Analyzing Non-Orthogonal Multiple Access (NOMA) in Downlink Poisson Cellular Networks

Konpal Shaukat Ali*, Hesham ElSawy*, Anas Chaaban*, Martin Haenggi†, and Mohamed-Slim Alouini*

***Abstract*—Non-orthogonal multiple access (NOMA) is a spectrum reutilization technique that superposes messages in the power domain allowing multiple users to be served in the same time-frequency resource block. Successive interference cancellation (SIC) techniques are used for decoding NOMA messages. A network model is considered where Poisson distributed base stations transmit to N NOMA users each. We present a signal-to-interference-and-noise-ratio analysis for the coverage of the typical user. Due to SIC, coverage implies the ability to decode the messages of all weaker users in the SIC chain. An efficient algorithm for finding a feasible resource allocation that maximizes the cell sum rate $\mathcal{R}_{\rm tot}$ subject to a minimum rate constraint $\mathcal{T}$ on the individual UEs is provided for general N. We show the existence of an optimum N that maximizes $\mathcal{R}_{\rm tot}$ given a set of network parameters. It is also shown that NOMA outperforms orthogonal multiple access if the residual intracell interference is below a certain level. The results highlight the importance in choosing network parameters N and $\mathcal{T}$ to balance $\mathcal{R}_{\rm tot}$ and fairness.**

I. Introduction

The available spectrum is a scarce resource, and many new technologies to be incorporated into 5G focus on reusing the spectrum more efficiently to improve data rates. Traditionally, temporal, spectral, or spatial[1] orthogonalization techniques, referred to as orthogonal multiple access (OMA), are used to avoid interference among users in a cell. They allow only one UE per time-frequency resource block in a cell. A promising candidate for more efficient spectrum reuse in 5G is non-orthogonal multiple access (NOMA), which allows multiple users to share the same time-frequency resource block. This is achieved by having messages multiplexed either in the power domain or in the code domain. NOMA is therefore a special case of superposition coding [1]. Decoding techniques using successive interference cancellation (SIC) [2] for multiple-access channels have been studied from an information theoretic perspective for several decades [3], and they were implemented on a software radio platform in [4]. The focus of our work is on NOMA where messages are superposed in the power domain, i.e., by transmitting them at different power levels. Hence, NOMA allows multiple users to transmit/receive messages in the same time-frequency resource block by multiplexing them at different power levels. SIC techniques are then used for decoding.

Promising results for NOMA as an efficient spectrum reuse technique have been shown [5], [6]. In [7], [8] power allocation (PA) schemes are investigated for universal fairness by achieving identical rates for NOMA users. The idea of cooperative NOMA is investigated [9], [10]. The works in [5]–[10] consider NOMA in single cells and therefore do not account for intercell interference which has a drastic negative impact on NOMA performance as shown in [11]. Stochastic geometry has succeeded in providing a unified mathematical paradigm to model large cellular networks and characterize their operation while accounting for intercell interference [12]–[14]. Using stochastic geometry-based modeling, a large uplink NOMA network is studied in [15], a large downlink NOMA network in [16], and a qualitative study on NOMA in large networks is carried out in [11]. However, [16] does not take into account the SIC chain in the signal to interference and noise ratio (SINR) analysis which overestimates coverage.

To the best of our knowledge, an analytical study of a large multi-cell downlink NOMA system that takes into account intercell interference and intracell interference, henceforth called intraference, and the SIC chain for a general number of UEs served by each BS does not exist. Hence, in this work, we aim at analyzing the performance of such a large network setup, based on stochastic geometry, with both perfect and imperfect SIC capability. Resource allocation (RA) schemes for maximizing rates with constraints are often for small NOMA clusters such as the two-user case [17], [18], though works such as [19] consider a general number of UEs in a NOMA cluster. However, works like [17]–[19] assume a single-cell setup and therefore do not account for intercell interference. Accordingly, we propose intercell interference-aware RA with the goal of maximizing the cell sum rate defined as the sum of the rates of all the UEs in a NOMA cluster of the cell, subject to a threshold minimum rate (TMR) constraint of $\mathcal{T}$ on the individual UEs. OMA is used to benchmark the gains attained by NOMA.

The contributions and findings of this paper can be summarized as follows:

- We study a Poisson cellular network employing NOMA in the downlink and account for intercell interference in the RA, SINR analysis, and optimization.
- We introduce a model where NOMA UEs are distributed uniformly in the largest disk centered at the BS that fits inside the cell. This model is realistic and tractable.
- We present an SINR analysis that considers error propagation in the SIC chain and the effects of imperfect SIC.

*The authors are with the Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Divison, King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia. (Email: {konpal.ali, hesham.elsawy, anas.chaaban, slim.alouini}@kaust.edu.sa)

† The author is with the Department of Electrical Engineering, University of Notre Dame, USA. (Email:mhaenggi@nd.edu)

[1]Spatial separation of UEs with MIMO can be used with either OMA or NOMA.

- We formulate an optimization problem where the cell sum rate is maximized subject to a TMR constraint. Since the problem is non-convex, an efficient algorithm is proposed to find a feasible but possibly suboptimum solution to the problem.
- We show that there exists an optimum NOMA cluster size for a given $\mathcal{T}$ and residual intraference (RI) factor β that maximizes the cell sum rate. We also show that for sufficiently small β, NOMA outperforms OMA. Our results shed light on how to choose the cluster size and $\mathcal{T}$ to balance the cell sum rate and user fairness. In particular, when $\mathcal{T}$ is small, large cluster sizes improve network performance at the expense of lower rates for individual UEs.

Notation: Vectors are denoted using bold text, $\|\mathbf{x}\|$ denotes the Euclidean norm of the vector $\mathbf{x}$. $\mathcal{L}_X(s) = \mathbb{E}[e^{-sX}]$ denotes the Laplace transform (LT) of the PDF of the random variable X. The ordinary hypergeometric function is denoted by ${}_2F_1$.

II. System Model

A. NOMA System Model

We consider a downlink cellular network where BSs transmit with a total power budget of P. The BSs use fixed rate transmissions, which are vulnerable to outage, for the transmission of each message leading to UE rates that are lower than the transmission target rate. Each BS serves N UEs in one time-frequency resource block by multiplexing the signals for each UE with different power levels; here N is called the cluster size. The BSs are distributed according to a homogeneous Poisson point process (PPP) Φ with intensity λ. Denote by ρ the distance between a BS and its nearest neighbor. The N UEs are distributed uniformly and independently in a disk centered at the serving BS with radius $\rho/2$, referred to as the in-disk. The in-disk is the largest disk centered at the serving BS that fits inside the Voronoi cell. UEs outside of this disk are relatively far from their BS and thus are better served in their own resource block (without sharing) or even using CoMP if they are near the cell edge. These UEs will not be discussed further in this work. To the network we add an extra BS at the origin $\mathbf{o}$, which, under expectation over the PPP Φ, becomes the typical BS serving UEs in the typical cell. In this work we study the performance of the typical cell. A realization of the cell at $\mathbf{o}$, its in-disk, and the surrounding cells is shown in Fig. 1. Note that since Φ does not include the BS at $\mathbf{o}$, Φ is the set of the interfering BSs for the UEs in the typical cell. The UEs of the typical cell are indexed with respect to their ascending ordered distance from $\mathbf{o}$; hence, the i^{th} closest UE is referred to as UE$_i$, for $1 \leq i \leq N$. Since the order of the UEs is known at the BS, we use order statistics for the PDF of R_i, the distance of UE$_i$ to its serving BS, given ρ. Accordingly, in the typical cell,

$$f_{R_i|\rho}(r\,|\,\rho)=\binom{N-1}{i-1}\frac{8rN}{\rho^2}\left(\frac{4r^2}{\rho^2}\right)^{i-1}\left(1-\frac{4r^2}{\rho^2}\right)^{N-i}, 0\leq r\leq\frac{\rho}{2},$$

where $\binom{c}{d}=\frac{c!}{d!(c-d)!}$. Note that this model guarantees that the nearest interfering BS from UE$_i$ is farther than $\rho - R_i$.

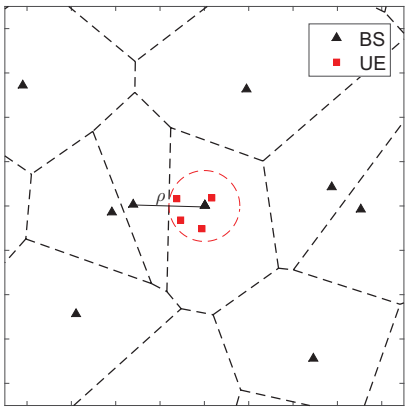

Fig. 1: A realization of the network with $N = 4$. The UEs and in-disk, marked by the dashed circle, for the cell at $\mathbf{o}$ are shown.

The power for the signal intended for UE$_i$ is denoted by P_i, hence $P = \sum_i P_i$. Without loss of generality, we set $P = 1$. A Rayleigh fading environment is assumed such that the fading coefficients follow a unit mean exponential distribution. It is also assumed that the channel gains are independent of one another. A power law path loss model is considered where the signal power decays at the rate $r^{-\eta}$ with distance r, where $\eta > 2$ denotes the path loss exponent and $\delta = \frac{2}{\eta}$.

SIC is employed for decoding at the UEs. According to the NOMA scheme, the PA and transmission target rate are designed such that the i^{th} strongest UE is able to decode the messages intended for all those UEs that have weaker channels than itself. SIC requires ordering UEs according to channel strength, which includes the effect of both path loss (and therefore link distance R_i) and fading. The impact of the large-scale path loss is more stable and dominant than the fading effect which varies on a much shorter time scale. For small values of N in particular, the path loss dominates the channel relative to fading; considering the strength of a channel to be based on the distance between a UE and its BS is therefore a reasonable approximation [15], [20], [21]. The analysis based on the actual instantaneous channel strength is complex and out of the scope of this work. Similarly a more efficient ordering scheme could include intercell interference in determining UE strength. However, including interference and fading would necessitate very high feedback overhead. Hence, we assume in our setup the strength of the channel corresponds to the proximity of a UE to the BS. To successfully decode its own message, UE$_i$ must therefore be able to decode the messages intended for UEs that lie farther from the BS than itself, i.e., UE$_{i+1}, \ldots,$ UE$_N$. This is achieved by allocating higher powers and/or lower target rates to the data streams of UEs farther than UE$_i$. Correspondingly, UE$_i$ is not able to decode any of the streams sent to UEs that are closer to the BSs due to their smaller powers and/or higher target rates. Assuming perfect SIC, the intraference experienced at UE$_i$, I_i°, is comprised of the powers from the messages intended for UE$_1, \ldots,$ UE$_{i-1}$. Since in practice SIC is not perfect, our mathematical model additionally considers a fraction $0 \leq \beta \leq 1$ of RI from the UEs farther than UE$_i$ in I_i°. When perfect SIC is assumed, $\beta = 0$ while $\beta = 1$ corresponds to no SIC at all. Additionally,

UE$_i$ also suffers from intercell interference, $I_i^{\emptyset}$, arising from the power received from all the other BSs in the network, and noise power σ^2. For the NOMA network $2N-1$ parameters are to be selected, namely N target rates and $N-1$ powers. In this work although UE ordering is agnostic to intercell interference and fading, our RA (choice of the $2N-1$ parameters) is not.

B. OMA System Model

We compare our NOMA model with a traditional OMA network where only one UE is served by a BS in a single time-frequency resource block. We focus on time division multiple access (TDMA), i.e., the time resources are split between the contending UEs. For a fair comparison with the NOMA system, the BS serves N UEs at distances R_i which follow the same distribution of the order statistics conditioned on ρ as in the NOMA setup. Each TDMA message is transmitted using full power $P=1$ for a duration T_i. Without loss of generality, a unit time duration is assumed for a NOMA transmission and therefore $\sum_i T_i = 1$. Consequently, $2N-1$ parameters are to be selected for the OMA network, namely N target rates and $N-1$ fractions of the time slot.

III. SINR Analysis

A. NOMA Network

In the NOMA network, the SINR of the j^{th} message at UE$_i$ of the typical cell for $i \le j \le N$ is

$$\mathrm{SINR}_j^i = \frac{h_i R_i^{-\eta} P_j}{\underbrace{h_i R_i^{-\eta}\left(\sum_{m=1}^{j-1} P_m + \beta \sum_{k=j+1}^{N} P_k\right)}_{I_i^{\circ}} + \underbrace{\sum_{\mathbf{x}\in\Phi} g_{\mathbf{y}_i}\|\mathbf{y}_i\|^{-\eta}}_{I_i^{\emptyset}} + \sigma^2},$$

where $\mathbf{y}_i = \mathbf{x} - \mathbf{u}_i$, $\mathbf{u}_i$ is the location of UE$_i$, h_i ($g_{\mathbf{y}_i}$) is the fading coefficient from the serving BS (interfering BS) located at $\mathbf{o}$ ($\mathbf{x}$) to UE$_i$. In order to decode its intended message, UE$_i$ needs to decode the messages intended for all UEs farther from the BS than itself. We use θ_j to denote the SINR threshold corresponding to the target rate associated with the message for UE$_j$. Hence, coverage at UE$_i$ is defined as the event

$$C_i = \bigcap_{j=i}^{N} \left\{\mathrm{SINR}_j^i > \theta_j\right\} = \bigcap_{j=i}^{N} \left\{ h_i > R_i^{\eta}(I_i^{\emptyset}+\sigma^2)\frac{\theta_j}{\tilde{P}_j} \right\}, \quad (1)$$

where

$$\tilde{P}_j = P_j - \theta_j \left(\sum_{m=1}^{j-1} P_m + \beta \sum_{k=j+1}^{N} P_k\right).$$

Hence, the impact of the intraference is that of a reduction in the effective transmit power to $\tilde{P}_j$; without intraference, $\tilde{P}_j$ in (1) would be replaced by P_j. This reduction and thus $\tilde{P}_j$ is dependent on the target-rate of the message to be decoded.

We introduce the notion of *NOMA necessary condition* for coverage, which is coverage when only intraference, arising from NOMA UEs within a cell, is considered. By definition we can write the signal-to-intraference ratio ($\mathring{\mathrm{SIR}}$) of message j at UE$_i$ as

$$\mathring{\mathrm{SIR}}_j^i = \frac{h_i R_i^{-\eta} P_j}{\frac{h_i}{R_i^{\eta}}\left(\sum_{m=1}^{j-1} P_m + \beta \sum_{k=j+1}^{N} P_k\right)} = \frac{P_j}{\sum_{m=1}^{j-1} P_m + \beta \sum_{k=j+1}^{N} P_k}. \quad (2)$$

From (2), the $\mathring{\mathrm{SIR}}$ of message j is independent of the UE (i.e., UE$_i$) it is being decoded at; hence, it can be rewritten as $\mathring{\mathrm{SIR}}_j$. In order for the j^{th} message to satisfy the NOMA necessary condition for coverage, we require

$$\mathring{\mathrm{SIR}}_j > \theta_j \quad \Leftrightarrow \quad \tilde{P}_j > 0. \quad (3)$$

If this condition is not satisfied, the j^{th} message cannot be decoded since $\mathring{\mathrm{SIR}}_j$ can be viewed as an upper bound on SINR_j^i, $j \ge i$. Consequently UE$_i$ will be in SINR-outage as $\tilde{P}_j$ will not be positive. Note that for the particular case of UE$_1$ with perfect SIC (i.e., $\beta = 0$), there is no intraference and $\mathring{\mathrm{SIR}}_1 = \infty$ implying UE$_1$ always satisfies the NOMA necessary condition for coverage when SIC is perfect; equivalently, when $\beta = 0$, $\tilde{P}_1 = P_1$. Hence, failing to satisfy the NOMA necessary condition for coverage guarantees SINR outage for all UEs that need to decode that message simply because the target rate is too high for the given PA. This shows the importance of RA in terms of PA and target rate choice.

Upon taking the expectation over the BS PPP and UE locations, the UEs in the cell with the BS at $\mathbf{o}$ become the typical UEs, from UE$_1$ to UE$_N$.

Lemma 1: The LT of $I_i^{\emptyset}$ at the typical UE$_i$ conditioned on R_i and ρ, where $u = \rho - R_i$, is approximated as

$$\mathcal{L}_{I_i^{\emptyset}|R_i,\rho}(s) \approx \exp\left(-\frac{2\pi\lambda s}{(\eta-2)u^{\eta-2}}\,{}_2F_1\left(1, 1-\delta; 2-\delta; \frac{-s}{u^{\eta}}\right)\right) \times \frac{1}{1+s\rho^{-\eta}} \quad (4)$$

$$\overset{\eta=4}{=} e^{-\pi\lambda\sqrt{s}\tan^{-1}\left(\frac{\sqrt{s}}{u^2}\right)} \frac{1}{1+s\rho^{-4}}. \quad (5)$$

Proof: We can rewrite $I_i^{\emptyset}$ as

$$I_i^{\emptyset} = \sum_{\substack{\mathbf{x}\in\Phi \\ \|\mathbf{x}\|>\rho}} g_{\mathbf{y}_i}\|\mathbf{y}_i\|^{-\eta} + \sum_{\substack{\mathbf{x}\in\Phi \\ \|\mathbf{x}\|=\rho}} g_{\mathbf{y}_i}\|\mathbf{y}_i\|^{-\eta}. \quad (6)$$

The first term of the LT accounts for the first term in (6) corresponding to the non-nearest interferers from $\mathbf{o}$ lying at a distance at least u away from UE$_i$. It is obtained from employing Slivnyak's theorem, the probability generating functional of the PPP, and MGF of $g_{\mathbf{y}_i} \sim \exp(1)$. However, this does not include the BS at distance ρ from $\mathbf{o}$; which is accounted for by the second term in (6) using the MGF of $g_{\mathbf{y}_i}$. Denote by z the distance between this interferer and the typical UE$_i$. The exact expression is $\mathbb{E}_z\left[(1+sz^{-\eta})^{-1}\right]$. For simplicity we approximate it using the approximate mean of this distance. Since the average position of the typical UE$_i$ is $\mathbf{o}$, $\mathbb{E}[z] \approx \rho$. ■

Note: The first term of the LT is pessimistic since the interference guard zone in our model u is smaller than the

actual one. For the second term, an exact evaluation shows that the difference between $\mathbb{E}[z]$ and ρ is less than 3.2%.

Theorem 1: If $\tilde{P}_j > 0$, the coverage probability of the typical UE_i is approximated as

$$\mathbb{P}(C_i) \approx \int_0^{\infty} \int_0^{x/2} e^{-r^{\eta}\sigma^2 M_i} \mathcal{L}_{I_i^{\varnothing}|R_i,\rho}(r^{\eta}M_i) f_{R_i|\rho}(r|x) dr\, f_{\rho}(x) dx, \quad (7)$$

where $M_i = \max\limits_{i \le j \le N} \frac{\theta_j}{\tilde{P}_j}$, $f_{\rho}(x) = 2\pi\lambda x e^{-\pi\lambda x^2}$, $x \ge 0$, and the LT of $I_i^{\varnothing}$ conditioned on R_i and ρ is approximated in (4).

Proof: C_i can be rewritten as $h_i > R_i^{\eta}(I_i^{\varnothing} + \sigma^2)M_i$. Hence,

$$\mathbb{P}(C_i) \overset{(a)}{=} \mathbb{E}_{\rho}\left[\mathbb{E}_{R_i}\left[e^{-R_i^{\eta}\sigma^2 M_i}\mathbb{E}\left[e^{-\left(R_i^{\eta}M_i\right)I_i^{\varnothing}} \mid R_i, \rho\right]\right]\right],$$

where (a) follows from $h_i \sim \exp(1)$. The inner expectation is the conditional LT of $I_i^{\varnothing}$ (given R_i and ρ). The PDF of ρ is due to the BSs forming a PPP. From this we obtain (7). ∎

For a given SINR threshold θ_i, corresponding to a target (normalized) rate of $\log(1+\theta_i)$, the rate of the typical UE_i is

$$\mathcal{R}_i = \mathbb{P}(C_i)\log(1+\theta_i). \quad (8)$$

The cell sum rate is $\mathcal{R}_{\text{tot}} = \sum\limits_{i=1}^{N} \mathcal{R}_i$.

B. OMA Network

The SINR for UE_i of the typical cell is

$$\text{SINR}_{\text{OMA}}^{i} = \frac{h_i R_i^{-\eta}}{\underbrace{\sum\limits_{\mathbf{x}\in\Phi} g_{\mathbf{y}_i}\|\mathbf{x}-\mathbf{u}_i\|^{-\eta}}_{I_i^{\varnothing}} + \sigma^2}.$$

where $\mathbf{u}_i$ is the location of UE_i, h_i ($g_{\mathbf{y}_i}$) is the fading coefficient from the serving BS (interfering BS) located at $\mathbf{o}$ ($\mathbf{x}$) to UE_i. Coverage at UE_i is defined as $\tilde{C}_i = \left\{\text{SINR}_{\text{OMA}}^{i} > \theta_i\right\}$.

Lemma 2: In the OMA network, the coverage probability of the typical UE_i is approximated as

$$\mathbb{P}(\tilde{C}_i) \approx \int_0^{\infty} \int_0^{x/2} e^{-\theta_i r^{\eta}\sigma^2} \mathcal{L}_{I_i^{\varnothing}|R_i,\rho}(\theta_i r^{\eta}) f_{R_i|\rho}(r|x) dr\, f_{\rho}(x) dx, \quad (9)$$

where $\mathcal{L}_{I_i^{\varnothing}|R_i,\rho}(s)$ and $f_{\rho}(x)$ are given in Lemma 1.

Proof: Using the exponential distribution of h_i and the LT of $I_i^{\varnothing}$ conditioned on R_i and ρ we obtain (9). ∎

For a given SINR threshold θ_i and corresponding target (normalized) rate $\log(1+\theta_i)$, the rate of the typical UE_i is

$$\mathcal{R}_i = T_i\, \mathbb{P}(\tilde{C}_i)\log(1+\theta_i), \quad (10)$$

where T_i is the fraction of the time slot allotted to UE_i.

IV. NOMA Optimization

A. Problem Formulation

The objective of NOMA is to provide coverage to multiple UEs in the same time-frequency resource block. Naturally we are interested in maximizing the cell sum rate to cater to the ever growing demands. It is well known that the cell sum rate is maximized by allocating all resources (power in the NOMA network) to the best UE [22]. However, this comes at the price of losing the essence of NOMA, which is to cater to multiple UEs in this time-frequency resource block. Hence, we constrain the objective of maximizing cell sum rate to ensure multiple UEs are served. Specifically, the goal is to maximize the cell sum rate given that the typical UEs are able to achieve at least a TMR $\mathcal{T}$. Formally this is stated as follows:

- Maximum cell sum rate $\mathcal{R}_{\text{tot}}$ subject to the TMR $\mathcal{T}$:

$$\max_{(P_i,\theta_i)_{i=1,\ldots,N}} \quad \mathcal{R}_{\text{tot}}$$
$$\text{subject to} \quad \sum_{i=1}^{N} P_i = 1 \text{ and } \mathcal{R}_i \ge \mathcal{T}.$$

Because this problem is non-convex, an optimum solution cannot be found using standard methods. However, from the rate-region for static channels we know that a RA that results in all UEs achieving the TMR $\mathcal{T}$, while all of the remaining power being allocated to the nearest UE, i.e., UE_1, to maximize its rate is the optimum solution for that problem. An example of this for the two-user case is presented in [17].

The same objective holds for OMA networks. The constrained resource allocated to the UEs, however, is time for TDMA instead of power for NOMA, i.e., $\sum_i T_i = 1$. The OMA UEs enjoy full power in their transmissions. Optimization over target rate is done similarly to NOMA.

B. Efficient Algorithm

Since standard techniques cannot be employed for general N, the optimum solution to the problem in Section IV-A can only be found exhaustively by searching over all combinations of power and target rate for each of the N NOMA UEs. This, however, is an extremely tedious approach, particularly as N increases. In this subsection we propose an efficient algorithm based on intuition which, while not guaranteeing an optimum solution, finds a feasible solution, i.e., a solution that satisfies the constraints (but there is no guarantee that the cell sum rate is close to the global maximum).

Since UE_1 has a better channel on average than the other UEs, given a certain power, it is able to achieve the largest gains from this resource. It therefore makes sense to solve the problem in Section IV-A by first ensuring that all non-nearest UEs achieve TMR with the smallest powers possible. This will leave the largest P_1 for UE_1. UE_1 can then maximize the cell sum rate by maximizing $\mathcal{R}_1$ with this power by finding the appropriate target-rate.

We tackle this problem by decoupling the choice of power and target rate; our algorithm finds the minimum possible

power and corresponding smallest target rate[2] that achieve $\mathcal{T}$ for UE_2 to UE_N and allocates the remaining power to UE_1. We start with UE_N and work backwards until UE_2, as SIC only requires a UE to decode the messages of UEs weaker than itself. UE_1 then optimizes its target rate (and therefore θ_1) with the remaining power to maximize its rate. It should be noted that if a UE cannot attain $\mathcal{T}$, the available power is insufficient and the algorithm is unable to find a feasible solution as the cluster size (N) is too large to attain this TMR for all UEs. This can be remedied by either decreasing $\mathcal{T}$ or the number of NOMA UEs N. Formally, we state the working of the algorithm as follows:

Algorithm 1 Efficient algorithm for RA of a feasible solution

```
Begin with UE_N, i = N, P = [ ], θ = [ ]
while i > 0 do
  if i > 1 then
    if R_i ≥ T using 0 ≤ P_i ≤ 1 − Σ_{k=i+1}^{N} P_k then
      Find minimum P_i and corresponding minimum θ_i
      able to achieve R_i = T according to (7) and (8)
      Update: P = [P_i; P]; θ = [θ_i; θ]; i = i − 1
    else
      TMR cannot be met for all UEs; exit
    end if
  else
    P_1 = 1 − Σ_{k=2}^{N} P_k
    if P_1 > 0 then
      Find θ_1 to maximize R_1 using (7) and (8)
      if R_1 ≥ T then
        Update: i = i − 1
      else
        TMR cannot be met for all UEs; exit
      end if
    else
      TMR cannot be met for all UEs; exit
    end if
  end if
end while
```

The same algorithm is employed for OMA, except that rates are calculated using (9) and (10), and the contending resource is T instead of P. Note that since our problem includes intercell interference, our RA is intercell interference-aware.

V. Results

We first show that the approximation in Theorem 1 is tight. To this end, we simulate using BS intensity $\lambda = 10$ BSs/km^2, noise power $\sigma^2 = -90$ dB and $\eta = 4$. The simulations are repeated 10,000 times. The results are shown in Fig. 2 which considers a system with $N = 3$ and a fixed PA scheme where $P_1 = 1/6$, $P_2 = 1/3$ and $P_3 = 1/2$. For clarity of presentation we choose the same target rate for all three UEs and plot coverage of each UE against the corresponding SINR threshold. The figure verifies the accuracy of our SINR analysis as the coverage expressions match the simulation closely. As expected, increasing β decreases the coverage for any UE except UE_N since it does not have any RI.

[2]For an $i \in \{1, \ldots, N\}$, the function $\mathcal{R}_i(\theta_i)$ is monotonically increasing from zero and then monotonically decreasing to zero, with a unique maximum at a finite $\theta_i > 0$. This is because for small θ_i, $\mathbb{P}(C_i)$ is close to 1, hence $\mathcal{R}_i$ increases linearly with $\log(1+\theta_i)$, while for large θ_i, $\mathbb{P}(C_i)$ goes to zero more quickly than $\log(1+\theta_i)$ grows. Hence, each $\mathcal{R}_i$ (except the maximum) can be satisfied by two θ_i values. Unless TMR occurs at the optimum θ_i, we select the smaller value since it increases the coverage probability for all UEs that are to decode the i^{th} message.

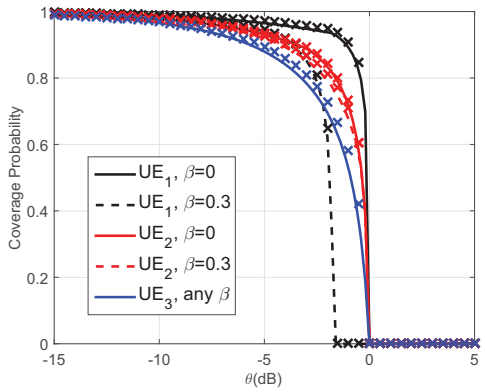

Fig. 2: SINR Coverage vs. θ (identical target rate for all UEs) for $N = 3$ employing a fixed PA $P_1 = 1/6$, $P_2 = 1/3$ and $P_3 = 1/2$. Lines show the analysis and markers show the Monte Carlo simulations.

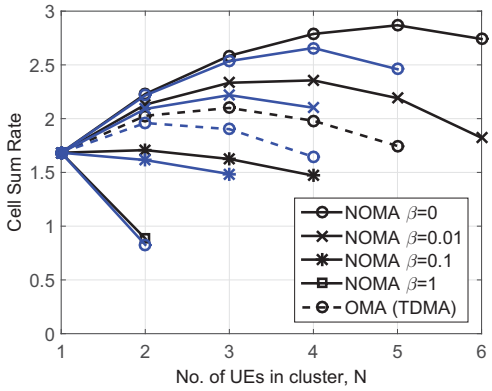

Fig. 3: $\mathcal{R}_{\text{tot}}$ vs. N for different TMR constraints using OMA and NOMA with different β values. Black lines are for $\mathcal{T} = 0.3$, while blue lines are for $\mathcal{T} = 0.4$. The curves end at the largest N that can be supported given the TMR constraint and β.

RA for the remaining figures is done according to Algorithm 1. Fig. 3 is a plot of the cell sum rate against the number of UEs in a NOMA cluster, N. We have included $N = 1$ in these plots which has the same $\mathcal{R}_{\text{tot}}$ for all the curves since it only has one UE in a resource block ($\because$ independent of β) which maximizes its rate ($\because$ independent of $\mathcal{T}$). Given $\mathcal{T}$ and β, there exists an optimum N that maximizes $\mathcal{R}_{\text{tot}}$. When β is large we observe that using NOMA may not necessarily be more beneficial in terms of $\mathcal{R}_{\text{tot}}$. Otherwise, for small N, increasing N enhances $\mathcal{R}_{\text{tot}}$ because interference cancellation is efficient in this regime, and more users are covered. Also, increasing N decreases R_1 on average which enhances $\mathcal{R}_1$ given a P_1. This in turn enhances $\mathcal{R}_{\text{tot}}$ which receives the largest contribution from $\mathcal{R}_1$. However, increasing N beyond the optimum leaves

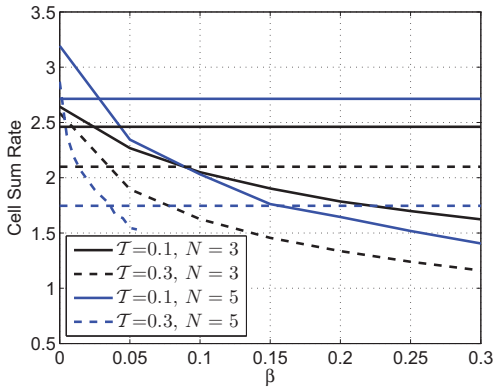

Fig. 4: $\mathcal{R}_{\text{tot}}$ vs. β for different values of N and $\mathcal{T}$. Curves represent NOMA and horizontal lines (independent of β) represent OMA.

too little power for UE_1 to boost $\mathcal{R}_1$ with. For a given $\mathcal{T}$ and β, the resources are only sufficient to support a maximum cluster size; after this N (discontinuation of the plots), not all of the UEs are able to achieve $\mathcal{T}$. Increasing $\mathcal{T}$ or β results in a decrease in the maximum cluster size that can be supported. NOMA outperforms OMA significantly if β is small and can support the same number of UEs or more.

Fig. 4 plots the cell sum rate against β for different N and $\mathcal{T}$. Since OMA does not use SIC, the corresponding $\mathcal{R}_{\text{tot}}$ plots are horizontal lines independent of β. The decrease of $\mathcal{R}_{\text{tot}}$ as a function of β is steeper for larger N and $\mathcal{T}$ highlighting their increased susceptibility to RI. The figure shows the existence of a maximum β value until which a NOMA system with a particular N and $\mathcal{T}$ is able to outperform the corresponding OMA system.

From these results we observe that using larger N and small $\mathcal{T}$ when β is low achieves higher $\mathcal{R}_{\text{tot}}$ than smaller N. However, smaller N, which can support larger $\mathcal{T}$, can improve individual UE rate at the expense of $\mathcal{R}_{\text{tot}}$.

VI. Conclusion

In this paper a large cellular network that employs NOMA in the downlink is studied. Our SINR analysis considers coverage at a UE as the event of being able to decode all messages before the UE in the SIC chain. It also takes into account RI from imperfections in SIC. We focus on the non-convex problem of maximizing the cell sum rate $\mathcal{R}_{\text{tot}}$ subject to a TMR constraint $\mathcal{T}$. Since the optimum solution for RA requires an exhaustive search, an efficient algorithm for a general NOMA cluster size N that gives a feasible solution for intercell interference-aware PA and choice of target rate is proposed. We show the existence of an optimum NOMA cluster size that maximizes $\mathcal{R}_{\text{tot}}$ given a $\mathcal{T}$ and RI factor β. It is also shown that NOMA outperforms OMA provided β is below a certain value that is a function of N and $\mathcal{T}$. The results highlight the importance and impact of choosing network parameters such as N and $\mathcal{T}$ depending on the network objective. In particular, using small $\mathcal{T}$ with large N improves $\mathcal{R}_{\text{tot}}$ at the expense of lower individual UE performance; however increasing $\mathcal{T}$ and reducing N will improve individual UE performance at the expense of $\mathcal{R}_{\text{tot}}$.

References

[1] T. M. Cover and J. A. Thomas, "Elements of information theory," *NJ: John Wiley*, 2006.

[2] D. Tse and P. Viswanath, "Fundamentals of wireless communication," *Cambridge University Press*, 2004.

[3] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE J. Selec. Areas Commun.*, vol. 12, no. 5, pp. 796–807, Jun. 1994.

[4] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2628–2639, Jul. 2012.

[5] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. of IEEE 77th Vehicular Technology Conference (VTC Spring 2013)*, Jun. 2013, pp. 1–5.

[6] 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 Proc. Letters*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

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

[8] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Comm. Letters*, vol. 20, no. 10, pp. 2055–2058, Oct. 2016.

[9] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Comm. Letters*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[10] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Select. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[11] K. S. Ali, H. Elsawy, A. Chaaban, and M. S. Alouini, "Non-orthogonal multiple access for large-scale 5G networks: Interference aware design," *IEEE Access*, vol. 5, pp. 21 204–21 216, 2017.

[12] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[13] H. ElSawy, A. Sultan-Salem, M. S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Commun. Surveys and Tutorials*, vol. 19, no. 1, pp. 167–203, 2017.

[14] W. Lu and M. D. Renzo, "Stochastic geometry modeling of cellular networks: Analysis, simulation and experimental validation," *CoRR*, vol. abs/1506.03857, 2015. [Online]. Available: http://arxiv.org/abs/1506.03857

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

[16] Z. Zhang, H. Sun, R. Q. Hu, and Y. Qian, "Stochastic geometry based performance study on 5G non-orthogonal multiple access scheme," in *Proc. of IEEE Global Communications Conference (GLOBECOM16)*, Dec. 2016, pp. 1–6.

[17] C. L. Wang, J. Y. Chen, and Y. J. Chen, "Power allocation for a downlink non-orthogonal multiple access system," *IEEE Wireless Comm. Letters*, vol. 5, no. 5, pp. 532–535, Oct. 2016.

[18] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[19] 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. Selec. Areas Commun.*, vol. PP, no. 99, pp. 1–1, 2017.

[20] G. Geraci, M. Wildemeersch, and T. Q. S. Quek, "Energy efficiency of distributed signal processing in wireless networks: A cross-layer analysis," *IEEE Trans. Signal Proc.*, vol. 64, no. 4, pp. 1034–1047, Feb. 2016.

[21] M. Wildemeersch, T. Q. S. Quek, M. Kountouris, A. Rabbachin, and C. H. Slump, "Successive interference cancellation in heterogeneous networks," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4440–4453, Dec 2014.

[22] J. Choi, "NOMA: principles and recent results," *CoRR*, vol. abs/1706.08805, 2017. [Online]. Available: http://arxiv.org/abs/1706.08805