Cooperative NOMA with Incremental Relaying: Performance Analysis and Optimization

Guoxin Li, Deepak Mishra, *Member, IEEE*, and Hai Jiang, *Senior Member, IEEE*

***Abstract*—In conventional cooperative non-orthogonal multiple access (NOMA) networks, spectral efficiency loss occurs due to a half-duplex constraint. To address this issue, we propose an incremental cooperative NOMA (ICN) protocol for a two-user downlink network. In particular, this protocol allows the source to adaptively switch between a direct NOMA transmission mode and a cooperative NOMA transmission mode according to a 1-bit feedback from the far user. We analytically prove that the proposed ICN protocol outperforms the conventional cooperative NOMA protocol. In addition, an optimal power allocation strategy at the source is studied to minimize the asymptotic system outage probability. Finally, numerical results validate our theoretical analysis, present insights, and quantify the enhancement achieved over the benchmark scheme.**

***Index Terms*—Diversity order, incremental relaying, non-orthogonal multiple access, optimal power allocation, outage probability.**

I. Introduction

Due to the ability to serve multiple users simultaneously in a single resource block, non-orthogonal multiple access (NOMA) is a viable solution to fulfill the fifth-generation (5G) wireless networks' requirements of high spectrum efficiency (SE) and massive connectivity [1]. Accordingly, NOMA has been included in the study item on 5G new radio (NR) by 3GPP in its Release 15 [2].

A typical scenario of NOMA is that, when a source needs to send signals to two users (e.g., in a downlink cellular system), it sends both signals simultaneously as a superimposed signal. The user with better channel condition (the strong user) first decodes the weak user's signal, and then performs successive interference cancellation (SIC) and decodes its own signal. The weak user decodes its own signal directly. Since the strong user decodes the weak user's signal first, the work in [3] proposes a cooperative NOMA protocol in which the strong user works as a half-duplex (HD) relay to help the weak user. This conventional cooperative NOMA (CCN) protocol [3] promises to improve the weak user's performance by introducing a diversity gain. However, since the HD relay (the strong user) needs half of its time to forward information, the CCN protocol makes inefficient use of the degrees of freedom (DoF) of the channel and may cause a loss of SE (compared to a non-cooperative NOMA network). To efficiently exploit the DoF of the channel in a two-user downlink NOMA (TUDN) network, the work in [4] proposes a new cooperative protocol, termed as relaying with NOMA backhaul (R-NB). In this protocol, the source can adaptively adjust the time durations of NOMA transmission and relay transmission based on global instantaneous channel state information (CSI). However, global instantaneous CSI at the source may be difficult or costly to obtain in practice. This observation motivates us to propose a new and practically viable cooperative protocol for a TUDN network to improve SE of the CCN protocol.

Recall that in conventional cooperative networks, the incremental relaying (IR) protocol [5] is widely adopted since it can achieve higher SE by introducing a negligible 1-bit-feedback overhead. Specifically, the IR protocol invokes a relay for cooperation only when the source-to-destination channel gain is below a predetermined threshold. Inspired by this feature, in this correspondence we propose an incremental cooperative NOMA (ICN) protocol for a TUDN network with only statistical CSI at the source. In this protocol, the strong user works as a HD relay only when the weak user broadcasts a 1-bit negative feedback. The main contributions of this correspondence can be summarized as follows. 1) We propose a new and practical cooperative protocol for TUDN networks. To the best of our knowledge, the proposed ICN is the first time that the IR protocol is introduced into NOMA networks. 2) For the proposed ICN protocol, we derive exact or tightly approximated closed-form expressions of the outage probability (OP) of each user and the overall system. We prove that the ICN protocol outperforms the CCN protocol in terms of each user's OP and the system OP (SOP). 3) Asymptotic outage behavior of the ICN protocol is studied to derive the diversity order of each user and the optimal power allocation (OPA) strategy that minimizes the SOP. 4) Valuable insights regarding the ICN protocol are provided through detailed theoretical analysis and numerical results.

II. System Model

As shown in Fig. 1, we consider a TUDN scenario with a source (S) and two users: user 1 (U_1) is the near user while user 2 (U_2) is the far user. Similar to [6], [7], the two users are ordered according to their distance to S. Thus, U_1 and U_2 are treated as the strong user and the weak user, respectively. All the channels suffer Rayleigh fading. Let h_1, h_2 and h_3 denote the channel coefficients from S to U_1, S to U_2, and U_1 to U_2, respectively, where $h_i \sim \mathcal{CN}(0, \Omega_i)$ $(i = 1, 2, 3)$. We assume that channel coefficients remain unchanged during one transmission block, but may vary from one transmission block to another. Next we introduce the proposed ICN protocol in details.

A. Incremental Cooperative NOMA Protocol

At the beginning of each transmission block, S broadcasts a pilot signal to U_1 and U_2. Based on the received pilot signal, U_2 performs channel estimation of h_2 and compares it with a predefined threshold. If U_2 judges that it can correctly decode its desired message through direct transmission, it feedbacks a 1-bit positive acknowledgement (ACK) to S and U_1. After receiving the ACK feedback, S adopts a *direct NOMA transmission (DNT) mode*, i.e., it sends the superimposed signal

Manuscript received May 30, 2018; accepted August 29, 2018. The review of this paper was coordinated by Dr. K. Adachi (Corresponding author: Hai Jiang.)

G. Li and H. Jiang are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada (e-mail: guoxin@ualberta.ca, hai1@ualberta.ca).

D. Mishra is with the Division of Communication Systems, Department of Electrical Engineering (ISY), Linköping University, Linköping 58183, Sweden (e-mail: deepak.mishra@liu.se).

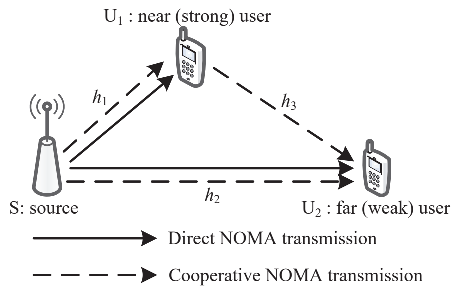

Fig. 1. System model.

to U_1 and U_2 within the whole transmission block. If U_2 finds that it is unable to decode its desired message without U_1's cooperation, it feedbacks a 1-bit negative acknowledge (NACK) to S and U_1. Upon hearing the NACK feedback, S adopts a *cooperative NOMA transmission (CNT) mode*, i.e., it broadcasts the superimposed signal in the first half of the transmission block, and then U_1 decodes U_2's message and forwards it in the second half of the transmission block.

To identify the difference between our proposed ICN and the CCN protocols, here we briefly review the CCN protocol [3]. In the CCN protocol, the transmission block is divided into two phases with equal duration. During the first phase, S sends the superimposed signal to U_1 and U_2, and U_1 decodes U_2's message and forwards it in the second phase. Compared to the CCN protocol, our proposed ICN protocol is essentially an adaptive protocol which can adaptively switch between the DNT mode and the CNT mode based on a 1-bit indicator.[1]

B. Signal Model

1) DNT Mode: S sends a superimposed signal to U_1 and U_2, which occupies the whole transmission block. The resulted signal at U_n is defined by

$$y_n = \sqrt{\alpha_1 P_s} h_n x_1 + \sqrt{\alpha_2 P_s} h_n x_2 + w_n, \ n = 1, 2, \quad (1)$$

where P_s is the transmit power of S, x_n denotes the message for U_n, α_n is the power allocation (PA) factor for x_n with $\alpha_1 + \alpha_2 = 1$, and w_n is the additive white Gaussian noise (AWGN) at U_n with zero mean and variance σ^2.

According to the NOMA principle, U_n first decodes x_2 upon observing y_n. Denote $\gamma_{n,2}$ as the received signal-to-interference-pulse-noise ratio (SINR) at U_n to decode x_2, and then $\gamma_{n,2}$ is given by $\gamma_{n,2} = \frac{\alpha_2 \rho_s |h_n|^2}{\alpha_1 \rho_s |h_n|^2 + 1}$, where $\rho_s = P_s/\sigma^2$ denotes the transmit signal-to-noise ratio (SNR) of S. After U_1 successfully decodes x_2 and performs SIC, the received SNR to detect x_1 at U_1, denoted by $\gamma_{1,1}$, is $\gamma_{1,1} = \alpha_1 \rho_s |h_1|^2$.

2) CNT Mode: Here the entire transmission block consists of two phases with equal duration. In the first phase, the received signal at U_n is the same as defined in (1), and the received SINR at U_n for message x_2 is also given as $\gamma_{n,2}$ defined in the DNT mode. If U_1 successfully decodes x_2 and performs SIC in the first phase, its received SNR to detect x_1 is given as $\gamma_{1,1}$ defined in the DNT mode. Then, in the second phase, U_1 forwards the re-encoded x_2 to U_2. The corresponding received signal at U_2 in the second phase can be expressed as $y_2' = \sqrt{P_r} h_3 x_2 + w_2$, where P_r is the transmit power of U_1. Finally, U_2 combines the observed signals y_2 and y_2' using the maximal ratio combining (MRC), and thus, the received SINR at U_2 to decode x_2 after MRC is given by $\gamma_{2,2}^{\mathrm{MRC}} = \frac{\alpha_2 \rho_s |h_2|^2}{\alpha_1 \rho_s |h_2|^2 + 1} + \rho_r |h_3|^2$, where $\rho_r = P_r/\sigma^2$ is U_1's transmit SNR.

III. Outage Performance Analysis and Optimization

For each user, an outage event happens when the received SINR (or SNR) is below a pre-determined decoding threshold. Note that the decoding thresholds of the DNT and the CNT modes are different. In the DNT mode, the decoding threshold is $\gamma_{\mathrm{th}} = 2^R - 1$ with R being the target rate of x_1 and x_2. In the CNT mode, the threshold is $\gamma'_{\mathrm{th}} = 2^{2R} - 1$.

A. Outage Probability Analysis

1) Near User: According to the ICN protocol, the OP of U_1 can be expressed as

$$\begin{aligned} P_1^{\mathrm{ICN}} = 1 &- \Pr\left\{\gamma_{2,2} \geq \gamma_{\mathrm{th}}, \gamma_{1,2} \geq \gamma_{\mathrm{th}}, \gamma_{1,1} \geq \gamma_{\mathrm{th}}\right\} \\ &- \Pr\left\{\gamma_{2,2} < \gamma_{\mathrm{th}}, \gamma_{1,2} \geq \gamma'_{\mathrm{th}}, \gamma_{1,1} \geq \gamma'_{\mathrm{th}}\right\}, \end{aligned} \quad (2)$$

where $\Pr\{\cdot\}$ means probability of an event, $\gamma_{2,2} \geq \gamma_{\mathrm{th}}$ indicates that the system works in the DNT mode, and $\gamma_{2,2} < \gamma_{\mathrm{th}}$ indicates that the system works in the CNT mode. As $\gamma_{2,2}$ is independent from $\gamma_{1,2}$ and $\gamma_{1,1}$, (2) can be rewritten as

$$\begin{aligned} P_1^{\mathrm{ICN}} = 1 &- \underbrace{\Pr\{\gamma_{2,2} \geq \gamma_{\mathrm{th}}\}}_{Q_1} \underbrace{\Pr\{\gamma_{1,2} \geq \gamma_{\mathrm{th}}, \gamma_{1,1} \geq \gamma_{\mathrm{th}}\}}_{Q_2} \\ &- \underbrace{\Pr\left\{\gamma_{2,2} < \gamma_{\mathrm{th}}\right\}}_{\bar{Q}_1} \underbrace{\Pr\left\{\gamma_{1,2} \geq \gamma'_{\mathrm{th}}, \gamma_{1,1} \geq \gamma'_{\mathrm{th}}\right\}}_{Q_3}, \end{aligned} \quad (3)$$

where $\bar{Q}_1 = 1 - Q_1$. It is easy to verify that $Q_1 = Q_2 = 0$ for $\frac{1}{1+\gamma_{\mathrm{th}}} \leq \alpha_1 < 1$, and $Q_3 = 0$ for $\frac{1}{1+\gamma'_{\mathrm{th}}} \leq \alpha_1 < 1$. Thus, $P_1^{\mathrm{ICN}} = 1$ for $\frac{1}{1+\gamma_{\mathrm{th}}} \leq \alpha_1 < 1$. When $0 < \alpha_1 < \frac{1}{1+\gamma_{\mathrm{th}}}$, Q_2 is given by

$$\begin{aligned} Q_2 &= \Pr\left\{|h_1|^2 \geq \frac{\gamma_{\mathrm{th}}}{\rho_s\left(\alpha_2 - \gamma_{\mathrm{th}}\alpha_1\right)}, \ |h_1|^2 \geq \frac{\gamma_{\mathrm{th}}}{\alpha_1 \rho_s}\right\} \\ &= \Pr\left\{|h_1|^2 \geq \frac{\gamma_{\mathrm{th}}}{\rho_s \Theta}\right\} = e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_1 \Theta}}, \end{aligned} \quad (4)$$

where $\Theta \triangleq \min\{\theta, \alpha_1\}$ and $\theta \triangleq \alpha_2 - \gamma_{\mathrm{th}}\alpha_1$. Q_2 is derived using the fact that $|h_i|^2$ $(i = 1, 2, 3)$ follows exponential distribution with mean Ω_i. Following similar steps, we have $Q_1 = e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_2 \theta}}$ for $0 < \alpha_1 < \frac{1}{1+\gamma_{\mathrm{th}}}$, and $Q_3 = e^{-\frac{\gamma'_{\mathrm{th}}}{\rho_s \Omega_1 \Theta'}}$ for $0 < \alpha_1 < \frac{1}{1+\gamma'_{\mathrm{th}}}$, where $\Theta' \triangleq \min\{\theta', \alpha_1\}$ and $\theta' \triangleq \alpha_2 - \gamma'_{\mathrm{th}}\alpha_1$. Substituting the results of Q_1, Q_2 and Q_3 into (3), a closed-form expression of U_1's OP is given by

$$P_1^{\mathrm{ICN}} = \begin{cases} 1 - e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s}\left(\frac{1}{\Omega_1 \Theta} + \frac{1}{\Omega_2 \theta}\right)} - e^{-\frac{\gamma'_{\mathrm{th}}}{\rho_s \Omega_1 \Theta'}} & \\ \quad + e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_2 \theta}} e^{-\frac{\gamma'_{\mathrm{th}}}{\rho_s \Omega_1 \Theta'}}, & 0 < \alpha_1 < \frac{1}{1+\gamma'_{\mathrm{th}}}, \\ 1 - e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s}\left(\frac{1}{\Omega_1 \Theta} + \frac{1}{\Omega_2 \theta}\right)}, & \frac{1}{1+\gamma'_{\mathrm{th}}} \leq \alpha_1 < \frac{1}{1+\gamma_{\mathrm{th}}}, \\ 1, & \frac{1}{1+\gamma_{\mathrm{th}}} \leq \alpha_1 < 1. \end{cases} \quad (5)$$

[1] In the CCN protocol, both S and U_1 need to send pilot signals, for channel estimation at the receiver side(s). In the ICN protocol, only S sends a pilot signal in the DNT mode, while both S and U_1 send pilot signals in the CNT mode. Thus, the signaling overhead of the two protocols are comparable to each other.

2) Far User: The OP of U_2 with the ICN protocol is given by

$$\begin{aligned} P_2^{\mathrm{ICN}} &= \Pr\{\gamma_{2,2}<\gamma_{\mathrm{th}},\gamma_{1,2}<\gamma'_{\mathrm{th}}\} \\ &\quad + \Pr\{\gamma_{2,2}<\gamma_{\mathrm{th}},\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{2,2}^{\mathrm{MRC}}<\gamma'_{\mathrm{th}}\} \\ &= \underbrace{\Pr\{\gamma_{2,2}<\gamma_{\mathrm{th}}\}}_{\bar{Q}_1}\underbrace{\Pr\{\gamma_{1,2}<\gamma'_{\mathrm{th}}\}}_{Q_4} \\ &\quad + \underbrace{\Pr\{\gamma_{1,2}\geq\gamma'_{\mathrm{th}}\}}_{\bar{Q}_4}\underbrace{\Pr\{\gamma_{2,2}<\gamma_{\mathrm{th}},\gamma_{2,2}^{\mathrm{MRC}}<\gamma'_{\mathrm{th}}\}}_{Q_5}, \end{aligned} \tag{6}$$

where $\bar{Q}_4=1-Q_4$. Similar to U_1's OP, the OP of U_2 is also segmented regarding α_1 as follows.

When $\frac{1}{1+\gamma_{\mathrm{th}}}\leq\alpha_1<1$, we have $P_2^{\mathrm{ICN}}=1$ since $\bar{Q}_1=Q_4=1$.

When $\frac{1}{1+\gamma'_{\mathrm{th}}}\leq\alpha_1<\frac{1}{1+\gamma_{\mathrm{th}}}$, we have $Q_4=1$ and thus, $P_2^{\mathrm{ICN}}=\bar{Q}_1=1-e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s\Omega_2\theta}}$, which is an increasing function of α_1.

Now we derive P_2^{ICN} over the region $\alpha_1\in\left(0,\frac{1}{1+\gamma'_{\mathrm{th}}}\right)$, where $Q_4=1-e^{-\frac{\gamma'_{\mathrm{th}}}{\rho_s\Omega_1\theta'}}$ and Q_5 can be derived as

$$\begin{aligned} Q_5 &= \Pr\left\{\frac{\alpha_2\rho_s|h_2|^2}{\alpha_1\rho_s|h_2|^2+1}<\gamma_{\mathrm{th}},\rho_r|h_3|^2+\frac{\alpha_2\rho_s|h_2|^2}{\alpha_1\rho_s|h_2|^2+1}<\gamma'_{\mathrm{th}}\right\} \\ &= \int_0^{\frac{\gamma_{\mathrm{th}}}{\rho_s\theta}} F_{|h_3|^2}\left(\frac{\gamma'_{\mathrm{th}}}{\rho_r}-\frac{\alpha_2\rho_s x}{\rho_r(\alpha_1\rho_s x+1)}\right)f_{|h_2|^2}(x)\,dx \\ &= 1-e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s\Omega_2\theta}}-\underbrace{\int_0^{\frac{\gamma_{\mathrm{th}}}{\rho_s\theta}} e^{-\frac{1}{\rho_r\Omega_3}\left(\gamma'_{\mathrm{th}}-\frac{\alpha_2\rho_s x}{\alpha_1\rho_s x+1}\right)}\frac{1}{\Omega_2}e^{-\frac{x}{\Omega_2}}\,dx}_{Q_6}. \end{aligned} \tag{7}$$

Here $F_Z(\cdot)$ and $f_Z(\cdot)$ are cumulative distribution function and probability density function of random variable Z. Though it is difficult to derive a closed-form expression for Q_6, we can obtain an approximation for it. By replacing the variable $x=\frac{\gamma_{\mathrm{th}}}{2\rho_s\theta}(t+1)$ in Q_6 and using Gaussian-Chebyshev quadrature [8, Eq. 25.4.38], we have

$$\begin{aligned} Q_6 &= \frac{\gamma_{\mathrm{th}}}{2\rho_s\Omega_2\theta}\int_{-1}^{1} e^{-\frac{g(t)}{\rho_r\Omega_3}}e^{-\frac{\gamma_{\mathrm{th}}(t+1)}{2\rho_s\Omega_2\theta}}dt \\ &\approx \frac{\gamma_{\mathrm{th}}}{2\rho_s\Omega_2\theta}\frac{\pi}{K}\sum_{k=1}^{K}\sqrt{1-\xi_k^2}\,e^{-\frac{g(\xi_k)}{\rho_r\Omega_3}}e^{-\frac{\gamma_{\mathrm{th}}(\xi_k+1)}{2\rho_s\Omega_2\theta}}, \end{aligned} \tag{8}$$

where K is a parameter to balance accuracy and complexity, $\xi_k=\cos\left(\frac{2k-1}{2K}\pi\right)$, and $g(x)=\gamma'_{\mathrm{th}}-\frac{\gamma_{\mathrm{th}}(x+1)\alpha_2}{\gamma_{\mathrm{th}}(x+1)\alpha_1+2\theta}$. Substituting (8) into (7), we can obtain an approximation of Q_5.

Combining the results for Q_1, Q_4 and Q_5, and after some algebraic manipulations, a closed-form expression of approximated P_2^{ICN} over the region $\alpha_1\in\left(0,\frac{1}{1+\gamma'_{\mathrm{th}}}\right)$ is given by

$$P_2^{\mathrm{ICN}}\approx 1-e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s\Omega_2\theta}}-e^{-\frac{\gamma'_{\mathrm{th}}}{\rho_s\Omega_1\theta'}}Q_6, \tag{9}$$

where Q_6 is given by (8).

From the above derivations, we know that P_1^{ICN} and P_2^{ICN} are both equal to 1 when $\frac{1}{1+\gamma_{\mathrm{th}}}\leq\alpha_1<1$. Thus, in the sequel we only focus on the remaining region, i.e., $0<\alpha_1<\frac{1}{1+\gamma_{\mathrm{th}}}$.

3) Overall System: Similar to [3], the system outage is defined as the event when one user or both users in the system are in outage. Thus, the SOP with the ICN protocol can be expressed as

$$\begin{aligned} P_{1\&2}^{\mathrm{ICN}} &= 1-\Pr\{\gamma_{2,2}\geq\gamma_{\mathrm{th}},\gamma_{1,2}\geq\gamma_{\mathrm{th}},\gamma_{1,1}\geq\gamma_{\mathrm{th}}\} \\ &\quad -\Pr\{\gamma_{2,2}<\gamma_{\mathrm{th}},\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{1,1}\geq\gamma'_{\mathrm{th}},\gamma_{2,2}^{\mathrm{MRC}}\geq\gamma'_{\mathrm{th}}\}. \end{aligned} \tag{10}$$

Following similar procedures to those in the derivations of P_1^{ICN} and P_2^{ICN}, a closed-form approximation of the SOP can be given as

$$P_{1\&2}^{\mathrm{ICN}}=\begin{cases} 1-e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s}\left(\frac{1}{\Omega_1\Theta}+\frac{1}{\Omega_2\theta}\right)}-Q_6 e^{-\frac{\gamma'_{\mathrm{th}}}{\rho_s\Omega_1\Theta'}}, & 0<\alpha_1<\frac{1}{1+\gamma'_{\mathrm{th}}}, \\ 1-e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s}\left(\frac{1}{\Omega_1\Theta}+\frac{1}{\Omega_2\theta}\right)}, & \frac{1}{1+\gamma'_{\mathrm{th}}}\leq\alpha_1<\frac{1}{1+\gamma_{\mathrm{th}}}, \end{cases} \tag{11}$$

where Q_6 is given by (8). Comparing the expressions of P_1^{ICN} and $P_{1\&2}^{\mathrm{ICN}}$ given in (5) and (11), respectively, we find that the OP of U_1 is identical to the SOP when $\alpha_1\in\left[\frac{1}{1+\gamma'_{\mathrm{th}}},\frac{1}{1+\gamma_{\mathrm{th}}}\right)$. In other words, when the overall system is in outage, it also means that U_1 is in outage. This is due to the following two facts: 1) The system works in the DNT mode only when U_2 can correctly decode its desired information (which means that U_2 has no outage). In this case, U_1 in outage also leads to an outage of the overall system. 2) When U_2 requests cooperation (which indicates that the target rate of U_2 cannot be achieved in the DNT mode), if $\alpha_1\in\left[\frac{1}{1+\gamma'_{\mathrm{th}}},\frac{1}{1+\gamma_{\mathrm{th}}}\right)$, we have $\gamma_{1,2}<\gamma'_{\mathrm{th}}$, i.e., U_1 fails to decode x_2, which results in an outage at both U_1 and U_2.

B. Outage Performance Comparison with the CCN protocol

We denote the OP of U_1, U_2, and the overall system in the CCN protocol by P_1^{CCN}, P_2^{CCN} and $P_{1\&2}^{\mathrm{CCN}}$, respectively. Following the CCN protocol details from [3] along with the expressions of P_1^{ICN}, P_2^{ICN} and $P_{1\&2}^{\mathrm{ICN}}$ given in (2), (6) and (10), respectively, we have

$$\begin{aligned} P_1^{\mathrm{ICN}} &< 1-\Pr\{\gamma_{2,2}\geq\gamma_{\mathrm{th}},\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{1,1}\geq\gamma'_{\mathrm{th}}\} \\ &\quad -\Pr\{\gamma_{2,2}<\gamma_{\mathrm{th}},\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{1,1}\geq\gamma'_{\mathrm{th}}\} \\ &= 1-\Pr\{\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{1,1}\geq\gamma'_{\mathrm{th}}\}=P_1^{\mathrm{CCN}}, \end{aligned} \tag{12}$$

$$\begin{aligned} P_2^{\mathrm{ICN}} &< \Pr\{\gamma_{2,2}<\gamma'_{\mathrm{th}},\gamma_{1,2}<\gamma'_{\mathrm{th}}\} \\ &\quad +\Pr\{\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{2,2}^{\mathrm{MRC}}<\gamma'_{\mathrm{th}}\}=P_2^{\mathrm{CCN}}, \end{aligned} \tag{13}$$

and

$$\begin{aligned} P_{1\&2}^{\mathrm{ICN}} &< 1-\Pr\{\gamma_{2,2}\geq\gamma_{\mathrm{th}},\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{1,1}\geq\gamma'_{\mathrm{th}},\gamma_{2,2}^{\mathrm{MRC}}\geq\gamma'_{\mathrm{th}}\} \\ &\quad -\Pr\{\gamma_{2,2}<\gamma_{\mathrm{th}},\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{1,1}\geq\gamma'_{\mathrm{th}},\gamma_{2,2}^{\mathrm{MRC}}\geq\gamma'_{\mathrm{th}}\} \\ &= 1-\Pr\{\gamma_{1,2}\geq\gamma'_{\mathrm{th}},\gamma_{1,1}\geq\gamma'_{\mathrm{th}},\gamma_{2,2}^{\mathrm{MRC}}\geq\gamma'_{\mathrm{th}}\}=P_{1\&2}^{\mathrm{CCN}}. \end{aligned} \tag{14}$$

Therefore, it can be concluded that the ICN protocol outperforms the CCN protocol in terms of each user's OP and the SOP.

C. SOP Minimization and Diversity Order Analysis

In this subsection, we first investigate the asymptotic outage performance of the ICN protocol when $\rho_s\to\infty$ and $\rho_r=\lambda\rho_s$ with $0<\lambda\leq 1$. Based on the asymptotic analysis, an OPA strategy that minimizes the SOP is developed, and the diversity order of each user is derived as well.

1) SOP minimization: As $\rho_s \to \infty$, we have $\gamma_{2,2}^{\mathrm{MRC}} \to \frac{\alpha_2}{\alpha_1} + \rho_r |h_3|^2 > \gamma'_{\mathrm{th}}$ for $0 < \alpha_1 < \frac{1}{1+\gamma'_{\mathrm{th}}}$, which indicates that $\Pr\left\{\gamma_{2,2}^{\mathrm{MRC}} > \gamma'_{\mathrm{th}}\right\} \to 1$, and thus, $P_{1\&2}^{\mathrm{ICN}}$ converges to P_1^{ICN} based on (2) and (10). Together with the fact that U_1's OP is identical to the SOP when $\alpha_1 \in \left[\frac{1}{1+\gamma'_{\mathrm{th}}}, \frac{1}{1+\gamma_{\mathrm{th}}}\right)$, it can be concluded that the SOP converges to U_1's OP as $\rho_s \to \infty$. Noting this key observation, in the following we focus on the minimization of U_1's OP.

When $\rho_s \to \infty$, applying $e^{-x} \overset{x\to 0}{\simeq} 1-x$ into (5), we can derive the asymptotic OP of U_1 as

$$P_{1,\mathrm{asy}}^{\mathrm{ICN}} \simeq \begin{cases} \frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_1 \Theta} + \frac{\gamma_{\mathrm{th}} \gamma'_{\mathrm{th}}}{\rho_s^2 \Omega_1 \Omega_2 \theta \Theta'}, 0 < \alpha_1 < \frac{1}{1+\gamma'_{\mathrm{th}}}, \\ \frac{\gamma_{\mathrm{th}}}{\rho_s} \left(\frac{1}{\Omega_1 \Theta} + \frac{1}{\Omega_2 \theta}\right), \frac{1}{1+\gamma'_{\mathrm{th}}} \le \alpha_1 < \frac{1}{1+\gamma_{\mathrm{th}}}. \end{cases} \tag{15}$$

Substituting the expressions of θ, Θ, and Θ' into (15), $P_{1,\mathrm{asy}}^{\mathrm{ICN}}$ can be further expressed as

$$P_{1,\mathrm{asy}}^{\mathrm{ICN}} \simeq \begin{cases} \frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_1} f_1(\alpha_1), 0 < \alpha_1 < \frac{1}{2+\gamma'_{\mathrm{th}}}, \\ \frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_1} f_2(\alpha_1), \frac{1}{2+\gamma'_{\mathrm{th}}} \le \alpha_1 < \min\left\{\frac{1}{2+\gamma_{\mathrm{th}}}, \frac{1}{1+\gamma'_{\mathrm{th}}}\right\}, \\ \frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_1} f_3(\alpha_1), \min\left\{\frac{1}{2+\gamma_{\mathrm{th}}}, \frac{1}{1+\gamma'_{\mathrm{th}}}\right\} \le \alpha_1 < \frac{1}{1+\gamma'_{\mathrm{th}}}, \\ \frac{\gamma_{\mathrm{th}}}{\rho_s} f_4(\alpha_1), \frac{1}{1+\gamma'_{\mathrm{th}}} \le \alpha_1 < \max\left\{\frac{1}{2+\gamma_{\mathrm{th}}}, \frac{1}{1+\gamma'_{\mathrm{th}}}\right\}, \\ \frac{\gamma_{\mathrm{th}}}{\rho_s} f_5(\alpha_1), \max\left\{\frac{1}{2+\gamma_{\mathrm{th}}}, \frac{1}{1+\gamma'_{\mathrm{th}}}\right\} \le \alpha_1 < \frac{1}{1+\gamma_{\mathrm{th}}}, \end{cases} \tag{16}$$

in which we have

$$f_1(\alpha_1) = \frac{1}{\alpha_1} + \frac{\gamma'_{\mathrm{th}}}{\rho_s \Omega_2 (1 - \alpha_1 (1+\gamma_{\mathrm{th}})) \alpha_1}, \tag{17}$$

$$f_2(\alpha_1) = \frac{1}{\alpha_1} + \frac{\gamma'_{\mathrm{th}}}{\rho_s \Omega_2 (1-\alpha_1(1+\gamma_{\mathrm{th}}))(1-\alpha_1(1+\gamma'_{\mathrm{th}}))}, \tag{18}$$

$$f_3(\alpha_1) = \frac{1}{1-\alpha_1(1+\gamma_{\mathrm{th}})} \left[1 + \frac{\gamma'_{\mathrm{th}}}{\rho_s \Omega_2 (1-\alpha_1(1+\gamma'_{\mathrm{th}}))}\right], \tag{19}$$

$$f_4(\alpha_1) = \frac{1}{\Omega_1 \alpha_1} + \frac{1}{\Omega_2 (1 - \alpha_1 (1+\gamma_{\mathrm{th}}))}, \tag{20}$$

$$f_5(\alpha_1) = \frac{1}{1 - \alpha_1 (1+\gamma_{\mathrm{th}})} \left(\frac{1}{\Omega_1} + \frac{1}{\Omega_2}\right). \tag{21}$$

For $f_1(\alpha_1)$: It can be shown that $\frac{1}{\alpha_1(1-\alpha_1(1+\gamma_{\mathrm{th}}))}$ monotonically decreases with $\alpha_1 \in \left(0, \frac{1}{2+2\gamma_{\mathrm{th}}}\right)$. Thus, $f_1(\alpha_1)$ is a decreasing function over $\alpha_1 \in \left(0, \frac{1}{2+\gamma'_{\mathrm{th}}}\right)$ since $\frac{1}{2+\gamma'_{\mathrm{th}}} \le \frac{1}{2+2\gamma_{\mathrm{th}}}$.

For $f_2(\alpha_1)$: $f_2(\alpha_1)$ is a convex function of α_1 due to the facts that $\frac{1}{\alpha_1}$, $\frac{1}{1-\alpha_1(1+\gamma_{\mathrm{th}})}$ and $\frac{1}{1-\alpha_1(1+\gamma'_{\mathrm{th}})}$ are convex functions of α_1 and that the sum of convex functions is still a convex function. The first-order derivative of $f_2(\alpha_1)$ is given by

$$\frac{\mathrm{d} f_2(\alpha_1)}{\mathrm{d}\alpha_1} = -\frac{1}{\alpha_1^2} + \frac{a(b(1-c\alpha_1) + c(1-b\alpha_1))}{((1-b\alpha_1)(1-c\alpha_1))^2}, \tag{22}$$

where $a = \frac{\gamma'_{\mathrm{th}}}{\rho_s \Omega_2}$, $b = 1 + \gamma_{\mathrm{th}}$ and $c = 1 + \gamma'_{\mathrm{th}}$. From (22), we can easily verify that $\frac{\mathrm{d} f_2(\alpha_1)}{\mathrm{d}\alpha_1}|_{\alpha_1 \to 0} < 0$ and $\frac{\mathrm{d} f_2(\alpha_1)}{\mathrm{d}\alpha_1}|_{\alpha_1 \to \frac{1}{1+\gamma'_{\mathrm{th}}}} > 0$. Since $f_2(\alpha_1)$ is a convex function, the critical point of $f_2(\alpha_1)$, denoted as δ, must lie in the interval $\left(0, \frac{1}{1+\gamma'_{\mathrm{th}}}\right)$, and is the root of $\frac{\mathrm{d} f_2(\alpha_1)}{\mathrm{d}\alpha_1} = 0$ that falls in $\left(0, \frac{1}{1+\gamma'_{\mathrm{th}}}\right)$.[2] Thus, for $\frac{1}{2+\gamma'_{\mathrm{th}}} \le \alpha_1 < \min\left\{\frac{1}{2+\gamma_{\mathrm{th}}}, \frac{1}{1+\gamma'_{\mathrm{th}}}\right\}$, the minimal point of $P_{1,\mathrm{asy}}^{\mathrm{ICN}}$ is at $\alpha_1 = \beta_1$ with $\beta_1 \triangleq \max\left\{\frac{1}{2+\gamma'_{\mathrm{th}}}, \min\left\{\delta, \frac{1}{2+\gamma_{\mathrm{th}}}\right\}\right\}$.

For $f_3(\alpha_1)$: $f_3(\alpha_1)$ is an increasing function of α_1.

For $f_4(\alpha_1)$: Like $f_2(\alpha_1)$, $f_4(\alpha_1)$ is also a convex function of α_1, whose critical point can be obtained as $\alpha_1 = \frac{1}{1+\psi+\gamma_{\mathrm{th}}}$, where $\psi = \sqrt{\frac{\Omega_1(1+\gamma_{\mathrm{th}})}{\Omega_2}}$. Thus, for $\frac{1}{1+\gamma'_{\mathrm{th}}} \le \alpha_1 < \max\left\{\frac{1}{2+\gamma_{\mathrm{th}}}, \frac{1}{1+\gamma'_{\mathrm{th}}}\right\}$, the minimal point of $P_{1,\mathrm{asy}}^{\mathrm{ICN}}$ is at $\alpha_1 = \beta_2$ with $\beta_2 \triangleq \max\left\{\frac{1}{1+\psi+\gamma_{\mathrm{th}}}, \frac{1}{1+\gamma'_{\mathrm{th}}}\right\}$.

For $f_5(\alpha_1)$: $f_5(\alpha_1)$ is an increasing function of α_1.

Combing all above observations, we conclude that $P_{1,\mathrm{asy}}^{\mathrm{ICN}}$ achieves its global minimum value at $\alpha_1 = \beta_1$ if $\frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_1} f_2(\beta_1) < \frac{\gamma_{\mathrm{th}}}{\rho_s} f_4(\beta_2)$, or at $\alpha_1 = \beta_2$ otherwise.

2) Diversity order of each user: From (15), we can observe that the diversity order of U_1 is 1, which is the full diversity order for U_1.

As $\rho_s \to \infty$, the asymptotic OP of U_2 over the region $\alpha_1 \in \left[\frac{1}{1+\gamma'_{\mathrm{th}}}, \frac{1}{1+\gamma_{\mathrm{th}}}\right)$ can be easily derived as $P_{2,\mathrm{asy}}^{\mathrm{ICN}} = \bar{Q}_1 \simeq \frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_2 \theta}$, which illustrates that the diversity order of U_2 in this region is 1. The reason for the diversity loss is that in this region of α_1, U_1 cannot work in the cooperative mode since $\gamma_{1,2} < \gamma'_{\mathrm{th}}$, and thus, it fails to provide assistance to U_2.

Now we focus on the derivation of $P_{2,\mathrm{asy}}^{\mathrm{ICN}}$ when $0 < \alpha_1 < \frac{1}{1+\gamma'_{\mathrm{th}}}$. As $\rho_s \to \infty$, Q_6 in (7) can be approximated as

$$\begin{aligned} Q_6 &\overset{\text{(i)}}{\simeq} \int_0^{\frac{\gamma_{\mathrm{th}}}{\rho_s \theta}} \left(1 - \frac{1}{\rho_r \Omega_3}\left(\gamma'_{\mathrm{th}} - \frac{\alpha_2 \rho_s x}{(\alpha_1 \rho_s x + 1)}\right)\right) \frac{1}{\Omega_2} e^{-\frac{x}{\Omega_2}} dx \\ &\overset{\text{(ii)}}{\simeq} \left(1 - e^{-\frac{\gamma_{\mathrm{th}}}{\rho_s \Omega_2 \theta}}\right)\left(1 - \frac{\gamma'_{\mathrm{th}}}{\rho_r \Omega_3}\right) + \frac{\gamma_{\mathrm{th}}}{2\rho_s \rho_r \Omega_2 \Omega_3 \theta} \frac{\pi}{K} \\ &\times \sum_{k=1}^{K} \sqrt{1-\xi_k^2} \frac{\alpha_2 \gamma_{\mathrm{th}} (\xi_k + 1)}{\alpha_1 \gamma_{\mathrm{th}} (\xi_k + 1) + 2\theta} e^{-\frac{\gamma_{\mathrm{th}}(\xi_k+1)}{2\rho_s \Omega_2 \theta}}, \end{aligned} \tag{23}$$

where step (i) is obtained by using $e^{-x} \overset{x\to 0}{\simeq} 1-x$, and step (ii) is achieved by applying the Gaussian-Chebyshev quadrature. Now substituting (23) into (7) and applying $e^{-x} \overset{x\to 0}{\simeq} 1-x$ again, we have $Q_5 \simeq \frac{\gamma_{\mathrm{th}} \Xi}{\lambda \rho_s^2 \Omega_2 \Omega_3 \theta}$, where Ξ is given by

$$\Xi = \gamma'_{\mathrm{th}} - \frac{\pi}{2K} \sum_{k=1}^{K} \sqrt{1-\xi_k^2} \frac{\alpha_2 \gamma_{\mathrm{th}} (\xi_k + 1)}{\alpha_1 \gamma_{\mathrm{th}} (\xi_k + 1) + 2\theta}. \tag{24}$$

In addition, an approximation of $\bar{Q}_1 Q_4$ in (6) can be easily obtained as $\bar{Q}_1 Q_4 \simeq \frac{\gamma_{\mathrm{th}} \gamma'_{\mathrm{th}}}{\rho_s^2 \Omega_1 \Omega_2 \theta \theta'}$. To this end, by combining the approximate results for $\bar{Q}_1 Q_4$ and Q_5, the asymptotic OP of U_2 over the region $\alpha_1 \in \left(0, \frac{1}{1+\gamma'_{\mathrm{th}}}\right)$ is given by

$$P_{2,\mathrm{asy}}^{\mathrm{ICN}} \simeq \frac{1}{\rho_s^2} \left(\frac{\gamma_{\mathrm{th}} \gamma'_{\mathrm{th}}}{\Omega_1 \Omega_2 \theta \theta'} + \frac{\gamma_{\mathrm{th}} \Xi}{\lambda \Omega_2 \Omega_3 \theta}\right). \tag{25}$$

According to (25), it is clear that in region $\alpha_1 \in \left(0, \frac{1}{1+\gamma'_{\mathrm{th}}}\right)$, U_2 achieves its full diversity order of two.

[2] Note that $\frac{\mathrm{d} f_2(\alpha_1)}{\mathrm{d}\alpha_1}$ can be transformed to a quartic function of α_1, and the procedures in [9] can be used to find closed-form roots of $\frac{\mathrm{d} f_2(\alpha_1)}{\mathrm{d}\alpha_1} = 0$.

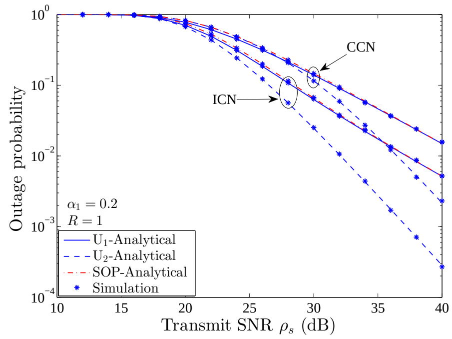

Fig. 2. Outage performance of the ICN and CCN protocols.

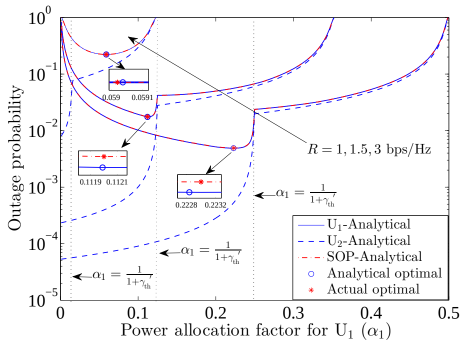

Fig. 3. Outage performance of the ICN protocol for varying α_1 ($\rho_s = 40$dB).

IV. Numerical Results

Now numerical investigation is carried out to verify the analytical results and present some non-trivial design insights. Unless otherwise specified, the following parameters are used: $\Omega_1 = \Omega_3 = 0.1$, $\Omega_2 = 0.01$, $\rho_s = \rho_r$, and $K = 10$.

Fig. 2 compares outage performance of the proposed ICN protocol against the CCN protocol.[3] A close match between the analytical and simulation results in Fig. 2 verifies the accuracy of our analysis. Fig. 2 also shows that both the ICN and CCN protocols achieve a full diversity order for each user. Further, we can observe that the proposed ICN protocol is superior to the CCN protocol in terms of each user's OP and the SOP, which is consistent with our analysis in Section III-B.

We define *performance gain* of the ICN protocol relative to the CCN protocol as $\mathrm{G}\,(\%) = 100 \times \left(1 - \frac{P_\Delta^{\mathrm{ICN}}}{P_\Delta^{\mathrm{CCN}}}\right)$, where $\Delta \in \{1, 2, 1\&2\}$. In our numerical results with $\alpha_1 = 0.2$, $R = 1$ bps/Hz, and $\rho_s = 30$dB, performance gains of U_1, U_2, and the system are $(12.3, 17.7, 11.9)$ when $\Omega_2 = 0.001$, $(46.6, 68.5, 46.8)$ when $\Omega_2 = 0.005$, and $(55.0, 78.8, 55.2)$ when $\Omega_2 = 0.01$. It is obvious that U_2 has the highest performance gain, while the performance gains of U_1 and the system are almost the same. Note that this observation is also verified by Fig. 2. All the performance gains shrink as Ω_2 decreases, because S in the ICN protocol tends to transmit information in the CNT mode as the channel from S to U_2 deteriorates.

[3]Here we compare our ICN protocol with the CCN protocol as only statistical CSI is needed in both protocols. If global instantaneous CSI is available, better outage performance can be achieved (e.g., the R-NB protocol with optimal block length allocation in [4]).

Fig. 3 investigates the impact of power allocation factor α_1 on the outage performance of the network. It can be observed that the OP of U_2 increases with α_1, while the OP of U_1 first decreases and then increases with α_1. The reasons are as follows. With a higher α_1, α_2 is lower, and thus, the chance that U_2 can successfully decode its information in the DNT mode is lower. Further, in the CNT mode, a lower α_2 means the chance that U_1 correctly decodes U_2's message is lower, and thus, the chance that U_1 can help U_2 to achieve U_2's target rate is lower. Therefore, the OP of U_2 increases with α_1. The OP of U_1 is affected by two factors as follows. Factor 1: A higher α_1 means more power for U_1's signal, which tends to decrease its OP. Factor 2: As aforementioned, a higher α_1 also means the chance that U_1 correctly decodes U_2's message is lower, or in other words, the chance that U_1 performs SIC is lower, which tends to increase U_1's OP. When α_1 is low, Factor 1 dominates, and thus, U_1's OP decreases with α_1. When α_1 increases beyond a point, Factor 2 dominates, and thus, U_1's OP increases with α_1. From Fig. 3, we can see that the analytical approximation of the optimal α_1 (which minimizes $P_{1,\mathrm{asy}}^{\mathrm{ICN}}$) is close to the actual optimal value (which is the point of α_1 that minimizes the SOP). It is worth noticing that when $R = 3\mathrm{bps/Hz}$, the optimal α_1 lies in the region $\left[\frac{1}{1+\gamma'_{\mathrm{th}}}, \frac{1}{1+\gamma_{\mathrm{th}}}\right)$, which indicates that to minimize SOP, the system should stay in the DNT mode in this case. When $R = 1\mathrm{bps/Hz}$ and $R = 1.5\mathrm{bps/Hz}$, the optimal α_1 is smaller than $\frac{1}{1+\gamma'_{\mathrm{th}}}$, and thus, the best system outage performance is achieved by adaptively switching its transmission mode according to the quality of direct link to U_2.

V. Conclusion

We have proposed a cooperative protocol for TUDN networks. We have analytically proved that the proposed ICN protocol outperforms the CCN protocol. Numerical results have validated our analysis and demonstrated valuable insights.

References

[1] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[2] 3GPP, "New study item proposal: Study on non-orthogonal multiple access for NR," 3GPP, Dubrovnik, Croatia, Tech. Rep. RP-170829, Mar. 2017.

[3] 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.

[4] Y. Hu, M. C. Gursoy, and A. Schmeink, "Efficient transmission schemes for low-latency networks: NOMA vs. relaying," in *Proc. IEEE PIMRC*, Oct. 2017, pp. 1–6.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

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

[7] N. T. Do, D. B. D. Costa, T. Q. Duong, and B. An, "A BNBF user selection scheme for NOMA-based cooperative relaying systems with SWIPT," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 664–667, Mar. 2017.

[8] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, NY, USA: Dover, 1972.

[9] S. Neumark, *Solution of Cubic and Quartic Equations*. Oxford, U.K.: Pergamon Press, 1965.