^*}}. \quad (14)$$

Proof. For the equal-rate allocation, the NOMA scheme overcomes the OMA configuration when

$$\log_2(1 + \gamma_1) > \frac{1}{2} \log_2(1 + \varphi_1^*)$$
$$\gamma_1 > \sqrt{1 + \varphi_1^*} - 1. \quad (15)$$

Now, substituting (13) into (15), and after some algebraic manipulations we attain the condition given in (14). $\square$

Based on (14) notice that NOMA performance probably overcomes OMA's if φ_1^* and φ_2^* are sufficiently different values. On the other hand, when $\varphi_2^* = \varphi_1^*$ OMA will perform better. Therefore, when scheduling users to work on the same channel with NOMA they should have sufficiently different topological characteristics, e.g., different $\frac{r_i^{-\alpha}}{\sum_{j \in \Phi \setminus \{b_0\}} r_{j,i}^{-\alpha}}$, and/or QoS requirements, e.g., different ϵ_i, as shown in (8). This is corroborated in Fig. 3b where we can notice that the gap between both schemes increases when $\varphi_2^* - \varphi_1^*$ also

increases. Fig. 3b also corroborates Corollary 2 and notice that when evaluating (14) with $\varphi_1^* = 0.4$, $\varphi_2 = 0.6$ we obtain that the required values of μ for which NOMA performs worse than OMA are $\mu \geq 1.53$, which are infeasible since $\mu \leq 1$, thus, for that setup NOMA will always be the better choice.

B. Maximum Sum-rate allocation

The previous subsection addressed the problem of finding the power allocation profile and scheduling order such that both UEs can operate with the same maximum possible rate. This is equivalent to find $\max_{\beta} \min(\gamma_1, \gamma_2)$, since increasing γ_1 conduces always to decrease γ_2 and vice versa. Herein we are going to focus on finding the allocation strategy for maximum sum-rate, thus, $\max_{\beta} \mathcal{R}$, where

$$\mathcal{R} = R_1 + R_2 = \log_2(1 + \gamma_1) + \log_2(1 + \gamma_2)$$
$$= \log_2\left((1 + \gamma_1)(1 + \gamma_2)\right) = \log_2 \tilde{\gamma}, \quad (16)$$

where $\tilde{\gamma} = (1 + \gamma_1)(1 + \gamma_2)$, thus, $\max_{\beta} \mathcal{R} = \max_{\beta} \tilde{\gamma}$.

Notice that solving that problem could be extremely difficult, if not impossible, due to the tangled dependence of $\tilde{\gamma}$ on β, as shown next

$$\tilde{\gamma} = (1 + \gamma_1)(1 + \gamma_2)$$
$$= \frac{(1 + \varphi_1^*)(1 + \varphi_2 - \beta\varphi_2^*(1 - \mu))}{(1 + (1 - \beta)\varphi_1^*)(1 + \beta\mu\varphi_2^*)}. \quad (17)$$

Only for $\mu = 0$ the problem solution comes easy as follows

$$\tilde{\gamma}|_{\mu=0} = (1 + \varphi_1^*)\frac{1 + \varphi_2^*(1 - \beta)}{1 + \varphi_1^*(1 - \beta)}, \quad (18)$$

$$\frac{d\tilde{\gamma}|_{\mu=0}}{d\beta} = (1 + \varphi_1^*)\frac{\varphi_1^* - \varphi_2^*}{(1 + \varphi_1^*(1 - \beta))^2}, \quad (19)$$

thus, $\tilde{\gamma}|_{\mu=0}$ is a strictly decreasing (increasing) function for $\varphi_1^* < \varphi_2^*$ ($\varphi_1^* > \varphi_2^*$) and any value of $\beta \in [0, 1]$.

Remark 2. *Therefore, if $\mu = 0$ then the optimal performance for the NOMA setup, in terms of maximum sum-rate, is obtained by setting*

$$\beta^* = \begin{cases} 1/2, & \text{if } \varphi_2^* \geq \varphi_1^* \to \tilde{\gamma}|_{\mu=0} = \frac{(1+\varphi_1^*)(2+\varphi_2^*)}{2+\varphi_1^*} \\ 1, & \text{if } \varphi_2^* < \varphi_1^* \to \tilde{\gamma}|_{\mu=0} = 1 + \varphi_1^* \end{cases}. \quad (20)$$

Of course, it is desirable having/setting $\varphi_2^* > \varphi_1^*$ while using $\beta = 1/2$, otherwise the optimum performance is when $\beta = 1$ which matches the OMA setup when only one user is being served at the time. Also, notice that $\beta^* = 1/2$ conduces to a greater value of $\tilde{\gamma}|_{\mu=0}$ than when $\beta = 1$.

For the general case when $\mu > 0$ we propose using the following result henceforth.

Proposition 1. *With $\bar{\gamma} = \gamma_1 + \gamma_2$, $\tilde{\gamma}$ approximates to*

$$\tilde{\gamma} \approx \left[1 + \frac{1}{2}\bar{\gamma}\right]^2, \quad (21)$$

Proof. See Appendix F.

Finally, we provide two approximate results characterizing the optimum performance of NOMA and the region for which NOMA outperforms OMA for the general case of $\mu \geq 0$.
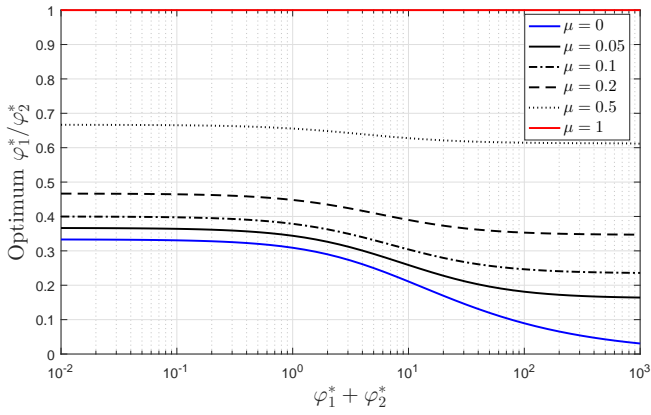
Fig. 4. Optimum ratio φ_1^*/φ_2^* as a function of $\varphi_1^* + \varphi_2^*$ for $\mu \in \{0, 0.05, 0.1, 0.2, 0.5, 1\}$ and $\beta = 1/2$.

Theorem 4. *The optimal power allocation profile for maximum sum-rate in the NOMA setup is*

$$\beta^* = \begin{cases} 1/2, & \text{if } \mu < \frac{2(\varphi_2^* - \varphi_1^*) + \varphi_1^*(\varphi_2^* - 2)}{\varphi_1^* \varphi_2^*(1 + \varphi_1^*)} \\ 1, & \text{otherwise} \end{cases} . \quad (22)$$

Proof. See Appendix G.

Notice that when $\beta = 1/2$ we have

$$\bar{\gamma} = \frac{\varphi_1^*}{2 + \varphi_1^*} + \frac{\varphi_2^*}{2 + \mu \varphi_2^*}., \quad (23)$$

and for $\beta = 1 \to \bar{\gamma} = \varphi_1^*$.

As done in the previous subsection for the optimum equal-rate allocation, herein we show in Fig. 4 the optimum ratio φ_1^*/φ_2^* as a function of $\varphi_1^* + \varphi_2^*$ for the maximum sum-rate allocation problem. Again, the trend is that the larger $\varphi_1^* + \varphi_2^*$ and/or smaller μ, the more contrasting should be the values of φ_i^* for optimum performance. However, in this case the dependence is much weaker on $\varphi_1^* + \varphi_2^*$ than for the case of equal-rate allocation which may facilitate the design and scheduling tasks in a practical system.

As in the previous subsection, herein we also consider OMA with equal partition of spectrum resources, either in time or frequency, thus, the maximum transmission rate for each UE_i signal is $\frac{1}{2} \log_2(1 + \varphi_i^*)$. Then, the following result holds.

Theorem 5. *The NOMA setup with $M = 2$ outperforms always OMA in terms of maximum sum-rate when $\mu = 0$, while almost surely[5] when*

$$\mu < \frac{\sqrt{2}(\varphi_1^* + 2)^2 \sqrt{2 + \varphi_1^* + \varphi_2^*}}{\varphi_1^{*2}(2\varphi_1^* + 3) + 2\varphi_2^*(\varphi_1^* + 2)^2} + $$
$$- \frac{4\varphi_1^{*3} + 6\varphi_1^{*2} + \varphi_1^{*2}\varphi_2^* + 6\varphi_1^*\varphi_2^* + 8\varphi_2^*}{\varphi_2^*(\varphi_1^{*2}(2\varphi_1^* + 3) + 2\varphi_2^*(\varphi_1^* + 2)^2)} \quad (24)$$

with $\varphi_2^* > \varphi_1^*$.

Proof. See Appendix H.

---

[5]For $\mu > 0$ the proof takes advantage of result in Proposition 1 since using the exact expression of $\bar{\gamma}$ is cumbersome. Therefore, (24) is expected to hold given the accuracy of (21), hence the term *almost surely*.
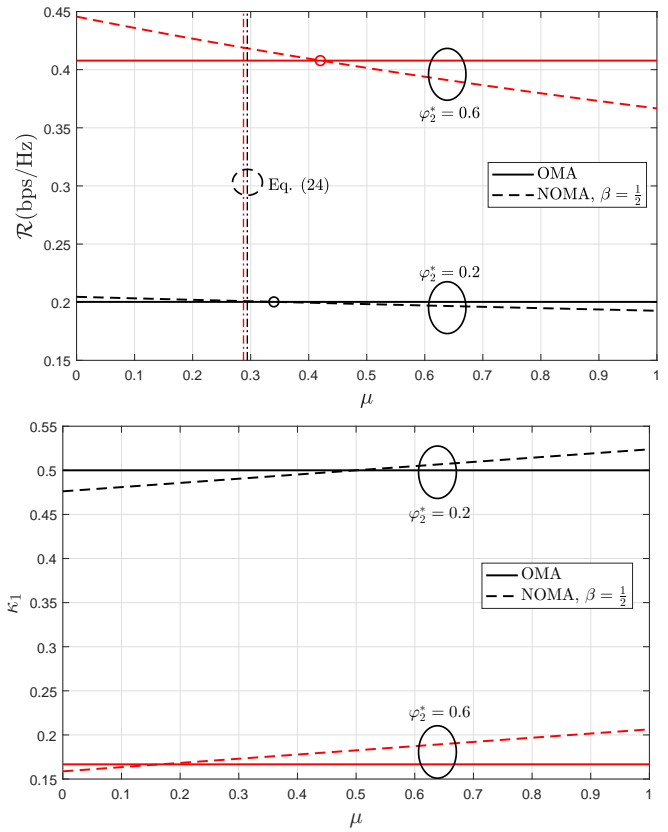

Fig. 5. (a) Transmission rate in bps/Hz (top) and (b) Fairness coefficient (bottom), as a function of μ, for NOMA and OMA, $\varphi_1^* = 0.1$ and $\varphi_2^* \in \{0.2, 0.6\}$.

See Fig. 5a for a comparison between NOMA and OMA schemes and notice that (24) holds in each case. Also, and as shown in the figure, we can assure that NOMA always overcomes OMA when $\mu \to 0$, as it was claimed by Theorem 5. Finally, a comparison in terms of fairness when allocating the transmission rate is required. While the power allocation scheme proposed in the previous subsection allows reaching maximum fairness, this is not longer the case when the goal is maximizing the sum-rate. As a measure of fairness let us consider the ratio between the SIR thresholds of both UEs, since this is analytically easier to handle than the transmission rate itself. We define the fairness coefficient of UE_i with respect to UE_j, with $i \neq j$ and $i, j \in \{1, 2\}$, as $\kappa_i^{\text{oma}} = \varphi_i^*/\varphi_j^*$ and $\kappa_i^{\text{noma}} = \gamma_i/\gamma_j$, and the closer κ_i is from 1, the fairer the rate allocation. For optimality we set $\varphi_1^* \leq \varphi_2^*$ according to Theorem 5, then, focusing on κ_1 we have that $\kappa_1^{\text{oma}} = \varphi_1^*/\varphi_2^* \leq 1$ and

$$\kappa_1^{\text{noma}} \overset{(a)}{=} \frac{\frac{\frac{1}{2}\varphi_1^*}{1 + \frac{1}{2}\varphi_1^*}}{\frac{\frac{1}{2}\varphi_2^*}{1 + \frac{1}{2}\mu\varphi_2^*}} = \frac{\varphi_1^*(2 + \mu\varphi_2^*)}{\varphi_2^*(2 + \varphi_1^*)} = \kappa_1^{\text{oma}} \frac{2 + \mu\varphi_2^*}{2 + \varphi_1^*}, \quad (25)$$

where (a) comes from setting $\beta = 1/2$ in (5) as this was proved to be the optimum when NOMA provides greater sum-rate than OMA. Notice that as long as $\mu < \varphi_1^*/\varphi_2^*$, OMA is fairer as also illustrated in Fig. 5b, however this is not advantageous when heterogeneous services need to be provided. In practice, when considering the entire set of

frequency-time resources, previous results are useful when making the pairing of those UEs that will be scheduled with the same spectrum resources.

V. Algorithmic and Performance Remarks

Section III focused on the optimum allocation of the transmission rates of concurrent downlink NOMA users under practical reliability constraints and given their preassigned transmit powers. Meanwhile, Section IV explained how to find these transmit powers for optimum performance in the 2-user NOMA scenario. Next, we summarize how to integrate both rate and power allocation schemes for attaining the optimum performance. This is discussed in Subsection V-A, while Subsection V-B shows how to calculate the average rate performance for NOMA and OMA setups as it will be used for illustrating numerical results in Section VI.

A. Optimum Rate-Power Allocation

By combining the results from Section III and IV, the optimum rate-power allocation algorithm for a 2-user downlink NOMA setup without instantaneous CSI at the BS is given as:

1) Compute φ^* according to (8) for each UE.
2) Set the UE with the lowest φ^* as first in the decoding order, while the other becomes second. This is grounded on Corollary 1 and Theorem 5 results.
3) Compute $P_1 = \beta P_T$ and $P_2 = (1 - \beta)P_T$, where β is given by (3) and (22) for equal-rate and maximum sum-rate allocation, respectively.
4) Compute γ_i, $i = 1, 2$ according to (5) with $M = 2$. For the case of equal-rate allocation it is also possible using (13), which yields the same result.
5) Set P_i and $\log_2(1 + \gamma_i)$ as transmit power and rate, respectively, of the signal intended to UE_i.

B. Average performance of NOMA and OMA

For each UE_i, the distribution of its SIR threshold γ_i with respect to the PPP spatial randomness was found in Subsection III-A for a fixed power allocation profile. When BSs allocate their transmit power based on the network deployment, as it is the case of the power control scheme for 2-user NOMA proposed in Section IV for equal-rate and maximum sum-rate problems, expressions (9) and (10) in Subsection III-A do not hold anymore. This is because P_i is now another random variable in the PPP. In such cases characterizing the distribution of γ_i, hence, the distribution of the allocated rate, seems analytically intractable because of the tangled dependence of β on φ_i^* according to (12), and (22) for $\mu > 0$, respectively. Only for the case of perfect SIC in a system where the goal is maximizing the sum-rate, such distribution can be obtained since the optimum β is deterministically $1/2$ according to (22), thus, (9) holds and (11) becomes

$$z_i(\theta) = \left(\frac{P_T}{2\theta} - \frac{\mu P_T}{2} \sum_{j=1}^{i-1} 1 - \frac{P_T}{2} \sum_{j=i+1}^{2} 1 \right) \frac{\epsilon_i}{P_T}$$
$$= \frac{\epsilon_i}{2} \left(\frac{1}{\theta} + (1 - \mu)(i - 1) - 1 \right). \quad (26)$$
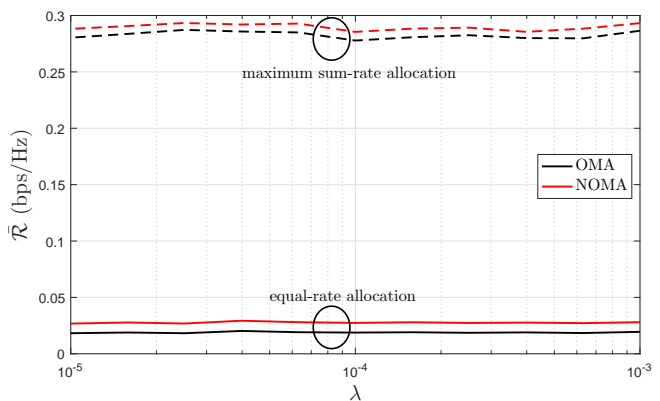

Fig. 6. Average rate for OMA and NOMA setups operating with equal-rate and maximum sum-rate configurations and $\epsilon_i = 10^{-2}$.

For a comparison between OMA and NOMA in terms of average rate performance, $\bar{\mathcal{R}}$, over the PPP, following expressions hold

- equal-rate allocation

$$\bar{\mathcal{R}} = \begin{cases} \mathbb{E}\left[\frac{1}{2} \log_2\left(1 + \min\left(\varphi_1^*, \varphi_2^*\right)\right) \Big| \Phi \right] : & \text{OMA} \\ \mathbb{E}\left[\log_2\left(1 + \gamma_1\right) \Big| \Phi \right] : & \text{NOMA} \end{cases}, \quad (27)$$

- maximum sum-rate allocation

$$\bar{\mathcal{R}} = \begin{cases} \mathbb{E}\left[\frac{1}{2} \log_2\left(1 + \varphi_1^*\right)\left(1 + \varphi_2^*\right) \Big| \Phi \right] : & \text{OMA} \\ \mathbb{E}\left[\log_2 \tilde{\gamma} \Big| \Phi \right] : & \text{NOMA} \end{cases}, \quad (28)$$

where $\tilde{\gamma}$ is given in (17).

These expressions are used in Section VI for numerical analysis. Notice also that as a direct consequence of the discussions in Subsection III-A, the average transmission rate performance of NOMA and OMA setups does not depend on the network density when randomizing the network topology through the PPP. This is corroborated in Fig. 6 where the average is taken over the network realizations. The values of the system parameters are those given at the beginning of the next section.

VI. Numerical Results

Herein we compare numerically the performance of NOMA and OMA setups under the proposed distributed rate control methodology. We simulate completely random scenarios by generating 50000 instances (Monte Carlo runs) of a PPP and a sufficiently large area such that 1000 BSs are placed on average. Since the average performance of NOMA and OMA setups do not depend on λ, we are going to use a generic value of $\lambda = 10^{-4}/\text{m}^2$ (100/km²). Unless stated otherwise, results are obtained by setting $\alpha = 4$ and $\mu = 0.1$. We utilize the optimum values of β given in Section IV-A and Section IV-B for equal-rate and maximum sum-rate allocations, respectively.

Additionally, the performance when perfect CSI and knowledge of instantaneous levels of interference at the UEs are fully available at the BS side, is illustrated as benchmark. As in [21], the decoding order is now established for holding

TABLE I
AVERAGE TRANSMISSION RATE FOR BENCHMARK OMA AND NOMA SCHEMES.

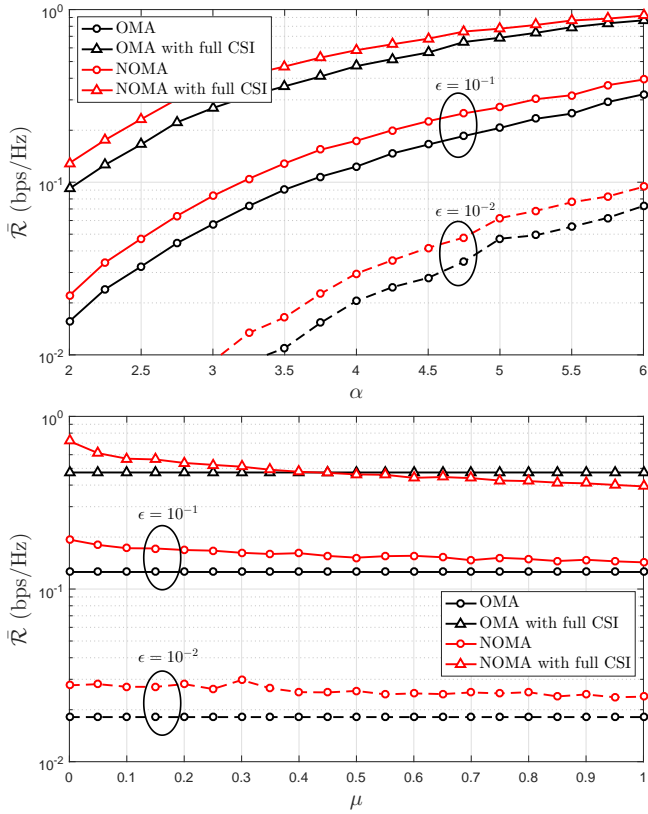| Scheme	Equal-rate	Maximum sum-rate
OMA	$\mathbb{E}\left[\frac{1}{2}\log_2\left(1+\min(h_1 r_1^{-\alpha} P_T/I_1, h_2 r_2^{-\alpha} P_T/I_2)\right)\Big|\Phi\right]$	$\mathbb{E}\left[\frac{1}{2}\log_2(1+h_1 r_1^{-\alpha} P_T/I_1)(1+h_2 r_2^{-\alpha} P_T/I_2)\Big|\Phi\right]$
NOMA	$\mathbb{E}\left[\log_2(1+\mathrm{SIR}_1) \text{ with } 1/2 \le \beta \le 1: \mathrm{SIR}_1 = \mathrm{SIR}_2\Big|\Phi\right]$	$\mathbb{E}\left[\max_{1/2 \le \beta \le 1}\log_2(1+\mathrm{SIR}_1)(1+\mathrm{SIR}_2)\Big|\Phi\right]$

Fig. 7. Average rate for OMA and NOMA setups as a function of (a) α (top), (b) μ (bottom), for equal-rate allocation and $\epsilon_i \in \{10^{-1}, 10^{-2}\}$.

$h_1 r_1^{-\alpha}/I_1 < h_2 r_2^{-\alpha}/I_2$. Since h_i and I_i are deterministic values in such scenario, the rate can be allocated directly from (4) for each channel realization as $2^{\mathrm{SIR}_i} - 1$. According to Shannon, this rate allows operating with an arbitrarily small error probability as the blocklength goes to infinity [33][6]. Finally, the average allocated transmission rates under the benchmark NOMA scheme, and also under its OMA counterpart, appear shown in Table I for equal-rate and maximum sum-rate problems.

A. Equal-rate allocation

Fig. 7 shows the average rate for OMA and NOMA setups when operating with equal-rate allocation, and as a function of α (Fig. 7a) and μ (Fig. 7b). Notice that for our CSI-free schemes the average performance of NOMA overcomes

always the OMA's, and the gap increases for smaller SIC imperfection μ and/or path-loss exponent[7] α. Meanwhile, NOMA under full CSI at the BS is more sensitive to μ, and for $\mu > 0.4$ its OMA counterpart performs better. These benchmark schemes provide an upper-bound performance not only because full CSI is considered at the BS, but also because only when transmitting over an infinite blocklength their results hold. Therefore and as expected, it is shown a large gap between these bounds and the performance of the OMA and NOMA schemes discussed in this work, and increases as the reliability constraint becomes more stringent. For all the schemes the performance improves when the path-loss exponent increases, which is expected according to our discussions in Subsection III-A. In a nutshell, since interfering signals endure larger distances, as α increases their impact is affected more than the receive power of the desired signal itself, hence, larger values of SIR are more likely. Additionally, the performance of OMA does not depend on μ, thus, it appears as a straight line in Fig. 7b, and remarkably even in the worst case of $\mu = 1$ NOMA attains a greater average performance than OMA when operating without CSI at the BS. Notice that $\mu = 1$ implies only that the interference coming from UEs that come first in the decoding order cannot be canceled. Therefore, this affects only the denominator of (4), and the UE being currently decoded has still chances of succeeding. Obviously, the OMA setup could be optimized such that different partitions of time/frequency resources are allocated to each user, thus improving its performance and even overcoming NOMA. However, notice that this is extremely cumbersome since involves synchronization tasks if the split resource is time, and it could be even unfeasible if the resource is the frequency bandwidth. Meanwhile, NOMA only requires to use a different power level for the signal of each UE.

B. Maximum sum-rate allocation

Fig. 8 shows the average rate for OMA and NOMA setups when operating with maximum sum-rate allocation, and as a function of α (Fig. 8a) and μ (Fig. 8b). For this case the performance of OMA improves faster than the NOMA's when α increases, although when reliability constraint becomes more stringent both ideal CSI-free schemes tend to have the same improvement rate with respect to α. Notice that it is required that $\mu \lesssim 0.1$ ($\lesssim 10\%$ of the co-interference is not canceled) so that NOMA is preferable to OMA. This is

---

[6]The trade-off between the blocklength and the error probability was analytically investigated in [33] but herein we assume the asymptotic scenario for which the blocklength is infinity, thus $\epsilon_i \to 0$ if the transmission rate is set to be $2^{\mathrm{SIR}_i} - 1$ for each channel realization.

[7]This behavior can be corroborated by checking that indeed $\frac{d}{d\alpha}\left(\frac{\varphi_2^*}{\varphi_1^*}\right) < 0$ in most of the PPP realizations for $\epsilon_1 = \epsilon_2$. Notice that as $\frac{\varphi_2^*}{\varphi_1^*}$ decreases NOMA loses its advantages over OMA as discussed after Corollary 2.
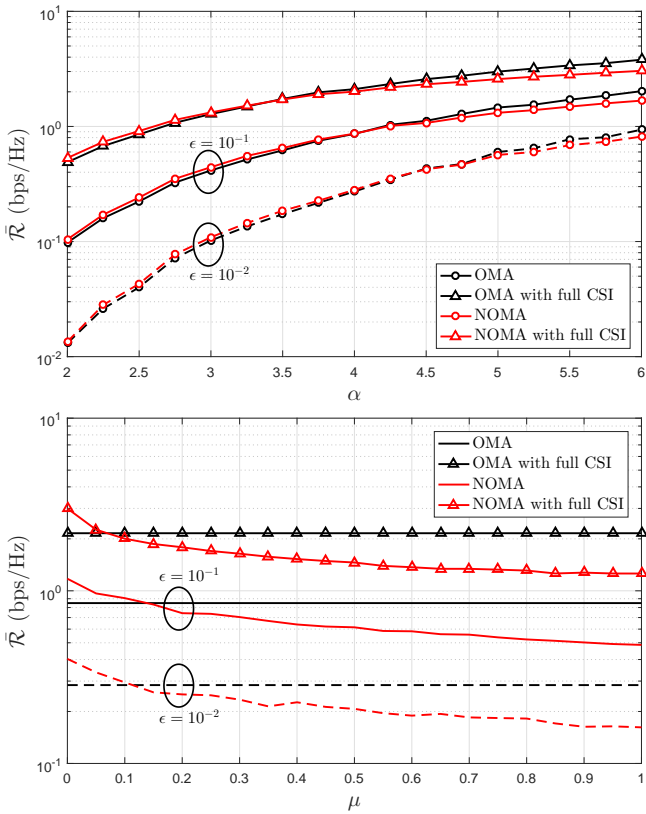
Fig. 8. Average rate for OMA and NOMA setups as a function of (a) α (top), (b) μ (bottom), when operating with maximum sum-rate allocation and $\epsilon_i \in \{10^{-1}, 10^{-2}\}$.

illustrated in Fig. 8b, where the gains of NOMA over OMA are of 39%, 38% and 42% when $\mu \to 0$ for the benchmark and CSI-free schemes with $\epsilon_i = 10^{-1}$ and $\epsilon_i = 10^{-2}$, respectively; while when $\mu \to 1$, OMA offers a performance gain around of 72%, 81% and 71%. Therefore, failing to efficiently eliminate the co-interference is critical for NOMA, and can be a challenging issue when implementing it in practice. Finally and as commented in the Section VI-A, benchmark schemes provide a loose upper-bound performance which is impossible to attain for any practical communication system.

VII. CONCLUSION

In this paper, we proposed a rate allocation scheme for a downlink NOMA cellular system operating with reliability constraints. The allocated rate depends on: *i)* the average receive power from the desired signal and from the interfering BSs; and *ii)* the reliability constraint. This information is usually easy to get and therein lies the main advantage of the scheme. Additionally, we characterize analytically the distribution of the allocated SIR threshold in Poisson cellular networks in order to meet the link's reliability constraint. We derive and analyze the main conditions that are necessary so that NOMA overcomes the OMA alternative, while we discuss the optimum allocation strategies, e.g., rate and power allocation profile and optimum decoding order, for the 2-UEs NOMA setup with equal-rate and maximum sum-rate goals.

We compare numerically the performance of our allocation scheme with its ideal counterpart requiring full CSI at the BSs and infinitely long blocklength, and show how the gap increases as the reliability constraint becomes more stringent. Also, results highlight the benefits of NOMA over OMA when the co-interference can be efficiently canceled, e.g., with an efficiency of 90%, specially when we aim to maximize the sum-rate.

APPENDIX A
PROOF OF LEMMA 1

Setting the conditional link success probability to $1 - \epsilon_i$, we have

$$1 - \epsilon_i = \mathbb{P} \left(\frac{h_i r_i^{-\alpha} P_i}{h_i r_i^{-\alpha} \left[\mu \sum_{j=1}^{i-1} P_j + \sum_{j=i+1}^{M} P_j \right] + I_i} > \gamma_i \middle| \Phi \right)$$

$$= \mathbb{P} \left(h_i > \frac{I_i r_i^{\alpha}}{\frac{P_i}{\gamma_i} - \mu \sum_{j=1}^{i-1} P_j - \sum_{j=i+1}^{M} P_j} \middle| \Phi \right)$$

$$\overset{(a)}{=} \mathbb{E} \left[\mathbb{P} \left(h_i > \frac{r_i^{\alpha} P_T \sum_{j \in \Phi \setminus \{b_0\}} g_j r_{j,i}^{-\alpha}}{\frac{P_i}{\gamma_i} - \mu \sum_{j=1}^{i-1} P_j - \sum_{j=i+1}^{M} P_j} \right) \middle| \Phi \right]$$

$$\overset{(b)}{=} \mathbb{E} \left[\exp \left(-\frac{r_i^{\alpha} P_T \sum_{j \in \Phi \setminus \{b_0\}} g_j r_{j,i}^{-\alpha}}{\frac{P_i}{\gamma_i} - \mu \sum_{j=1}^{i-1} P_j - \sum_{j=i+1}^{M} P_j} \right) \middle| \Phi \right]$$

$$\overset{(c)}{=} \mathbb{E} \left[\prod_{j \in \Phi \setminus \{b_0\}} \exp \left(-\varphi_i r_i^{\alpha} g_j r_{j,i}^{-\alpha} \right) \middle| \Phi \right]$$

$$\overset{(d)}{=} \prod_{j \in \Phi \setminus \{b_0\}} \frac{1}{1 + \varphi_i r_i^{\alpha} r_{j,i}^{-\alpha}}, \tag{29}$$

where (a) comes from using (3), (b) from using the complementary CDF (CCDF) of h_i, while (c) from using the algebraic transformation $\exp \left(\sum_{i=1}^{n} k_i \right) = \prod_{i=1}^{n} \exp(k_i)$ and setting

$$\varphi_i = \frac{P_T}{\frac{P_i}{\gamma_i} - \mu \sum_{j=1}^{i-1} P_j - \sum_{j=i+1}^{M} P_j}. \tag{30}$$

Finally, (d) comes from averaging over the fading g_j, attaining (6). By isolating γ_i in (30) we reach (5). Notice that (6) may have up to $|\Phi \setminus \{b_0\}|$ solutions; however $\gamma_i > 0$ then φ_i^* has to be also real positive according to (5). Now, the left term of (5) is monotonically increasing on φ_i and both left and right terms share the same domain $(0, \infty)$ $\forall \varphi_i > 0, 0 < \epsilon_i < 1$, hence (5) has only one real positive solution φ_i^*. $\qquad \square$

APPENDIX B
PROOF OF THEOREM 1

Let's consider that the number of interfering BSs is finite and equal to n and we sort in ascending order the interfering
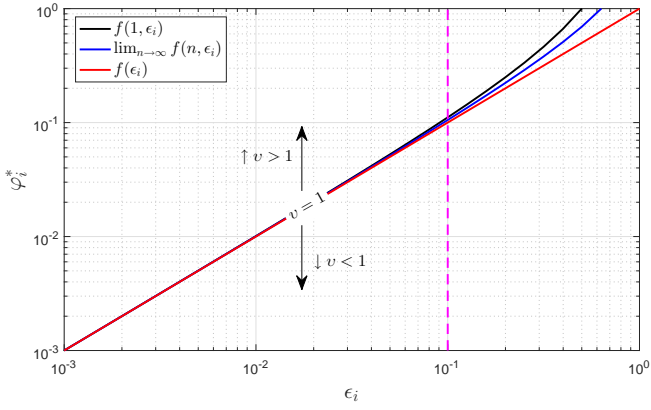
Fig. 9. φ_i^* as a function of ϵ_i and according to (33). We set $\upsilon = r_i^{-\alpha}/\sum_{j=1}^{n} r_{j,i}^{-\alpha}$.

BSs according to their distance to UE_i. Hence, $r_{j,i}, j \geq 1$, denotes the distance from UE_i to its $j-$nearest interfering BS. Notice that by letting $n \to \infty$ we are also able to model an infinite network deployment. Now, by using the relation between the geometric and the arithmetic mean we have that

$$\left(\prod_{j=1}^{n}(1+\varphi_i r_i^\alpha r_{j,i}^{-\alpha})\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{j=1}^{n}(1+\varphi_i r_i^\alpha r_{j,i}^{-\alpha})$$

$$\prod_{j=1}^{n}(1+\varphi_i r_i^\alpha r_{j,i}^{-\alpha}) \leq \left[\frac{1}{n}\Big(n+\varphi_i r_i^\alpha \sum_{j=1}^{n} r_{j,i}^{-\alpha}\Big)\right]^n$$

$$\prod_{j=1}^{n}(1+\varphi_i r_i^\alpha r_{j,i}^{-\alpha}) \leq \left[1+\frac{\varphi_i r_i^\alpha}{n}\sum_{j=1}^{n} r_{j,i}^{-\alpha}\right]^n. \quad (31)$$

By using the right term of (31) instead of the left one when solving (6) for φ_i yields

$$\varphi_i^* \approx \frac{n r_i^{-\alpha}}{\sum_{j=1}^{n} r_{j,i}^{-\alpha}}\left[\Big(\frac{1}{1-\epsilon_i}\Big)^{\frac{1}{n}}-1\right]. \quad (32)$$

By looking at (6) notice that for small ϵ_i, e.g., $\epsilon_i \leq 10^{-1}$, $\frac{1}{1-\epsilon_i}$ becomes closer to 1 and since each of the terms $(1+\hat{\varphi}_i r_i^\alpha r_{j,i}^{-\alpha})$ are strictly greater that 1, we can expect that each of them approximates heavily to the unity, hence, all of these terms are very similar between each other. Since geometric and arithmetic means match when all the operating terms are equal, we can be sure that for small ϵ_i, e.g., $\epsilon_i \leq 10^{-1}$, the approximation given in (32) is accurate.

Now we can write φ_i^* as

$$\varphi_i^* \approx \frac{r_i^{-\alpha}}{\sum_{j=1}^{n} r_{j,i}^{-\alpha}} f(n,\epsilon_i), \quad (33)$$

where $f(n,\epsilon_i) = n\big((1-\epsilon_i)^{-\frac{1}{n}}-1\big)$ is a decreasing function of n since $\frac{\partial f(n,\epsilon_i)}{\partial n} = \frac{1}{(1-\epsilon_i)^n}+\frac{\ln(1-\epsilon_i)}{n(1-\epsilon_i)^n}-1 < 0$, $\forall \epsilon_i : 0 < \epsilon_i < 1$. Therefore, its maximum and minimum values are when $n=1$ and $n \to \infty$, respectively, for which $f(1,\epsilon_i) = \epsilon_i/(1-\epsilon_i)$ and $\lim_{n\to\infty} f(n,\epsilon_i) = -\ln(1-\epsilon_i)$. Interestingly, when $\epsilon_i \leq 10^{-1}$ both results are very close to each other, which leads to

the conclusion that $f(n,\epsilon_i)$ depends very weakly on n in that region. In fact, following result holds

$$f(n,\epsilon_i) \approx f(\epsilon_i) = \epsilon_i \quad (34)$$

hence φ_i^* can be written as in (8). The accuracy of (34) when $\epsilon_i \leq 10^{-1}$ can be easily appreciated in Fig. 9. Notice that the value of $\upsilon = r_i^{-\alpha}/\sum_{j=1}^{n} r_{j,i}^{-\alpha}$ only causes a vertical shift of the curve φ_i^* vs ϵ_i. □

APPENDIX C
PROOF OF THEOREM 2

Using (8) we proceed as follows

$$F_{\varphi_i^*}(\omega) = \mathbb{P}\left(\frac{r_i^{-\alpha}}{\sum_{j\in\Phi\setminus\{b_0\}} r_{j,i}^{-\alpha}}\epsilon_i < \omega\right)$$

$$= \mathbb{P}\left(\sum_{j\in\Phi\setminus\{b_0\}}\Big(\frac{r_i}{r_{j,i}}\Big)^\alpha > \frac{\epsilon_i}{\omega}\right)$$

$$= \mathbb{P}\left(\sum_{x\in\Phi\setminus\{x:||x||<r_i\}}\Big(\frac{r_i}{||x||}\Big)^\alpha > \frac{\epsilon_i}{\omega}\right)$$

$$= 1 - F_\Psi\left(\frac{\epsilon_i}{\omega}\right), \quad (35)$$

where $\Psi = \sum_{x\in\Phi\setminus\{x:||x||<r_i\}}\Big(\frac{r_i}{||x||}\Big)^\alpha$. Similar to [34, Def. 1] we define the squared relative distance process (SRDP) of the point process Φ as $\Theta = \{x \in \Phi\setminus\{x : ||x|| < r_i\} : \frac{||x||^2}{r_i^2}\}$, then

$$\Psi = \sum_{x\in\Theta} x^{-1/\delta}, \quad (36)$$

with $\delta = 2/\alpha$. Now, the probability generating functional (PGFL) of the SRDP is given by

$$G_\Theta[f] = \mathbb{E}\left[\prod_{x\in\Theta} f(x)\right] = \mathbb{E}\left[\prod_{x\in\Phi} f\Big(\frac{||x||^2}{r_i^2}\Big)\right]$$

$$\overset{(a)}{=} \int_0^\infty \exp\left(-2\lambda\pi\int_{r_i}^\infty \eta(1-f(\eta^2/r_i^2))\mathrm{d}\eta\right)f_{r_i}(r_i)dr_i$$

$$\overset{(b)}{=} 2\pi\lambda\int_0^\infty r_i \exp\left(-\lambda\pi r_i^2\Big(1+\int_1^\infty(1-f(y))\mathrm{d}y\Big)\right)dr_i$$

$$\overset{(c)}{=} \frac{1}{1+\int_1^\infty(1-f(y))dy}, \quad (37)$$

where (a) follows from the PGFL of the PPP, (b) is obtained by the substitution $\eta^2/r_i^2 \to y$ and using the distribution of the nearest neighbor distance, while (c) follows from solving the outer integral. Then,

$$\mathcal{L}_\Psi(s) = \frac{1}{1+\int_1^\infty\Big(1-\exp\big(-sx^{-1/\delta}\big)\Big)dy}$$

$$= \frac{1}{{}_1F_1(-\delta,1-\delta,-s)}, \quad (38)$$

$$F_\Psi(x) = \mathcal{L}^{-1}\left\{\frac{1}{s}\mathcal{L}_\Psi(s)\right\}(x)$$
$$= \mathcal{L}^{-1}\left\{\frac{1}{s\ _1F_1(-\delta, 1-\delta, -s)}\right\}(x), \qquad (39)$$

while plugging (39) into (35) and then using the variable transformation given in (30) we attain (9). Notice that although (9) is general, it is not in closed-form. However, by noticing that φ_i^* in (8) has the form of the SIR of a typical link in a Poisson network without fading, but scaled by the reliability constraint ϵ_i, we can take advantage of [35, Eq. 6] to write

$$F_{\varphi_i^*}(\omega) = 1 - \text{sinc}(\delta)\left(\frac{\omega}{\epsilon_i}\right)^{-\delta}, \qquad \text{for } \omega \geq \epsilon_i, \qquad (40)$$

which also works as a lower-bound when $\omega < \epsilon_i$, and for the special case of $\alpha = 4$ yields $F_{\varphi_i^*}(\omega) = 1 - 1/\left(\Gamma(1.5)\sqrt{\pi\omega/\epsilon_i}\right)$. Finally, using (40) along with the variable transformation given in (30) we attain (10). $\qquad \square$

APPENDIX D
PROOF OF THEOREM 3

Using (5) and the fact that $\gamma_1 = \gamma_2$ we have that

$$\frac{\varphi_1^*\beta}{1 + (1-\beta)\varphi_1^*} - \frac{\varphi_2^*(1-\beta)}{1 + \beta\mu\varphi_2^*} = 0$$
$$\varphi_1^*\varphi_2^*(1-\mu)\beta^2 - (\varphi_1^* + \varphi_2^* + 2\varphi_1^*\varphi_2^*)\beta + \varphi_2^*(1+\varphi_1^*) = 0, \quad (41)$$

and (12) comes easy by solving this quadratic equation for β while discarding the solution that leads to $\beta > 1$. The reader can check that β in (12) always lies in $[0, 1]$. However, the other necessary condition for feasibility is that $\beta \geq 1/2$, thus

$$\frac{\varphi_1^* + \varphi_2^* + 2\varphi_1^*\varphi_2^*}{2(1-\mu)\varphi_1^*\varphi_2^*} +$$
$$-\frac{\sqrt{(\varphi_1^* + \varphi_2^*)^2 + 4\varphi_1^*\varphi_2^*(\varphi_1^* + \mu\varphi_2^* + \mu\varphi_1^*\varphi_2^*)}}{2(1-\mu)\varphi_1^*\varphi_2^*} \geq \frac{1}{2}$$
$$\varphi_1^*\varphi_2^* - 2\mu\varphi_1^*\varphi_2^* + \mu^2\varphi_1^*\varphi_2^* - 2\varphi_1^* + 2\mu\varphi_1^* + 2\varphi_2^* - 2\mu\varphi_2^* \geq 0$$
$$\frac{2\varphi_1^*(1-\mu)}{2 - 2\mu + \varphi_1^* - 2\mu\varphi_1^* + \mu^2\varphi_1^*} \leq \varphi_2^*$$
$$\frac{2\varphi_1^*(1-\mu)}{2(1-\mu) + \varphi_1^*(1-\mu)^2} \leq \varphi_2^*$$
$$\frac{2\varphi_1^*}{2 + \varphi_1^*(1-\mu)} \leq \varphi_2^*. \qquad (42)$$

Notice that when μ increases, the required φ_2^* to ensure similar rate performance for both UEs, increases. This is, the worse the SIC performs, the stringent the resource reservation for UE_2. This is also aligned with the fact that β in (12) is a decreasing function of μ.

Continuing with the proof, φ_1^* and φ_2^* are independent since each of them only depends on the distances from the UEs to the transmitting and interfering BSs, path-loss exponent and reliability constraints of the UEs (See (6)). Let the UEs with decoding order 1 and 2, be denoted as A and B, respectively, thus, $\varphi_1^* = \varphi_A^*$ and $\varphi_2^* = \varphi_B^*$. If this assignment already satisfies (42), e.g., $\varphi_B^* \geq \frac{2\varphi_A^*}{2+\varphi_A^*(1-\mu)}$, then, the power allocation profile given in (12) is feasible and both UEs can

attain the same rate performance. In the opposite case we have that

$$\varphi_B^* < \frac{2\varphi_A^*}{2 + \varphi_A^*(1-\mu)}$$
$$\varphi_A^* > \frac{2\varphi_B^*}{2 - \varphi_B^*(1-\mu)} \geq \frac{2\varphi_B^*}{2 + \varphi_B^*(1-\mu)}. \qquad (43)$$

Therefore, if we scheduled the UEs such that $\varphi_1^* = \varphi_B^*$ and $\varphi_2^* = \varphi_A^*$ we can also ensure the same rate performance for both UEs. $\qquad \square$

APPENDIX E
PROOF OF COROLLARY 1

By setting $S = \varphi_1^* + \varphi_2^*$ and $P = \varphi_1^*\varphi_2^*$, which are fixed-value terms and independent of the decoding order, we can write (13) as

$$\gamma_1 = \frac{\sqrt{S^2 + 4Pk} - S}{2k}, \qquad (44)$$

where $k = \varphi_1^* + \mu\varphi_2^* + \mu P$. Now we have that

$$\frac{d\gamma_1}{dk} = \frac{S\sqrt{S^2 + 4Pk} - (S^2 + 2PK)}{2k^2\sqrt{S^2 + 4Pk}} < 0, \qquad (45)$$

which is negative since

$$4P^2k^2 > 0$$
$$4P^2k^2 + S^4 + 4PS^2k > S^4 + 4PS^2k$$
$$(S^2 + 2Pk)^2 > S^2(S^2 + 4Pk)$$
$$S^2 + 2Pk > S\sqrt{S^2 + 4Pk}, \qquad (46)$$

therefore, γ_1 is a decreasing function of k and gets the greater value for the smaller k, which occurs when we choose $\varphi_2^* \geq \varphi_1^*$. Also, notice that with such scheduling the operation is guaranteed since (42) is satisfied $\left(\varphi_2^* \geq \varphi_1^* \geq \frac{2\varphi_1^*}{2+\varphi_1^*(1-\mu)}\right)$. $\square$

APPENDIX F
PROOF OF COROLLARY 1

We proceed as follows

$$\tilde{\gamma} = (1+\gamma_1)(1+\gamma_2) \overset{(a)}{\lessgtr} \left[\frac{(1+\gamma_1) + (1+\gamma_2)}{2}\right]^2$$
$$= \left[1 + \frac{\gamma_1 + \gamma_2}{2}\right]^2, \qquad (47)$$

where (a) comes from using the inequality between arithmetic and geometric means. By using (47) as an approximation and setting $\bar{\gamma} = \gamma_1 + \gamma_2$ we attain (21).

Now let's present some insights on the accuracy of (21). The relative error function

$$\xi(\%) = 100 \times \frac{\left[1 + \frac{1}{2}\bar{\gamma}\right]^2 - \tilde{\gamma}}{\tilde{\gamma}}$$
$$= 25 \times \left[\frac{1+\gamma_1}{1+\gamma_2} + \frac{1+\gamma_2}{1+\gamma_1}\right] - 50, \qquad (48)$$

measures the error when adopting the approximation given in (21), and it is plotted in Fig. 10. As expected, the approximation given in (21) is more accurate when $\gamma_1 \approx \gamma_2$ and/or $\gamma_1, \gamma_2 \ll 1$. The latter condition is typical of scenarios where
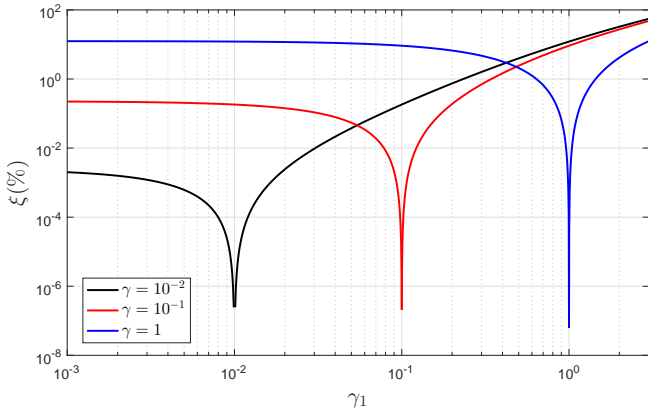
Fig. 10. Relative error function, $\xi(\%)$, as a function of γ_1 with $\gamma_2 \in \{10^{-2}, 10^{-1}, 1\}$.

devices are with low transmission rates because of the short length of their packets and/or stringent reliability constraints, e.g., MTC setups. In those scenarios $\max_{\beta} \tilde{\gamma} \approx \max_{\beta} \bar{\gamma}$. □

APPENDIX G
PROOF OF THEOREM 4

We have that

$$
\begin{aligned}
\bar{\gamma} &= \gamma_1 + \gamma_2, \\
\frac{d\bar{\gamma}}{d\beta} &= \frac{d\gamma_1}{d\beta} + \frac{d\gamma_2}{d\beta} \\
&= \frac{\varphi_1^*(1+\varphi_1^*)}{(1+\varphi_1^*(1-\beta))^2} - \frac{\varphi_2^*(1+\mu\varphi_2^*)}{(1+\mu\beta\varphi_2^*)^2}, \quad (49) \\
\frac{d^2\bar{\gamma}}{d\beta^2} &= \frac{d^2\gamma_1}{d\beta^2} + \frac{d^2\gamma_2}{d\beta^2} \\
&= \frac{2\varphi_1^{*2}(1+\varphi_1^*)}{(1+(1-\beta)\varphi_1^*)^3} + \frac{2\mu\varphi_2^{*2}(1+\mu\varphi_2^*)}{(1+\mu\beta\varphi_2^*)^3}. \quad (50)
\end{aligned}
$$

Since $\frac{d^2\bar{\gamma}}{d\beta^2} > 0$ $\forall\beta$, then $\bar{\gamma}$ is convex on β and a global minimum exists. That means that the maximum lies on the interval extremes, $\beta^* = 1/2$ or $\beta^* = 1$. Thus,

$$
\begin{aligned}
\max_{\beta} \bar{\gamma} &= \max\left(\bar{\gamma}|_{\beta=\frac{1}{2}}, \bar{\gamma}|_{\beta=1}\right) \\
&= \max\left(\frac{\frac{1}{2}\varphi_1^*}{1+\frac{1}{2}\varphi_1^*} + \frac{\frac{1}{2}\varphi_2^*}{1+\frac{1}{2}\mu\varphi_2^*}, \varphi_1^*\right) \\
&= \max\left(\frac{\varphi_1^*}{2+\varphi_1^*} + \frac{\varphi_2^*}{2+\mu\varphi_2^*}, \varphi_1^*\right). \quad (51)
\end{aligned}
$$

Then,

$$
\begin{aligned}
\frac{\varphi_1^*}{2+\varphi_1^*} + \frac{\varphi_2^*}{2+\mu\varphi_2^*} &\overset{\beta^*=1/2}{\underset{\beta^*=1}{\gtrless}} \varphi_1^* \\
\frac{2(\varphi_2^*-\varphi_1^*)+\varphi_1^*(\varphi_2^*-2)}{\varphi_1^*\varphi_2^*(1+\varphi_1^*)} &\overset{\beta^*=1/2}{\underset{\beta^*=1}{\gtrless}} \mu, \quad (52)
\end{aligned}
$$

which implies (22). □

APPENDIX H
PROOF OF THEOREM 5

For the maximum sum-rate setup, the NOMA scheme overcomes the OMA configuration when

$$
\begin{aligned}
\log_2 \tilde{\gamma} &> \frac{1}{2}\log_2(1+\varphi_1^*) + \frac{1}{2}\log_2(1+\varphi_2^*) \\
&= \log_2 \sqrt{(1+\varphi_1^*)(1+\varphi_2^*)} \\
\tilde{\gamma} &> \sqrt{(1+\varphi_1^*)(1+\varphi_2^*)}. \quad (53)
\end{aligned}
$$

We know that $\varphi_2^* \geq \varphi_1^*$, and first we are going to work on that to proof the statement of the theorem for $\mu = 0$. We proceed as follows

$$
\begin{aligned}
\varphi_2^* &\geq \varphi_1^* \quad |\times \varphi_1^*, \quad (\varphi_1^* > 0) \\
\varphi_1^*\varphi_2^* &\geq \varphi_1^{*2} \\
(4+2\varphi_2^*+4\varphi_1^*+\varphi_1^*\varphi_2^*)+\varphi_1^*\varphi_2^* &\geq (4+2\varphi_2^*+4\varphi_1^*+\varphi_1^*\varphi_2^*)+\varphi_1^{*2} \\
2(1+\varphi_1^*)(2+\varphi_2^*) &\geq (2+\varphi_1^*+\varphi_2^*)(2+\varphi_1^*) \\
\frac{(1+\varphi_1^*)(2+\varphi_2^*)}{2+\varphi_1^*} &\geq \frac{2+\varphi_1^*+\varphi_2^*}{2} \\
\frac{(1+\varphi_1^*)(2+\varphi_2^*)}{2+\varphi_1^*} &\geq \frac{(1+\varphi_1^*)+(1+\varphi_2^*)}{2}. \quad (54)
\end{aligned}
$$

Notice that the left term matches the expression for $\tilde{\gamma}|_{\mu=0}$ given in (20) for $\beta = 1/2$ and using the inequality between the arithmetic and geometric means we have that $\frac{(1+\varphi_1^*)+(1+\varphi_2^*)}{2} \geq \sqrt{(1+\varphi_1^*)(1+\varphi_2^*)}$, thus,

$$
\tilde{\gamma}|_{\mu=0} \geq \sqrt{(1+\varphi_1^*)(1+\varphi_2^*)}, \quad (55)
$$

where the right term matches the performance of the OMA setup. Therefore, for $\mu = 0$ the NOMA scheme is the best choice.

For $\mu > 0$ we perform some simplifications to gain in mathematical tractability. We can establish that the NOMA scheme will overcome *almost surely* the OMA configuration if

$$
\begin{aligned}
\left[1 + \frac{1}{2}\bar{\gamma}\right]^2 &> 1 + \frac{\varphi_1^* + \varphi_2^*}{2} \\
\left[1 + \frac{\frac{1}{2}\varphi_1^*}{2+\varphi_1^*} + \frac{\frac{1}{2}\varphi_2^*}{2+\mu\varphi_2^*}\right]^2 &> 1 + \frac{\varphi_1^* + \varphi_2^*}{2}, \quad (56)
\end{aligned}
$$

where the first line comes from using (21) and using the inequality relating the arithmetic and geometric means for the right term ($\sqrt{(1+\varphi_1^*)(1+\varphi_2^*)}$, which is the achievable rate with OMA, is greater than or equal to $1 + \frac{\varphi_1^*+\varphi_2^*}{2}$), and the second line comes from using (23). Now, solving (56) as a function of μ we attain (24). □

REFERENCES

[1] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *IEEE Veh. Tech. Conf. (VTC Spring)*, June 2013, pp. 1–5.

[2] R. Razavi, M. Dianati, and M. A. Imran, "Non-orthogonal multiple access (NOMA) for future radio access," in *5G Mobile Communications*. Springer, 2017, pp. 135–163.

[3] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185–191, February 2017.

[4] Q. Wang, R. Zhang, L. L. Yang, and L. Hanzo, "Non-orthogonal multiple access: A unified perspective," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 10–16, April 2018.

[5] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. s. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surv. and Tut.*, vol. 19, no. 2, pp. 721–742, Secondquarter 2017.

[6] H. Zhang, F. Fang, J. Cheng, K. Long, W. Wang, and V. C. M. Leung, "Energy-efficient resource allocation in NOMA heterogeneous networks," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 48–53, April 2018.

[7] D. Wan, M. Wen, F. Ji, H. Yu, and F. Chen, "Non-orthogonal multiple access for cooperative communications: Challenges, opportunities, and trends," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 109–117, April 2018.

[8] M. S. Ali, E. Hossain, and D. I. Kim, "Coordinated multipoint transmission in downlink multi-cell NOMA systems: Models and spectral efficiency performance," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 24–31, April 2018.

[9] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1501–1505, Dec 2014.

[10] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. on Veh. Tech.*, vol. 65, no. 8, pp. 6010–6023, Aug 2016.

[11] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink nonorthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 458–461, March 2016.

[12] Z. Chen, Z. Ding, X. Dai, and R. Zhang, "An optimization perspective of the superiority of NOMA compared to conventional OMA," *IEEE Trans. on Sig. Process.*, vol. 65, no. 19, pp. 5191–5202, Oct 2017.

[13] K. Kang, Z. Pan, J. Liu, and S. Shimamoto, "A game theory based power control algorithm for future MTC NOMA networks," in *IEEE Annual Cons. Commun. Net. Conf. (CCNC)*, Jan 2017, pp. 203–208.

[14] Y. Xu, C. Shen, T. Chang, S. Lin, Y. Zhao, and G. Zhu, "Energy-efficient non-orthogonal transmission under reliability and finite blocklength constraints," in *2017 IEEE Globecom Workshops (GC Wkshps)*, Dec 2017, pp. 1–6.

[15] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic, "On the performance of non-orthogonal multiple access in short-packet communications," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 590–593, March 2018.

[16] Y. Huang, C. Zhang, J. Wang, Y. Jing, L. Yang, and X. You, "Signal processing for MIMO-NOMA: Present and future challenges," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 32–38, April 2018.

[17] H. Tabassum, E. Hossain, and J. Hossain, "Modeling and analysis of uplink non-orthogonal multiple access in large-scale cellular networks using Poisson cluster processes," *IEEE Trans. on Commun.*, vol. 65, no. 8, pp. 3555–3570, Aug 2017.

[18] O. L. A. López, H. Alves, P. H. J. Nardelli, and M. Latva-aho, "Aggregation and resource scheduling in machine-type communication networks: A stochastic geometry approach," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 7, pp. 4750–4765, July 2018.

[19] ——, "Hybrid resource scheduling for aggregation in massive machine-type communication networks," *Ad Hoc Networks*, vol. 94, p. 101932, 2019.

[20] B. Di, L. Song, Y. Li, and G. Y. Li, "Non-orthogonal multiple access for high-reliable and low-latency V2X communications in 5G systems," *IEEE J. on Selec. Areas in Commun.*, vol. 35, no. 10, pp. 2383–2397, Oct 2017.

[21] Z. Zhang, H. Sun, R. Q. Hu, and Y. Qian, "Stochastic geometry based performance study on 5G non-orthogonal multiple access scheme," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec 2016, pp. 1–6.

[22] Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, and J. A. McCann, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE J. on Selec. Areas in Commun.*, vol. 35, no. 12, pp. 2667–2680, Dec 2017.

[23] M. Salehi, H. Tabassum, and E. Hossain, "Meta distribution of SIR in large-scale uplink and downlink NOMA networks," *IEEE Transactions on Communications*, vol. 67, no. 4, pp. 3009–3025, April 2019.

[24] S. S. Kalamkar and M. Haenggi, "Distributed rate control for high reliability in poisson bipolar networks," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec 2017, pp. 1–6.

[25] O. L. A. López, H. Alves, and M. Latva-Aho, "Rate control under finite blocklength for downlink cellular networks with reliability constraints," in *Int. Symp. on Wireless Commun. Syst. (ISWCS)*, Aug 2018, pp. 1–6.

[26] O. L. A. López, H. Alves, and M. Latva-aho, "Joint power control and rate allocation enabling ultra-reliability and energy efficiency in SIMO wireless networks," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5768–5782, Aug 2019.

[27] M. Haenggi, "On distances in uniformly random networks," *IEEE Tran. on Inf. Theory*, vol. 51, no. 10, pp. 3584–3586, Oct 2005.

[28] H. Sun, B. Xie, R. Q. Hu, and G. Wu, "Non-orthogonal multiple access with SIC error propagation in downlink wireless MIMO networks," in *IEEE Veh. Tech. Conf. (VTC-Fall)*, Sept 2016, pp. 1–5.

[29] J. Abate and W. Whitt, "A unified framework for numerically inverting Laplace transforms," *INFORMS Journal on Computing*, vol. 18, no. 4, pp. 408–421, 2006.

[30] M. Haenggi, "The meta distribution of the SIR in Poisson bipolar and cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 4, pp. 2577–2589, April 2016.

[31] Z. Zhang, H. Sun, and R. Q. Hu, "Downlink and uplink non-orthogonal multiple access in a dense wireless network," *IEEE J. on Selec. Areas in Commun.*, vol. 35, no. 12, pp. 2771–2784, Dec 2017.

[32] F. Liu, P. Mähönen, and M. Petrova, "Proportional fairness-based power allocation and user set selection for downlink NOMA systems," in *IEEE Int. Conf. on Commun. (ICC)*, May 2016, pp. 1–6.

[33] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Tran. on Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[34] R. K. Ganti and M. Haenggi, "SIR asymptotics in poisson cellular networks without fading and with partial fading," in *IEEE Int. Conf. on Commun. (ICC)*, May 2016, pp. 1–5.

[35] X. Zhang and M. Haenggi, "The performance of successive interference cancellation in random wireless networks," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6368–6388, Oct 2014.