Energy Efficient Full-Duplex Cooperative Non-Orthogonal Multiple Access

Zhongxiang Wei, *Member, IEEE,* Xu Zhu, *Senior Member, IEEE,* Sumei Sun, *Fellow, IEEE,* Jingjing Wang, *Student Member, IEEE,* Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Full-duplex (FD) cooperative non-orthogonal multiple access (NOMA) achieves superior throughput over conventional half-duplex (HD) cooperative NOMA, where the strong users (SUs) with good channel conditions can act as an FD relay node for the weak users (WUs) with poor channel conditions. However, the energy efficiency (EE) of cooperative NOMA may be degraded due to additional power consumption incurred at the SUs. We are therefore motivated to investigate the EE maximization problem of an FD cooperative NOMA system. More importantly, we investigate the "signal-to-inference-noise ratio (SINR) gap reversal" problem of cooperative NOMA systems, which imposes successive interference cancellation (SIC) performance degradation at the SUs. This problem has not been documented in the exiting cooperative NOMA literature. A low-complexity algorithm is proposed for maximizing the system's EE while guaranteeing successful SIC operation. Our numerical results show that the proposed algorithm achieves both higher EE and throughput over the existing HD cooperative NOMA and non-adaptive FD cooperative NOMA. More importantly, the proposed scheme guarantees a successful SIC operation at the SUs.**

***Index Terms*—Cooperative non-orthogonal multiple access, full-duplex, energy efficiency**

I. Introduction

Non-orthogonal multiple access (NOMA) techniques have received considerable attention in the context of next-generation communications as a benefit of their increased sum-rate. The key concept of NOMA is that of allowing multiple users to occupy the same frequency-, time- or code-resource [1]. The user having better channel conditions performs successive interference cancellation (SIC). Explicitly, since the weaker signal imposes limited interference on the stronger signal, the latter is detected first. The resultant symbols are remodulated and they are subtracted from the composite signals, which results in the decontaminated weaker signal. This refreshed signal can then for example be relayed to the distant user. In this spirit, the throughput of cooperative NOMA was investigated by Kim [2]. As a further development, the outage probability was provided in [3], where the base station (BS) broadcasts superimposed signals to both the stronger user (SU) and to the weaker user (WU) in the first time slot, while the SU forwards the WU's refreshed signal to the WU in the second time slot. However, the aforementioned cooperative NOMA assume that the SU operates in half-duplex (HD) mode and hence an additional time slot is required, which halves the throughput. As a remedy, full-duplex (FD) can be invoked by cooperative NOMA to avoid halving the rate by HD. By applying self-interference (SI) mitigation at the SU, the throughput of FD cooperative NOMA potentially doubles that of HD cooperative NOMA [4] [5] [6] [7] [8].

Nevertheless, there are more fundamental challenges to be addressed in cooperative NOMA systems: (a) The signal-to-interference plus noise-ratio (SINR) experienced at the WU can indeed be enhanced. However, the WU's SINR observed at the SU is reduced by the residual SI caused by the FD operation. As a result, the SINR difference between the SU and the WU is reduced and may even become negative. Then the original SU becomes the new WU and hence fails to perfectly decode the original WU's signal. We refer to this undesired phenomenon as "SINR gap reversal", which becomes more severe when the relaying power at the SU keeps on increasing, as shown in Fig. 1. (b) In cooperative NOMA systems, extra transmission power plus extra transmit circuit power are required for relaying the WU's signal. Hence, achieving a high energy efficiency (EE) remains challenging. In contrast to conventional single-component throughput optimization, it is more beneficial to strike a trade-off between the throughput and EE. EE maximization [9] and power minimization [10] have been analyzed in non-cooperative NOMA, where the overwhelming majority of the transmission power is assigned to the distant WU for maintaining its SINR target. As a result, low throughput and EE are achieved due to the poor channel conditions between the BS and the WU. (c) The FD cooperative NOMA systems of [5] [6] assumed non-adaptive power allocation schemes, which are unaware of the users' channel conditions and suffer from significant throughput loss. More explicitly, given a large distance between the SU and WU, assigning high transmission power to the SU may not substantially improve the WU's throughput, while significantly degrading the EE of the SU.

Motivated by the above open issues, we propose a novel EE-oriented FD relay protocol for the downlink of NOMA

Z. Wei and X. Zhu are with the University of Liverpool, Liverpool, UK. (e-mail:{hszwei,xuzhu}@liverpool.ac.uk)

S. Sun is with the Institute for Infocom Research, A*STAR, Singapore. (e-mail:sunsm@i2r.a-star.edu.sg)

J. Wang is with the Department of Electronic Engineering, Tsinghua University, Beijing, China. (e-mail:chinaeephd@gmail.com)

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton, UK. L. Hanzo would like to acknowledge the ERC's financial support of his Advanced Fellow Grant QuantCom. (e-mail: lh@ecs.soton.ac.uk)

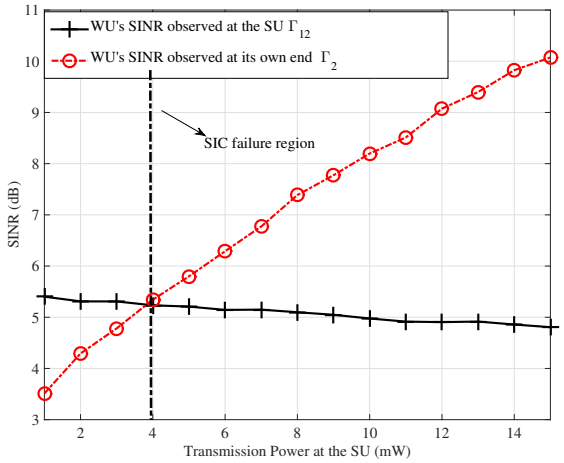

Fig. 1. The illustration of the "SINR gap reversal" in cooperative NOMA. Specifically, when the relaying power at the SU is higher than 4 mW, the SU fails to perfectly decode the WU's signal due to the "SINR gap reversal".

systems. This treatise has the following novel contributions:

a) It is the first contribution addressing the "SINR gap reversal" issue of cooperative NOMA systems, so that successful SIC operation is guaranteed at the SU resulting in a reduced SIC performance degradation compared to [5] and [6]. Additionally, we derive the closed form expression of the cooperative transmission power required at the SU.

b) We strike a compelling trade-off between the throughput and power consumption. This is in explicit juxtaposition to the existing work on throughput maximization [2] alone or to the total power consumption minimization alone [11]. A range of practical aspects are taken into account in the optimization, such as the residual SI at the FD SU. Furthermore, the power amplifier's (PA) dissipated circuit power and SI-mitigation power are all included in our power consumption model.

c) A low-complexity algorithm is proposed for EE maximization, where the power allocation at the BS and the SU is adaptively controlled rather than being fixed as in [5] and [6]. As a result, a more beneficial EE vs throughput is struck compared to the designs of [5] and [6].

II. System Model

We consider a NOMA system supporting two users by a BS. The users have different channel conditions on the same frequency, which is a common scenario in NOMA systems [3]. The BS, SU and WU are equipped with a single antenna for the sake of having a low hardware complexity. In particular, the SU acts as a decode-and-forward relay, which cooperates with the WU. FD operation is enabled at the SU, which can transmit and receive signals at the same time based on the state-of-the-art shared antenna technique of [12]. The Channel State Information (CSI) is obtained by channel estimation in the training phase based on the channels reciprocity [9]. Explicitly, during the training phase, the BS estimates the downlink CSI by inferring it from the channel quality indicator feedback received from the SU and WU. Hence, the BS can regularly update the CSI of h_1 and h_2. Furthermore, since the SU is aware of its own channel quality indicator and of the WU, the CSI of h_r and h_{12} can also be reported to the BS. Since a three-node single-antenna scenario is assumed, based on the state-of-the-art in overhead research, the signaling overhead (power, bandwidth, and throughput) maintains low [13] [14]. Hence, hereby we only focus our attention on the subsequent transmission stage.

Let us define the transmission power allocated to the SU and to the WU at the BS by p_1 and p_2, where the the maximum transmission power available at the BS is upper bounded by $p_{BS,max}$. Let us furthermore define the transmission power consumed at the SU by p_3, which is upper bounded by the maximum available transmission power constraint $p_{3,max}$. Furthermore, as discussed in Section I, to guarantee that the SU can successfully decode the WU's signal and subtract it by SIC, we introduce the constraint of $\Gamma_{12} \geq \Gamma_2$, where Γ_{12} and Γ_2 respectively represent the WU' SINR observed at the SU and the WU's signal at its own end. In addition, to guarantee the WU's quality-of-service (QoS), we have $\Gamma_2 \geq \Gamma_{2,req}$, where $\Gamma_{2,req}$ is the minimum SINR requirement of the WU.

A. Problem Formulation

Let us define EE (in bits/Joule/Hz) as the ratio of the system's throughput T_{total} to the total power consumption P_{total}. To maximize the system's EE, we jointly optimize the transmission power p_1 p_2 and p_3. Accordingly, the EE maximization problem is formulated as

$$\begin{aligned} P1: &\underset{p_1,p_2,p_3}{\text{argmax}} \frac{T_{\text{total}}}{P_{\text{total}}}, \\ \text{s.t } &(C1): p_1+p_2 \leq p_{BS,max}, (C2): p_3 \leq p_{3,max}, \\ &(C3): 0 \leq p_1, 0 \leq p_2, 0 \leq p_3, \\ &(C4): \Gamma_{12} \geq \Gamma_2, \ (C5): \Gamma_2 \geq \Gamma_{2,req}. \end{aligned} \tag{1}$$

Constraint $(C1)$ indicates that the transmission power allocated at the BS should be lower than the available transmission power $p_{BS,max}$. Constraint $(C2)$ means that the transmission power of the SU should be lower than its total available transmission power $p_{3,max}$. Constraint $(C3)$ implies that all the transmission power should be non-negative. Constraint $(C4)$ guarantees that the SU can decode the WU's signal, while $(C5)$ suggests that the WU's QoS should be satisfied.

III. Throughput and Power Consumption Analysis

The optimization problem involves both the throughput and power consumption, which will be discussed in Subsections III-A and III-B, respectively.

A. Throughput Analysis

At the WU, the received signal arrives both from the BS and SU, which is given by

$$r_2[i] = (\sqrt{p_1}x_1[i] + \sqrt{p_2}x_2[i])h_2 + \sqrt{p_3}x_2[i-\tau]h_{12} + n[i], \tag{2}$$

where $x_1[i]$ and $x_2[i]$ are the i-th data symbols intended for the SU and the WU, respectively. $n[i]$ is the additive White Gaussian noise (AWGN). The symbol delay τ is caused by the processing delay at the SU and h_2 is the channel response from the BS to the WU, while h_{12} is the channel response from the SU to the WU. All propagation channels capture the effects of large-scale and small-scale fading. Since the phase of h_2 and h_{12} is different, the desired signal replicas arriving from the BS and the SU are mis-aligned in phase. Hence, the phase of the transmitted signal at the SU has to be shifted to align the pair of received signals. The channel spanning from the BS to the WU can be represented as $h_2 = |h_2|e^{j\theta_2}$, where $|h_2|$ and $e^{j\theta_2}$ are the magnitude and phase of the channel h_2, respectively. Similarly, the channel spanning from the SU to the WU is represented as $h_{12} = |h_{12}|e^{j\theta_{12}}$, where $|h_{12}|$ and $e^{j\theta_{12}}$ are the magnitude and phase of the channel h_{12}, respectively. Therefore, the phase difference between the two channels is calculated as $\phi = \theta_2 - \theta_{12}$. To align the two signals, the signal transmitted from the SU is pre-processed as $(\sqrt{p_3}e^{j\phi})x_2$. As a result, the signal received at the WU becomes

$$\begin{aligned} r_2[i] = \big(&\sqrt{p_3}|h_{12}|e^{j(\theta_{12}+\phi)}x_2[i-\tau]+ \\ &\sqrt{p_2}|h_2|e^{j\theta_2}x_2[i]\big) + \sqrt{p_1}|h_2|e^{j\theta_2}x_1[i] + n[i]. \end{aligned} \tag{3}$$

As seen, the pair of desired signals arriving from the BS and the SU are now well aligned at the WU[1], and the SINR of the WU at its own end is given by

$$\Gamma_2 = \frac{p_2|h_2|^2 + p_3|h_{12}|^2}{p_1|h_2|^2 + \sigma^2}. \tag{4}$$

The signal received at the SU is given by

$$\begin{aligned} r_1[i] = &(\sqrt{p_1}x_1[i] + \sqrt{p_2}x_2[i])h_1 + \\ &(\sqrt{p_3}e^{j(\phi+\theta_r)}x_2[i-\tau])\frac{h_r}{\sqrt{\alpha}} + n[i], \end{aligned} \tag{5}$$

where h_1 and h_r represent the channel spanning from the BS to the SU and the SI leakage channel from the SU's transmitter to its receiver, respectively. Still referring to (5), α is the SI reduction factor defined as the ratio of the SI powers before and after SI suppression. Therefore, the WU's SINR observed at the SU is given by

$$\Gamma_{12} = \frac{p_2|h_1|^2}{p_3|\widetilde{h_r}|^2 + p_1|h_1|^2 + \sigma^2}, \tag{6}$$

where $\widetilde{h_r}$ is the residual SI channel. To guarantee successful SIC at the SU, we have to ensure that $\Gamma_{12} \geq \Gamma_2$. Based on (4) and (6), the inequality $\Gamma_{12} \geq \Gamma_2$ is further derived into $\frac{p_2|h_1|^2}{p_3|\widetilde{h_r}|^2+p_1|h_1|^2+\sigma^2} \geq \frac{p_2|h_2|^2+p_3|h_{12}|^2}{p_1|h_2|^2+\sigma^2}$. Since $|h_2|^2 \leq |h_1|^2$, we have $\Delta = (p_1|h_1|^2|h_{12}|^2 + |h_{12}|^2\sigma + p_2|h_2|^4)^2 - 4|h_{12}|^2|h_2|^2p_2(|h_2|^2 - |h_1|^2)\sigma^2 \geq 0$, As a result, the feasible domain of p_3 can be represented in closed form

$$p_3 \in [0, \frac{-B + \sqrt{\Delta}}{2A}], \tag{7}$$

where we have $B = p_1|h_1|^2|h_{12}|^2 + |h_{12}|^2\sigma + p_2|h_2|^4$ and $A = |h_{12}|^2|h_2|^2$. Finally, the SU's SINR after SIC is calculated as

$$\Gamma_1 = \frac{p_1|h_1|^2}{p_3|\widetilde{h_r}|^+\sigma^2}. \tag{8}$$

B. Power Consumption Analysis

The total power consumption mainly consists of the PA power, circuit power of the transmit/receive chains, and the power consumed by SI cancellation [15].

a) The PA power is closely related to the radiated transmit power and to the drain efficiency of the PA, which is given by $\frac{1}{\eta}(p_1+p_2+p_3)$ [15], where η is the drain efficiency of the PAs. Without loss of generality, we assume that all PAs of the BS and of the users have the same drain efficiency performance.

b) The circuit power consumption of the transmit/receive chains is proportional to the number of active transmit and receive chains, including the power consumed by the digital-to-analog converter, by the filter and synthesizer, *etc*. The dynamic power consumption can be calculated as $2(p_{c,r} + p_{c,t})$, where $p_{c,t}$ and $p_{r,t}$ denote the circuit power consumed by the receive and transmit chains of the BS, of the SU and of the WU. The multiplier 2 indicates that we have two transmit chains and two receive chains in the system.

c) For the SU acting as a FD relay, additional power p_{SI} is required for SI cancellation, which can be modeled by a constant. Generally speaking, complex SI cancellation schemes dissipate higher power by their digital-to-analog converter, transmit radio unit and adders [15].

d) The fixed power consumption p_{fix} represents power consumed by the power supply, by the active cooling system, *etc*. This part is independent of the state of the transmit/receive chains [15]. Finally, the total power consumption is given by

$$P_{\text{total}} = \frac{p_1 + p_2 + p_3}{\eta} + 2(p_{c,r} + p_{c,t}) + p_{SI} + p_{fix}. \tag{9}$$

IV. Energy Efficiency Performance Optimization

For simplicity, we collect the transmission power into a vector $\boldsymbol{p} = [p_1, p_2, p_3]^T$ and define $\{\boldsymbol{\Theta}\}$ as the feasible domain confined by the constraints. Therefore, the EE maximization problem is re-formulated as:

$$\begin{aligned} P2: &\underset{\boldsymbol{p}\in\{\boldsymbol{\Theta}\}}{\text{argmax}} \frac{\log_2(1 + \frac{\boldsymbol{a_1 p}|h_1|^2}{\boldsymbol{a_3 p}|\widetilde{h_r}|^+\sigma^2}) + \log_2(1 + \frac{\boldsymbol{a_2 p}|h_2|^2+\boldsymbol{a_3 p}|h_{12}|^2}{\boldsymbol{a_1 p}|h_2|^2+\sigma^2})}{\frac{1}{\eta}(||\boldsymbol{p}||_1) + 2(p_{c,r} + p_{c,t}) + p_{fix} + p_{SI}}, \\ \text{s.t } &(\hat{C1}): \boldsymbol{a_1 p} + \boldsymbol{a_2 p} \leq p_{BS,max}, \\ &(\hat{C2}): \boldsymbol{a_3 p} \leq p_{3,max}, (\hat{C3}): \boldsymbol{0} \preceq \boldsymbol{p}, \\ &(C4): \Gamma_{12} \geq \Gamma_2,\ (C5): \Gamma_2 \geq \Gamma_{2,req}, \end{aligned} \tag{10}$$

where $\boldsymbol{a_1} = [1, 0, 0]$, $\boldsymbol{a_2} = [0, 1, 0]$ and $\boldsymbol{a_3} = [0, 0, 1]$. The operator $||\cdot||_1$ represents the 1-norm of vector. It can been seen that the objective function is non-convex and $(C4)$ $(C5)$ represent quadratic constraints, which impose a challenge in

[1]The signal received at the WU has a low time delay, which can be mitigated by an equalizer or sequence detector. Hereby, we assume that the signal from the BS and the SU can be readily combined by the WU [5].

terms of obtaining a globally optimal result within polynomial time. To strike an attractive balance between the performance and complexity, a low-complexity near-optimal solution is desirable. We observe that the numerator of the objective function is the difference between two convex functions. To handle this challenge, we first reformulate the objective function of (10) as $\underset{\boldsymbol{p}\in\{\boldsymbol{\Theta}\}}{\operatorname{argmin}} \dfrac{f(\boldsymbol{p})-g(\boldsymbol{p})}{\frac{1}{\eta}||\boldsymbol{p}||_1+2(p_{c,r}+p_{c,t})+p_{fix}+p_{SI}}$, where $f(\boldsymbol{p}) = -\log_2\big(\boldsymbol{a_3}\boldsymbol{p}|\widetilde{h_r}|^2+\boldsymbol{a_1}\boldsymbol{p}|h_1|^2+\sigma^2\big) - \log_2\big(\boldsymbol{a_1}\boldsymbol{p}|h_2|^2+\boldsymbol{a_2}\boldsymbol{p}|h_2|^2+\boldsymbol{a_3}\boldsymbol{p}|h_{12}|^2+\sigma^2\big)$ and $g(\boldsymbol{p}) = -\log_2\big(\boldsymbol{a_1}\boldsymbol{p}|h_2|^2+\sigma^2\big) - \log_2\big(\boldsymbol{a_3}\boldsymbol{p}|\widetilde{h_r}|^+\sigma^2\big)$. Therefore, the Frank-Wolfe method of [16] can be adopted, which approximates $g(\boldsymbol{p})$ by its first-order Taylor series, and iteratively updates the first-order Taylor approximation along the specific direction that approaches the original function. The first-order Taylor approximation of $g(\boldsymbol{p})$ at the i-th iteration is given by $g^{(n)}(\boldsymbol{p}) = -\log_2\big(\boldsymbol{a_1}\boldsymbol{p}^{(n)}|h_2|^2+\sigma^2\big) - \log_2\big(\boldsymbol{a_3}\boldsymbol{p}^{(n)}|\widetilde{h_r}|^+\sigma^2\big) - \frac{|h_2|^2\boldsymbol{a_1}(\boldsymbol{p}-\boldsymbol{p}^{(n)})}{\ln 2(\boldsymbol{a_1}\boldsymbol{p}^{(n)}|h_2|^2+\sigma^2)} - \frac{|h_r|^2\boldsymbol{a_3}(\boldsymbol{p}-\boldsymbol{p}^{(n)})}{\ln 2(\boldsymbol{a_3}\boldsymbol{p}^{(n)}|h_r|^2+\sigma^2)}$, where $\boldsymbol{p}^{(n)}$ is the value of $\boldsymbol{p}$ at the n-th iteration. Now, the overall throughput is the difference between a convex function $f(\boldsymbol{p})$ and an affine function $g^{(n)}(\boldsymbol{p})$. Let us state Theorem 1 to solve the problem.

Theorem 1: The reformulated problem, namely $\dfrac{f(\boldsymbol{p})-g^{(n)}(\boldsymbol{p})}{P_{\text{total}}(\boldsymbol{p})}$, is jointly quasi-convex with respect to the vector variables $\boldsymbol{p}$ in the feasible domain.

Proof: Let us define the sublevel set of $\omega(\boldsymbol{p})$ as $S_\delta = \{\boldsymbol{p}\in\boldsymbol{\Theta}|\omega(\boldsymbol{p})\le\delta\}$. Recall from [15] that $\omega(\boldsymbol{p})$ is jointly quasi-convex with respect to the variables in $\boldsymbol{p}$, if S_δ is convex for any real number δ. For $\delta\le 0$, we have no physical interpretation. By contrast, for $\delta\ge 0$, S_δ is equivalent to $S_\delta = \{[f(\boldsymbol{p})-g^{(n)}(\boldsymbol{p})]-\delta P_{\text{total}}(\boldsymbol{p})\le 0|\boldsymbol{p}\in\boldsymbol{\Theta}\}$. According to our analysis above, $P_{\text{total}}(\boldsymbol{p})$ is affine with respect to the variables, while $[f(\boldsymbol{p})-g^{(n)}(\boldsymbol{p})]$ is strictly jointly convex with respect to the variables. Therefore, the summation is strictly convex with respect to the variables, and $\omega(\boldsymbol{p})$ is quasi-convex with respect to the variables in $\boldsymbol{p}\in\{\boldsymbol{\Theta}\}$. ■

Theorem 1 confirms the optimality of the re-formulated objective function. For the fractional structured quasi-convex problem of (10), $\beta = \frac{f(\boldsymbol{p})-g^{(n)}(\boldsymbol{p})}{P_{\text{total}}(\boldsymbol{p})}$ can be associated with a subtract programming formulation of $f(\boldsymbol{p})-g^{(n)}(\boldsymbol{p})-\beta P_{\text{total}}(\boldsymbol{p})$ [16]. Therefore, with the aid of the equivalent subtract programming, the problem is reformulated as that of solving $f(\boldsymbol{p})-g^{(n)}(\boldsymbol{p})-\beta P_{\text{total}}(\boldsymbol{p})$ with a given β.

Now we handle the constraints $(C4)$ and $(C5)$. Based on (7), $(C4)$ is equivalent to $\boldsymbol{p}^T\boldsymbol{A}\boldsymbol{p}+\boldsymbol{a_4}\boldsymbol{p}\le 0$, where $\boldsymbol{a_4} = \big[0,(|h_2|^2-|h_1|^2)\sigma^2,|h_{12}|^2\sigma^2\big]$ and $\boldsymbol{A} = \begin{bmatrix} 0 & 0 & |h_1|^2|h_{12}|^2/2 \\ 0 & 0 & |h_2|^2|h_r|^2/2 \\ |h_1|^2|h_{12}|^2/2 & |h_2|^2|h_r|^2/2 & |h_{12}|^2\sigma^2 \end{bmatrix}$. Since the matrix $\boldsymbol{A}$ is not a semi-positive definite matrix, its elements are not confined to a convex set. Therefore, we introduce the Schur complement [2] of $\boldsymbol{P}=\boldsymbol{p}\boldsymbol{p}^T$ to relax $(C4)$ into $(\hat{C4a})$: $\text{Tr}(\boldsymbol{A}\boldsymbol{P})+a_4^T\boldsymbol{p}\le 0$ and constraint $(\hat{C4b})$: $\begin{bmatrix}\boldsymbol{P} & \boldsymbol{p}\\ \boldsymbol{p}^T & 1\end{bmatrix}\succeq 0$. On the other hand, constraint $(C5)$ is equivalent to $(\hat{C5})$: $\boldsymbol{a_5}\boldsymbol{p}+\Gamma_{2,req}\sigma^2\le 0$, where we have $\boldsymbol{a_5} = \big[\Gamma_{2,req}\sigma^2,\,-|h_2|^2,\,-|h_{12}|^2\big]$. After a series of transformations, the problem becomes

$$\begin{aligned} P3: \underset{\boldsymbol{p}\in\{\boldsymbol{\Theta}\}}{\operatorname{argmin}}\ & f(\boldsymbol{p})-g^{(n)}(\boldsymbol{p})-\beta P_{\text{total}}(\boldsymbol{p}), \\ \text{s.t } & (\hat{C1})-(\hat{C3}),(\hat{C4a}),(\hat{C4b}) \text{ and } (\hat{C5}). \end{aligned} \tag{11}$$

The problem $P3$ in (11) now is a standard semi-definite programming (SDP) problem with a convex set, which can be readily solved by the CVX package of Matlab. Finally, a so-called "EE oriented FD cooperative NOMA" algorithm is proposed for optimizing $\boldsymbol{p}$, which solves the SDP problem $P3$ of (11) in the inner layer and updates β in the outer layer. To tighten the first-order Taylor approximation $g^{(n)}(\boldsymbol{p})$, we update the value of $\boldsymbol{p}^{(n)}$ in the specific direction that approaches the original function. Since $f(\boldsymbol{p})-g^{(n)}(\boldsymbol{p})-\beta P_{total}(\boldsymbol{p})$ serves as the upper bound of the original problem $f(\boldsymbol{p})-g(\boldsymbol{p})-\beta P_{total}(\boldsymbol{p})$, the value of the upper bound is reduced iteratively until convergence is reached, as suggested by Line 4 of the Algorithm. Furthermore, the value of β is squeezed by the bisection method in the outer layer according to the accuracy factor ϵ, and the optimal $\beta^\star$ can be found after a few updates. Hence, the tightness and convergence of the proposed algorithm is confirmed. The associated procedure is summarized in Algorithm 1.

Algorithm 1 EE Oriented FD Cooperative NOMA Algorithm

Input: Left/right bounds β_l and β_r, channel condition, *i.e.,* $h_1,h_2,h_{12},\widetilde{h_r}$, and power consumption parameters, *i.e.,* $\eta,p_{c,t},p_{r,t},p_{SI},p_{fix}$.
Output: Optimal transmission power vector to $\boldsymbol{p}^*$.
1: Set the accuracy factor to $\epsilon>0$, and assume that $\mathcal{F}(\beta)$ is the optimal value of $f(\boldsymbol{p})+g^{(n)}(\boldsymbol{p})-\beta P_{\text{total}}(\boldsymbol{p})$. Let us initialize the left bound β_l and the right bound β_r for ensuring that $\mathcal{F}(\beta_l)\cdot\mathcal{F}(\beta_r)<0$.
2: **while** $\beta_r-\beta_l>\epsilon$ **do**
3: $\beta=\frac{\beta_r+\beta_l}{2}$.
4: Solve the problem $P3$ using the Frank-Wolfe method until convergence.
5: **if** $\mathcal{F}(\beta_l)\cdot\mathcal{F}(\beta)<0$ **then**
6: $\beta_r=\beta$.
7: **else**
8: $\beta_l=\beta$.
9: **end if**
10: **end while**

Let us now consider the complexity of the algorithm. Let us assume that $\beta_1=\frac{\beta_r+\beta_l}{2}$ is the midpoint of the initial interval, and β_n is the midpoint of the interval in the n-th step. Then the difference between β_n and β^* is bounded by $|\beta_n-\beta^*|\le\frac{\beta_r-\beta_l}{2^n}$. Given a tolerance factor ϵ, the required number of iterations is given by $n\le log_2(\frac{\beta_r-\beta_l}{\epsilon})$. In the proposed algorithm, the left bound β_l can be set to 0. Then the value of $\mathcal{F}(\beta_l)$ is definitely positive. Furthermore, a sufficient large value of β_r can be chosen as the right bound for making the value of $\mathcal{F}(\beta_r)$ negative. Therefore, the function $\mathcal{F}(\cdot)$ has opposite signs at the two bounds and thus the classic bisection method readily leads to convergence. In the inner layer, the CVX solver invokes an interior-point method to solve the SDP problem, which belongs to the class of path-following methods and leads to a rapid convergence. Upon denoting the complexity of the inner layer by ξ, and the complexity order of the proposed algorithm becomes $\mathcal{O}(\log_2(\frac{\beta_r-\beta_l}{\epsilon})\cdot\xi)$.

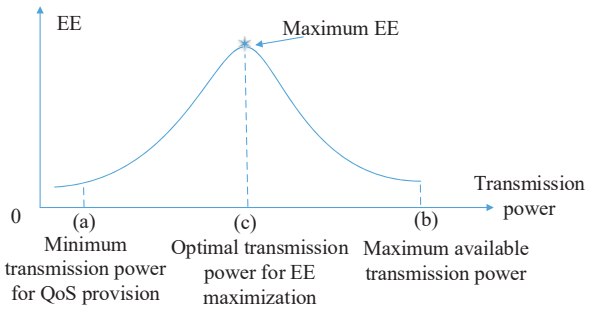

Fig. 2. The stylized relationship of EE maximization, power minimization and throughput maximization problems.

Remark 1: Generally speaking, a throughput maximization problem aims for achieving a high throughput regardless of the power consumption. Hence, the BS only has to satisfy the WU's SINR threshold, while allocating all the remaining power to the SU. Then the SU utilizes a high transmission power to cooperate the WU. By contrast, a power minimization problem aims for reducing the total power consumption and thus merely satisfies the WU's QoS constraint by a minimum transmission power. As a better choice, EE maximization improves the ratio between the throughput and power consumption, striking a trade-off between the throughput maximization and power minimization problems. As suggested by Theorem 1, $\omega(\boldsymbol{p})$ is quasi-convex (thus the re-formulated EE is quasi-concave) with respect to the variables in $\{\boldsymbol{\Theta}\}$. For a quasi-concave function, utilizing all transmission power may not lead to the most EE solution, because the power consumption in the denominator is also boosted, hence the EE is decreased. For illustration, let us consider a simple EE maximization problem in Fig. 2, where the EE curve is a quasi-concave function of transmission power. Neither the power minimization (point (a)) nor the throughput maximization solution (point (c)) achieves the optimal EE.

Remark 2: As suggested by Equations (4) (6) and (7), the BS-SU, BS-WU and SU-WU distances also affect the SIC failure region. Given a shorter BS-SU distance, the value of Γ_{12} is increased due to the reduced PL, and thus a higher cooperative power may be assigned to the SU hence guaranteeing $\Gamma_{12} \geq \Gamma_2$. A similar trend can be observed for a shorter BS-WU distance. By contrast, a longer SU-WU distance makes $\Gamma_{12} \geq \Gamma_2$ more likely. However, this also makes the cooperation between the SU-WU less energy efficient, since additional power is consumed at the SU without a reasonable throughput improvement. As expected, when the SU-WU distance tends to infinity, the cooperative power assigned to the SU tends to 0, as in non-cooperative NOMA systems.

Remark 3: Given a lower drain efficiency, the systems EE is reduced owing to the boosted PA power consumption. Furthermore, the power allocation of both the BS and of the SU becomes more conservative due to the low drain efficiency, since the throughput improvement attained by increasing the transmission power may not be in line with the increased power consumption.

TABLE I. Simulation Setup

Bandwidth	1.25 MHz [17]
AWGN power spectral density	-174 dBm/Hz
Drain efficiency of PA η	35% [18]
$p_{c,r}$ and $p_{c,t}$	100 mW
p_{SI} and p_{fix}	50 mW and 500 mW
$p_{BS,max}$ and $p_{3,max}$	100 mW and 20 mW
SI cancellation amount	80 dB
Distance between the BS and the WU	200 m

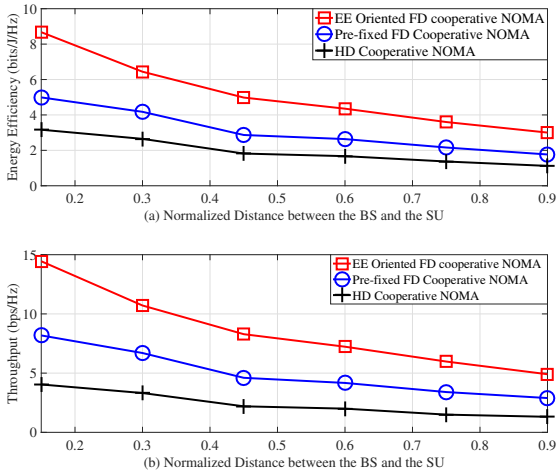

Fig. 3. Impact of the normalized distance between the BS and the SU on the value of EE and throughput, with $p_{3,max} = 20 \times 10^{-3}$ mW.

V. Numerical Results

Let us now discuss our numerical results to verify our analysis using the parameters given in TABLE I. The path loss (PL) model of $\mathrm{PL}(d) = 145.4 + 37.5\log_{10}(d/1000)$ [16] and log normal shadowing model are adopted, which is featured in the 3GPP LTE standards operating at 2 GHz. The small-scale fading is modeled by Rayleigh fading except for the SI channel at the SU, which is modeled as Rician fading having a Rician factor of 5 dB [12]. A pair of typical cooperative NOMA systems are selected for performance comparison. (a) In HD cooperative NOMA systems [6], the BS transmits its downlink signals to the two users (with transmission power $p_1 = \alpha p_{BS,max}, p_2 = (1-\alpha)p_{BS,max}$, and $\alpha = 0.2$) during the first half of the time slot, while the SU helps the WU by assigning its full transmission power $p_3 = p_{3,max}$ in the second half of the time slot. (b) In the pre-fixed FD cooperative NOMA systems [5], the SU acts as a FD relay node for simultaneously helping the WU by assigning its full transmission power $p_3 = p_{3,max}$, whereas the power allocation at the BS is pre-fixed as $p_1 = \alpha p_{BS,max}, p_2 = (1-\alpha)p_{BS,max}$, and $\alpha = 0.2$.

Fig. 3 (a) and (b) show the impact of the normalized distance between the BS and the SU on the EE and SE, respectively. As can be seen, the proposed algorithm outperforms the others in terms of its EE, whilst exhibiting a higher robust against the SU's location as well. This is because the benchmarking algorithms consume all the available transmission power, which degrades their EE. A further degrading factor

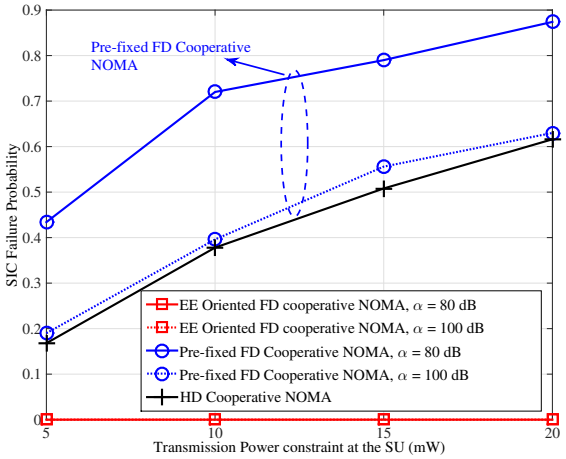

Fig. 4. Impact of the maximum constraint $p_{3,max}$ at the SU on the SIC's failure probability, where the SU is in the middle of the BS and the WU.

is constituted by their low throughput. In Fig. 3 (b), the proposed algorithm shows the highest throughput among the three algorithms. This is because the proposed algorithm supports the effective cooperation of the SU and of the WU, hence improving the WU's throughput. More importantly, no additional time slot is required for the cooperation phase, and thus the proposed algorithm significantly outperforms the HD cooperative NOMA. As for the pre-fixed FD cooperative NOMA, the power allocation is unaware of the relative distance among the communication nodes. Furthermore, the SU invokes its full transmission power to cooperate with the WU, which may result in the "SINR gap reversal" and may lead an unsuccessful SIC operation at the SU. As a result, the throughput of the pre-fixed FD cooperative NOMA is also inferior to the proposed algorithm. Finally, the throughput of all the algorithms degrades for a longer distance between the BS and the SU due to the high PL.

Fig. 4 shows the impact of the constraint $p_{3,max}$ on the SIC's failure probability, which is defined as the ratio of the times of unsuccessful SIC operation at the SU to the times of simulations. As seen, the proposed algorithm is shown to guarantee that the SU can successfully decode the WU's signal by SIC. This is because an additional constraint is imposed on the power control at the SU by the proposed algorithm. By comparison, the SIC failure probability of the pre-fixed FD cooperative NOMA systems increases rapidly with the transmission power constraint $p_{3,max}$. This is because for a higher transmission power used at the SU, the WU's SINR Γ_2 increases, while the WU's SINR Γ_{12} observed at the SU decreases due to the higher level of residual SI. As a result, a higher value of $p_{3,max}$ leads to a higher SIC failure probability at the SU. In case of poor SI cancellation, the SIC failure probability is substantially increased by the strong residual SI. As for the HD cooperative NOMA system, its SIC failure probability also increases upon imposing a higher transmission power constraint due to the "SINR gap reversal" problem. Since no residual SI is imposed by the HD operation, its SIC failure probability is lower than that of the FD operation [5]. However, its SIC failure probability still remains significantly higher than that of the proposed algorithm, because the HD cooperative NOMA neglects that Γ_2 may be higher than Γ_{12} owing to the cooperation between the SU and WU. However, its SIC failure probability is not affected by the SI cancellation factor α in the presence of HD operation.

VI. Conclusions

We have proposed an EE oriented algorithm for FD cooperative NOMA systems, where the transmission power of both the BS and of the SU is adaptively allocated, rather than being pre-fixed. More importantly, the "SINR gap reversal" issue of cooperative NOMA systems was solved by adaptively confining the transmission power at the SU. Our simulation results show that the proposed design demonstrates both significant EE and throughput enhancements over the HD cooperative NOMA [6] and the pre-fixed FD cooperative NOMA regime [5]. Furthermore, the proposed algorithm guarantees successful SIC operation at the SU, while the SIC failure probability of the cooperative NOMA schemes in [5] [6] is at a high level.

References

[1] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5G and beyond, " in *Proc. of IEEE*, vol. 105, no. 12, pp. 2347-2381, Dec. 2017.

[2] J. Kim and I. Lee, "Capacity analysis of cooperative relaying systems using non-orthogonal multiple access," *IEEE Commun. Lett*., vol. 19, no. 11, pp. 1949-1952, Nov. 2015.

[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] C. Zhong and Z. Zhang, "Non-orthogonal multiple access with cooperative full-duplex relaying," *IEEE Commun. Lett.,* vol. 20, no. 12, pp. 2478-2481, Dec. 2016.

[5] Z. Zhang, Z. Ma, M. Xiao, Z. Ding, and P. Fan, "Full duplex device-to-device aided cooperative non-orthogonal multiple access," to appear in *IEEE Trans. Veh. Technol*., DOI: 10.1109/TVT.2016.2600102.

[6] L. Lv, J. Chen, and Q. Ni, "Cooperative NOMA in cognitive radio," *IEEE Commun. Lett*., vol. 20, no. 10, pp. 2069-2062, Oct. 2016.

[7] G. Liu, X. Chen, Z. Ding, Z. Ma, F. R. Yu, "Hybrid Half-Duplex/Full-Duplex Cooperative Non-Orthogonal Multiple Access With Transmit Power Adaptation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 506-519, Nov. 2018.

[8] L. Zhang, J. Liu, M. Xiao, G. Wu, Y.-C. Liang, S. Li, "Performance Analysis and Optimization in Downlink NOMA Systems With Cooperative Full-Duplex Relaying," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2398-2412, Jul. 2017.

[9] F. Fang, H. Zhang, J. Chen, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.,* vol. 64, no. 9, pp. 3722-3733, Sep. 2016.

[10] J. Choi, "Minimum power multicast beamforming with superposition coding for multi resolution broadcast and application to NOMA system," *IEEE Trans. Commun*., vol. 63, no. 3, pp. 791-800, Mar. 2015.

[11] Q. Sun, S. Han, C. L. I and Z. Pan, "Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems," in *Proc. ICC*' 15, London, UK, pp. 3668-3674.

[12] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: challenges, solutions, and future research directions," in *Proc. of IEEE*, vol. 104, no. 7, pp. 1369-1409, Jul. 2016.

[13] B. Klaiqi, X. Chu, and J. Zhang, "Energy-efficient and low signaling overhead cooperative relaying with proactive relay subset selection," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1001-1015, Mar. 2016.

[14] Y. Xiao and L. J. Cimini, Jr., "Impact of overhead on spectral efficiency of cooperative relaying," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 5, pp. 2228-2239, May 2013.

[15] Z. Wei, X. Zhu, S. Sun, and Y. Huang, "Energy efficiency oriented cross-Layer resource allocation for multiuser full-duplex decode-and-forward indoor relay systems at 60 GHz," accepted by *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3366-3379, Dec. 2016.

[16] D. Nguyen, L. N. Tran, P. Pirinen, and M. L. Aho, "Precoding for FD multiuser MIMO systems: spectral and energy efficiency maximization," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4038-4050, Aug. 2013.
[17] J. Zyren, White Paper: Overview of the 3GPP Long Term Evolution Physical Layer, https://www.nxp.com/docs/en/white-paper/3GPPEVOLUTIONWP.pdf
[18] Z. Wei, X. Zhu, S. Sun, A. AL-Tahmeesschi, and Y. Jiang, "Energy-Efficiency of Millimeter-Wave Full-Duplex Relaying Systems: Challenges and Solutions," *IEEE Access,* vol. 4, pp. 4848-4860, Dec. 2016.