Unveiling the Importance of SIC in NOMA Systems: Part II: New Results and Future Directions

(Invited Paper)

Zhiguo Ding, *Fellow, IEEE*, Robert Schober, *Fellow, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

***Abstract*—In most existing works on non-orthogonal multiple access (NOMA), the decoding order of successive interference cancellation (SIC) is prefixed and based on either the users' channel conditions or their quality of service (QoS) requirements. A recent work on NOMA assisted semi-grant-free transmission showed that the use of a more sophisticated hybrid SIC scheme can yield significant performance improvements. This letter illustrates how the concept of hybrid SIC can be generalized and applied to different NOMA applications. We first use NOMA assisted mobile edge computing (MEC) as an example to illustrate the benefits of hybrid SIC, where new results for delay and energy minimization are presented. Then, future directions for generalizing hybrid SIC with adaptive decoding order selection as well as its promising applications are discussed.**

I. Introduction

Successive interference cancellation (SIC) is a key component of non-orthogonal multiple access (NOMA) systems, and is crucial for the performance of NOMA transmission [1]–[3]. In the first part of this two-part invited paper, we have explained that, in most existing works on NOMA, the design of the SIC decoding order is prefixed and based on either the users' channel state information (CSI) or their quality of service (QoS) requirements [2]–[4]. This is primarily due to the general perception that the use of more than one SIC decoding orders is trivial and unnecessary. In the first part of this paper, the recent work in [5] on a hybrid implementation of CSI- and QoS-based SIC has also been reviewed, where we showed that adaptively switching between CSI- and QoS-based SIC can avoid an outage probability error floor, which is inevitable with either of the two individual schemes.

The aim of the second part of this paper is to show that the findings in [5] can be generalized and can be applied to different NOMA communication scenarios. For illustration, we use NOMA assisted mobile edge computing (MEC) as an example [6]–[9]. Recall that the key idea of MEC is to ask users to offload their computationally intensive tasks to the base station, instead of computing these tasks locally. Compared to orthogonal multiple access (OMA) based MEC, the use of NOMA-MEC ensures that multiple users can offload their tasks simultaneously, which is beneficial for reducing the delay and energy consumption of MEC offloading. New results for NOMA-MEC are presented in this letter by applying hybrid SIC. In particular, the problem of joint energy and delay minimization is considered, in order to demonstrate that the findings in [5] are useful not only for performance analysis but also for resource allocation. The optimal solution for joint energy and delay minimization is obtained first, and then compared to OMA-MEC and the existing NOMA-MEC solution [8], [9]. Furthermore, future directions for the design of sophisticated SIC schemes as well as promising applications in different NOMA communication scenarios are presented.

Z. Ding and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA. Z. Ding is also with the School of Electrical and Electronic Engineering, the University of Manchester, Manchester, UK (email: zhiguo.ding@manchester.ac.uk, poor@princeton.edu). R. Schober is with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nurnberg (FAU), Germany (email: robert.schober@fau.de).

II. System Model

Consider a NOMA-MEC offloading scenario, where two users, denoted by U_m and U_n, respectively, offload their computationally intensive and inseparable tasks to the base station. U_i's channel gain and task deadline are denoted by h_i and D_i seconds, $i \in \{m, n\}$, respectively. It is assumed that the users' tasks contain the same number of nats, denoted by N. We note that unlike the first part of this paper which focuses on performance analysis, the second part of the paper concerns resource allocation, where the use of nats is more convenient than the use of bits. We further assume that $D_m < D_n$, i.e., U_m's task is more delay sensitive than U_n's, which means that, in OMA-MEC, U_m is served during the first D_m seconds and then U_n is served during the remaining $(D_n - D_m)$ seconds.

A. Basics of NOMA-MEC

Instead of allowing the first D_m seconds to be solely occupied by U_m, NOMA-MEC encourages that U_n offloads a part of its task during the first D_m seconds, and then the remainder of its task during the following T_n seconds, where $T_n \leq D_n - D_m$. Denote U_n's transmit powers during the two time slots by $P_{n,1}$ and $P_{n,2}$, respectively. The advantage of NOMA-MEC over OMA-MEC can be illustrated by considering the extreme case $(D_n - D_m) \to 0$. In this case, U_n's transmit power in OMA has to be infinity in order to deliver N nats in a short period, whereas this singular situation does not exist for NOMA-MEC since U_n can also use the first D_m seconds for offloading.

B. Existing NOMA-MEC Strategies

To ensure that the use of NOMA-MEC is transparent to U_m, QoS-based SIC has been used, i.e., U_n's signal is decoded before U_m's during the first D_m seconds, where U_n's data rate during the first D_m seconds needs to be constrained as $R_n = \ln\left(1 + \frac{P_{n,1}|h_n|^2}{P_m|h_m|^2+1}\right)$ and P_m denotes U_m's transmit power [8], [9]. Therefore, the problem of joint energy and delay minimization can be formulated as follows:

$$\min_{T_n, P_{n,1}, P_{n,2}} \quad D_m P_{n,1} + T_n P_{n,2} \tag{P1a}$$

$$\text{s.t.} \quad D_m R_n + T_n \ln\left(1 + |h_n|^2 P_{n,2}\right) \geq N \tag{P1b}$$

$$0 \leq T_n \leq D_n - D_m \tag{P1c}$$

$$P_{n,i} \geq 0, \quad i \in \{1, 2\}, \tag{P1d}$$

where constraints (P1b) and (P1c) ensure that U_n can finish its offloading within D_n seconds. We note that we omit the costs for the computation at the base station as well as the costs for downloading the computation results from the base station, similar to [6]–[9]. Following the same steps as in [9], we can show that the optimal solution of T_n is $T_n^* = D_n - D_m$, and the optimal power allocation solution is given by

$$\begin{cases} P_{n,1}^* = \frac{\left(1+P_m|h_m|^2\right)}{|h_n|^2}\left(e^{\frac{N-T_n\ln\left(1+P_m|h_m|^2\right)}{D_m+T_n}}-1\right) \\ P_{n,2}^* = \frac{\left(1+P_m|h_m|^2\right)e^{\frac{N-T_n\ln\left(1+P_m|h_m|^2\right)}{D_m+T_n}}-1}{|h_n|^2} \end{cases}, \quad (1)$$

if $P_m \leq |h_m|^{-2}\left(e^{\frac{N}{D_n-D_m}}-1\right)$, otherwise OMA is used.

III. New NOMA-MEC with Hybrid SIC

The aim of this section is to investigate whether there is any benefit in applying hybrid SIC, i.e., selecting the SIC orders in an adaptive manner, which means that the problem of joint energy and delay minimization can be formulated as follows:

$$\begin{aligned} \min_{T_n,P_{n,1},P_{n,2}} \quad & D_mP_{n,1}+T_nP_{n,2} && \text{(P2a)} \\ \text{s.t.} \quad & D_mR_{n,1}+T_n\ln(1+P_{n,2}|h_n|^2)\geq N && \text{(P2b)} \\ & D_m\ln\left(1+\frac{P_m|h_m|^2}{P_{n,1}|h_n|^2+1}\right)\geq \mathbf{1}_nN && \text{(P2c)} \\ & \text{(P1c), (P1d)}. && \text{(P2d)} \end{aligned}$$

where $R_{n,1} = \mathbf{1}_n\ln(1+P_{n,1}|h_n|^2) + (1-\mathbf{1}_n)\ln\left(1+\frac{P_{n,1}|h_n|^2}{P_m|h_m|^2+1}\right)$, $\mathbf{1}_n$ is the indicator function, i.e., $\mathbf{1}_n = 1$ if U_m's signal is decoded first during the first D_m seconds, otherwise $\mathbf{1}_n = 0$. We note that P2 is degraded to P1 if $\mathbf{1}_n = 0$. Therefore, in the remainder of the letter, we focus on the case of $\mathbf{1}_n = 1$:

$$\begin{aligned} \min_{T_n,P_{n,1},P_{n,2}} \quad & D_mP_{n,1}+T_nP_{n,2} && \text{(P3a)} \\ \text{s.t.} \quad & D_m\ln(1+P_{n,1}|h_n|^2) \\ & +T_n\ln(1+P_{n,2}|h_n|^2)\geq N && \text{(P3b)} \\ & D_m\ln\left(1+\frac{P_m|h_m|^2}{P_{n,1}|h_n|^2+1}\right)\geq N && \text{(P3c)} \\ & \text{(P1c), (P1d)}. && \text{(P3d)} \end{aligned}$$

The following lemma provides the optimal solution of P3.

Lemma 1. *Assume $P_m > |h_m|^{-2}\left(e^{\frac{N}{D_m}}-1\right)$. For P3, the optimal solution of T_n is given by $T_n^* = D_n - D_m$. The optimal power allocation solution is given by*

$$\begin{cases} P_{n,1}^* = |h_n|^{-2}\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1} - |h_n|^{-2} \\ P_{n,2}^* = |h_n|^{-2}e^{\frac{N}{D_n-D_m}-\frac{D_m}{D_n-D_m}\ln\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1}\right)} - |h_n|^{-2} \end{cases}, \quad (2)$$

if $|h_m|^{-2}\left(e^{\frac{N}{D_m}}-1\right) < P_m \leq |h_m|^{-2}e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}}-1\right)$, otherwise

$$P_{n,1}^* = P_{n,2}^* = \quad |h_n|^{-2}\left(e^{\frac{N}{D_n}}-1\right). \quad (3)$$

Proof. See Appendix A. □

Remark 1: Constraint (P3c) can be written as $P_{n,1}|h_n|^2 \leq P_m|h_m|^2\left(e^{\frac{N}{D_m}}-1\right)^{-1} - 1$. In order to ensure $P_{n,1} \neq 0$, the feasibility of the constraint needs the assumption $P_m > |h_m|^{-2}\left(e^{\frac{N}{D_m}}-1\right)$ or equivalently $D_m\ln(1+P_m|h_m|^2) > N$. Otherwise, OMA-MEC is used. In practice, this assumption can be justified if U_m is willing to increase its transmit power to help U_n. Also, if U_m applies a coarse-level power control, P_m has to be strictly larger than $|h_m|^{-2}\left(e^{\frac{N}{D_m}}-1\right)$ anyways.

Remark 2: The solutions of P1 and P3 share two common features. The first one is that they both outperform OMA, as shown in [9] and in the proof for Lemma 1 in this letter. The second one is that pure NOMA, i.e., $P_{n,2} = 0$, is never preferred. In particular, the solutions in (1), (2), and (3) correspond to the class of hybrid NOMA schemes, i.e., U_n uses NOMA during the first D_m seconds, and then OMA during the remaining $(D_n - D_m)$ seconds.

The optimal solution of P2 can be straightforwardly obtained by numerically comparing the energy consumption required for the closed-form solutions in (1) and (2) (or (3)), and selecting the most energy efficient solution. The solutions in (1) and (3) can be compared analytically, as shown in the following lemma.

Lemma 2. *Assume $e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}}-1\right) \leq \left(e^{\frac{N}{D_n-D_m}}-1\right)$. For the case of $|h_m|^{-2}e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}}-1\right) \leq P_m \leq |h_m|^{-2}\left(e^{\frac{N}{D_n-D_m}}-1\right)$, the new solution shown in* (3) *is more energy efficient than the existing one shown in* (1).

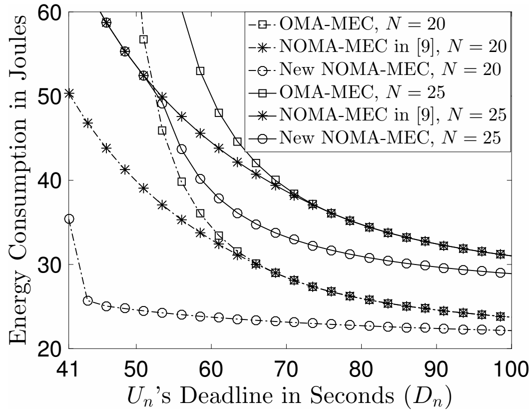
Proof. See Appendix B. □

Numerical Studies: In this section, the performance of different MEC strategies is studied by using computer simulations, where the users' average channel gains are assumed to be identical and normalized, a situation ideal for the application of QoS-based SIC. We will show that it is still beneficial to use hybrid SIC in this situation. In Fig. 1(a), the energy consumption of MEC offloading is shown as a function of D_n. As can be observed from the figure, the use of the new NOMA-MEC strategy can yield a significant reduction in energy consumption, compared to OMA-MEC and the existing NOMA-MEC solution proposed in [9], particularly when N is small.

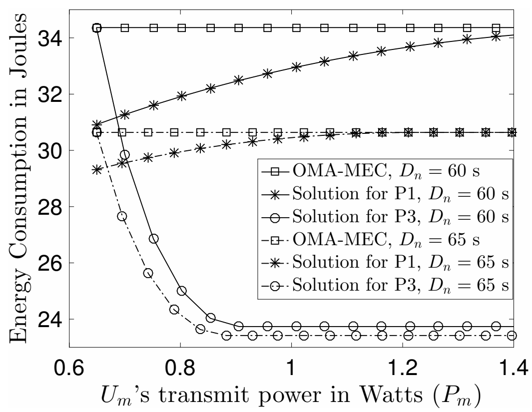
Fig. 1(a) also shows that there are instances when the new NOMA-MEC scheme achieves the same performance as the existing NOMA-MEC solution, which indicates that the solution of P1 can outperform the one of P3. Therefore, in Fig. 1(b), the solutions of P1 and P3 are compared in detail, where $P_m \geq |h_m|^{-2}\left(e^{\frac{N}{D_m}}-1\right)$ is considered. When P_m is small, the solution in (2) is used, and Fig. 1(b) shows that it is possible for the solution of P1 to outperform the one of P3. By increasing P_m, the solution in (3) becomes feasible, and Fig. 1(b) shows that the solution in (3) is more energy efficient than the one in (1), which confirms Lemma 2.

IV. Conclusions and Future Directions

In the second part of this invited paper, we have used NOMA-MEC as an example to illustrate how the new findings



(a) $P_m = 1$ W



(b) $N = 20$

Fig. 1. The impact of the NOMA strategies on the energy consumption required by MEC offloading. $|h_m|^2 = |h_n|^2 = 1$, and $D_m = 40$ s.

in [5] can be generalized. In particular, a hybrid SIC based optimal solution for joint energy and delay minimization was obtained and its superior performance compared to benchmark schemes was demonstrated. Some promising directions for future research on hybrid SIC with adaptive decoding order selection are listed in the following.

1) Fundamentals of hybrid SIC: For uplink NOMA, [5] showed the benefits of using hybrid SIC in two-user scenarios. When the number of users increases, the number of possible SIC orders increases significantly. Therefore, an important future direction is to design practical hybrid SIC schemes for striking a balanced tradeoff between system complexity and performance [10]. For downlink NOMA, it is still not known whether hybrid SIC is beneficial, but the duality between uplink and downlink suggests that the design of hybrid SIC for downlink NOMA is an important direction for future research.

2) Green communications: The initial results shown in Fig. 1(b) indicate that the use of hybrid SIC can significantly improve the energy efficiency of NOMA transmission. However, the energy reduction experienced by U_n is obtained at the price of increasing U_m's transmit power, which motivates a future study of user cooperation to improve the energy efficiency, which opens up a new dimension for the design of future green communication systems.

3) User clustering and resource allocation: For CSI-based SIC, it is preferable to group users with different channel conditions and encourage them to transmit/receive in the same subcarrier/time-slot. For QoS-based SIC, it is preferable to group users with different QoS requirements. These clear preferences provide simple guidances for the design of user clustering and resource allocation. However, hybrid SIC does not have these clear preferences, which makes a compact problem formulation difficult and results in a higher complexity, which is the price for the significant performance improvements. Therefore, designing low-complexity user clustering and resource allocation schemes for hybrid SIC is another important future research direction, where advanced tools, such as game theory and machine learning, can be useful.

4) Multiple-input multiple-output (MIMO) and intelligent reflecting surface (IRS) assisted NOMA: The use of hybrid SIC could be particularly useful in MIMO-NOMA systems. Recall that it is difficult to order MIMO users due to the fact that the users' channels are in vector/matrix form. Therefore, most existing MIMO-NOMA schemes simply rely on the prefixed SIC decoding order, whereas the use of hybrid SIC increases the degrees of freedom available for system design. Similarly, in the context of IRS-NOMA, the use of hybrid SIC avoids relying on a single SIC decoding order, and hence introduces more flexibility not only at the transceivers, but also at the IRS, which is helpful for improving the system performance.

5) Emerging applications of NOMA: Many emerging applications of NOMA will benefit from the use of hybrid SIC. For example, the delay and energy consumption of MEC offloading can be reduced, as shown by the initial results reported in this letter, but more rigorous studies from both the performance analysis and optimization perspectives are needed. In addition to MEC, wireless caching is another functionality to be supported by fog networking, where hybrid SIC can also be useful. Particularly, in addition to the users' channel conditions and QoS requirements, the type of file content can also be taken into account for the design of SIC. Similarly, in the context of NOMA assisted orthogonal time frequency space modulation (OTFS), hybrid SIC can be further extended by taking the users' heterogenous mobility profiles into account for selecting the SIC decoding order.

Appendix A
Proof for Lemma 1

A. Obtaining Possible Solutions for Optimal Power Allocation

We first find closed-form solutions for power allocation by fixing T_n. By recasting constraint (P3b) as $-D_m \ln(1 + P_{n,1}|h_n|^2) - T_n \ln(1 + P_{n,2}|h_n|^2) \leq -N$, it is straightforward to show that P3 is convex, and the optimal power allocation solution can be obtained by using the KKT conditions listed in the following:

$$\begin{cases} D_m - \lambda_3 \frac{D_m |h_n|^2}{1+P_{n,1}|h_n|^2} + \lambda_4 |h_n|^2 - \lambda_1 = 0 \\ T_n - \lambda_3 \frac{T_n |h_n|^2}{1+P_{n,2}|h_n|^2} - \lambda_2 = 0 \\ \lambda_3 (N - D_m \ln(1 + P_{n,1}|h_n|^2) \\ \qquad - T_n \ln(1 + P_{n,2}|h_n|^2)) = 0 \\ \lambda_4 \left(P_{n,1}|h_n|^2 - \frac{P_m |h_m|^2}{e^{\frac{N}{D_m}} - 1} + 1 \right) = 0 \\ \lambda_i P_{n,i} = 0, \quad i \in \{1, 2\} \\ \text{(P3b), (P3c), (P1c), (P1d)} \end{cases}, \tag{4}$$

where λ_i, $i \in \{1, \cdots, 4\}$, denote Lagrange multipliers.

Depending on the choices of the Lagrange multipliers, possible solutions are obtained as follows.

- The choice of $\lambda_1 \neq 0$ yields an OMA solution:

$$P^*_{n,1} = 0, \quad P^*_{n,2} = |h_n|^{-2}\left(e^{\frac{N}{T_n}} - 1\right). \tag{5}$$

- The choice of $\lambda_1 = 0$, $\lambda_2 = 0$, and $\lambda_4 \neq 0$ yields a possible hybrid NOMA solution:

$$\begin{cases} P^*_{n,1} = & |h_n|^{-2}\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1} - 1\right) \\ P^*_{n,2} = & |h_n|^{-2}\left(e^{\frac{N}{T_n} - \frac{D_m}{T_n}\ln\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1}\right)} - 1\right) \end{cases}, \tag{6}$$

if $e^{\frac{N}{D_m}}\left(e^{\frac{N}{D_m}} - 1\right) \geq P_m|h_m|^2 \geq e^{\frac{N}{D_m}} - 1$.

- The choice of $\lambda_1 = 0$, $\lambda_2 = 0$, and $\lambda_4 = 0$ yields another possible hybrid NOMA solution:

$$P^*_{n,1} = P^*_{n,2} = |h_n|^{-2}\left(e^{\frac{N}{D_m+T_n}} - 1\right), \tag{7}$$

if $P_m|h_m|^2 \geq e^{\frac{N}{D_m+T_n}}\left(e^{\frac{N}{D_m}} - 1\right)$.

- The choice of $\lambda_1 = 0$ and $\lambda_2 \neq 0$ yields a pure NOMA solution:

$$P^*_{n,1} = |h_n|^{-2}\left(e^{\frac{N}{D_m}} - 1\right), P^*_{n,2} = 0, \tag{8}$$

if $P_m|h_m|^2 \geq e^{\frac{N}{D_m}}\left(e^{\frac{N}{D_m}} - 1\right)$.

B. Optimizing T_n

Without loss of generality, take the power allocation solution in (6) as an example. The corresponding overall energy consumption is given by

$$\begin{aligned} E_{H1} =& D_m|h_n|^{-2}\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}} - 1} - 1\right) \\ &+ T_n|h_n|^{-2}\left(e^{\frac{N}{T_n} - \frac{D_m}{T_n}\ln\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1}\right)} - 1\right). \end{aligned} \tag{9}$$

By defining $N' = D_m \ln\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1}\right)$, the overall energy consumption can be simplified as follows:

$$E_{H1} = \frac{D_m}{|h_n|^2}\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}} - 1} - 1\right) + \frac{T_n}{|h_n|^2}\left(e^{\frac{N-N'}{T_n}} - 1\right). \tag{10}$$

Define $f(x) \triangleq x\left(e^{\frac{a}{x}} - 1\right)$ which is shown to be a monotonically decreasing function of x for $x \geq 0$, where a is a constant. The first order derivative of $f(x)$ is given by

$$f'(x) = \left(e^{\frac{a}{x}} - 1\right) - e^{\frac{a}{x}}\frac{a}{x}. \tag{11}$$

Further define $g(y) \triangleq (e^y - 1) - ye^y$. One can find that $g(y)$ is a monotonically decreasing function of y for $y \geq 0$, since

$$g'(y) = e^y - e^y - ye^y = -ye^y \leq 0. \tag{12}$$

Therefore, $f'(x)$ is a monotonically increasing function of x, which means $f'(x) \leq f'(\infty) = 0$, and hence $f(x)$ is indeed a monotonically decreasing function of x. Therefore, $T^*_n = D_n - D_m$ for the hybrid NOMA solution shown in (6). Similarly, $T^*_n = D_n - D_m$ also holds for the other power allocation solutions.

C. Comparison of the Solutions

1) Comparing the two hybrid NOMA solutions: For the case of $e^{\frac{N}{D_m}}\left(e^{\frac{N}{D_m}} - 1\right) \geq P_m|h_m|^2 \geq e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}} - 1\right)$, the two hybrid NOMA solutions are feasible, and we will show that the solution in (7) outperforms the one in (6).

By using the fact that $T^*_n = D_n - D_m$, the overall energy consumption for the solution in (7) is given by

$$E_{H2} = D_n|h_n|^{-2}\left(e^{\frac{N}{D_n}} - 1\right), \tag{13}$$

and the energy consumption of the solution in (6) is given by (10). In order to show $E_{H1} \geq E_{H2}$, it is sufficient to show that the following inequality holds

$$\begin{aligned} &D_n e^{\frac{N}{D_n}}\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}} - 1}\right)^{\frac{D_m}{D_n-D_m}} - D_m\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}} - 1}\right)^{\frac{D_n}{D_n-D_m}} \leq \\ &(D_n - D_m)e^{\frac{N}{D_n-D_m}}. \end{aligned} \tag{14}$$

To prove the inequality in (14), we define the following function

$$\phi(x) = D_n e^{\frac{N}{D_n}} x^{\frac{D_m}{D_n-D_m}} - D_m x^{\frac{D_n}{D_n-D_m}}, \tag{15}$$

where $e^{\frac{N}{D_n}} \leq x \leq e^{\frac{N}{D_m}}$. The first order derivative of $\phi(x)$ is given by

$$\phi'(x) = \frac{D_m D_n}{D_n - D_m} x^{\frac{D_m}{D_n-D_m}}\left(e^{\frac{N}{D_n}} x^{-1} - 1\right). \tag{16}$$

By using the fact that $x \geq e^{\frac{N}{D_n}}$, $\phi'(x)$ can be upper bounded as follows:

$$\phi'(x) \leq \frac{D_m D_n}{D_n - D_m} x^{\frac{D_m}{D_n-D_m}}\left(e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_n}}\right)^{-1} - 1\right) = 0, \tag{17}$$

which shows that $\phi(x)$ is a monotonically decreasing function of x for $e^{\frac{N}{D_n}} \leq x \leq e^{\frac{N}{D_m}}$. Therefore, we have the following inequality

$$\begin{aligned} &D_n e^{\frac{N}{D_n}}\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}} - 1}\right)^{\frac{D_m}{D_n-D_m}} - D_m\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}} - 1}\right)^{\frac{D_n}{D_n-D_m}} \\ &\leq \phi\left(e^{\frac{N}{D_n}}\right) = D_n e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_n}}\right)^{\frac{D_m}{D_n-D_m}} - D_m\left(e^{\frac{N}{D_n}}\right)^{\frac{D_n}{D_n-D_m}} \\ &= D_n e^{\frac{N}{D_n-D_m}} - D_m e^{\frac{N}{D_n-D_m}}. \end{aligned} \tag{18}$$

Therefore, the inequality in (14) is proved, i.e., $E_{H1} \geq E_{H2}$ for $e^{\frac{N}{D_m}}\left(e^{\frac{N}{D_m}} - 1\right) \geq P_m|h_m|^2 \geq e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}} - 1\right)$.

2) Comparison of hybrid NOMA and pure NOMA: For the case of $P_m|h_m|^2 \geq e^{\frac{N}{D_m}}\left(e^{\frac{N}{D_m}} - 1\right)$, the pure NOMA solution in (8) and the hybrid NOMA solution in (7) are feasible. The energy consumption required by the pure NOMA solution is given by

$$\begin{aligned} E_N =& D_m|h_n|^{-2}\left(e^{\frac{N}{D_m}} - 1\right) \\ \geq& D_n|h_n|^{-2}\left(e^{\frac{N}{D_n}} - 1\right) = E_{H2}, \end{aligned} \tag{19}$$

where the inequality follows from the fact that $f(x)$ is a monotonically decreasing function of x for $x \geq 0$.

3) Comparison of OMA and hybrid NOMA: By following the same steps as in the previous subsection, it is straightforward to show that the hybrid NOMA solution shown in (7) outperforms OMA. The comparison between OMA and the hybrid NOMA solution shown in (6) is challenging and will be focused on in the following.

Recall the energy consumption for OMA is $E_{OMA} = (D_n - D_m)|h_n|^{-2}\left(e^{\frac{N}{D_n-D_m}} - 1\right)$. In order to show $E_{H1} \leq E_{OMA}$, it is sufficient to prove the following inequality

$$\begin{aligned} D_m|h_n|^{-2}\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1} - 1\right) + (D_n - D_m)|h_n|^{-2} \\ \times\left(e^{\frac{N}{D_n-D_m}}\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1}\right)^{-\frac{D_m}{D_n-D_m}} - 1\right) \leq \\ (D_n - D_m)|h_n|^{-2}\left(e^{\frac{N}{D_n-D_m}} - 1\right), \end{aligned} \tag{20}$$

where $|h_m|^{-2}\left(e^{\frac{N}{D_m}} - 1\right) \leq P_m \leq |h_m|^{-2}e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}} - 1\right)$.

Eq. (20) is equivalent to the following inequality:

$$\begin{aligned} D_m\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1} - 1\right) + (D_n - D_m)e^{\frac{N}{D_n-D_m}} \\ \times\left[\left(\frac{P_m|h_m|^2}{e^{\frac{N}{D_m}}-1}\right)^{-\frac{D_m}{D_n-D_m}} - 1\right] \leq 0. \end{aligned} \tag{21}$$

In order to prove (20), we define the following function

$$\begin{aligned} \varphi(x) =& D_m(x-1) \\ &+ (D_n - D_m)e^{\frac{N}{D_n-D_m}}\left[x^{-\frac{D_m}{D_n-D_m}} - 1\right]. \end{aligned} \tag{22}$$

The inequality in (20) can be proved if $\varphi(x) \leq 0$, for $1 \leq x \leq e^{\frac{N}{D_n}}$, which is proved in the following. The first order derivative of $\varphi(x)$ is given by

$$\varphi'(x) = D_m - D_m e^{\frac{N}{D_n-D_m}} x^{-\frac{D_n}{D_n-D_m}}, \tag{23}$$

which shows that $\varphi'(x)$ is a monotonically increasing function of x. By using the fact that $x \leq e^{\frac{N}{D_n}}$, $\varphi'(x)$ can be lower bounded as follows:

$$\begin{aligned} \varphi'(x) \leq& \varphi'\left(e^{\frac{N}{D_n}}\right) \\ =& D_m - D_m e^{\frac{N}{D_n-D_m}}\left(e^{\frac{N}{D_n}}\right)^{-\frac{D_n}{D_n-D_m}} = 0. \end{aligned} \tag{24}$$

Therefore, $\varphi(x)$ is a monotonically decreasing function of x. Since $x \geq 1$, we have

$$\varphi(x) \geq \varphi(1) = 0, \tag{25}$$

which proves the inequality in (20), i.e., $E_{OMA} > E_{H1}$. Therefore, hybrid NOMA outperforms pure NOMA and OMA, when all of them are feasible. When both the hybrid solutions are feasible, the solution in (7) outperforms the one in (6). Thus, the proof is complete.

Appendix B
Proof for Lemma 2

For the case of $|h_m|^{-2}e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}} - 1\right) \leq P_m \leq |h_m|^{-2}\left(e^{\frac{N}{D_n-D_m}} - 1\right)$, both the two solutions in (1) and (3) are feasible. With some algebraic manipulations, the overall energy consumption realized by the solution in (1) is given by

$$\begin{aligned} E_0 = D_m|h_n|^{-2}\left(e^{\frac{N}{D_n}}\left(1 + P_m|h_m|^2\right)^{\frac{D_m}{D_n}} - P_m|h_m|^2 - 1\right) \\ + (D_n - D_m)|h_n|^{-2}\left(e^{\frac{N}{D_n}}\left(1 + P_m|h_m|^2\right)^{\frac{D_m}{D_n}} - 1\right). \end{aligned}$$

The overall energy consumption with the solution in (7) is given by (13). In order to show that $E_0 \geq E_{H2}$, it is sufficient to prove the following inequality:

$$-D_m P_m|h_m|^2 + D_n e^{\frac{N}{D_n}}\left(1 + P_m|h_m|^2\right)^{\frac{D_m}{D_n}} \geq D_n e^{\frac{N}{D_n}}. \tag{26}$$

In order to prove (26), we define the following function

$$\psi(x) = -D_m x + D_n e^{\frac{N}{D_n}}(1+x)^{\frac{D_m}{D_n}}, \tag{27}$$

where $e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}} - 1\right) \leq x \leq e^{\frac{N}{D_n-D_m}} - 1$. The first order derivative of $\psi(x)$ is given by

$$\psi'(x) = -D_m + D_m e^{\frac{N}{D_n}}(1+x)^{\frac{D_m-D_n}{D_n}}. \tag{28}$$

Because $D_m < D_n$, $\psi'(x)$ is a monotonically decreasing function of x. Given $x \leq e^{\frac{N}{D_n-D_m}} - 1$, we have

$$\begin{aligned} \psi'(x) \geq& \psi'\left(e^{\frac{N}{D_n-D_m}} - 1\right) \\ =& -D_m + D_m e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_n-D_m}}\right)^{\frac{D_m-D_n}{D_n}} = 0, \end{aligned} \tag{29}$$

which means that $\psi(x)$ is a monotonically increasing function of x for $x \leq e^{\frac{N}{D_n-D_m}} - 1$. Therefore,

$$\psi(x) \geq \psi\left(e^{\frac{N}{D_n}}\left(e^{\frac{N}{D_m}} - 1\right)\right) \geq \psi(0) = D_n e^{\frac{N}{D_n}}. \tag{30}$$

Thus, (26) holds, i.e., $E_0 \geq E_{H2}$. The proof is complete.

References

[1] B. Makki, K. Chitti, A. Behravan, and M. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open Journal of the Commun. Society*, vol. 1, pp. 179–189, 2020.

[2] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond.* Springer Press, 2019.

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

[4] F. Zhou, Y. Wu, Y. Liang, Z. Li, Y. Wang, and K. Wong, "State of the art, taxonomy, and open issues on cognitive radio networks with NOMA," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 100–108, 2018.

[5] Z. Ding, R. Schober, and H. V. Poor, "A new QoS-guarantee strategy for NOMA assisted semi-grant-free transmission," *IEEE Trans. Wireless Commun.*, (submitted) Available on-line at arXiv:2004.12997.

[6] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, Aug. 2018.

[7] M. Elbayoumi, M. Kamel, W. Hamouda, and A. Youssef, "NOMA-assisted machine-type communications in UDN: State-of-the-art and challenges," *IEEE Commun. Surveys Tuts.*, pp. 1–1, 2020.

[8] M. Zeng, N. Nguyen, O. A. Dobre, and H. V. Poor, "Delay minimization for NOMA-assisted MEC under power and energy constraints," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1657–1661, 2019.

[9] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, "Joint power and time allocation for NOMA-MEC offloading," *IEEE Wireless Commun. Lett.*, vol. 68, no. 6, pp. 6207–6211, 2019.

[10] P. D. Diamantoulakis, K. N. Pappi, Z. Ding, and G. K. Karagiannidis, "Wireless-powered communications with non-orthogonal multiple access," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 12, pp. 8422–8436, Dec. 2016.