Nonorthogonal Random Access for 5G Mobile Communication Systems

Jun-Bae Seo, *Member, IEEE*,
Bang Chul Jung, *Senior Member, IEEE*,
and Hu Jin, *Member, IEEE*

***Abstract*—This correspondence paper proposes two nonorthogonal *random* access (NORA) techniques for 5G mobile communication networks, where user equipments (UEs) make use of the *channel inversion* technique such that their received power at the base station (BS) can be one of the two target values. It enables the BS to decode two packets simultaneously with the successive interference cancelation (SIC) technique if a different power level is chosen. We propose two NORA systems; that is, UEs choose one of the two target power levels based on the channel gain or the region where they are. The performance of the proposed systems is analyzed in terms of access delay, throughput, and energy efficiency. Through analysis and extensive computer simulations, we show that the maximum throughput of the proposed NORA techniques can exceed 0.7, which is a significant improvement compared to the maximum throughput of conventional random access 0.368.**

***Index Terms*—5G mobile communications, uplink NOMA, random access, successive interference cancellation.**

I. Introduction

To improve the spectral efficiency (SE) for the 5th generation (5G) mobile communication systems, non-orthogonal multiple access (NOMA) has been proposed [1], in which the receiver separates the super-imposed signals via successive interference cancellation (SIC) technique. When it is used for the downlink, a base station (BS) constructs the super-imposed signal for a group of users in the same radio resource and allocates different transmission powers to each user equipment (UE). The multiplexed signal experiences the same (small-scale) fading and path-loss *collectively* over the downlink, and then it can be successfully separated by each UE with SIC technique if the BS properly allocates different levels of powers to the UEs. For the uplink, in contrast, the BS receives the super-imposed signals from different UEs, each of which may experience *independent* fading and path-loss due to their different locations.

Manuscript received October 13, 2017; revised February 22, 2018; accepted April 4, 2018. Date of publication April 12, 2018; date of current version August 13, 2018. This work was supported in part by Grant RP03514G and in part by the National Research Foundation of Korea Grant funded by the Korea Government (MSIT) under Grant NRF-2018R1C1B6008126 and Grant NRF-2016R1A2B4014834. The review of this paper was coordinated by Dr. L. Dai. *(Corresponding author: Hu Jin.)*

J.-B. Seo is with the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India (e-mail: jbseo@iitd.ac.in).

B. C. Jung is with the Department of Electronics Engineering, Chungnam National University, Daejeon 34134, South Korea (e-mail: bcjung@cnu.ac.kr).

H. Jin is with the Division of Electrical Engineering, Hanyang University, Ansan 15588, South Korea (e-mail: hjin@hanyang.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2825462

As prior work for the uplink NOMA, the outage probability and sum-rate were investigated [2], [3], in which UEs utilize the transmit power control (TPC) technique to compensate the path-loss, but only two [2] or three UEs [3] are considered. The uplink NOMA was also analyzed when UEs are deployed based on Poisson point process [4] or a clustered point process [5], respectively. In [2]–[5], dynamics of users' retransmissions and *channel inversion* [6] were not considered.

In [7], [8], the SIC technique was integrated with *splitting algorithm* for contention resolution. In particular, [8] proposed a dual power multiple access (DPMA) system, where UEs exploit channel inversion; that is, UEs transmit their packet so that their received power can be one of two power levels and the BS decodes the received packets with the SIC.

In contrast with [2], [3], this paper considers a *random access* network with M UEs, where the BS adopts the SIC technique to separate the received packets from multiple UEs in power domain, called *non-orthogonal random access (NORA)*. As in DPMA [8], in our proposed technique, UEs utilize the channel inversion as well. However, the proposed NORA fundamentally is different: DPMA allows UEs to target at one of two power levels randomly, while our scheme asks UEs to opportunistically choose their target power level based on their channel gains which may further improve the energy efficiency (EE).

As main contribution, we propose two NORA systems and analyze their throughput (packets/slot), the average access delay, and EE (packets/slot/joule), where Rayleigh fading and path-loss are taken into account.

II. System Model

Suppose an *uplink* time division duplex (TDD) wireless network, where a BS is at the center of a circular coverage area with radius R and M UEs share the wireless uplink. Time is divided into slots of a constant size; each slot is equal to one packet transmission time. We assume that each UE can hold only one packet and that at each slot the probability of a new packet arrival to UE is σ. Once a UE has a packet to send, it measures its channel gain $Y = hr^{-\alpha}$ through the downlink reference signal with channel reciprocity of TDD. Here, h indicate a short-term fading, which is assumed to be exponentially distributed random variable with unit mean, whereas r and α denote the distance from the BS to the UE, and the path-loss exponent, respectively.

In NORA systems, UE makes the received power at the BS either P_1 or P_2 for $P_1 > P_2$ by utilizing channel inversion. If $P_{T,i}$ denotes the transmission power of the UE which targets at P_i, we have $P_{T,i} = \frac{P_i}{hr^{-\alpha}} = \frac{P_i}{Y}$ for $i \in \{1, 2\}$. With the target received powers P_1 and P_2, we consider the SIC-based receiver at the BS. It is always successful for the BS to decode one packet, if it receives only one packet with P_i. When the BS receives more than one packets including the one with P_1, it can decode the packet with P_1 successfully, if

$$\frac{P_1}{kP_2 + N_0} \geq \gamma \quad \text{for} \quad k = 0, 1, \ldots, \tag{1}$$

where N_0 and γ denote noise power and SINR threshold for the successful decoding, respectively. In (1), k indicates the number of received packets with P_2. We assume that depending on target power P_i, the reference signal for the BS to perform channel estimation is differently located, which also facilitates the BS's detecting and decoding the packets received. Moreover, the channel estimation at the BS is assumed to be perfect. Let k^* be the maximum number of the packets with P_2 so that the packet with P_1 is successfully decoded. Using (1), we can get $k^* \geq 1$ as

$$k^* = \lfloor (P_1/\gamma - N_0)/P_2 \rfloor = \lfloor (\mathsf{SNR}_1 - \gamma)/(\gamma \mathsf{SNR}_2) \rfloor, \quad (2)$$

where $\mathsf{SNR}_i \triangleq P_i/N_0$ for $i = 1$ and 2. For the packet with P_2 to be decoded successfully, we have two cases. As mentioned before, the BS receives a single packet with P_2, where $\mathsf{SNR}_2 \geq \gamma$ is satisfied. The second one is when two packets are received at the BS, each of which targets at P_1 and P_2, respectively, i.e., $k = 1$ in (1). After the packet with P_1 is decoded, it is removed with the SIC technique and the packet with P_2 is also successfully decoded provided that $\mathsf{SNR}_2 \geq \gamma$. Note that if there exist more than one packets with P_1, no packets can be decoded. We assume that the BS notifies the outcome of a packet transmission to the UE just before the beginning of the next slot. The UEs that do not have the feedback shall regard themselves as backlogged. It is important to note that instead of allowing UEs to target at P_1 or P_2 randomly and independently, a higher throughput is expected if the system increases a joint probability (or correlation) that one UE targets at P_1 and the other at P_2 in a *distributed and energy-efficient* way. To do this, we consider two systems, say NORA-A and -B: In NORA-A, if the UE finds $Y \geq \beta_1$, it sends its packet with probability μ by adjusting its transmission power so that the received power at the BS is equal to P_1. If $\beta_2 \leq Y < \beta_1$, it transmits its packet with probability μ as well, but the BS receives the packet with power P_2. The UE does not (re)transmit the packet if $Y < \beta_2$, where β_2 is called *outage threshold*. On the other hand, in NORA-B, the coverage area is divided into two regions: One is a small circular region with radius r_c, say region 1, and the other is the region with r for $r_c < r \leq R$, say region 2. In the two regions, we assume that M_1 and M_2 UEs are randomly deployed in each region, respectively. At each slot, a new packet is generated at UEs in region 1 and 2 with probability σ_1 and σ_2, respectively. Unlike NORA-A, UEs in NORA-B have a specific threshold β_i for region i. If the UE in region i finds $Y \geq \beta_i$ for $i = 1, 2$, it transmits the packet at the next slot with probability μ_i by adjusting its transmission power such that the received power at the BS is equal to P_i.

III. Analysis and Design

Before analyzing the throughput, average access delay, and EE of NORA-A and -B, we examine the probability that a channel gain exceeds the threshold in both systems.

Lemma 1: In NORA-A, the probability that a UE finds $Y \geq y$ for $\alpha = 4$, i.e., a typical value of pathloss exponents for urban area, is expressed as

$$\Pr[Y \geq y] = 1 - F_Y(y) = \frac{1}{2R^2}\sqrt{\frac{\pi}{y}}\mathrm{erf}\left(\sqrt{y}R^2\right), \quad (3)$$

where $F_Y(y)$ is the cumulative distribution function (CDF) of Y, and $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$.

Proof. Since $F_Y(y) = \Pr[Y \leq y]$, we can write it as

$$F_Y(y) = \Pr\left[hr^{-\alpha} \leq y\right] = \int_0^R \left(1 - e^{-yr^\alpha}\right) f_R(r)dr$$
$$= 1 - \int_0^R e^{-yr^\alpha} f_R(r)dr, \quad (4)$$

where $f_R(r)$ is the probability density function (pdf) of a UE being randomly in the coverage area of NORA-A, which is $f_R(r) = \frac{2r}{R^2}$ for $0 \leq r \leq R$. For $\alpha = 4$, we have $F_Y(y) = 1 - \frac{2}{R^2}\int_0^R e^{-yr^4} r dr = 1 - \frac{1}{2R^2}\sqrt{\frac{\pi}{y}}\mathrm{erf}\left(\sqrt{y}R^2\right)$. We then have (3) using $\Pr[Y \geq y] = 1 - F_Y(y)$. ■

Corollary 1: For NORA-B, let $q_1(\beta_1)$ denote the probability that a UE in the region 1 transmits its packet by targetting at P_1 if finding its channel gain $Y \geq \beta_1$. For $\alpha = 4$, we have

$$q_1(\beta_1) = \Pr[Y \geq \beta_1] = \frac{1}{2r_c^2}\sqrt{\frac{\pi}{\beta_1}}\mathrm{erf}\left(\sqrt{\beta_1}r_c^2\right). \quad (5)$$

Proof. With slight abuse of notation, let $F_{Y_1}(y)$ be the CDF of Y in the region 1, whereas $f_{R_1}(r)$ denotes the pdf of a UE in the region 1, i.e., $f_{R_1}(r) = \frac{2r}{r_c^2}$ for $0 \leq r \leq r_c$. We can get (5) by replacing $f_R(r)$ with $f_{R_1}(r)$ in (4). ■

Corollary 2: Let $q_2(\beta_2)$ be the probability that a UE in the region 2 transmits its packet by targetting at P_2 when $Y \geq \beta_2$. For $\alpha = 4$, it is obtained as

$$q_2(\beta_2) = \frac{\sqrt{\pi}\left[\mathrm{erf}(\sqrt{\beta_2}R^2) - \mathrm{erf}(\sqrt{\beta_2}r_c^2)\right]}{2(R^2 - r_c^2)\sqrt{\beta_2}}. \quad (6)$$

Proof. Let $F_{Y_2}(y)$ be the CDF of Y of a UE in the region 2. In (4), by replacing $f_{R_1}(r)$ with $f_{R_2}(r) = 2r/(R^2 - r_c^2)$ for $r_c \leq r \leq R$, we can get $F_{Y_2}(y)$ and $q_2(\beta_2) = 1 - F_{Y_2}(\beta_2)$ ■

Now, let us characterize the performance of NORA-A. Note that the number of backlogged UEs n changes at each slot boundary, the system state can be captured by a discrete-time Markov chain. To do this, let $\boldsymbol{\phi} = [\phi_n]$ for $n \in \{0, 1, \ldots, M\}$ be the (steady) state probability row vector of length $1 + M$, where ϕ_n denotes the probability that the system has n backlogged UEs in steady state. Furthermore, S denotes the state transition probability matrix whose element $s_{n,m}$ is the state transition probability that state n at time t changes into m at time $t+1$. Then, based on theory of discrete-time Markov process, we can get $\boldsymbol{\phi}$ as $\boldsymbol{\phi} = \boldsymbol{\phi} \cdot S$ and $\boldsymbol{\phi} \cdot \boldsymbol{e} = 1$, where $\boldsymbol{e}$ is a column vector of all ones, whose length corresponds to $\boldsymbol{\phi}$. We can find first $s_{0,m} = \mathcal{B}_m^M(\sigma)$, where a binomial distribution, $\mathcal{B}_m^n(x) = \binom{n}{m}x^m(1-x)^{n-m}$, but it is zero if $m < 0$ or $m > n$. For $m \geq \max(0, n-2)$ and $n \geq 1$, we have

$$s_{n,m} = \mathcal{B}_1^2(\vartheta)\mathcal{B}_2^n(p_o)\mathcal{B}_{m-(n-2)}^{M-n}(\sigma) + \left[\sum_{k=3}^{k^*+1}\mathcal{B}_1^k(\vartheta)\mathcal{B}_k^n(p_o) + \mathcal{B}_1^n(p_o)\right]$$
$$\times \mathcal{B}_{m-(n-1)}^{M-n}(\sigma) + \mathcal{B}_{m-n}^{M-n}(\sigma)\left[\mathcal{B}_0^n(p_o) + \sum_{k=k^*+2}^{n}\mathcal{B}_k^n(p_o)\right.$$
$$\left. + \sum_{k=3}^{k^*+1}\left(1 - \mathcal{B}_1^k(\vartheta)\right)\mathcal{B}_k^n(p_o) + (1 - \mathcal{B}_1^2(\vartheta))\mathcal{B}_2^n(p_o)\right], \quad (7)$$

where p_o is the probability that a backlogged UE (re)transmits its packet. Using Lemma 1, we can get $p_o = \mu \Pr[Y \geq \beta_2] = \frac{\mu}{2R^2}\sqrt{\frac{\pi}{\beta_2}}\mathrm{erf}\left(\sqrt{\beta_2}R^2\right)$. In (7), ϑ indicates the probability that the UE with $Y \geq \beta_2$ will (re)transmit its packet by targeting at P_1, i.e., $\vartheta = \frac{r \Pr[Y \geq \beta_1]}{p_o}$. (7) can be read as follows: For instance, the first term shows that the transition from n backlogged UEs (i.e., $M - n$ nonbacklogged UEs) to m occurs in the system, if two out of n backlogged UEs transmit successfully with probability $\mathcal{B}_1^2(\vartheta)\mathcal{B}_2^n(p_o)$, whereas $m - (n-2)$ out of $M - n$ UEs join the backlogged UEs newly with probability $\mathcal{B}_{m-(n-2)}^{M-n}(\sigma)$. The first term in the first bracket indicates that the packet with P_1 can be successfully decoded as long as the number of packets targetting at P_2 is less than $k^\star$. The second term in the bracket indicates that one pakcet successfully is received with either P_1 or P_2 if only one is transmitted. Due to lack of space, we explain the last term in (7), which means that two UEs access with probability $\mathcal{B}_2^n(p_o)$ and both of them choose either P_1 or P_2 with probability $1-\mathcal{B}_1^2(\vartheta)$. If $m-n$ out of $M-n$ UEs join, the system has m backlogged UEs at the next time.

To get the throughput (packets/slot) of NORA-A, let τ_a denote the throughput, which can be obtained as

$$\tau_a = \sum_{n=1}^{M}\left(2\mathcal{B}_1^2(\vartheta)\mathcal{B}_2^n(p_o) + \sum_{k=3}^{k^*+1}\mathcal{B}_1^k(\vartheta)\mathcal{B}_k^n(p_o) + \mathcal{B}_1^n(p_o)\right)\phi_n. \tag{8}$$

The first term considers the case that two out of n backlogged UEs transmit with probability p_o. In this, both of them make a successful transmission if one chooses P_1 with probability ϑ and the other does P_2 with probability $1-\vartheta$. The second term means that if there are accessing UEs more than two, the one with P_1 can transmit its packet successfully if the number of UEs with P_2 is less than or equal to k^*. The last term implies that if only one UE transmits (whether it chooses P_1 or P_2), it will make a successful transmission. Using Little's result, we obtain the average access delay of NORA-A as $\overline{d}_a = \overline{N}_a/\tau_a$, and $\overline{N}_a = \sum_{n=0}^{N} n\phi_n$.

Let us consider EE of NORA-A, defined as $\mathcal{E}_a = \tau_a/\overline{\mathcal{P}}_a$, where $\overline{\mathcal{P}}_a$ is the average transmission power consumption of NORA-A. We can write $\overline{\mathcal{P}}_a = \mu(\overline{P}_{a,1} + \overline{P}_{a,2})\overline{N}_a$, where $\overline{P}_{a,i}$ denotes the average transmission power consumption of a UE aiming at P_i for $i=1,2$.

Lemma 2: In NORA-A, $\overline{P}_{a,i}$ for $\alpha = 4$ is obtained as $\overline{P}_{a,1} = P_1\mathcal{Y}_a(\beta_1)$ and $\overline{P}_{a,2} = P_2(\mathcal{Y}_a(\beta_2) - \mathcal{Y}_a(\beta_1))$, where $\mathcal{Y}_a(y)$ is given in the proof.

Proof. Based on $P_{T,i} = \frac{P_i}{Y}$, we can write $\overline{P}_{a,1}$ as

$$\begin{aligned}\overline{P}_{a,1} &= \mathsf{E}[P_{T,1}] = P_1\int_{\beta_1}^{\infty}\frac{1}{y}f_Y(y)dy = P_1\mathcal{Y}_a(\beta_1)\\ &= P_1\int_{\beta_1}^{\infty}\left[\frac{\sqrt{\pi}}{4R^2\sqrt{y^5}}\mathrm{erf}\left(\sqrt{y}R^2\right) - \frac{1}{2y^2}e^{-R^4y}\right]dy,\end{aligned} \tag{9}$$

where we have $f_Y(y) = \frac{2}{R}\int_0^R r^5 e^{-yr^4}dr$ from (4). Similiarly, we get $\overline{P}_{a,2} = P_2\int_{\beta_2}^{\beta_1}\frac{1}{y}f_Y(y)dy = P_2\left(\mathcal{Y}_a(\beta_2) - \mathcal{Y}_a(\beta_1)\right)$.

Let us move onto the performance of NORA-B. The system state can be captured by (n_1, n_2), where n_1 and n_2 denote the number of backlogged UEs in region 1 and 2 at time t, respectively. Thus, its state space is $\{(n_1,n_2)|n_1\in\{0,1,\ldots,M_1\}, n_2\in\{0,1,\ldots,M_2\}\}$. Let $\boldsymbol{\pi} = [\pi_{n_1,n_2}]$ be the steady state probability row vector of length $(1+M_1)\times(1+M_2)$, where π_{n_1,n_2} is the steady state probability that the system has n_1 and n_2 backlogged UEs in region 1 and 2, respectively. If the state transition probability matrix Q is obtained, whose element $q_{(n_1,n_2),(m_1,m_2)}$ denotes the state transition probability that state (n_1, n_2) at time t changes into (m_1, m_2) at time $t+1$, we can get $\boldsymbol{\pi}$ as $\boldsymbol{\pi} = \boldsymbol{\pi}\cdot Q$ and $\boldsymbol{\pi}\cdot\boldsymbol{e} = 1$, where the length of $\boldsymbol{e}$ corresponds to that of $\boldsymbol{\pi}$. For $n_1 = n_2 = 0$, we can have $q_{(0,0),(m_1,m_2)} = \mathcal{B}_{m_1}^{M_1}(\sigma_1)\mathcal{B}_{m_2}^{M_2}(\sigma_2)$. For $m_1 \geq n_1 - 1$, $n_1 \geq 1$ and $n_2 = 0$, we have

$$\begin{aligned}q_{(n_1,0),(m_1,m_2)} = \mathcal{B}_{m_2}^{M_2}(\sigma_2)\Big[&\mathcal{B}_1^{n_1}(\theta_1)\mathcal{B}_{m_1-(n_1-1)}^{M_1-n_1}(\sigma_1)\\ &+ (1-\mathcal{B}_1^{n_1}(\theta_1))\mathcal{B}_{m_1-n_1}^{M_1-n_1}(\sigma_1)\Big],\end{aligned} \tag{10}$$

where $\theta_i = \mu_i q_i(\beta_i)$ for $i = 1,2$ is the probability that a backlogged UE in region i (re)transmits its packet. In (10), the region 2 has m_2 backlogged UEs from zero with probability $B_{m_2}^{M_2}(\sigma)$. Meanwhile, the region 1 has m_1 backlogged UEs from n_1, if one out of n_1 UEs transmits successfully and $m_1-(n_1-1)$ newly join the backlogged. Otherwise, it has m_1 backlogged UEs, if $m_1 - n_1$ newly join the backlogged. Similarly, for $m_2 \geq n_2 - 1$, $n_2 \geq 1$ and $n_1 = 0$, we get

$$\begin{aligned}q_{(0,n_2),(m_1,m_2)} = \mathcal{B}_{m_1}^{M_1}(\sigma_1)\Big[&\mathcal{B}_1^{n_2}(\theta_2)\mathcal{B}_{m_2-(n_2-1)}^{M_2-n_2}(\sigma_2)\\ &+ (1-\mathcal{B}_1^{n_2}(\theta_2))\mathcal{B}_{m_2-n_2}^{M_2-n_2}(\sigma_2)\Big].\end{aligned} \tag{11}$$

For $m_1 \geq n_1 - 1$, $m_2 \geq n_2 - 1$, $n_1 \geq 1$, $n_2 \geq 1$, $q_{(n_1,n_2),(m_1,m_2)}$ is given in (16) shown at the bottom of the next page.

The system throughput of NORA-B is then

$$\tau_b = \sum_{j=1}^{2}\sum_{n_1=0}^{M_1}\sum_{n_2=0}^{M_2}\pi_{n_1,n_2}\tau_{b,j,n_1,n_2}, \tag{12}$$

where $\tau_{b,1,n_1,n_2} = \mathcal{B}_1^{n_1}(\theta_1)\sum_{i=0}^{k^*}\mathcal{B}_i^{n_2}(\theta_2)$, and $\tau_{b,2,n_1,n_2} = \mathcal{B}_1^{n_2}(\theta_2)\sum_{i=0}^{1}\mathcal{B}_i^{n_1}(\theta_1)$ denote the expected number of packets successfully transmitted in region 1 and 2, respectively.

Let $\overline{d}_{b,i}$ (slots) denote the average access delay of a UE in the region i in NORA-B. As in NORA-A, we also get $\overline{d}_{b,i} = \overline{N}_b/\tau_{b,i}$, where $\overline{N}_b = \sum_{n_1=0}^{M_1}\sum_{n_2=0}^{M_2}(n_1+n_2)\pi_{b,n_1,n_2} = \overline{N}_1 + \overline{N}_2$. We can get EE of NORA-B as $\mathcal{E}_b = \tau_b/\overline{\mathcal{P}}_b$, in which we have the average transmission power consumption $\overline{\mathcal{P}}_b = \sum_{n_1=1}^{M_1}\sum_{n_2=1}^{M_2}\left(r_1\overline{P}_{b,1}n_1 + r_2\overline{P}_{b,2}n_2\right)\pi_{n_1,n_2}$.

Corollary 3: Let $\overline{P}_{b,i}$ be the average transmission power consumption of a UE aiming at P_i for $i = 1,2$ in NORA-B. For $\alpha = 4$, we have

$$\overline{P}_{b,1} = P_1\int_{\beta_1}^{\infty}\left[\frac{\sqrt{\pi}}{4r_c^2\sqrt{y^5}}\mathrm{erf}\left(\sqrt{y}r_c^2\right) - \frac{1}{2y^2}e^{-r_c^4 y}\right]dy, \tag{13}$$

and

$$\begin{aligned}\overline{P}_{b,2} = \frac{P_2}{R^2-r_c^2}\int_{\beta_2}^{\infty}\Bigg[&\frac{\sqrt{\pi}}{4\sqrt{y^5}}\left(\mathrm{erf}\left(\sqrt{y}R^2\right) - \mathrm{erf}\left(\sqrt{y}r_c^2\right)\right)\\ &- \frac{1}{2y^2}\left(R^2e^{-R^4y} - r_c^2e^{-r_c^4y}\right)\Bigg]dy.\end{aligned} \tag{14}$$

Proof. We can write $\overline{P}_{b,1}$ as

$$\overline{P}_{b,1} = P_1\int_{\beta_1}^{\infty}\frac{1}{y}f_{Y_1}(y)dy = P_1\mathcal{Y}_{b,1}(\beta_1), \tag{15}$$

in which we can get $f_{Y_1}(y) = \frac{dF_{Y_1}(y)}{dy}$ from (5). For $\alpha = 4$, we can rewrite $\mathcal{Y}_{b,1}(\beta_1)$ in (15) as

$$\mathcal{Y}_{b,1}(\beta_1) = \int_{\beta_1}^{\infty}\frac{1}{y}f_{Y_1}(y)dy = \frac{1}{r_c^2}\int_{\beta_1}^{\infty}\frac{1}{\sqrt{y^5}}\int_0^{\sqrt{y}r_c^2}z^2e^{-z^2}dzdy$$

so that we have (13). Likewise, the average transmission power of UEs for P_2 is expressed as $\overline{P}_{b,2} = P_2\int_{\beta_2}^{\infty}\frac{1}{y}f_{Y_2}(y)dy = P_2\mathcal{Y}_{b,2}(\beta_2)$, where $f_{Y_2}(y)$ is obtained from (6). ■

We now discuss how the system parameters can be selected. First, to maximize the throughput of the system in *congestion*, we consider retransmission probabilities for UEs to use in both systems. Suppose that NORA-A has n backlogged UEs for $n \gg 2$, say *congested* at time t. Then, the throughput is maximized when we maximize the term in the parenthesis in (8) with respect to μ. However, even for $k^* = 1$, it is not easy to find a closed form of μ. Therefore, let us focus on maximizing the term $2\mathcal{B}_1^2(\vartheta)\mathcal{B}_2^n(p_o)$ and we have a maximizer $\mu^* = \min\left(\frac{2}{n\Pr[Y \ge \beta_2]}, 1\right)$. Now consider NORA-B when it has n_1 and n_2 backlogged UEs in region 1 and 2 at time t. To maximize the system throughput, UEs in region i should use μ_i of maximizing $\tau_{b,1,n_1,n_2} + \tau_{b,2,n_1,n_2}$, whose solution can not be obtained as a closed form. When focusing on the term $2\mathcal{B}_1^{n_1}(\theta_1)\mathcal{B}_1^{n_2}(\theta_2)$, the (re)transmission probability employed in region i is $\mu_i^* = \min\left(\frac{1}{n_i \Pr[Y \ge \beta_i]}, 1\right)$. We call μ^* and μ_i^* a sub-optimal retransmission probability under congestion, which later help us to estimate the throughput limit.

Let us discuss how to set other system parameters such as the thresholds for channel gain, and target received powers. In determining β_i, we can consider three methods as follows. The *first* one is to restrict access opportunity based on channel gain such that $\Pr[hr^{-\alpha} \ge \beta_i] = q_i(\beta_i) = \epsilon_i$. If $\epsilon_i = 1$, UEs can (re)transmit their packet with probability μ or μ_i regardless of channel gain. Such a β_i can be numerically obtained as $\beta_i = q_i^{-1}(\epsilon_i)$. Notice that a higher ϵ_i can give more frequent access opportunities in expense of the transmission power consumption. For $\epsilon_1 = \epsilon_2$, it can be said that UEs in region 1 and 2 have *equal access opportunities*, but different transmission power consumptions. The *second* method is that once we set β_2 as the outage threshold of region 2, we set β_1 such that either the throughput, or EE can be maximized. The *last* one is to constrain the average transmission power consumption by a value δ_i, i.e., $\mathcal{Y}_i(\beta_i) = \delta_i$. Given δ_i, numerically we can find $\beta_i = \mathcal{Y}_i^{-1}(\delta_i)$. If $\delta_1 = \delta_2$, UEs in the region 1 and 2 have equal transmission power consumption, but unequal access opportunity. Accordingly, a trade-off is expected between access opportunity and the average transmission power consumption. Finally, let us consider how P_1 and P_2 can be chosen. Given a bandwidth B, we can set P_2 as $\mathsf{SNR}_2 = P_2/N_0 = (2^{R_2/B} - 1) \ge \gamma$, where R_2 is the required data rate of UEs aiming at P_2. Then, P_1 can be set as $P_1 = \theta P_2$ for $\theta > 1$. For $\mathsf{SNR}_2 = \gamma$, i.e., the minimum power P_2 to meet R_2, we write (2) as $k^* = \lfloor \gamma^{-1}(\theta - 1) \rfloor$ for $\theta > 1$.

IV. Numerical Results

Throughout this section, unless otherwise stated, we use the parameters as $r_c = 100$, $R = 200$, and $\alpha = 4$, whereas $\gamma = 1.5$, $\theta = 3.5$ and $N_0 = 1$ in (2), which gives $k^* = 1$. Notice that the symbols in each figure denote simulation results, while the lines indicate analysis.

In Figs. 1 and 2, our analysis is verified against simulations, where the throughputs, EE and access delay of NORA-A and -B are presented. For NORA-A, we set $\sigma = 25$, while $M = M_1 + M_2$, $M_1 = M_2$, and $\sigma_1 = \sigma_2 = 0.025$ for NORA-B. Note that the product $M\sigma$ can be called the average traffic intensity to the system. When we increase M and at the same time reduce σ by keeping $M\sigma$ constant, the same throughput as shown in Fig. 1 can be observed. We set two thresholds $\beta_1 = 3.278 \times 10^{-8}$, and $\beta_2 = 1.9344 \times 10^{-9}$ for both systems as follows: For a given β_2 (arbitrarily selected here, but determined with cell coverage in practice), we choose β_1 such that EE of NORA-A for $M = 25$ and $\mu = 0.15$ can be maximized. This implies in NORA-B that UEs in region 1 have the access opportunity $\epsilon_1 = q_1(\beta_1) = 0.4844$ and $\epsilon_2 = 0.35$ in

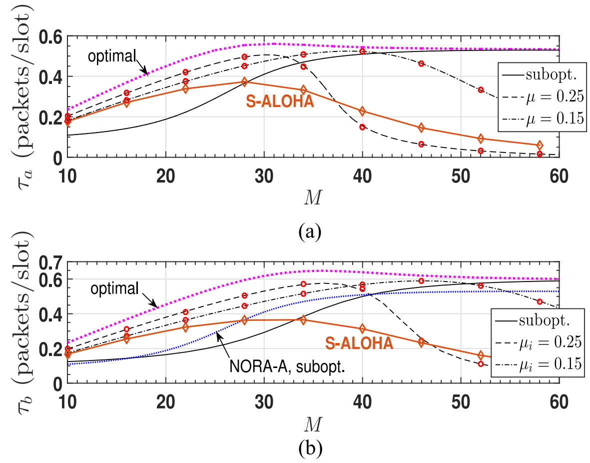

Fig. 1. Throughput of NORA-A and -B: $k^* = 1$.

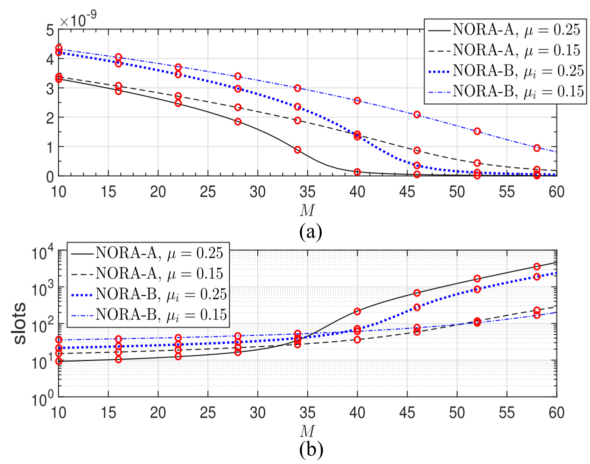

Fig. 2. EE and average access delay of NORA-A and -B: $k^* = 1$.

$$q_{(n_1,n_2),(m_1,m_2)} = \mathcal{B}_1^{n_1}(\theta_1)\mathcal{B}_{m_1-(n_1-1)}^{M_1-n_1}(\sigma_1)\left[\mathcal{B}_{m_2-n_2}^{M_2-n_2}(\sigma_2)\sum_{i=0,i\neq 1}^{k^*}\mathcal{B}_i^{n_2}(\theta_2) + \mathcal{B}_1^{n_2}(\theta_2)\mathcal{B}_{m_2-(n_2-1)}^{M_2-n_2}(\sigma_2)\right] + \mathcal{B}_{m_1-n_1}^{M_1-n_1}(\sigma_1)$$
$$\times\left\{\left(\sum_{i=2}^{n_1}\mathcal{B}_i^{n_1}(\theta_1) + \mathcal{B}_1^{n_1}(\theta_1)\sum_{i=k^*+1}^{n_2}\mathcal{B}_i^{n_2}(\theta_2)\right)\mathcal{B}_{m_2-n_2}^{M_2-n_2}(\sigma_2) + \mathcal{B}_0^{n_1}(\theta_1)\left[\mathcal{B}_1^{n_2}(\theta_2)\mathcal{B}_{m_2-(n_2-1)}^{M_2-n_2}(\sigma_2) + (1-\mathcal{B}_1^{n_2}(\theta_2))\mathcal{B}_{m_2-n_2}^{M_2-n_2}(\sigma_2)\right]\right\}. \quad (16)$$

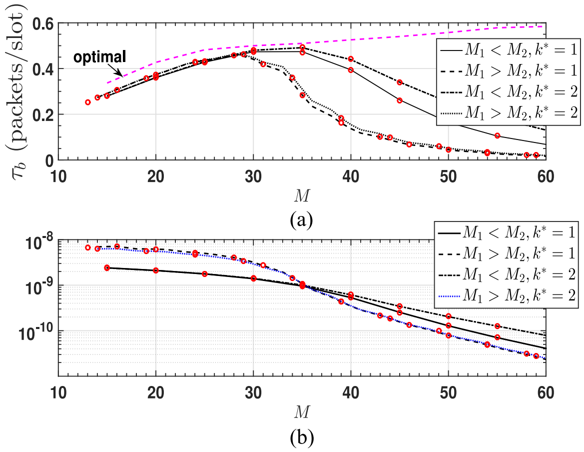

Fig. 3. Throughput and EE of NORA–B: $k^* = 1, 2$ and $\mu_i = 0.25$.

region 2. First of all, in Figs. 1 and 2, it is demonstrated that our analysis agrees well with simulation, while τ_a and τ_b very similarly change upon the increase in M. One difference with μ^* and μ_i^* is that τ_b is slightly higher than τ_a for a larger M, and vice versa for a smaller M. This is because two UEs can choose P_1 or P_2 rather independently with μ^*, which makes both UEs target at P_1 or P_2 accidentally. This leads to more collisions than in NORA-B.

To find the maximum throughput supported by each system (marked as 'optimal' in Fig. 1), we numerically find (exhaustive search) the optimal retransmission probability of maximizing the term in the parenthesis in (8) for NORA-A and $\tau_{b,1,n_1,n_2} + \tau_{b,2,n_1,n_2}$ for NORA-B, respectively. For $k^* = 1$, the maximum throughput with the optimal retransmission probability is 0.56 and 0.65 in NORA-A and -B, respectively. Using the suboptimal retransmission probabilities under congestion, we can achieve 0.6 throughput of NORA-B. When θ is increased up to 7 (not shown here), which yields $k^* = 4$, the optimal retransmission probability shows 0.7177 throughput. As a comparison, we depict throughput of the system without SIC, i.e., S-ALOHA, for $\mu = 0.15$ (or $\mu_i = 0.15$), where no packets can be decoded upon transmissions of more than one packets. While NORA is better, it shows also that ill-designed retransmission probability can make it even worse than S-ALOHA. As well known, the maximum throughput of S-ALOHA is 0.368 which can also be observed from Fig. 1 approximately. As M increases, it can be seen that the throughput with the suboptimal retransmission probability meets the optimal one. Since a higher retransmission probability yields a higher throughput for the system with a smaller population size, a dynamic retransmission probability (depending on the backlog size) is needed in order to achieve and maintain the maximum throughput over time.

Fig. 2 shows EE and access delay with two retransmission probabilities used in Fig. 1. As expected, as M increases, it is shown that EE decreases due to more collisions. Furthermore, NORA-B shows better EE than NORA-A, which is due to its higher throughput. Notice that the access delay of NORA-A is smaller than that of NORA-B for $M \leq 35$. In Figs. 1 and 2, it can be concluded that NORA-B is better than NORA-A if NORA-B has the same average traffic intensity of region 1 and 2. While it is simple to run NORA-A in practice, for NORA-B it is needed to make UEs aware of two regions so that they can realize channel inversion with β_i in each region. Notice that although not presented here, the equal access opportunities, $\epsilon_1 = \epsilon_2$ shows the same average access delay of UEs in two regions, i.e., $\overline{d}_1 = \overline{d}_2$ in NORA-B.

Fig. 3 illustrates the throughput and EE of two NORA-B systems: One is that the number of UEs deployed in region 1, M_1, is four times larger than M_2; the other is that M_1 is one quarter of M_2, and $M = M_1 + M_2$. The throughput of both systems in Fig. 3 is found lower than NORA-B in Fig. 1, where M_1 is equal to M_2. Moreover, the throughput of the system with $M_1 = 4M_2$ drops drastically as M exceeds 30 in comparison with the other system while it is slightly higher for $M < 30$. This shows that UEs in region 1 can have access priority for a lightly loaded system, but deteriorate the overall system if they are dominantly active or backlogged. EE shows similar behavior. When θ is increased from 3.5 to 4.5, which gives $k^* = 2$, the throughput of the system having a larger M_2 is much more improved than that of the other system. Such throughput gains result from that the system of employing a higher P_1 is more robust to the interference from accessing UEs with P_2, which gives a higher success probability for UEs with P_1. Notice that even if EE of the system with $M_1 < M_2$ is much lower than the other system for $M < 30$, it becomes improved along with the enhanced throughput for $M > 30$. The 'optimal' throughput shown in Fig. 3(a) is obtained for the system with $M_1 > M_2$ and the optimal retransmission.

V. Conclusion

We proposed two NORA techniques: NORA-A and NORA-B, where UEs aim at two levels of target received powers according to their channel gain or location. It has been shown that NORA-B is better than NORA-A in general, and it can particularly achieve the maximum throughput above 0.7 with propor retransmission control. In other words, NORA-B can make use of power-domain NOMA fully over the uplink if only one user in each region is allowed to access by retransmission control and targets at a specific power level associated with each region. We leave a dynamic retransmission scheme in NORA to maximize the system throughput and NORA with multilevel target powers as future work.

References

[1] S. M. R. Islam, N. Avazo, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tut.*, vol. 19, no. 2, pp. 721–741, Second Quarter 2017.

[2] 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, Mar. 2016.

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

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

[5] 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.

[6] A. J. Goldsmith, *Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press 2005.

[7] Y. Yu and G. B. Giannakis, "High-throughput random access using successive interference cancellation in a tree algorithm," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4628–4639, Aug. 2007.

[8] R. Yim, N. B. Mehta, A. F. Molish, and J. Zhang, "Dual power multiple access with multipacket reception using local CSI," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4078–4088, Aug. 2009.