Energy-Efficiency Maximization for a WPT-D2D Pair in a MISO-NOMA Downlink Network

Shiyu Jiao, Fang Fang, Member, IEEE and Zhiguo Ding, Fellow, IEEE

Abstract—The combination of non-orthogonal multiple access (NOMA) and wireless power transfer (WPT) is a promising solution to enhance the energy efficiency of Device-to-Device (D2D) enabled wireless communication networks. In this paper, we focus on maximizing the energy efficiency of a WPT-D2D pair in a multiple-input single-output (MISO)-NOMA downlink network, by alternatively optimizing the beamforming vectors of the base station (BS) and the time switching coefficient of the WPT assisted D2D transmitter. The formulated energy efficiency maximization problem is non-convex due to the highly coupled variables. To efficiently address the non-convex problem, we first divide it into two subproblems. Afterwards, an alternating algorithm based on the Dinkelbach method and quadratic transform is proposed to solve the two subproblems iteratively. To verify the proposed alternating algorithm’s accuracy, partial exhaustive search algorithm is proposed as a benchmark. We also utilize a deep reinforcement learning (DRL) method to solve the non-convex problem and compare it with the proposed algorithm. To demonstrate the respective superiority of the proposed algorithm and DRL-based method, simulations are performed for two scenarios of perfect and imperfect channel state information (CSI). Simulation results are provided to compare NOMA and orthogonal multiple access (OMA), which demonstrate the superior performance of energy efficiency of the NOMA scheme.

Index Terms—Energy Efficiency, non-orthogonal multiple access (NOMA), wireless power transfer (WPT), Device-to-Device (D2D), convex optimization, deep reinforcement learning (DRL)

I. INTRODUCTION

With the development of wireless communication from fifth-generation (5G) to sixth-generation (6G), the demand for massive machine-type communications (mMTC) is raised to ultra-mMTC (umMTC) [1]. The emergence of new usage scenarios and applications, such as the Internet of Everything (IoE), dramatically drove this upgrade. However, simultaneously serving massive devices by utilizing the limited spectrum resource is challenging. In the meanwhile, the ultra-dense networks formed by massively connected devices lead to huge power consumption, which significantly increases the operating cost of wireless communication networks. Thus, a spectrum and energy efficient solution that enables ultra-dense networks is urgent and critical.

To support ultra-dense networks in 5G and 6G, D2D communication has still been regarded as a promising scheme and will be gradually appended to existing cellular networks [2]–[4]. D2D communication was introduced in 4G LTE as a kind of peer-to-peer short wireless transmission between two devices without relaying by base stations (BS) or access points (AP) [5], [6], which can mitigate the load on the BSs. Generally, D2D communication is classified into two categories: Inband D2D and Outband D2D [6]. Inband D2D communication utilizes the same licensed spectrum in cellular networks with cellular devices such as mobile phones. For outband D2D communication, it occurs in Ad-hoc networks such as Wi-Fi, Bluetooth etc., which is out of the scope of this paper. In terms of licensed spectrum utilizing, there are two ways to assign the spectrum to D2D devices, namely Underlay and Overlay [7], [8]. The Underlay type allows the licensed spectrum to be shared with both D2D devices and original cellular devices while the Overlay type divides the licensed spectrum into two parts and allocates them to cellular devices and D2D devices respectively. Although D2D has been widely studied in existing works [9]–[11], it still has many challenges that demand prompt solutions [12], [13]. For example, if deploying battery powered D2D pairs in an existing cellular network, resource allocation, interference controlling and energy efficiency improvement, etc., are required to enhance the network’s performance and prolong the D2D pair’s battery life.

To further improve the spectrum efficiency of the D2D pair, non-orthogonal multiple access (NOMA) can be applied in cellular networks [12]–[14]. In 6G, NOMA remains in the spotlight and is ever-evolved in academia and industry. In recent studies on using NOMA, the authors of [15] and [16] demonstrated its enormous potential in 6G and tremendous benefit for 6G. Furthermore, the spectral efficiency of NOMA enabled IoT network for 6G was further improved by [17]. Different from conventional multiple access techniques, including frequency-division multiple access (FDMA), time-division multiple access (TDMA), code-division multiple access (CDMA) and orthogonal frequency division multiple access (OFDMA) for previous generations of cellular communications, NOMA allows all users to share the same frequency band and channel coding at the same time. By applying NOMA, high mutual interference will be introduced when the NOMA users are decoding signals. Thus, successive interference cancellation (SIC) is applied at the receiver [18].

From the energy efficiency perspective, wireless power transfer (WPT) has been regarded as one of the promising technology to improve energy efficiency of the D2D pair, especially for the near-field scenarios in 6G [19]. The authors in [20] further discussed the application of WPT in 6G, especially in the scenarios of short-distance communications and massive battery-less devices deployed IoT networks. Generally, there are two types of WPT, i.e., time switching (TS) and power splitting (PS) WPT respectively [21]. In particular, for TS-WPT, the receiver periodically switches between harvesting energy mode and transmitting signal or decoding information mode [13], [22], [23], whereas the PS-WPT receiver splits
the received signal into two power level streams and then assigns them to the energy harvesting receiver and information decoding receiver respectively \[24\], \[25\]. Note that, in this paper, we consider TS-WPT into the D2D pair in a NOMA downlink network.

A. Related Works

In order to address the spectral and energy challenges in the 6G ultra-dense networks as aforementioned, we intuitively apply NOMA and WPT technologies to the D2D deployed networks, which is the theme of this paper. In literature, D2D, NOMA and WPT were combined and studied in pairs or all together for different scenarios. The authors in \[26\] analysed the performance of a NOMA uplink network consisting of a single non-energy constrained device and multiple energy-constrained WPT supported devices, which provides the research directions for WPT-NOMA. The authors in \[27\] maximized the uplink sum rate of multiple WPT-assisted devices in a single user downlink NOMA network, where the time switching coefficient and power allocation were alternatively optimized. In \[28\], the energy efficiency of a downlink SWIPT-enabled NOMA system with TS-based terminals was maximized by jointly optimizing the time switching coefficients of terminals and the power allocation strategy of the BS. The authors in \[29\] obtained the optimal power allocation scheme for a single-carrier single-uplink-user NOMA-enabled network by using convex optimization, where one D2D transmitter and two D2D receivers are taken into account. \[30\] optimized the resource allocation and channel assignment scheme in a NOMA downlink cellular network, where multiple D2D devices are deployed. A recent work \[31\] applied WPT to two NOMA uplink users group to improve the energy efficiency and spectrum efficiency, where users in these two groups perform energy harvesting and signal transmission alternatively.

B. Motivation and Contributions

Motivated by the aforementioned concept of Underlay type D2D and spectrum sharing policy of NOMA (i.e., allowing multiple devices to be multiplexed on the same frequency band), in this paper, we combine NOMA with Inband-Underlay D2D communication. Additional, considering that most D2D devices are battery-powered, solving the energy-saving problems to prolong battery life is imminent. Furthermore, due to the green communications attribute of WPT \[32\]–\[34\], this paper assumes that the D2D pair is WPT supported. In summary, appending WPT supported D2D pairs to a NOMA cellular network is a spectrum and energy efficient scheme while increasing the connection density.

Although the combination of NOMA and D2D communication can bring benefits such as higher spectrum efficiency and energy efficiency, there still are challenges that need to be overcome \[5\]. Considering the spectrum sharing of NOMA, severe co-channel interference to cellular users will be introduced when the D2D devices are appended to cellular networks. Therefore, interference control has to be carried out to guarantee the original cellular users’ quality of service (QoS) when D2D devices are deployed. On the other hand, massive-MIMO has been certainly applied to cellular networks, and the accuracy of beamforming design is indispensable.

Different from \[35\] maximizing the energy efficiency of a D2D pair in a single-antenna BS NOMA uplink network, this paper considers a WPT supported D2D pair in a multiple-input single-output (MISO) NOMA downlink system. The energy efficiency of a D2D pair is maximized by alternatively optimizing the beamforming vectors and the time switching coefficient under the premise of ensuring the QoS of NOMA downlink users. The main contributions are summarised as follows:

- In this paper, we consider a scenario where a WPT supported D2D pair (including a transmitter and a receiver) is inserted into a multi-user MISO-NOMA downlink network. Assume that the D2D transmitter adopts the harvested energy from the receivers before transmitting information strategy. This scenario is bound to appear in the process of future cellular networks upgrades to ultra-dense networks. The use of WPT further improves energy efficiency.
- The formulated energy efficiency maximization problem is non-convex as the two variables (beamforming vectors of the BS and time switching coefficient of the D2D pair) are highly coupled in both the fractional objective function and constraints. In order to address the challenges mentioned above, the problem is first analysed then transformed to a simpler form. Specifically, in the considered scenario, the constraints for the D2D transmitter energy harvesting stage can be removed, according to proposition 1. Subsequently, to tackle the non-convex problem, we split it into two subproblems and recast them into tractable convex forms. For time switching coefficient optimization, we use the Dinkelbach method to decouple the fractional objective function and convert constraints to linear form. For beamforming designing, the quadratic transform is applied for this multi-dimension case. Then an alternating algorithm based on the Dinkelbach method and quadratic transform is proposed to iteratively optimize beamforming vectors and time switching coefficient. Simulation results reveal that the proposed algorithm can converge perfectly.
- To show the superior performance of the proposed algorithm, especially to verify its accuracy, this paper proposes a partial exhaustive search algorithm which can avoid the alternating operation. Additionally, this paper contrasts the conventional convex optimization theory based algorithm (i.e., the proposed algorithm) with a deep reinforcement learning (DRL) approach (i.e., deep deterministic policy gradient (DDPG)). Simulation results reveal that the proposed algorithm can provide better performance when channel estimation is accurate, while the DDPG-based algorithm has the capability to mitigate the adverse impact caused by channel estimation error.
- The same energy efficiency maximization problem in the orthogonal multiple access (OMA) system is optimized by applying the proposed algorithm. Simulation results illustrate that the proposed algorithm is also applicable
to OMA. On the other hand, integrating WPT supported D2D pair with NOMA is superior to integrating with OMA.

C. Organization

The rest of this paper is arranged as follows. Section II describes the WPT supported D2D pair deployed MISO-NOMA downlink network as well as the energy harvesting and information transmission strategy. The problem formulation and preliminary handling are also discussed in section II. In section III, solutions to two subproblems, the proposed alternating algorithm and the partial exhaustive search for time switching coefficient are provided. Section IV introduces the DDPG algorithm and the application to the original problem. In section V, we discuss the scenario when channel estimation error exists. In section VI, simulation results are provided and analysed. In the end, we conclude this paper in section VII.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Consider a MISO-NOMA downlink network with one pair of WPT assisted D2D devices as shown in Fig. 1. This network consists of a base station (BS), K downlink users and a pair of D2D devices which are denoted by \( D_t \) (the D2D signal transmitter) and \( D_r \) (the D2D signal receiver) respectively. Assume that \( D_t \) can support the WPT and apply the harvest-energy-then-transmit strategy. Thus, \( D_t \) performs the energy harvesting then stores in the battery during the first \( \tau T \) seconds by utilizing the BS transmitted downlink signal. During the rest \((1 - \tau)T\) seconds, \( D_t \) sends its signal \( s_D \) to \( D_r \) by using the harvested energy, as shown in Fig. 2. \( \tau \) represents the time-switching coefficient \((0 \leq \tau \leq 1)\) and \( T \) is the duration of one time slot. For simplicity, we set \( T = 1 \) in this paper. Assume that downlink users receive signals \( s_k^{(1)} \) and \( s_k^{(2)} \) in \( \tau \) and \( 1 - \tau \) seconds respectively. According to the principle of Inband-Underlay, the WPT-D2D pair can share the same communication resource with NOMA downlink users. Assume that the BS is equipped with \( M \) antennas while D2D devices and users are equipped with a single antenna.

During the first stage (i.e., \( \tau \) seconds), the BS transmits the superposition signal \( s^{(1)} = \sum_{k=1}^{K} \omega_k s_k^{(1)} \) to all K downlink users while the WPT supported D2D transmitter \( D_t \) harvests energy from the BS with broadcast signal \( s_D \). The \( U_k \)'s received signal at this stage is

\[
y^{(\tau)}_k = h_k^H \sum_{k=1}^{K} \omega_k s_k^{(1)} + n_k,
\]

where \( h_k \in \mathbb{C}^{M \times 1} \) is the channel vector from the BS to the \( U_k \), \( \omega_k \in \mathbb{C}^{M \times 1} \) is the complex beamforming vector for the \( U_k \), \( s_k^{(1)} \) is the \( U_k \)'s required signal at the first stage and \( n_k \sim \mathcal{CN}(0, \sigma^2) \) denotes the additive White Gaussian noise (AWGN) at user \( U_k \). Assume that \( E\{|s_k^{(1)}|^2\} = 1 \).

According to the SIC principle, a stronger user (who has better channel gain) can decode the signal of weaker users (who has worse channel gain). Define that \( U_i \)'s channel is stronger than \( U_h \)'s. The data rate of \( U_i \) to decode \( k \) weaker users can be defined as

\[
P_{k \rightarrow i}^{(c)} = \tau \log(1 + \frac{|h_i^H \omega_k|^2}{\sum_{j=k+1}^{K} |h_i^H \omega_j|^2 + \sigma^2}),
\]

After removing the weaker users’ signal, \( U_i \) can decode its own signal by simply treating strong users’ signal as interference. Therefore, the data rate that \( U_i \) to decode its own signal in this stage is given by

\[
P_{i \rightarrow i}^{(c)} = \tau \log(1 + \frac{|h_i^H \omega_i|^2}{\sum_{j=1}^{K} |h_i^H \omega_j|^2 + \sigma^2}).
\]

In this paper, we assume that \( U_1 \) is the weakest user whereas \( U_K \) is the strongest user (i.e., \( 1 \leq k < t \leq K \) for \( (2) \)). In other words, the channel gains are sorted as \(|h_1|^2 \leq |h_2|^2 \leq \cdots \leq |h_K|^2 \). In the considered MISO-NOMA downlink system, assume that all the energy beams can be harvested. Therefore, the received power at \( D_t \) is given by

\[
P_t = \sum_{k=1}^{K} |h_{D_t}^H \omega_k|^2,
\]

where \( h_{D_t} \) is the channel vector from the BS to \( D_t \). Denote the BS maximum transmit power by \( P_{max} \). We have

![Fig. 1: System model.](image1)

![Fig. 2: Block structure.](image2)
\[ \sum_{k=1}^{K} |\omega_k|^2 \leq P_{\text{max}}. \] Assume that the harvested energy will be totally used to transmit the signal, whereas the circuit needed energy is provided by the battery. Therefore, the transmit power of \( D_i \) can be represented by

\[ P_t = \frac{\eta \tau P_c}{1 - \tau}, \tag{5} \]

where \( \eta \) is the RF energy conversion coefficient.

During the second stage (i.e., the rest \( 1 - \tau \) seconds), downlink users receive \( s^{(2)} \) while \( D_i \) transmits its signal to \( D_r \). The transmitted signal from \( D_i \) to \( D_r \) is denoted by \( s_D \).

Since the D2D pair shares the same frequency band, the signal \( s_D \) sent by \( D_i \) will interfere with downlink users to receive the signal \( s^{(2)} \). Therefore, the \( U_k \)'s received signal in the second stage is given by

\[ y_{(1-\tau)} = \sqrt{P_t} h_{dk}s_D + h^H_{i} \sum_{k=1}^{K} \omega_k s^{(2)}_k + n_k, \tag{6} \]

where \( h_{dk} \) is the channel between the \( D_i \) and the \( U_k \). For the D2D pair, the received signal at the \( D_i \) is given by

\[ y_{D_i}^{(1-\tau)} = \sqrt{P_t} h_{dd}s_D + h^H_{D_i} \sum_{k=1}^{K} \omega_k s^{(2)}_k + n_D, \tag{7} \]

where \( h_{dd} \) denotes the channel gain from \( D_i \) to \( D_r \). \( h_{D_i} \) is the channel vector between the BS and \( D_r \), and \( n_D \) is the AWGN at the D2D receiver.

Since the \( D_i \)'s transmitted signal is not superimposed in the BS transmitted signal and is unexpected introduced signal, the BS simply regards it as interference for downlink users. From Proposition 1, it can be observed that the constraint (14b) is equivalent to the following two constraints:

\[ R_{k\rightarrow t} \geq R_{\text{min}}, \tag{15} \]

\[ R_{kightarrow k} \geq R_{\text{min}}. \tag{16} \]

**Proposition 1.** \( R_{kightarrow k} \geq R_{\text{min}} \) is equivalent to \( \frac{1}{1-\tau}P_{t\rightarrow k}^{(1-\tau)} \geq R_{\text{min}}, \) and (16) can also be recast in the same way.

**Proof.** Define \( A = \log(1 + \frac{|h^H_{i}\omega_k|^2}{\sum_{j=k+1}^{K} |h^H_{i}\omega_j|^2 + \sigma^2}) \), \( B = \log(1 + \frac{|h^H_{i}\omega_k|^2}{\sum_{j=k+1}^{K} |h^H_{i}\omega_j|^2 + \sigma^2}) \), and \( R = R_{\text{min}} \). Substitute (3) and (9) into (10). We have

\[ \tau A + (1 - \tau)B \geq R, \tag{17} \]

which is equivalent to

\[ \tau(A - R) \geq (\tau - 1)(B - R). \tag{18} \]

It can be observed that \( \tau - 1 \leq 0 \) and \( A \geq B \) are always held. Therefore, (18) can always be satisfied if \( B \geq R \) is held.

Hence, the constraint (14b) is equivalent to

\[ \frac{1}{1-\tau}P_{t\rightarrow k}^{(1-\tau)} \geq R_{\text{min}} \]

\[ \Rightarrow \log(1 + \frac{|h^H_{i}\omega_k|^2}{P_t|h_{dd}|^2 + \sum_{j=k+1}^{K} |h^H_{i}\omega_j|^2 + \sigma^2}) \geq R_{\text{min}} \tag{19} \]

and

\[ \frac{1}{1-\tau}P_{k\rightarrow k}^{(1-\tau)} \geq R_{\text{min}} \]

\[ \Rightarrow \log(1 + \frac{|h^H_{i}\omega_k|^2}{P_t|h_{dd}|^2 + \sum_{j=k+1}^{K} |h^H_{i}\omega_j|^2 + \sigma^2}) \geq R_{\text{min}}. \tag{20} \]
Define \( \bar{\tau} \triangleq \frac{1}{\tau} \), then [13] can be recast as \( E_c = \bar{\tau}(\eta P_r + P_c) + P_c \). According to the Proposition 1 and after some simple manipulation, the problem P1 can be further reduced to

\[
P_2: \max_{\{\tau, \omega\}} \log(1 + \frac{\bar{\tau} \eta P_r |h_{dk}|^2}{\tau (\eta P_r + P_c) + P_c}) \quad (21a)
\]

s.t.

\[
\tau \eta P_r |h_{dk}|^2 + \sum_{k=1}^{K} |h_k^H \omega_k|^2 \geq \gamma_{\min},
\]

\[
1 \leq k < t \leq K, \quad |h_k^H \omega_k|^2 \geq \gamma_{\min}
\]

\[
1 \leq k \leq K, \quad \sum_{k=1}^{K} |\omega_k|^2 \leq P_{\text{max}}, \quad 1 \leq k \leq K
\]

\[
\bar{\tau} \geq 0,
\]

where \( \gamma_{\min} = R_{\min} - 1 \). P2 is still non-convex due to the non-convex objective function (21a) and two non-convex constraints (21b) and (21c) because the time switching dependent variable \( \bar{\tau} \) and the beamforming \( \omega \) are highly coupled in the problem, which is difficult to obtain the optimal solution directly. In the next section, P2 is divided into two subproblems corresponding \( \bar{\tau} \) and \( \omega \) respectively. The proposed solutions to the two subproblems and an alternating algorithm will be introduced in the next section.

### III. Time Switching Coefficient and Beamforming Optimization

In order to solve the problem discussed in the last section, an alternating algorithm is proposed in this section based on the Dinkelbach method [37] and multi-dimension quadratic transform [38] for fractional programming. Specifically, the problem is divided into two subproblems, one is the time switching coefficient \( \tau \) optimization by applying the Dinkelbach method, and the other is beamforming vectors designing by applying the multi-dimension quadratic transform. Each subproblem is analysed and converted from a non-concave form to a concave form. Afterwards, these tractable subproblems can be solved by using convex optimization tools such as CVX and fmincon on Matlab.

#### A. Time Switching Coefficient Optimization

For given beamforming vectors \( \omega = [\omega_1, \cdots, \omega_k], 1 \leq k \leq K \), the problem P1 can be reduce to

\[
P_3: \max \quad \log(1 + \frac{\tau A}{\tau B + C}) \quad (22a)
\]

s.t.

\[
\tau D_k + E_{k,t} \leq 0, \quad 1 \leq k < t \leq K,
\]

\[
\tau F_k + G_k \leq 0, \quad 1 \leq k \leq K,
\]

\[
\bar{\tau} \geq 0,
\]

where

\[
A = \frac{\eta P_r |h_{dk}|^2}{\sum_{k=1}^{K} |h_k^H \omega_k|^2 + \sigma^2}, \quad B = \eta P_r + P_c, \quad C = P_c,
\]

\[
D_k = \eta P_r |h_{dt}|^2, \quad E_{k,t} = \sum_{j=k+1}^{K} |h_k^H \omega_j|^2 + \sigma^2 - |h_k^H \omega_k|^2 \gamma_{\min},
\]

\[
F_k = \eta P_r |h_{dk}|^2 \quad \text{and} \quad G_k = \frac{\sum_{j=k+1}^{K} |h_k^H \omega_j|^2 + \sigma^2 - |h_k^H \omega_k|^2}{\gamma_{\min}}.
\]

The second-order derivative of (22a)’s numerator is given by

\[
\frac{\partial^2 \log(1 + A \bar{\tau})}{\partial \bar{\tau}^2} = - \frac{A^2}{\ln(2(1 + A \bar{\tau}))^2} \leq 0.
\]

Therefore, the numerator of (22a) is concave with respect to \( \bar{\tau} \). On the other hand, the denominator of (22a) is a linear function with respect \( \bar{\tau} \). It can be easily found that (22a) is a single-ratio concave-convex function of \( \bar{\tau} \), and hence Dinkelbach method can be applied to transform it into a concave function [37].

#### Proposition 2. The maximum EE can be achieved when \( F(q^*) = 0 \), where \( F(q) \) is defined as follows

\[
P_4: \log(1 + \bar{\tau} A) - q^*(\bar{\tau} B + C) \quad (24a)
\]

s.t.

\[
22(b) - 22(c),
\]

where

\[
q^* = \frac{\log(1 + \bar{\tau} A)}{\bar{\tau} B + C}.
\]

Proof. Please refer to [37].

For a given \( q \) the objective function (24a) is a log-concave function minus a convex function with respect to \( \bar{\tau} \), which yields a concave maximization problem. Therefore, the problem is reduced to a linear constraints concave problem P4 and can be solved by convex optimization tools. Finally, the optimized time switching coefficient \( \tau^* \) can be obtained by

\[
\tau^* = \frac{\bar{\tau}^*}{\bar{\tau}^* + \tau}.
\]

#### B. Beamforming Vectors Optimization

The last subsection developed Dinkelbach method to optimized the time switching coefficient \( \tau \) under fixed beamforming vectors \( \omega \). This section focuses on optimizing the beamforming vectors by regarding \( \tau \) as a constant.

For a given time switching coefficient \( \tau \), the problem P2 can be recast as follows

\[
P_5: \max_{\{\omega\}} \frac{\log(1 + \frac{\bar{\tau} \eta P_r |h_{dk}|^2}{\tau (\eta P_r + P_c) + P_c})}{\frac{\tau \eta P_r |h_{dt}|^2 + \sum_{j=k+1}^{K} |h_k^H \omega_j|^2 + \sigma^2}{\gamma_{\min}}} \quad (26a)
\]

s.t.

\[
\frac{h_k^H \omega_k)^H (h_k^H \omega_k)}{\tau \eta |h_{dt}|^2 |h_k^H \omega_j|^2 + \sigma^2} \geq \gamma_{\min},
\]

\[
1 \leq k < t \leq K,
\]

\[
\frac{h_k^H \omega_k)^H (h_k^H \omega_k)}{\tau |h_{dk}|^2 |h_k^H \omega_j|^2 + \sigma^2} \geq \gamma_{\min},
\]

\[
1 \leq k \leq K,
\]

\[
\sum_{k=1}^{K} |\omega_k|^2 \leq P_{\text{max}}, \quad 1 \leq k \leq K,
\]

where \( P_c = \frac{\sum_{k=1}^{K} |h_k^H \omega_k|^2}{\gamma_{\min}} \). The problem P5 is not a convex optimization problem as the existing of the non-convex objective function (26a) and the two non-convex constraints (26b) and (26c) with respect to \( \omega \). Note that the Dinkelbach’s method is no longer applicable as the objective function is not a concave-convex fractional form. In order to transform P5 to
a tractable convex optimization problem, the multi-dimension quadratic transform \cite{38} is applied.

**Lemma 1.** For $A_m \in \mathbb{C}^{x \times b}$ and $B_m \in \mathbb{C}^{b \times b}$, $m \in \mathbb{N}_+$, we have

$$
\sum_{m=1}^{M} A_m^H(x)B_m^{-1}(x)A_m(x)
= \max_z \sum_{m=1}^{M} (2\text{Re}\{z_m^H A_m(x)\} - z_m^H B_m(x)z_m),
$$

(27)

where $z_m$ are introduced auxiliary variables.

**Proof.** Define $f(z_m) = 2\text{Re}\{z_m^H A_m(x)\} - z_m^H B_m(x)z_m$. Note that $f(z_m)$ is a linear function minus a quadratic function with respect to $z_m$ (i.e., a concave function in terms of $z_m$). Therefore, the maximum value of $f(z_m)$ can be achieved when $\frac{\partial f(z_m)}{\partial z_m} = 0$ is satisfied. The optimal $z_m^*$ that can maximize $f(z_m)$ is $z_m^* = B_m(x)^{-1}A_m(x)$. Hence, the maximum value of $f(z_m)$ can be obtained by substituting $z_m^*$ into $f(z_m)$ (i.e., $f(z_m^*) = A_m(x)^H B_m(x)A_m(x)$). The equivalence of (27) now is established.

**Proposition 3.** The objective function \( (26a) \) is equivalent to the following concave form

$$
f_{qq}(W, y, z) = 2y\log(1 + \bar{\tau} \eta) h_{dd}^2 \sum_{k=1}^{K} (2\text{Re}\{z_k^H h_k^H \omega_k\} - z_k^H B_k(x)z_k)
= y^2(\bar{\tau} \eta \sum_{k=1}^{K} [h_k^H \omega_k]^2 + (1 + \bar{\tau}) P_r)
= y^2(\bar{\tau} \eta \sum_{k=1}^{K} [h_k^H \omega_k]^2 + (1 + \bar{\tau}) P_r),
$$

(28)
in terms of $\omega$, if the introduced auxiliary variables $y$ and $z = \{z_1, \ldots, z_K\}$ can satisfy \( (29) \) and \( (29) \) respectively.

$$
z_k^* = \frac{h_k^H \omega_k}{P_r + \sigma^2}, 1 \leq k \leq K.
$$

(29)

$$
y^* = \sqrt{\frac{R(W)}{E(W)}},
$$

(30)

where $R(W) = \log(1 + \bar{\tau} \eta h_{dd}^2 P_r)$ and $E(W) = \bar{\tau} \eta P_r + (1 + \bar{\tau}) P_r$. $W$ refers to the collection of $\{\omega_k\}$.

**Proof.** In order to decouple the numerator and denominator of \( (26a) \), the single-ratio quadratic transform \cite{38} is first applied.

$$
f_q(W, y) = 2yR(W)^{1/2} - y^2E(W),
$$

(31)

(31) is equivalent to \( (26a) \) if $f_q(W, y)$ can achieve the maximum value with $y$. The first-order derivative of (31) with respect to $y$ is $\frac{\partial f_q(W)}{\partial y} = 2\sqrt{R(W)} - 2yE(W)$. Since (31) is a quadratic function of $y$, its optimal $y^*$ can be obtained by letting $\frac{\partial f_q(W)}{\partial y} = 0$, which yields (30). Substitute (30) into (31) the objective function \( (26a) \) is recovered. In terms of $W$ in (31), note that $-E(W)$ is concave due to its minus quadratic form with respect $\omega$. However, \( (31) \) is still non-concave in terms of $W$ because the concavity of $R(W)$ is unprovable.

To restore the concavity of $R(W)$, Lemma 1 is used to each term of the SINR of $R(W)$ (i.e., each term of $\sum_{k=1}^{K} [h_k^H \omega_k]^2$).

Thus, $R(W)$ can be recast to

$$
R(W) = \log(1 + \bar{\tau} \eta) h_{dd}^2 \left(2\text{Re}\{z_k^H h_k^H \omega_k\} - z_k^H B_k(x)z_k\right),
$$

(32)

Similarly, (32) is equivalent to the numerator of \( (26a) \) when \( (32) \) can achieve its maximum value with optimal $z^*$ (i.e., when $z_k$ satisfies \( (29) \)), where $z$ denotes the collection of $z_k$. In this way, the term inside log function of \( (32) \) is converted to the summation of linear functions minus quadratic functions with respect to $\omega$. The form of \( (32) \) can be also called *log-concave* \cite{39}. Furthermore, it is worthy to mention that the function $f(x) = x^2$ is concave and non-decreasing. Hence, by substituting \( (32) \) into \( (31) \), the concavity of the equation \( (28) \) with respect to $\omega$ is obtained.

Similarly, Lemma 1 is also applicable to attain the concavity of constraints \( (26b) \) and \( (26c) \). They can be transformed into two concave sets with $\omega$ which are shown as follows:

$$
\frac{|h_k^H \omega_k|^2}{\bar{\tau} \eta h_{dd}^2 P_r + \sum_{j=k+1}^{K} |h_j^H \omega_j|^2 + \sigma^2}
= \max_{\nu_{t,k}} 2\text{Re}\{\nu_{t,k}^H h_k^H \omega_k\} - \nu_{t,k}^H \omega_k V_{t,k}^H V_{t,k}
\geq \gamma_{min}, 1 \leq t < k \leq K
$$

$$
\frac{|h_k^H \omega_k|^2}{\bar{\tau} \eta h_{dd}^2 P_r + \sum_{j=k+1}^{K} |h_j^H \omega_j|^2 + \sigma^2}
= \max_{\mu_k} 2\text{Re}\{\mu_k^H h_k^H \omega_k\} - \mu_k^H \omega_k \beta_k \mu_k
\geq \gamma_{min}, 1 \leq k \leq K
$$

where $\alpha_{t,k} = |h_{dd}^2 \bar{\tau} \eta P_r + \sum_{j=k+1}^{K} |h_j^H \omega_j|^2 + \sigma^2$ and $\beta_k = |h_k^H |^2 \bar{\tau} \eta P_r + \sum_{j=k+1}^{K} |h_j^H \omega_j|^2 + \sigma^2$. $\nu_{t,k}$ and $\mu_k$ are two introduced auxiliary variables whose optimal values are given by

$$
\nu_{t,k}^* = \frac{h_k^H \omega_k}{\alpha_{t,k}}, (1 \leq t < k \leq K),
$$

(35)

$$
\mu_{k}^* = \frac{h_k^H \omega_k}{\beta_k}, (1 \leq k \leq K).
$$

(36)

Define the collection of $\{\nu_{t,k}\}$ and $\{\mu_k\}$ as $\nu$ and $\mu$ respectively. By using \( (28) \), \( (33) \) and \( (34) \), the problem $P5$ can be reformulated as

**P6:** \( \max_{W, y, z, \nu, \mu} f_{qq}(W, y, z) \)

s.t.

$$
2\text{Re}\{\nu_{t,k}^H h_k^H \omega_k\} - \nu_{t,k}^H \omega_k V_{t,k}^H V_{t,k}
\geq \gamma_{min}, 1 \leq t < k \leq K.
$$

(37b)

$$
2\text{Re}\{\mu_k^H h_k^H \omega_k\} - \mu_k^H \omega_k \beta_k \mu_k
\geq \gamma_{min}, 1 \leq k \leq K,
$$

(37c)

$$
\sum_{k=1}^{K} |\omega_k|^2 \leq P_{max}, 1 \leq k \leq K.
$$

(37d)
Algorithm 1: Proposed quadratic transform and Dinkelbach method based alternative algorithm

1: **Initialization:** Initialize $W$ and $\tau$ to a feasible value.
2: repeat
3:   Update $z_k$ by using (29).
4:   Update $y$ by using (30).
5:   Update $\nu_{t,k}$ by using (35).
6:   Update $\mu_k$ by using (36).
7:   With fixed $z_k$, $y$, $\nu_{t,k}$ and $\mu_k$, solve the problem P6 and obtain the optimized $W$.
8:   With optimized $W$, update $q$ by using (25).
9:   With fixed $q$, solve the problem P4 and obtain the optimized $\tau$
10: until The value of $\theta$ is convergent.

Algorithm 2: Partial Exhaustive Search for $\tau$

1: **Initialization:** Initialize $W$ and $\tau$ to a feasible value. Initialize the step size $\xi$.
2: for $\tau = 0.001 : \xi : 0.999$ do
3:   repeat
4:     Update $z_k$ by using (29).
5:     Update $y$ by using (30).
6:     Update $\nu_{t,k}$ by using (35).
7:     Update $\mu_k$ by using (36).
8:     With fixed $z_k$ and $y$, solve the problem P6 and obtain the optimized $W$.
9:   until The value of $\theta$ is convergent.
10: end for
11: Select the $\tau$ corresponding to the maximum $\theta$.

For given $y$, $z$, $\nu$, and $\mu$, the (37a) is a concave function and constraints (37b), (37c) are all convex set in regard to $\omega$. Hence, problem P6 is a convex optimization problem [39], and therefore can be solved by convex optimization tools such as CVX or Matlab fmincon. Up to now, the original problem P1 has been solved by alternatively solving the subproblem P4 and P6. The quadratic transform and Dinkelbach method based alternating algorithm to maximize the energy efficiency of a WPT-D2D pair in a MISO-NOMA downlink network is summarised in Algorithm 1.

C. A Partial exhaustive search based algorithm for time switching coefficient optimization

This subsection provides a simple algorithm based on a partial exhaustive search for $\tau$ that can avoid the alternating process to further verify the accuracy of the proposed algorithm. As discussed in the last subsection, the energy maximization problem can be transformed to a convex problem with respect to beamforming vectors $\omega$ for a given time-switching coefficient $\tau$. Therefore, the solution can be obtained by solving P6 for all $\tau$ and selecting the one that corresponding to the maximum energy efficiency. The partial exhaustive search algorithm is summarised in Algorithm 2.

IV. A REINFORCEMENT LEARNING BASED APPROACH TO MAXIMIZE THE ENERGY EFFICIENCY

To compare with the proposed conventional convex optimization theory based algorithm, a reinforcement learning based algorithm, DDPG, is first introduced in this part. Afterwards, the neural networks’ structure and training procedure are provided. At the end of this section, we discuss the application of DDPG to the proposed problem including the setup of action, state, reward, as well as constraints handling.

A. A brief introduction to DDPG:

Reinforcement learning (RL) is neither like supervised learning uses an external supervisor labelled data set to learn, nor like unsupervised learning which aims to find the hidden structure in the unlabelled collections [40]. RL learns through the way that letting the agent interacts with the environment. Specifically, in RL, the agent decides what actions should be taken according to the current observation (also termed state) and then obtains the corresponding reward. Macroscopically, RL aims to find an optimal action that maximizes the reward. RL can generally be divided into two types which are value-based and policy-based respectively. Q-learning and state-action-reward-state’-action’ (SARSA) are two typical value-based learning. They only solve the problem with low-dimension discrete actions. Policy gradient (PG), as a policy-based RL, can solve the problems with continuous actions. However, PG usually convergents at a local optimal and evaluates a policy inefficient. The combination of Q-learning and deep neural networks (DNN) derives deep Q network (DQN), which is applicable to the problems when the state space and the discrete action space are enormous. To handle the problems with high-dimension continuous actions, DDPG is proposed by integrating DQN and GP [41].

B. Training neural networks:

Different from other RL methods, DDPG has four neural networks:

- An actor network $\mu(s|\theta_\mu)$: input current state $s$ then output action $a$.
- A critic network $Q(s,a|\theta_q)$: input current $a$ and $s$ then output Q-value.
- A target actor network $\mu'(s'|\theta_\mu')$: input state $s'$ then output target action $a'$.
- A target critic network $Q'(s',a'|\theta_q')$: input $a'$ and $s'$ then output target Q-value.

$\theta$ represents parameters of the corresponding neural network. The same as DQN, DDPG applies experience replay as well. Specifically, all transitions $\{a, s, r, s'\}$ are first stored into the experience replay buffer, then a certain number of transitions are randomly sampled to train those networks. The mathematical expression of the training process is as follows. To train the actor network, the gradient ascend and chain rule are used for the Q-value function

$$\nabla_{\theta_\mu} J = \frac{1}{N_B} \sum_{t=1}^{N_B} \nabla_a Q(s_t, \mu(s_t|\theta_\mu)|\theta_q) \nabla_{\theta_\mu} \mu(s_t|\theta_\mu), \tag{38}$$
where \( N_B \) is the number of sampled transitions (also termed mini-batch). The critic network is trained by minimizing the loss between the current Q-value and target state-value

\[
L = (y - Q(s, a|\theta_q))^2,
\]

where \( y \) is the target value for the previous state-value which is given by

\[
y = r + \gamma Q'(s', \mu'(s'|\theta_q'|\theta_p')).
\]

where \( r \) represents the reward and \( \gamma \) denotes the discount factor. To update the two target networks that have the same framework as their corresponding evaluation networks, soft updating is applied.

\[
\theta'_\mu = \xi \theta_\mu + (1 - \xi) \theta'_\mu, \quad \theta'_q = \xi \theta_q + (1 - \xi) \theta'_q.
\]

where \( \xi \) denotes the soft updating coefficient.

C. Application DDPG to the problem:

In this paper, the original problem \( P1 \) is solved directly by DDPG. Here we suppose that the BS is the agent and it can observe the CSI and downlink users’ data rate.

1) Action Space: As the optimization needed variables are beamforming vectors and time switching coefficient, naturally the action is defined as

\[
[\tau, \text{Re}\{\omega_1\}, \ldots, \text{Re}\{\omega_k\}, \text{Im}\{\omega_1\}, \ldots, \text{Im}\{\omega_k\}].
\]

Note that complex numbers have to be split to real part and the imaginary part.

2) State Space: For the considered system model, we define the state as follows

\[
[|h_1|^2, \ldots, |h_k|^2, |h_{D_1}|^2, |h_{D_2}|^2, |h_{dd}|^2, |h_{d_1}|^2, \ldots, |h_{d_k}|^2, R_1, \ldots, R_k, R_{t,k}, \omega_1^2, \ldots, \omega_k^2],
\]

where \( R_{t,k} \) denotes the collection of the data rate of user \( t \) to decode user \( k \), \( 1 \leq k < t \leq K \).

3) Reward: Our aim is to maximize the energy efficiency which can fit the goal of the DDPG algorithm to maximize the reward. Therefore, the objective function \((44a)\) is naturally used as the reward.

\[
r = \frac{R_D}{E_c}.
\]

To ensure the constraint \((44b)\) is satisfied, we introduce a punishment mechanism to the reward for each step as follows: 1) if all data rates of downlink users (including \(9\) and \(8\)) can satisfy the QoS requirement, the reward is given by \((44)\) and stored into the replay buffer. 2) if there exist data rates of downlink that are less than \( R_{\text{min}} \), the reward is given by 0. This operation helps neural networks to be trained in the correct direction.

Lemma 2. Optimal beam vectors are obtained if and only if the equality of \((14c)\) is established.

Proof. See appendix. \(\square\)

\(^1\)There is no any processing to the variables highly coupled non-convex problem including the objective function and constraints.

According to lemma 2, we apply normalization to the output beam vectors in each step to guarantee the power constraint \((14c)\) can be satisfied. First, the total transmit power of the beam vectors output by the actor in each step can be calculated by

\[
P_o = \sum_{k=1}^{K} |\omega_k|^2,
\]

where \( \omega^o \) represents the output beamforming vectors and hence \( P_o \) is the \( \omega^o \) determined transmit power. Therefore, keep the direction of \( \omega^o \) fixed, and reallocate the power according to the power ratio \( \frac{|\omega_k|^2}{P_o} \), the optimized beam vector of the \( k \)-th user is given by

\[
\omega^*_k = \omega^o_k \sqrt{\frac{P_{\text{max}}}{P_o}}.
\]

Finally, the total power of the optimized beam vectors can satisfy the constraint \((14c)\).

V. ENERGY EFFICIENCY MAXIMIZATION FOR D2D PAIR WITH IMPERFECT CSI

Channel estimation is one of the most important stages in real wireless communications. However, obtaining the perfect CSI in practice remains a challenge \([42]\). Therefore, this section investigates the energy efficiency maximization problem under considering the channel estimation error.

Assume that channel estimation is performed only for downlink signal transmission and D2D communication, and hence the channel estimation error exists in those corresponding channels. The channel between the BS and the \( k \)-th downlink user with estimation error can be modelled as follows \([43]\):

\[
\hat{h}_{k} = h_k + \epsilon_k,
\]

where \( \epsilon_k \) denotes the channel estimation error which follows the complex Gaussian distribution with zero mean and \( \sigma^2 \) variance. Similarly, the channel between the D2D transmitter and the receiver is given by

\[
\hat{h}_{dd} = h_{dd} + \epsilon_{dd}.
\]

The received signal of the \( k \)-th user during the \( 1 - \tau \) seconds is

\[
y_{k}^{(1-\tau)} = \sqrt{P_t h_{dk} s_D} + \sum_{k=1}^{K} \omega_k s_k^{(2)} + \epsilon_k \sum_{k=1}^{K} \omega_k s_k^{(2)} + n_k.
\]

The received signal of the D2D receiver during the \( 1 - \tau \) seconds is

\[
y_{D_e}^{(1-\tau)} = \sqrt{P_t h_{dd} s_D} + \sqrt{P_t \epsilon_{dd} s_D} + \sum_{k=1}^{K} \omega_k s_k^{(2)} + n_D.
\]

The data rate that the \( k \)-th downlink user to decode its own signal is denoted by \([44]\)

\[
R_{k_{\text{self}}-k}^{(1-\tau)} = (1 - \tau) \log(1 + \frac{|h_k^H \omega_k|^2}{P_1|h_{dk}|^2 + \sum_{j=k+1}^{K} |h_{dk}^j \omega_j|^2 + \sum_{k=1}^{K} |\epsilon_k^H \omega_k|^2 + \sigma^2}).
\]

(51)
The date rate that the $t$-th strong user to decode the $k$-th weak user’s signal is given by
\[
\hat{R}_{k\rightarrow t}^{(1-\tau)} = (1 - \tau) \log(1 + \frac{|h_{t}^{H} \omega_{k}|^2}{P_{t}|h_{dl}|^2 + \sum_{j=k+1}^{K} |h_{j}^{H} \omega_{j}|^2 + \sum_{k=1}^{K} |e_{k}^{H} \omega_{k}|^2 + \sigma^2}),
\]
(52)
The data rate of the D2D receiver is
\[
\hat{R}_{D} = (1 - \tau) \log(1 + \frac{P_{t}|h_{dl}|^2}{\sum_{k=1}^{K} |h_{Dk}^{H} \omega_{k}|^2 + P_{t}|e_{dd}|^2 + \sigma^2}).
\]
(53)
For the imperfect CSI scenario, the problems P1 becomes
\[
P7: \max_{\{\tau, \omega\}} \frac{\hat{R}_{D}}{E_c} \quad \text{s.t.} \quad \min\{\hat{R}_{k\rightarrow t}, \hat{R}_{k\rightarrow k}\} \geq R_{min}, 1 \leq k \leq t \leq K
\]
(54a)
\[
\sum_{k=1}^{K} |\omega_{k}|^2 \leq P_{max}, 1 \leq k \leq K
\]
(54b)
\[
0 \leq \tau \leq 1.
\]
(54d)
The problem P7 can also be solved by following the same processes and algorithms as discussed in the last section.

VI. SIMULATION RESULTS
In this section, we present all simulation results to analyse the performance of the proposed algorithm. The proposed algorithm is compared with the partial exhaustive search and DDPG. We also compared the D2D pair’s energy efficiency when they are deployed in a NOMA system and an OMA system.

In simulations, we assume that positions of the WPT supported D2D pair are fixed and the short-range communication is performed. The downlink users are assumed randomly distributed outdoor. We define a two dimensional horizontal Cartesian coordinate systems and the BS is located at the original point $(x_0, y_0) = (0, 0)$, and the positions of the $D_t$ and $D_r$ are $(x_t, y_t) = (0, 9)$ and $(x_r, y_r) = (0, 10)$ respectively. The positions of downlink users are randomly distributed in a certain range (i.e., $x_k, y_k \in [3, 8]$). For all simulations, channels are assumed to be the Rayleigh fading and the path loss is also considered. Therefore, the channels can be expressed as
\[
h_{xim} = \frac{h_{Ray}}{\sqrt{d^\alpha}},
\]
(55)
where $h_{Ray}$ represents Rayleigh channel coefficient vector, and $d$ and $\alpha$ are the corresponding distance and path loss coefficients, respectively. We set the path loss coefficient between the BS and outdoor downlink users to $\alpha_0 = 2.5$, the path loss coefficient between $D_t$ and $D_r$ is $\alpha_1 = 2$, $\alpha_2 = \alpha_3 = 3$ are the path loss coefficient between the BS and $D_1$, and between $D_2$ and downlink users, respectively. The number of downlink users is $K = 4$ and the noise power is set to $\sigma^2 = -94$ dBm.

Fig. 3 shows the convergence of the proposed algorithm 1 for both the NOMA and OMA schemes. In the simulation setting, we uses the same randomly generated channel vectors for NOMA and OMA, and the parameters are set as follows: transmit power $P_{max} = 20$ dBm, number of antennas $M = 10$ and number of downlink users $K = 4$. It can be observed that the proposed algorithm can converge for both NOMA and OMA schemes. In particular, the maximum value of the energy efficiency can be achieved within 7 iterations. On the other hand, this figure preliminary demonstrates the superiority of NOMA on the energy efficiency.

Fig. 4 presents the energy efficiency versus transmit power with different schemes. In this figure, we assume that the perfect CSI can be obtained, $K = 4$ and $M = 10$. The minimum target data rate is set as $R_{min} = 0.1$ bps/Hz and the RF energy conversion coefficient is $\eta = 0.1$. The randomness caused by the randomly generated positions and channels is averaged by performing Monte Carlo simulations. It is shown that regardless of what algorithm or multiple access is used, the energy efficiency of the WPT supported D2D pair increases...
Fig. 5: Energy efficiency versus transmit power with different M.

with the increase of the BS’s transmit power. However, the performances that different schemes can provide are significantly different. In this simulation, we choose the step size of exhaustive search $\xi = 0.1$. It can be seen that the performance of the partial exhaustive search for $\tau$ is slightly worse than the proposed algorithm. On the other aspect, the comparison with the partial exhaustive search algorithm further verifies the accuracy of the proposed algorithm. The gap between the partial exhaustive search and the proposed algorithm becomes larger when the transmit power increases. This figure also shows the comparison between the conventional convex optimizations and the DDPG based optimization. In DDPG based optimization, the fixed channel environment is used to train and the experience that has the maximum reward from all steps of all episodes is selected, instead of using the average reward (note that the energy efficiency is set as the reward). Even so, the result shows the proposed algorithm is better than the DDPG based algorithm, and the performance gap becomes larger with the increase of the BS’s transmit power. However, DDPG demonstrates its superiority when CSI is imperfect, which is shown later in Fig. 8. Furthermore, Fig. 4 shows under the same algorithm optimisation and network framework, the energy efficiency of the D2D pair in NOMA system outperforms in OMA system significantly. This benefits from the characteristic of the NOMA system that allows all communication resources to be shared.

Fig. 5 shows the energy efficiency versus the BS’s transmit power under the different numbers of antennas deployed for both NOMA and OMA schemes. In this figure, the parameters are set as follows: $K = 4$, $\eta = 0.1$ and $R_{\text{min}} = 0.1$ bps/Hz. It can be observed that the NOMA scheme shows superior performance in this network no matter how many antennas are equipped. Benefit from the spatial diversity, deploying more antennas results in higher D2D pair energy efficiency. Furthermore, by increasing the same number of antennas, the improvement of NOMA is more significant, compared to the OMA scheme. Additional, the enhancement of energy efficiency from 10 antennas to 50 antennas is more significant than from 50 antennas to 100 antennas. On the other hand, it can be observed that if massive antennas are deployed, the rate of energy efficiency increase becomes flat in the high transmit power range for both NOMA and OMA. The reason is that although more antennas bring better spatial diversity gain, each channel introduces fading as well. Hence, the trade-off between the cost and the performance improvement is crucial and needs to be considered when designing the system.

Fig. 6 shows the energy efficiency of the D2D pair versus the BS’s transmit power under different $\eta$ assumptions, where $K = 3$, $M = 10$ and $R_{\text{min}} = 0.1$ bps/Hz. Consistent with intuition, the higher RF energy conversion efficiency the hardware can provide, the better energy efficiency can be obtained.

Fig. 7 illustrates the relationship between the D2D pair’s energy efficiency and the downlink users’ QoS requirement $R_{\text{min}}$. In this experiment, the number of downlink users and
the transmit power of the BS are set to $K = 4$ and $P_{\text{max}} = 20$ dBm respectively. It can be observed that the energy efficiency is slightly decreasing with the increased $R_{\text{min}}$. This is because, in the considered NOMA system, the signal transmitted by $D_t$ will interfere with the reception of downlink users’ signals. Therefore, the transmit power of $D_t$ needs to be restricted more tight if the downlink users’ QoS requirement $R_{\text{min}}$ is higher.

Fig. 8 illustrates the different performances that can provide by the proposed algorithm and DDPG when the channel estimation accuracy is variant. In these simulations, we set $K = 4$, $M = 10$, $\eta = 0.1$ and $R_{\text{min}} = 0.1$ bps/Hz. Channels (i.e., $h_k$ and $h_{dd}$) and channel estimation errors (i.e., $e_k$ and $e_{dd}$) are used for both the proposed algorithm and DDPG. It can be seen that if channel estimation is perfect or only has slight errors (i.e., Fig. 8(a) and Fig. 8(b)), the proposed algorithm outperforms DDPG-based algorithm. In contrast, when channel estimation error is severe (i.e., Fig. 8(c) and Fig. 8(d)), the DDPG-based algorithm can demonstrate its superiority to mitigate the error caused performance degradation. On the other hand, unlike the scenario where channel estimation is perfect, the energy efficiency of the D2D pair is first slightly increasing and then decreases with the increase of BS’s transmit power when channel estimation error exists. This is because when the BS’s transmit power increases, the power of channel estimation error caused interference increases as well. Therefore, communication resources will be allocated to downlink users more to guarantee their QoS, which hinders the improvement of the energy efficiency of the D2D pair.

VII. CONCLUSION

This paper investigated the energy efficiency maximization problem of a WPT supported D2D pair in a MISO-NOMA downlink system, where beamforming vectors of the BS and time switching coefficient of the D2D transmitter were alternatively optimized. To solve the non-convex problem with variables highly coupled fractional objective function and constraints, the Dinkelbach method and multi-dimension quadratic transform based alternating algorithm were proposed. In addition, to demonstrate the accuracy of the proposed alternating algorithm, the partial exhaustive search was provided and compared. Moreover, the DDPG algorithm was directly used to solve the original non-convex problem. Simulation results illustrated that the proposed alternating algorithm outperforms the partial exhaustive search. Although the proposed algorithm outperforms DDPG when CSI can be obtained perfectly, DDPG provided much better performance when channel estimation error is severe. On the other hand, comparing the use of NOMA with OMA, the former is a better scheme for this system. Our future work will consider the multi-pair D2D scenarios for both the NOMA downlink and uplink networks.

APPENDIX

Assume that the optimal beam vector of the $k$-th user is $\omega_k^*$ and the total transmit power satisfies

$$\sum_{k=1}^{K} |\omega_k^*|^2 \leq P_{\text{max}}.$$  

(56)

Hence, the harvested power can be denoted by $P_r = \sum_{k=1}^{K} |h_{D_t}\omega_k^*|^2$. With all other parameters are fixed, we give an offset to each beam vector $\omega_k = \omega_k^* + \Delta_k$ such that $\sum_{k=1}^{K} |\omega_k|^2 = P_{\text{max}}$. Therefore, the received signal at the D2D transmitter is

$$y_{D_t} = h_{D_t}^H \sum_{k=1}^{K} \omega_k^* s_k^{(2)} + h_{D_t}^H \sum_{k=1}^{K} \Delta_k s_k^{(2)} + n_{D_t}.$$  

(57)

Therefore, the energy that can be received by the D2D transmitter is given by

$$P_r = \sum_{k=1}^{K} |h_{D_t}\omega_k^*|^2 + \sum_{k=1}^{K} |h_{D_t}\Delta_k|^2.$$  

(58)

Similarly, the power of interference from the BS to the D2D receiver is given by

$$P_i = \sum_{k=1}^{K} |h_{D_t}\omega_k^*|^2 + \sum_{k=1}^{K} |h_{D_t}\Delta_k|^2.$$  

(59)

Due to $P_r(\omega_k)$ and $P_i(\omega_k)$ to be positive real number, there exist a positive parameter such that $P_i = \xi P_r$ holds. Therefore, the objective function (14a) can be represented as

$$R_d(P_r) = \frac{\log(1 + \frac{\tau\eta P_r |h_{dd}|^2}{\bar{\tau} P_r})}{\tau\eta P_r + (\tau + 1)P_c}.$$  

(60)

The first derivative of $R_d$ and $E_c$ with respect to $P_r$ are

$$R'_d = \frac{\frac{\tau\eta}{\tau\eta} \frac{\xi P_r^2 + \sigma^2 P_r}{\sigma^2} \ln 2 \frac{\sigma^2}{|h_{dd}|^2 \sigma^2} + \ln 2 \frac{\sigma^2}{|h_{dd}|^2 \sigma^2}}{\sigma^2},$$  

$$E'_c = \frac{\tau\eta}{\tau\eta}.$$  

(61)

(62)

Denote $a = \frac{\ln 2 (\xi P_r + \sigma^2)}{|h_{dd}|^2 \sigma^2}$ and $b = \frac{\ln 2 (\xi P_r + \sigma^2)}{|h_{dd}|^2 \sigma^2}$ (Note that $a$ and $b$ are both positive), and hence we have

$$R'_d - E'_c = -a \frac{\tau\eta}{a\tau\eta + b}.$$  

(63)
is positive if and only if $\tau_\eta \in \left(0, \frac{1}{1-P}\right)$, which yields the objective function is monotonically decreasing and hence the maximum value is achieved when $P_r = 0$ (i.e., D2D transmitter does not harvest power), which is infeasible. Under this conclusion and the fact that $P_r > P_r^*$, $\tau_\eta(K_{F_1}) > \tau_\eta(K_{F_2})$ which contradicts the assumption that the solution (56) is optimal. Therefore, the optimal beam vectors should satisfy $\sum_{k=1}^K |\omega_k^*|^2 = P_{\text{max}}$. The lemma is proved.

REFERENCES

[1] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, “6G wireless networks: Vision, requirements, architecture, and key technologies,” IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 28–41, 2019.

[2] M. F. Hashim and N. I. Abdul Razak, “Ultra-dense networks: Integration with device to device (D2D) communication,” Wireless Personal Communications, vol. 106, no. 2, pp. 911–925, 2019.

[3] S. Gomalla, A. I. Azmi, M. R. Salim, M. F. Abdallah, F. Iqbal, W. A. Mabrouk, M. Othman, A. Y. Ashyap, and A. S. M. Supa’at, “Survey on Device to Device (D2D) Communication for 5G/6G Networks: Concept, Applications, Challenges, and Future Directions,” IEEE Access, 2022.

[4] T. Q. Duong, X. Chu, and H. A. Suraweera, Ultra-dense Networks for 5G and beyond: Modelling, Analysis, and Applications. John Wiley & Sons, 2019.

[5] S. Zhang, J. Liu, H. Guo, M. Qi, and N. Kato, “Envisioning device-to-device communications in 6G,” IEEE Network, vol. 34, no. 3, pp. 86–91, 2020.

[6] S. Jayakumar et al., “A review on resource allocation techniques in D2D communication for 5G and B5G technology,” Peer-to-Peer Networking and Applications, vol. 14, no. 1, pp. 243–269, 2021.

[7] H. ElSawy, E. Hessain, and M.-S. Alouini, “Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks,” IEEE Transactions on Communications, vol. 62, no. 11, pp. 4147–4161, 2014.

[8] Y. Pei and Y.-C. Liang, “Resource allocation for device-to-device communications overlaying two-way cellular networks,” IEEE Transactions on Wireless Communications, vol. 12, no. 7, pp. 3611–3621, 2013.

[9] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, “An overview of 3GPP device-to-device proximity services,” IEEE Communications Magazine, vol. 52, no. 4, pp. 40–48, 2014.

[10] A. Asadi, Q. Wang, and V. Mancuso, “A survey on device-to-device communication in cellular networks,” IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 1801–1819, 2014.

[11] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, and M. A. Javed, “A survey on NOMA in 6G networks: Future vision and research opportunities for next generation multiple access,” arXiv preprint arXiv:2103.02334, 2021.

[12] Y. Yuan, S. Wang, Y. Wu, H. V. Poor, Z. Ding, X. You, and L. Hanzo, “NOMA for next-generation massive IoT: Performance potential and technology directions,” IEEE Communications Magazine, vol. 59, no. 7, pp. 115–121, 2021.

[13] W. U. Khan, J. Liu, F. Jameel, V. Sharma, R. Jain, and Z. Han, “Spectral efficiency optimization for next generation NOMA-enabled IoT networks,” IEEE Transactions on Vehicular Technology, vol. 69, no. 12, pp. 15 284–15 297, 2020.

[14] K. Saito, A. Benjebbour, Y. Kishiyama, Y. Okumura, and T. Nakamura, “Performance and design of SIC receiver for downlink NOMA with open-loop SU-MIMO,” in 2015 IEEE International Conference on Communication Workshop (ICCW). IEEE, 2015, pp. 1161–1165.

[15] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, M. F. Imani, and V. C. El Dor, “Near-Field Wireless Power Transfer for 6G Internet of Everything Mobile Networks: Opportunities and Challenges,” IEEE Communications Magazine, vol. 56, no. 3, pp. 12–18, 2019.

[16] F. Tariq, M. R. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, “A speculative study on 6G,” IEEE Wireless Communications, vol. 27, no. 4, pp. 118–125, 2020.

[17] R. Zhang and C. K. Ho, “MIMO broadcasting for simultaneous wireless information and power transfer,” IEEE Transactions on Wireless Communications, vol. 12, no. 5, pp. 1989–2001, 2013.

[18] K. W. Choi, L. Ginting, A. A. Aziz, D. Setiawan, J. H. Park, S. I. Hwang, D. S. Kang, M. Y. Chung, and D. J. Kim, “Toward realization of long-range wireless-powered sensor networks,” IEEE Wireless Communications, vol. 26, no. 4, pp. 184–192, 2019.

[19] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, “Wireless networks with RF energy harvesting: A contemporary survey,” IEEE Communications Surveys & Tutorials, vol. 17, no. 2, pp. 757–789, 2014.

[20] Y. Cao and J. A. A. Qaboob, “Analysis and evaluation of input power splitting method between multiple transmitters for maximum wireless power transfer,” in 2018 IEEE Applied Power Electronics Conference and Exposition (APEEC). IEEE, 2018, pp. 959–962.

[21] Y. Ye, Y. Li, D. Wang, and G. Lu, “Power splitting protocol design for the cooperative NOMA with SWIPT,” in 2017 IEEE International Conference on Communications (ICC). IEEE, 2017, pp. 1–5.

[22] Z. Ding, “Harvesting devices’ heterogeneous energy profiles and QoS requirements in IoT: WPT-NOMA vs BAC-NOMA,” IEEE Transactions on Communications, vol. 69, no. 5, pp. 2837–2850, 2021.

[23] M. B. Goktas and Z. Ding, “A Wireless Power Transfer Assisted NOMA Transmission Scheme for 5G and Beyond mMTC,” IEEE Wireless Communications Letters, 2022.

[24] J. Tang, J. Lao, M. Liu, D. K. So, E. Alussa, G. Chen, K.-K. Wong, and J. A. Chambers, “Energy efficiency optimization for NOMA with SWIPT,” IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 3, pp. 452–466, 2019.

[25] S. Yu, W. U. Khan, X. Zhang, and J. Liu, “Optimal power allocation for NOMA-enabled wireless communication with imperfect SIC decoding,” Physical Communication, vol. 46, p. 101296, 2021.

[26] Y. Pan, C. Pan, Z. Yang, and M. Chen, “Resource allocation for D2D communications underlaying a NOMA-based cellular network,” IEEE Wireless Communications Letters, vol. 7, no. 1, pp. 130–133, 2018.

[27] A. Khazali, D. Tarchi, M. G. Shayesteh, H. Kalbkhani, and A. Bozorgchian, “Energy Efficient Uplink Transmission in Cooperative mmWave NOMA Networks with Wireless Power Transfer,” IEEE Transactions on Vehicular Technology, 2021.

[28] J. Xue, S. Wu, Z. Wang, H. Wang, and Q. Hu, “Research on energy transmission strategy based on MEC in green communication,” Multimedia Tools and Applications, pp. 1–21, 2022.

[29] N. Zhao, D. B. da Costa, J. Tang, X. Wang, and J. A. Chambers, “Special Issue on Unmanned Aerial Vehicle (UAV)-Enabled Green Communications and Networking,” IEEE Transactions on Green Communications and Networking, vol. 5, no. 3, pp. 1232–1235, 2021.

[30] H. Allgoewer, “Wireless power transfer (WPT) fundamentals with resonant frequency-dependent parameters, energy transfer efficiency, and green technology applications,” in 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC). IEEE, 2021, pp. 0036–0040.

[31] L. Pei, Z. Yang, C. Pan, W. Huang, M. Chen, M. Elkashlan, and A. Nallanathan, “Energy-efficient D2D communications underlaying NOMA-based networks with energy harvesting,” IEEE Communications Letters, vol. 22, no. 5, pp. 914–917, 2018.

[32] X. Chen, X. Wang, and X. Chen, “Energy-efficient optimization for wireless information and power transfer in large-scale MIMO systems employing energy beamforming,” IEEE Wireless Communications Letters, vol. 2, no. 6, pp. 667–670, 2013.

[33] W. Dinkelbach, “On nonlinear fractional programming,” Management Science, vol. 13, no. 7, pp. 492–498, 1967.

[34] K. Shen and W. Yu, “Fractional programming for communication systems—Part II: Power control and beamforming,” IEEE Transactions on Signal Processing, vol. 66, no. 10, pp. 2616–2630, 2018.
[39] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[40] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[41] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning.” *Computer ence*, 2015.

[42] N. Shaik and P. K. Malik, “A comprehensive survey 5G wireless communication systems: open issues, research challenges, channel estimation, multi carrier modulation and 5G applications,” *Multimedia Tools and Applications*, vol. 80, no. 19, pp. 28 789–28 827, 2021.

[43] T. Yoo and A. Goldsmith, “Capacity and power allocation for fading MIMO channels with channel estimation error,” *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, 2006.

[44] M. R. Zamani, M. Eslami, M. Khorramizadeh, and Z. Ding, “Energy-efficient power allocation for NOMA with imperfect CSI,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 1009–1013, 2018.