Cooperative NOMA Short-Packet Communications in Flat Rayleigh Fading Channels

Xiazhi Lai, Qi Zhang, *Member, IEEE*, and Jiayin Qin

***Abstract*—A two-user downlink cooperative non-orthogonal multiple access (NOMA) short-packet communication network is investigated in flat Rayleigh fading channels, where a central user cooperatively relays the signals intended to a cell-edge user. To derive the average block error rate (BLER) of the central user, unlike the work in the literature where the product term is neglected, a new method that works well for the whole range of power allocation factor is proposed. Furthermore, the average BLER of the cell-edge user is theoretically derived. Numerical results demonstrate that while remaining almost the same average BLER for the central user, the cell-edge user in our proposed cooperative NOMA network achieves lower average BLER than those in the conventional NOMA network and the cooperative orthogonal multiple access network.**

***Index Terms*—Block error rate (BLER), non-orthogonal multiple access (NOMA), short-packet communications.**

I. Introduction

Non-orthogonal multiple access (NOMA) is promising for future wireless communications [1]–[5]. In [1], the performance of NOMA systems with randomly deployed users was investigated. In [4], [5], the cooperative NOMA transmission scheme was proposed, where a central user acts like a cooperative relay to help the signal transmission to a cell-edge user.

In [1]–[5], the infinite blocklength of transmission is assumed, which enables the system to achieve the upper bound of channel capacity. However, for the Internet-of-Things (IoT), the ultra-reliable and low latency communications, i.e., short-packet communications, are required, where the impact of blocklength on the system performance is ineligible [6]–[8]. For short-packet communications, a new performance metric, i.e., block error rate (BLER) was proposed in [6]. Inspired by [6], NOMA systems with short-packet communications [9], [10] were studied. Specifically, Sun *et al.* solved the power allocation and code rate optimization problem in [9] and Yu *et al.* proposed a closed-form power allocation solution to achieve the minimum blocklength in [10].

In this paper, we evaluate the BLER performance of a two-user downlink cooperative NOMA network with finite blocklength, where a central user helps a cell-edge user by relaying the signals intended to the cell-edge user. Different from the cooperative NOMA network with infinite blocklength, perfect decoding in short-packet communications is not satisfied, and error propagation will happen at different levels of certainty with respect to the received signal. For the average BLER derivations of the central user in flat Rayleigh fading channels, unlike the work in [10] where the product term in derivations is neglected, we propose a new method to deal with the product term. The new method works well for whole range of power allocation factor. Furthermore, we also theoretically derive the average BLER performance of the cell-edge user when selective combining is used. When maximum ratio combining is employed, we obtain the tight lower bound.

Notations: $\mathcal{CN}(0,\sigma^2)$ denotes the distribution of a circularly symmetric complex Gaussian random variable with zero mean and variance σ^2. The probability density function (PDF) and cumulative density function (CDF) for a random variable γ are denoted as $f_\gamma(\cdot)$ and $F_\gamma(\cdot)$, respectively.

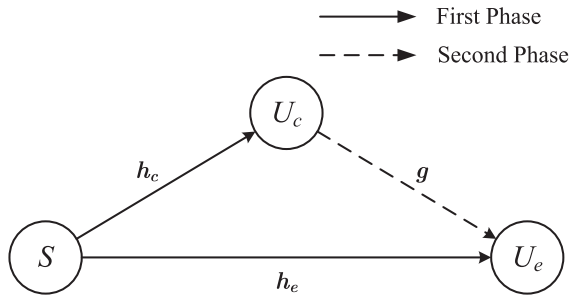

Fig. 1. System model of the two-user downlink cooperative NOMA network.

II. System Model

Consider a two-user downlink cooperative NOMA network in flat Rayleigh fading environment.[1] Two users are chosen by employing the user pairing scheme to lower the implementation complexity while maintain the benefit of NOMA as suggested in [2]–[4]. The work on the downlink cooperative NOMA network with more than two users may be a challenging future work. As depicted in Fig. 1, the network comprises one source S, one central user U_c and one cell-edge user U_e. The central user U_c is closer to the source than the cell-edge user U_e. The source S is aiming to convey information of N_c bits to U_c and information of N_e bits to U_e. The downlink transmission is divided into two phases. During the first phase, the source S broadcasts signals to U_c and U_e by employing the NOMA scheme and during the second phase, U_c cooperatively relays the signals intended to U_e.

Specifically, during the first phase, the source S sends a superposition signal x_S with blocklength m to both U_c and U_e with transmit power P_S. The signal x_S is expressed as $x_S = \sqrt{\alpha_c}x_c + \sqrt{\alpha_e}x_e$, where α_c and α_e are the power allocation factors with $\alpha_c + \alpha_e = 1$ and $\alpha_c < \alpha_e$, x_c and x_e are the signals intended to U_c and U_e, respectively.

[1] As reported in [11], the flat Rayleigh fading model is applicable for the non-light-of-sight environment with many scattering objects present (e.g., urban areas). Thus, the flat Rayleigh fading environment has been widely adopted in practice [1], [3], [4], [7], [8], [10].

Manuscript received November 17, 2018; revised March 12, 2019; accepted April 14, 2019. Date of publication April 22, 2019; date of current version June 18, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61672549, in part by the Guangdong Natural Science Foundation under Grant 2018B0303110016, and in part by the Guangzhou Science and Technology Program under Grant 201607010098 and Grant 201804010445. The review of this paper was coordinated by Prof. K. Le. *(Corresponding author: Qi Zhang.)*

X. Lai and Q. Zhang are with the School of Electronics and Information Technology, Sun Yat-sen University Guangzhou 510006, China, and also with the Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University), Ministry of Education China (e-mail: laixzh@mail2.sysu.edu.cn; zhqi26@mail.sysu.edu.cn).

J. Qin is with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, 510006, China, and with the Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University), Ministry of Education, China, and also with the Xinhua College of Sun Yat-sen University, Guangzhou 510520, China (e-mail: issqjy@mail.sysu.edu.cn).

Digital Object Identifier 10.1109/TVT.2019.2912391

The received signals at U_c and U_e are

$$y_c = h_c\sqrt{P_S}x_S + n_c \text{ and } y_{e,1} = h_e\sqrt{P_S}x_S + n_{e,1}, \quad (1)$$

respectively, where $h_c \sim \mathcal{CN}(0,\lambda_c)$ and $h_e \sim \mathcal{CN}(0,\lambda_e)$ are the channel parameters for $S \to U_c$ and $S \to U_e$ links, respectively; $n_c \sim \mathcal{CN}(0,\sigma^2)$ and $n_{e,1} \sim \mathcal{CN}(0,\sigma^2)$ are the additive Gaussian noises at U_c and U_e during the first phase, respectively. Without loss of generality, we assume $\sigma^2 = 1$ here. Using successive interference cancellation (SIC), U_c first decodes x_e with signal-to-noise-plus-interference ratio (SINR)

$$\gamma_{c,e} = \frac{P_S\alpha_e|h_c|^2}{P_S\alpha_c|h_c|^2+1}. \quad (2)$$

According to [6], the instantaneous block error rate (BLER) of decoding x_e at U_c is

$$\epsilon_{c,e} \approx \Psi(\gamma_{c,e}, N_e, m) \quad (3)$$

where $\Psi(\gamma,N,m) \triangleq Q((C(\gamma)-N/m)/\sqrt{V(\gamma)/m})$, $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}dt$, $C(\gamma) = \log_2(1+\gamma)$, and $V(\gamma) = (\log_2 e)^2\cdot(1-(1+\gamma)^{-2})$. In (3), the approximation holds well in short-packet communications when the blocklength $m > 100$ [6].

If x_e is successfully decoded at U_c, U_c then decodes x_c with signal-to-noise ratio (SNR)

$$\gamma_{c,c} = P_S\alpha_c|h_c|^2. \quad (4)$$

The instantaneous BLER of decoding x_c at U_c is

$$\epsilon_c = \epsilon_{c,e} + (1-\epsilon_{c,e})\epsilon_{c,c} \quad (5)$$

where $\epsilon_{c,c} \approx \Psi(\gamma_{c,c}, N_c, m)$. For U_e, the received SINR to decode x_e during the first phase is

$$\gamma_{e,1} = \frac{P_S\alpha_e|h_e|^2}{P_S\alpha_c|h_e|^2+1}. \quad (6)$$

During the second phase, U_c cooperatively relays x_e to U_e with transmit power P_c. The received signal at U_e is

$$y_{e,2} = g\sqrt{P_c}x_e + n_{e,2} \quad (7)$$

where $g \sim \mathcal{CN}(0,\mu)$ is the channel parameter of $U_c \to U_e$ link; $n_{e,2} \sim \mathcal{CN}(0,1)$ is the additive Gaussian noise at U_e during the second phase. Accordingly, the received SINR to decode x_e during the second phase is

$$\gamma_{e,2} = P_c|g|^2. \quad (8)$$

Remark 1: In this paper, we assume the channel parameters h_c, h_e, and g are mutually independent random variables.

We consider to employ both maximal ratio combining (MRC) and selective combining (SC) to combine the received signals at U_e during the first and the second phases. For SC, we have the SINR to decode x_e as follows

$$\gamma_e = \max\{\gamma_{e,1}, \gamma_{e,2}\}. \quad (9)$$

For MRC, we have

$$\bar{\gamma}_e = \gamma_{e,1} + \gamma_{e,2}. \quad (10)$$

Thus, the instantaneous BLER of decoding x_e at U_e is

$$\epsilon_e = \epsilon_{c,e}\Psi(\gamma_{e,1}, N_e, m) + (1-\epsilon_{c,e})\Psi(\gamma_e, N_e, m) \quad (11)$$

for SC and

$$\bar{\epsilon}_e = \epsilon_{c,e}\Psi(\gamma_{e,1}, N_e, m) + (1-\epsilon_{c,e})\Psi(\bar{\gamma}_e, N_e, m) \quad (12)$$

for MRC.

Remark 2: In this paper, the half-duplex relaying is considered. Compared with the full-duplex relaying, the half-duplex relaying has much lower system complexity whereas consumes the extra time interval for signal relaying. In practice, the half-duplex relaying is still widely employed in the literature [4], [5], [8]. To employ the full-duplex relaying for the downlink cooperative NOMA network is an interesting future work.

III. Analytical Average BLER

In this section, we will derive the analytical average BLER to decode x_c at U_c and that to decode x_e at U_e. The analytical average BLER to decode x_c at U_c is obtained by

$$\mathbb{E}[\epsilon_c] = \mathbb{E}[\epsilon_{c,c}] + \mathbb{E}[\epsilon_{c,e}] - \mathbb{E}[\epsilon_{c,e}\epsilon_{c,c}]. \quad (13)$$

The first term in the right-hand of (13) is expressed as

$$\mathbb{E}[\epsilon_{c,c}] \approx \int_0^\infty \Psi(\gamma_{c,c}, N_c, m) f_{\gamma_{c,c}}(t)dt. \quad (14)$$

Since in (14), the integrand includes $\Psi(\gamma_{c,c}, N_c, m)$, the exact close-form of $\mathbb{E}[\epsilon_{c,c}]$ is difficult to obtain. We resort to a liner approximation of function $\Psi(\gamma,N,m) \approx Z_{N,m}(\gamma)$ which is given as follows [7], [10]

$$Z_{N,m}(\gamma) = \begin{cases} 1, & \gamma \le v_{N,m} \\ \frac{1}{2} - \delta_{N,m}\sqrt{m}(\gamma-\beta_{N,m}), & v_{N,m} < \gamma < u_{N,m} \\ 0, & \gamma \ge u_{N,m} \end{cases} \quad (15)$$

where $\beta_{N,m} = 2^{\frac{N}{m}} - 1, \delta_{N,m} = (2\pi(2^{\frac{2N}{m}}-1))^{-\frac{1}{2}}, v_{N,m} = \beta_{N,m} - \frac{1}{2}\delta_{N,m}^{-1}m^{-\frac{1}{2}}$, and $u_{N,m} = \beta_{N,m} + \frac{1}{2}\delta_{N,m}^{-1}m^{-\frac{1}{2}}$.

Substituting (15) into (14), we have [10]

$$\begin{aligned}\mathbb{E}[\epsilon_{c,c}] &\approx \delta_{N_c,m}\sqrt{m}\int_{v_{N_c,m}}^{u_{N_c,m}} F_{\gamma_{c,c}}(t)dt \\ &= 1 + \delta_{N_c,m}\sqrt{m}\lambda_c P_S\alpha_c\left(e^{-\frac{v_{N_c,m}}{\lambda_c P_S\alpha_c}} - e^{-\frac{u_{N_c,m}}{\lambda_c P_S\alpha_c}}\right)\end{aligned} \quad (16)$$

where in the derivation, $F_{\gamma_{c,c}}(t) = 1 - e^{-\frac{t}{\lambda_c P_S\alpha_c}}$ is used.

Similarly, we have [10]

$$\begin{aligned}\mathbb{E}[\epsilon_{c,e}] &\approx 1 - \delta_{N_e,m}\sqrt{m}\alpha_c^{-1}\left(e^{-\theta_v}\eta_v - e^{-\theta_u}\eta_u\right) \\ &\quad - \delta_{N_e,m}\omega e^{\frac{\omega}{\alpha_e}}\sqrt{m}\alpha_c^{-1}\left(\mathrm{Ei}\left(-\omega\eta_v^{-1}\right) - \mathrm{Ei}\left(-\omega\eta_u^{-1}\right)\right)\end{aligned} \quad (17)$$

where $\mathrm{Ei}(z) = -\int_{-z}^\infty \frac{e^{-t}}{t}dt$ is the exponential integral function and

$$\eta_v = [\alpha_e - \alpha_c v_{N_e,m}]^+, \eta_u = [\alpha_e - \alpha_c u_{N_e,m}]^+, \quad (18)$$

$$\theta_v = \frac{v_{N_e,m}}{\lambda_c P_S\eta_v},\ \theta_u = \frac{u_{N_e,m}}{\lambda_c P_S\eta_u},\ \omega = \frac{\alpha_e}{\lambda_c P_S\alpha_c}. \quad (19)$$

Note that in [10], the third term in the right-hand of (13), i.e., $\mathbb{E}[\epsilon_{c,e}\epsilon_{c,c}]$ is neglected. In this paper, we have the following proposition to deal with $\mathbb{E}[\epsilon_{c,e}\epsilon_{c,c}]$.

Proposition 1: $\mathbb{E}[\epsilon_{c,e}\epsilon_{c,c}]$ can be effectively approximated by $\min\{\mathbb{E}[\epsilon_{c,e}], \mathbb{E}[\epsilon_{c,c}]\}$.

Proof: We have

$$\begin{aligned}&\mathbb{E}[\epsilon_{c,e}\epsilon_{c,c}] \\ &\approx \int_0^\infty\int_0^\infty Z_{N_e,m}(t_1)Z_{N_c,m}(t_2)f_{\gamma_{c,e},\gamma_{c,c}}(t_1,t_2)dt_1dt_2.\end{aligned} \quad (20)$$

In the expression of $Z_{N_c,m}(\gamma_{c,c})$, from (15), we obtain $Z_{N_c,m}(\gamma_{c,c}) = 1$ when $\gamma_{c,c} \leq v_{N_c,m}$, i.e.,

$$|h_c|^2 \leq \frac{v_{N_c,m}}{P_S \alpha_c}. \tag{21}$$

Furthermore, in the expression of $Z_{N_e,m}(\gamma_{c,e})$, $Z_{N_e,m}(\gamma_{c,e}) > 0$ when $\gamma_{c,e} \leq u_{N_e,m}$, i.e.,

$$|h_c|^2 \leq \frac{u_{N_e,m}}{P_S \alpha_e - P_S \alpha_c u_{N_e,m}}. \tag{22}$$

Therefore, if

$$\frac{v_{N_c,m}}{P_S \alpha_c} \geq \frac{u_{N_e,m}}{P_S \alpha_e - P_S \alpha_c u_{N_e,m}}, \tag{23}$$

we have

$$Z_{N_e,m}(\gamma_{c,e}) Z_{N_c,m}(\gamma_{c,c}) = Z_{N_e,m}(\gamma_{c,e}). \tag{24}$$

Substituting $\alpha_c + \alpha_e = 1$ into (23), the condition (23) is equivalent to $\alpha_c \leq \zeta_1 = v_{N_c,m}/(v_{N_c,m} + u_{N_e,m} + v_{N_c,m} u_{N_e,m})$. Similarly, we have

$$Z_{N_e,m}(\gamma_{c,e}) Z_{N_c,m}(\gamma_{c,c}) = Z_{N_c,m}(\gamma_{c,c}) \tag{25}$$

when $\alpha_c \geq \zeta_2 = u_{N_c,m}/(u_{N_c,m} + v_{N_e,m} + u_{N_c,m} v_{N_e,m})$. When $N_c = N_e = m = 100$ which are typical values for short-packet communications, we have $\zeta_2 = 0.4122$ and $\zeta_1 = 0.2651$. The gap is only 0.1471. When $N_c = N_e = 100$ and $m = 150$ [7], we have $\zeta_2 = 0.4745$ and $\zeta_1 = 0.3067$. The gap is only 0.1678.

Form (24) and (25), we know

$$\mathbb{E}[\epsilon_{c,e}\epsilon_{c,c}] \approx \min\{\mathbb{E}[\epsilon_{c,e}], \mathbb{E}[\epsilon_{c,c}]\}. \tag{26}$$

Thus, we have Proposition 1. ■

Using Proposition 1, substituting (26) into (13), the analytical average BLER to decode x_c at U_c is

$$\mathbb{E}[\epsilon_c] \geq \max\{\mathbb{E}[\epsilon_{c,c}], \mathbb{E}[\epsilon_{c,e}]\}. \tag{27}$$

Remark 3: From (27), the average BLER of the central user, $\mathbb{E}[\epsilon_c]$, is lower bounded by the average BLERs to decode signals intended to the central and cell-edge users, i.e., x_c and x_e, during the SIC process. Therefore, the over-allocating power for signals intended to the central user may lead to the performance degradation at the central user.

For the SC scheme, the analytical average BLER to decode x_e at U_e is obtained by

$$\begin{aligned}\mathbb{E}[\epsilon_e] &= \mathbb{E}[\epsilon_{c,e}]\mathbb{E}[\Psi(\gamma_{e,1}, N_e, m)] \\ &\quad + (1 - \mathbb{E}[\epsilon_{c,e}])\mathbb{E}[\Psi(\gamma_e, N_e, m)].\end{aligned} \tag{28}$$

To derive $\mathbb{E}[\Psi(\gamma_e, N_e, m)]$, we have the following proposition.

Proposition 2: When m is sufficiently large, $\mathbb{E}[\Psi(\gamma_e, N_e, m)]$ can be effectively approximated by $\mathbb{E}[Z_{N_e,m}(\gamma_{e,1})]\mathbb{E}[Z_{N_e,m}(\gamma_{e,2})]$.

Proof: See Appendix A. ■

Using Proposition 2, we have

$$\mathbb{E}[\Psi(\gamma_e, N_e, m)] \approx \chi_1 \chi_2 \tag{29}$$

where $\chi_i = \int_0^\infty Z_{N_e,m}(\gamma_{e,i}) f_{\gamma_{e,i}}(t_i) dt_i$ for $i \in \{1, 2\}$. After integration, we obtain

$$\begin{aligned}\chi_1 &= 1 - \delta_{N_e,m}\sqrt{m}\alpha_c^{-1}\left(e^{-\hat{\theta}_v}\eta_v - e^{-\hat{\theta}_u}\eta_u\right) \\ &\quad - \delta_{N_e,m}\hat{\omega}e^{\frac{\hat{\omega}}{\alpha_e}}\sqrt{m}\alpha_c^{-1}\left(\mathrm{Ei}\left(-\hat{\omega}\eta_v^{-1}\right) - \mathrm{Ei}\left(-\hat{\omega}\eta_u^{-1}\right)\right)\end{aligned} \tag{30}$$

where

$$\hat{\theta}_v = \frac{v_{N_e,m}}{\lambda_e P_S \eta_v},\ \hat{\theta}_u = \frac{u_{N_e,m}}{\lambda_e P_S \eta_u},\ \hat{\omega} = \frac{\alpha_e}{\lambda_e P_S \alpha_c} \tag{31}$$

and

$$\chi_2 = \left[1 - \delta_{N_e,m}\sqrt{m}\mu P_c\left(e^{-\frac{v_{N_e,m}}{\mu P_c}} - e^{-\frac{u_{N_e,m}}{\mu P_c}}\right)\right]. \tag{32}$$

From the derivation of $\mathbb{E}[\Psi(\gamma_e, N_e, m)]$, in (28), we have

$$\mathbb{E}[\Psi(\gamma_{e,1}, N_e, m)] \approx \chi_1. \tag{33}$$

Substituting (29) and (33) into (28), we obtain the analytical average BLER to decode x_e at U_e for the SC scheme.

For the MRC scheme, the analytical average BLER to decode x_e at U_e is obtained by

$$\begin{aligned}\mathbb{E}[\bar{\epsilon}_e] &= \mathbb{E}[\epsilon_{c,e}]\mathbb{E}[\Psi(\gamma_{e,1}, N_e, m)] \\ &\quad + (1 - \mathbb{E}[\epsilon_{c,e}])\mathbb{E}[\Psi(\bar{\gamma}_e, N_e, m)].\end{aligned} \tag{34}$$

It is noted that $Z_{N,m}(t_1 + t_2) \geq Z_{N,m}(2t_1) Z_{N,m}(2t_2)$ [7]. From (10), we have

$$\mathbb{E}[\Psi(\bar{\gamma}_e, N_e, m)] \geq \chi_3 \chi_4 \tag{35}$$

where

$$\begin{aligned}\chi_3 &= 1 - \delta_{N_e,m}\sqrt{m}\alpha_c^{-1}\left(e^{-\frac{1}{2}\hat{\theta}_v}\eta_v - e^{-\frac{1}{2}\hat{\theta}_u}\eta_u\right) \\ &\quad - \frac{1}{2}\delta_{N_e,m}\hat{\omega}e^{\frac{\hat{\omega}}{2\alpha_e}}\sqrt{m}\alpha_c^{-1}\left(\mathrm{Ei}\left(-\frac{\hat{\omega}\eta_v^{-1}}{2}\right) - \mathrm{Ei}\left(-\frac{\hat{\omega}\eta_u^{-1}}{2}\right)\right)\end{aligned} \tag{36}$$

and

$$\chi_4 = \left[1 - 2\delta_{N_e,m}\sqrt{m}\mu P_c\left(e^{-\frac{v_{N_e,m}}{2\mu P_c}} - e^{-\frac{u_{N_e,m}}{2\mu P_c}}\right)\right]. \tag{37}$$

IV. Asymptotic Average BLER

The derivations of the analytical average BLERs in Section III involve the integrals of the form $\int_{v_{N,m}}^{u_{N,m}} f(t)dt$. When m is sufficiently large, the integral interval $[v_{N,m}, u_{N,m}]$ is generically small. Thus, we can employ the first order Riemann integral approximation

$$\int_{v_{N,m}}^{u_{N,m}} f(x)dx \approx (u_{N,m} - v_{N,m}) f\left(\frac{v_{N,m} + u_{N,m}}{2}\right). \tag{38}$$

Using (38), we have $\mathbb{E}[\epsilon_{c,c}] \approx 1 - \exp(-\frac{\beta_{N_c,m}}{\lambda_c P_S \alpha_c})$ and $\mathbb{E}[\epsilon_{c,e}] \approx 1 - \exp(-\frac{\beta_{N_e,m}}{\lambda_c P_S (\alpha_e - \alpha_c \beta_{N_e,m})})$. When $P_S \to \infty$, we have

$$\mathbb{E}[\epsilon_{c,c}] \approx \frac{\beta_{N_c,m}}{\lambda_c P_S \alpha_c}, \tag{39}$$

$$\mathbb{E}[\epsilon_{c,e}] \approx \rho = \frac{\beta_{N_e,m}}{\lambda_c P_S (\alpha_e - \alpha_c \beta_{N_e,m})}. \tag{40}$$

Thus, when $P_S \to \infty$,

$$\mathbb{E}[\epsilon_c] \approx \max\left(\frac{\beta_{N_c,m}}{\lambda_c P_S \alpha_c}, \rho\right). \tag{41}$$

Similarly, when $P_S \to \infty$, we have

$$\mathbb{E}[\epsilon_e] \approx \frac{\rho\lambda_c}{\lambda_e}\left(\rho + \frac{\beta_{N_e,m}}{\mu P_c}\right),\ \mathbb{E}[\bar{\epsilon}_e] \approx \frac{\rho\lambda_c}{\lambda_e}\left(\rho + \frac{\beta_{N_e,m}}{4\mu P_c}\right). \tag{42}$$

Remark 4: From (41) and (42), the diversity order of the central user is one whereas that of the cell-edge user is two. This is because the cooperative NOMA network improves the performance of the cell-edge user with the signal relaying from the central user. From (42), it is found that the increase of P_c improves the average BLER performance of the cell-edge user. However, the improvement is lower bounded by $\rho^2\lambda_c/\lambda_e$ when $P_c \to \infty$.

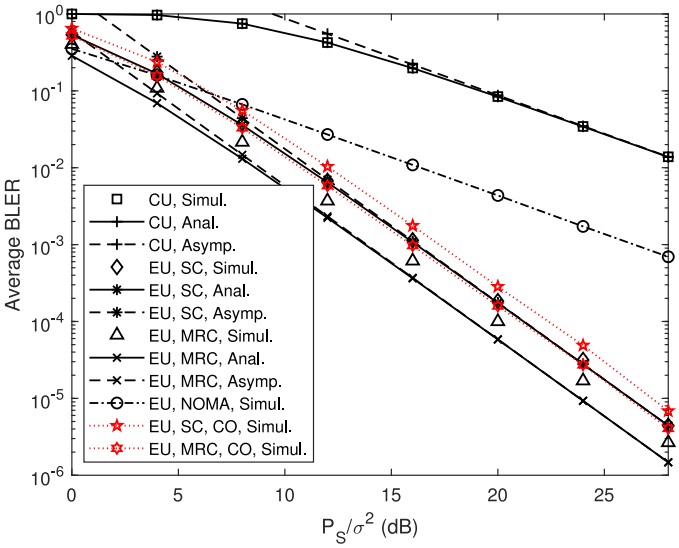

Fig. 2. Average BLER versus P_S/σ^2 where $m = 100$, $N_c = 300$ bits, $N_e = 100$ bits, $\alpha_c = 0.1$, and $\alpha_e = 0.9$.

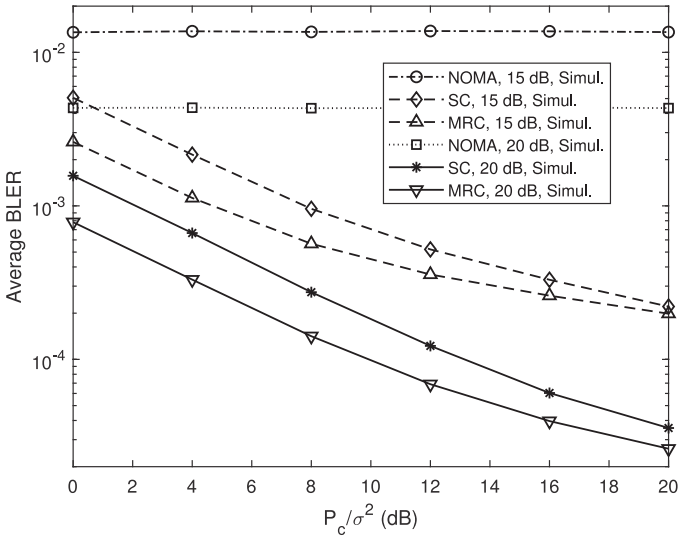

Fig. 3. Average BLER versus P_c/σ^2 where $m = 100$, $N_c = 300$ bits, $N_e = 100$ bits, $\alpha_c = 0.1$, and $\alpha_e = 0.9$.

V. Numerical Results

In the simulations, we assume that $\lambda_c = \mu = 8$, $\lambda_e = 1$, and $P_S = 10P_c$ [7]. In Fig. 2, we present the average BLER of the central and cell-edge users in the cooperative NOMA network where $m = 100$, $N_c = 300$ bits, $N_e = 100$ bits, $\alpha_c = 0.1$ and $\alpha_e = 0.9$, denoted as "CU" and "EU" in the legend, respectively. For comparison, we also present the average BLER of the cell-edge user in the conventional NOMA network, denoted as "NOMA". Since the cooperative NOMA network uses two phases for whole transmission whereas the conventional NOMA network uses one, we assume the blocklength of the conventional NOMA network, denoted as $\tilde{m}$, is $\tilde{m} = 2m$. Furthermore, the advantage of the cooperative NOMA network over the conventional NOMA network is that the former significantly improve the performance of the cell-edge user than the letter. When $N_c = 300$ bits, $N_e = 100$ bits, $\alpha_c = 0.1$, and $\alpha_e = 0.9$, from (41), we have $\mathbb{E}[\epsilon_c] \approx \frac{\beta_{N_c,m}}{\lambda_c P_S \alpha_c}$. For fairly comparison, we let average BLERs of the central user in the cooperative NOMA network and the conventional NOMA network be almost the same, i.e., $\beta_{N_c,m}/\alpha_c = \tilde{\beta}_{N_c,m}/\tilde{\alpha}_c$ where $\tilde{\alpha}_c$ and $\tilde{\alpha}_e = 1 - \tilde{\alpha}_c$ denote the power allocation factors in the conventional NOMA network and $\tilde{\beta}_{N_c,m} = 2^{\frac{N_c}{\tilde{m}}} - 1$. Thus, we have $\tilde{\alpha}_c = 0.0261$ and $\tilde{\alpha}_e = 0.9739$.

In Fig. 2, we also present the average BLER of the cell-edge user in the cooperative orthogonal multiple access (OMA) network [4], denoted as "CO", when the average BLER of the central user is almost the same as that in our proposed cooperative NOMA network.

From Fig. 2, it is observed that when the average BLERs of the central user are almost the same, the cell-edge user in our proposed cooperative NOMA network achieves lower average BLER than those in the conventional NOMA network and the cooperative OMA network. From Fig. 2, it is also found that in the proposed cooperative NOMA network, the analytical average BLER of the central user matches the simulation results. The analytical average BLER of the cell-edge user using SC is almost the same with the simulation results. The analytical average BLER of the cell-edge user using MRC is a tight lower bound of the simulation results.

In Fig. 3, given P_S/σ^2 specified in the legend, we present the average BLER versus P_c/σ^2, which illustrates the tradeoff for the transmit power consumption of the central user and the improvement of the BLER performance of the cell-edge user. From Fig. 3, it is shown that in our proposed cooperative NOMA network, the MRC scheme outperforms the SC scheme. The performance gap between the MRC and SC schemes decreases with the increase of P_c/σ^2.

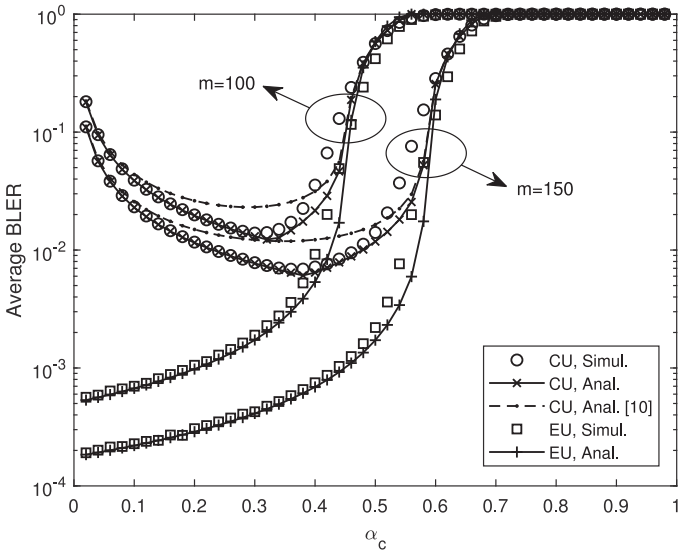

Fig. 4. Average BLER versus α_c where $P_S/\sigma^2 = 15$ dB and $N_c = N_e = 100$ bits.

In Section III, the derivations of analytical average BLER are based on two approximations, i.e., $\Psi(\gamma, N, m) \approx Z_{N,m}(\gamma)$ and $\mathbb{E}[\epsilon_{c,e}\epsilon_{c,c}] \approx \min\{\mathbb{E}[\epsilon_{c,e}], \mathbb{E}[\epsilon_{c,c}]\}$. To examine the two approximations, in Fig. 4, we present the average BLER versus α_c where $P_S/\sigma^2 = 15$ dB and $N_c = N_e = 100$ bits. For the cell-edge user, SC is used. From Fig. 4, it is observed that when $m = 150$ [7] and $\zeta_1(0.3067) \leq \alpha_c \leq \zeta_2(0.4745)$, the analytic average BLER is almost the same with the simulation results. The gap between the analytical and simulated results over $0.48 \leq \alpha_c \leq 0.60$ is mainly due to the approximation $\Psi(\gamma, N, m) \approx Z_{N,m}(\gamma)$. This also causes the gap over

$0.32 \leq \alpha_c \leq 0.46$ when $m = 100$. In Fig. 4, the analytical average BLERs of the central user derived in [10] are also shown.

VI. Conclusion

In this paper, we have proposed a downlink cooperative NOMA short-packet communication network, whose average BLER is investigated in flat Rayleigh fading channels. Numerical results show that when the average BLERs of the central user are almost the same, the cell-edge user in our proposed cooperative NOMA network achieves lower average BLER than that in the conventional NOMA network.

Appendix A
Proof of Proposition 2

Similar to the derivations of $\mathbb{E}[\epsilon_{c,c}]$ and $\mathbb{E}[\epsilon_{c,e}]$, we use the approximation $\mathbb{E}[\Psi(\gamma_e, N_e, m)] \approx \mathbb{E}[Z_{N_e,m}(\gamma_e)]$. From (9) and (15), we know

$$\mathbb{E}[Z_{N_e,m}(\gamma_e)] = \varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 + \varphi_5 \tag{43}$$

where

$$\varphi_1 = F_{\gamma_{e,1}}(v_{N_e,m}) F_{\gamma_{e,2}}(v_{N_e,m}), \tag{44}$$

$$\varphi_2 = \int_{v_{N_e,m}}^{u_{N_e,m}} Z_{N_e,m}(\gamma_{e,1}) f_{\gamma_{e,1}}(t) F_{\gamma_{e,2}}(v_{N_e,m}) dt, \tag{45}$$

$$\varphi_3 = \int_{v_{N_e,m}}^{u_{N_e,m}} Z_{N_e,m}(\gamma_{e,2}) f_{\gamma_{e,2}}(t) F_{\gamma_{e,1}}(v_{N_e,m}) dt, \tag{46}$$

$$\varphi_4 = \int_{v_{N_e,m}}^{u_{N_e,m}} \int_{v_{N_e,m}}^{t_1} Z_{N_e,m}(\gamma_{e,1}) f_{\gamma_{e,1}}(t_1) f_{\gamma_{e,2}}(t_2) dt_2 dt_1, \tag{47}$$

$$\varphi_5 = \int_{v_{N_e,m}}^{u_{N_e,m}} \int_{v_{N_e,m}}^{t_2} Z_{N_e,m}(\gamma_{e,2}) f_{\gamma_{e,1}}(t_2) f_{\gamma_{e,2}}(t_1) dt_1 dt_2. \tag{48}$$

Similarly, we have

$$\mathbb{E}[Z_{N_e,m}(\gamma_{e,1})]\mathbb{E}[Z_{N_e,m}(\gamma_{e,2})] = \varphi_1 + \varphi_2 + \varphi_3 + \varphi_6 \tag{49}$$

where

$$\varphi_6 = \int_{v_{N_e,m}}^{u_{N_e,m}} \int_{v_{N_e,m}}^{u_{N_e,m}} Z_{N_e,m}(\gamma_{e,1}) Z_{N_e,m}(\gamma_{e,2}) f_{\gamma_{e,1}}(t_1) f_{\gamma_{e,2}}(t_2) dt_1 dt_2. \tag{50}$$

The difference between $\mathbb{E}[Z_{N_e,m}(\gamma_e)]$ and $\mathbb{E}[Z_{N_e,m}(\gamma_{e,1})]\mathbb{E}[Z_{N_e,m}(\gamma_{e,2})]$ is only that $\varphi_4 + \varphi_5$ in (43) is replaced by φ_6 in (49). The expressions of φ_4, φ_5, and φ_6 contain $f_{\gamma_{e,1}}(t_1) f_{\gamma_{e,2}}(t_2)$ which decreases exponentially with the increases of $\gamma_{e,1}$ and $\gamma_{e,2}$. Furthermore, since

$$u_{N_e,m} - v_{N_e,m} = \sqrt{2\pi}\left(2^{\frac{2N_e}{m}} - 1\right)^{\frac{1}{2}} \Big/ \sqrt{m}, \tag{51}$$

φ_4, φ_5, and φ_6 decrease linearly with the increase of $\sqrt{m}$. When m is sufficiently large, $\mathbb{E}[\Psi(\gamma_e, N_e, m)]$ can be effectively approximated by $\mathbb{E}[Z_{N_e,m}(\gamma_{e,1})]\mathbb{E}[Z_{N_e,m}(\gamma_{e,2})]$. Since $Z_{N,m}(\max(t_1, t_2)) \geq Z_{N,m}(t_1) Z_{N,m}(t_2)$, we have $\mathbb{E}[\Psi(\gamma_e, N_e, m)] \geq \mathbb{E}[Z_{N_e,m}(\gamma_{e,1})]\mathbb{E}[Z_{N_e,m}(\gamma_{e,2})]$.

References

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

[2] 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. Tut.*, vol. 19, no. 2, pp. 721–742, Second Quarter 2016.

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

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

[5] Y. Li, M. Jiang, Q. Zhang, Q. Li, and J. Qin, "Cooperative non-orthogonal multiple access in multiple-input-multiple-output channels," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2068–2079, Mar. 2018.

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

[7] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of the incremental redundancy HARQ," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 529–532, Oct. 2014.

[8] Y. Gu, H. Chen, Y. Li, and B. Vucetic, "Ultra-reliable short-packet communications: Half-duplex or full-duplex relaying?" *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 348–351, Jun. 2018.

[9] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, Jul. 2018.

[10] 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, Mar. 2018.

[11] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321–331, Apr. 2003.