IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 18, 2021, accepted February 7, 2021, date of publication February 11, 2021, date of current version February 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3058677

Joint Resource Allocation and Mode Selection for Device-to-Device Communication Underlying Cellular Networks

YONGWEN DU, WENXIAN ZHANG, SHAN WANG, JINZONG XIA, AND HYTHEIM ALHAG MOHAMMAD

School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

Corresponding author: Yongwen Du (duyongwen@mail.lzjtu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61162066 and Grant 61163010, and in part by the Innovation Foundation of Colleges and Universities in Gansu Province under Grant 2020A-033.

ABSTRACT Device to device (D2D) communication has recently been established in the literature as an effective means to increase the frequency spectrum and enhance the efficiency of energy consumption in future cellular systems. However, certain issues, resulting from reusing resources in the same cell, have caused serious perturbations. We study issues pertaining to D2D communication, such as dual mode selection, channel allocation and power control, aiming at the maximization of the overall throughput of the system, while at the same time ensuring that the generated interference is kept minimized. This is an NP-Hard problem that decomposes the optimization problem into two layers: the inner layer, where the DQN algorithm is used as an indicator of the optimal transmission power that should be allocated to the D2D pairs in accordance with their mode of operation, and the outer layer, where strategic decisions, such as which communication mode to use and how to allocate the channels, are made. We have proved the superiority of the proposed scheme, in terms of both system throughput and performance, through simulating experiments involving different scenarios.

INDEX TERMS Device to device, mode selection, channel allocation, power control, throughput, NP hard.

I. INTRODUCTION

The architecture of the fifth-generation networks (5G) is expected to comprise heterogeneous networks that can be integrated with multiple advanced communication technologies to satisfy the quest for high-performance [1]. As of 2017, there were 8.4 billion connected devices across the world. It has been predicted that this number will surpass 75.4 billion by 2025 [2]. The growth rate is tremendous and would be increasing in the next decade, and thus bring Device-to-Device (D2D) resource allocation to be the hottest topics. [3]. It is expected that the reuse of identical wireless radio resources among users of the same cell (CU) will result in increasing the efficiency of the used frequency and saving large amounts of wireless bandwidth. However, with evitable mutual interference between the cell users and the D2D pairs, arising out this usage, reusing resources is a little more critical [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Dongxiao Yu.

A. RELATED WORKS

Recently, the question of how to allocate the resources of D2D networks has drawn the attention of researchers all over the world and many research works, such as literature works [4]–[9], [11], have been carried out (for the interested reader, a thorough coverage of the topic can be found in [4]). The work in [5] concentrated mostly on optimizing the throughput of D2D networks while ensuring that the cell priority registration constraints are satisfied. However, the dense nature of D2D networks renders this work impractical. An efficient practical resource allocation scheme, focusing mostly on the disturbances caused by resource reuse in D2D enabled networks, was presented in [6]. Resource allocation methods proposed in [5] and [6] are completely base-station controller. Due to the dense nature of D2D networks, this centralized control might lead to significant overhead [13]. Indeed, device-centralized architecture, in which a user's device can use its local information to guide his actions, are more suitable for dense D2D networks and facilitate distribution of control in the network [13].

A resource distribution scheme, that can be used for resource allocation, mainly by allowing ad-hoc D2D networks to organize themselves during the uplink transmission of the cell system, was introduced in [7]. Despite the relative improvement in the throughput of the system, the proper functioning of this method requires significant message passing. Literature work [4] has investigated the feasibility of using power control and model reuse selection criteria jointly and has demonstrated that this method has great potential for improving the performance of D2D-based systems. The method proposed in [8] used fractional programming to improve the usage of resources as well as energy consumption in a D2D network within a tractable iterative solution strategy framework. Moreover, a thorough survey on how to apply different game theory models to the D2D resource allocation problem was presented in [9]. Matching theory concepts were also thought of in [17], as a possible solution to the resource allocation problem and applied through a novel method that can be used to allocate power and channels accordingly, in addition to improving the overall cellular network throughput in D2D enabled systems. However, the works in [4], [7], [8], and [9] do not account for the presence of multiple D2D pairs on the same resource block, which can improve the overall system resource utilization, particularly in dense networks. In [11], author propose a polynomial time proportionally fair resource allocation scheme for D2D users that respect the rate requirements of the CUs.The proposed scheme can potentially work with any resource allocation scheme for CUs and can be adapted to the time and location varying channel conditions.

Due to the critical topic of mode selection in D2D communications, many research work on how to select the optimal mode of operation has been carried out. A mode selection method, focusing mostly on the preservation of the quality of both the communication between the D2D pairs and the cellular links, was proposed in [14] to lessen the interference caused by D2D communications. Mode selection in [15] considered network information such as link gain, noise level and signal-to-noise-and-interference ratio (SINR). In [5], an opportunistic algorithm that can be used for mode selection and sub-channel scheduling has been developed and applied in Orthogonal Frequency Division Multiple Access (OFDMA) based D2D systems.

Regarding channel allocation, in [10], by exploiting device-to-device (D2D) communication for enabling user collaboration and reducing the edge server's load. Author propose a one-to-one matching algorithm based on the Pareto improvement and swapping operations and extend the one-to-one matching algorithm to a many-to-one matching scenario. [12] study the downlink channel allocation in D2D-assisted small cell networks with heterogeneous spectrum bands. To derive the solution, the author decomposes the optimization problem into two games: a potential game and a coalition game. Then, a potential game-based scheme using an interference graph and a coalition scheme with D2D user transferring is proposed to solve these two games, respectively.

Regarding power distribution, in recent years, the machine learning (ML)-based approaches have been rapidly developed in power control [16]. These algorithms are usually model-free, and are compliant with optimizations in practical communication scenarios. Additionally, with developments of graphic processing unit (GPU) or specialized chips, the executions can be both fast and energy-efficient, which brings in solid foundations for massive applications. [17], author study resource allocation algorithm design is formulated as a non-convex optimization problem which jointly designs the power allocation, rate allocation, user scheduling, and successive interference cancellation (SIC) decoding policy for minimizing the total transmit power. To strike a balance between system performance and computational complexity, the author propose a suboptimal iterative resource allocation algorithm based on difference of convex programming. Joint mode selection and power control methods have also been investigated in [18], [19] to further improve the performance. The works in [20] has studied how to optimally allocate resources and control power for a single D2D pair and a single CU. [15] develops a kind of resource sharing algorithm based on the interference-aware algorithm. It is almost the most optimal, but the computational complexity is high. Interference mitigation can be achieved through reusing CU resources properly or exploiting the multiuser diversity inherent in cellular networks.

B. CONTRIBUTIONS

In this paper, we consider a scenario in which block level resource allocation for CUs has previously been done at the base station (BS), at each of its subframes. Because of massive downlink traffic in frequency-division duplex (FDD) based cellular systems, we will be focus only on the uplink resource blocks. We keep the QoS of each subframe for ensuring the minimum rate allocated to every single CU, and satisfying modest functional demand. If the received signal-to-noise ratio (SINR) at the BS is greater than the SINR needed by the CU to guarantee its proper functioning, the SINR gap can be utilized to allocate power to D2D users.

We consider joint mode selection, power control, and channel allocation to maximize the throughput of the entire system. It is proved that the proposed scheme blended of three patterns can reach a higher throughput with the increasement of users from a lot of experiments and simulations. Other evaluation indexes (distance between users and number of resource blocks) are used to illustrate the advantage of the proposed algorithm in improving system throughput compared with the algorithm considering only one or two modes. Finally, the validity of our algorithm is proved by using the Fairness index of Jain's Fairness index. The main contributions of this paper are as follows:

1) First of all, the selection of D2D model is done according to quality of link in order to alleviate the consequences of choosing unreliable D2D links; the method we propose here is applicable to the transmission mode selection, resource block allocation and power control

problems, as for guaranteeing that the delay and reliability requirements of D2D are met, and is different from the one proposed in [5], as which considers different transmission modes to approach the resource sharing among D2D pairs problem.

2) Furthermore, the proposed scheme uses three algorithms complementary to obtain a higher system throughput; a lightweight heuristic algorithm is to implement pattern selection among the proposal, and a Hungarian algorithm for optimal solutions and physical resource block (PRB) allocation problems; the last one is a power distribution algorithm which is based on DQN. It takes advantages of local observations, such as interference levels, large-scale channel quality and traffic loads, to make decisions accordingly.
3) Last of all, we have also carried out an analysis on the number of D2D pairs involved in communication, the maximum communication distance and outage threshold to assess their impact on the performance of system. The simulation results demonstrate that the proposed scheme is superior to contemporary algorithms.

The rest of this paper is arranged as follows. Section 2 is system model. Section 3 describes the optimization problem. Section 4 presents the three algorithms of this paper. Section 5 gives simulation results. Section 6 is the conclusion.

II. MODEL

A typical *LTE* cell has N_c active CU and N_d D2D pairs. The allocation of resources in the system is achieved by BS.

According to the LTE standard, we divide the time into subframes of $1ms$ duration and made allocation decisions in each subframe. Each physical resource block (PRB) has a frequency width of $180KHz$ and a duration of $0.5ms$. The two PRBs unite to form a resource block, the smallest resource unit that BS can assign to a user. We presume that there are a total of M uplink resource blocks available for CU. We designed a wireless backflow model for all users, where in each subframe, the BS can allocate all resource blocks to CU and D2D. A CU can get multiple resource blocks N in each subframe, but each resource block can be allocated to at most one CU. Given the resource block allocation of CU in each subframe, we believe that CU has the minimum rate requirement. Once a suitable channel is available, CU can share its resource block with D2D pair, while still satisfying their rate constraints.

The interference scenario is depicted in Figure 1, where CU c and D2D pair d share the same uplink resource blocks. Let dT and dR denote the transmitter and receiver of D2D pair d. The parameters corresponding to four possible link types, namely from CU c to the BS, from the transmitter to the receiver of D2D pair d, from the transmitter of D2D pair d to the BS, and CU c to the receiver of D2D pair d are differentiated through subscripts cB,$dTdR$,dTB and cdR respectively.

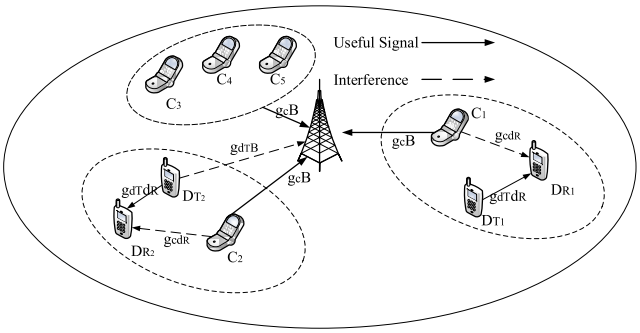

FIGURE 1. System scenario of the device-to-device underlaying communication.

A. CHANNEL MODEL

As showed in figure1,we considering a K-user downlink MIMO system, where the BS is equipped with $N_t \geq K$ antennas, and the users have a single antenna [25]. Denote the downlink channel for user k as h_k^H, where $h_k \in \mathbb{C}^{(N_t)}$ is a column vector and follows distribution $\mathcal{CN}(0,R_k)$, the $\mathbb{C}^{(N_t)}$ denote a complex number set consisting of N_t complex numbers. The channels between users are mutually independent. User k knows h_k perfectly, and the global statistics R_k is known by all the users.

Let $H = [h_1, h_2, \ldots, h_K] \in \mathbb{C}^{(N_t)\times K}$ be the channel matrix for all the users and $W \in \mathbb{C}^{(N_t)\times K}$ denote the recorder for the downlink transmission. The received signal $y = [y_1, y_2, \ldots, y_K]^T$ at the user side is:

$$y = H^H W x + n \tag{1}$$

where $x \in \mathbb{C}^k$ is the vector of transmission symbols that satisfies $\mathbb{E}\{xx^H\} = I_K$, the precoder W satisfies the sum power constraint $tr\{W^H W\} \leq P$, and $n \sim \mathcal{CN}(0, I_K)$ is the Gaussian noise.

In general, the desired precoder W is a function of the CSI(channel state information) H, which is initially available at the user side. To assist the precoding, each user can feedback B bits of CSI related information to the BS.

B. CSI EXCHANGE VIA D2D

In parallel to cellular communications, users exploit reliable D2D links to exchange the CSI directly. For example, D2D communication can be implemented in out-band mode with no interference to cellular communication, using existing technologies such as WiFi Direct, Bluetooth, and ZigBee.

Denote the CSI h_k of user k known by user j as $\hat{h}_k^{(j)}$, which is modeled as follows:

$$h_k = a_{jk}\hat{h}_k^{(j)} + \sqrt{1 - a_{jk}^2}\xi_k^{(j)} + h_k^{(j)\perp} \tag{2}$$

where $\alpha_{jk} \in [0, 1]$ is a parameter to capture the quality of the CSI obtained via D2D, $\xi_k^{(j)}$ is a zero mean random vector with distribution $\mathcal{CN}(0, \Xi_{kj})$ to model the noise due to quantization or transmission delays, and $h_k^{(j)\perp}$ is orthogonal to both $\hat{h}_k^{(j)}$ and $\xi_k^{(j)}$ to model the portion of CSI h_k that is not to be transmitted to user j. In particular, $\alpha_{jk} = 0$ means there is no D2D from

TABLE 1. List of notations.

Notations	Definitions	Notations	Definitions
N_c	numbers of CU	a_{jk}	parameter to capture the quality of the CSI obtained via D2D
N_d	numbers of D2D pairs	$p_k^{(1)}$	transmitting power K of D2D pair in cellular mode
dT	transmitter of D2D pair d	σ_N^2	power of the additive white gaussian noise
dR	receiver of D2D pair d	γ_0	terminal threshold
h_k^H	downlink channel for user k	P_0	allowable terminal probability
H	channel matrix for all the users	$p_k^{(2)}$	transmitting power of D2D pair k in dedicated mode
W	recorder for the downlink transmission	$p_{k,m}^{(3)}$	respectively the transmitting power of D2D pair k
y	received signal	$p_{k,m}^{(c)}$	CUm when D2D pair reuses CU
h^t	channel state information	C^t	downlink rate

user k to user j, and hence user j has no knowledge of h_k, whereas, $\alpha_{jk} = 1$ means there is perfect D2D, and user j knows perfectly $h_k - h_k^{(j)\perp}$.

After the exchange of CSI, user k has the imperfect global CSI $\hat{H}_k \in \mathbb{C}^{N_t \times K}$ given by:

$$\hat{H}_k = [\hat{h}_1^{(k)}, \hat{h}_2^{(k)}, \ldots, \hat{h}_{k-1}^{(k)}, \hat{h}_k, \hat{h}_{k+1}^{(k)}, \ldots, \hat{h}_K^{(k)}] \tag{3}$$

Consider to exchange the CSI via D2D using a limited amount of bits and the transmission delays are negligible. As proposed in [21], an efficient mechanism in correlated channels is based on signal subspace projection. Conceptually, if the channel subspaces of user k and j are partially overlapping, then only the portion of CSI that lies in the overlapping signal subspace is needed to be exchanged. The intuition is that if the two users have non-overlapping channel subspaces, they do not need to exchange the CSI because their preferable precoding vectors would not create interference to each other.

Therefore, let $R_k = V_k \Lambda_k V_k^H$ be the eigendecomposition, where Λ_k is an $M_k \times M_k$ diagonal matrix containing the nonzero eigenvalues of R_k sorted in descending order (with M_k being the rank), and V_k is an $N_t \times M_k$ semi-unitary matrix. The channel of user k can be written as $h_k = h_k^{(j)} + h_k^{(j)\perp}$ where $h_k^{(j)} = V_j V_j^H h_k$ and $h_k^{(j)\perp} = (I - V_j V_j^H) h_k$. As a result, $h_k^{(j)}$ contains all the necessary information for user j and is orthogonal to $h_k^{(j)\perp}$. Consider to quantize $h_k^{(j)}$ into $\hat{h}_k^{(j)}$ using B_d bits; the quantization error can be modeled by (2), where the parameters α_{kj} and Ξ can be computed using distortion-rate theories [21], [22].

Now, for the interference scenario depicted in Figure 1, the BS is exposed to interference from D2D transmitter dT, and CU c causes interference to D2D receiver dR. Thus for a subframe n, when resource block k of CU c is reused by D2D pair d, the received SINR of CU c at the BS is given by:

$$SINR_{[n]} = \frac{P_c H_{cB}^k[n]}{\sigma_N^2 + P_{dT}^k[n] H_{dTB}^k[n]} \tag{4}$$

III. D2D COMMUNICATION MODEL

A. CELLAR MODE

In the cellular mode, two D2D user(DU) as a conventional CU communicates through eNB(Base Station), and no D2D link is established. Intuitively, this mode is preferred if two users are too far away from each other, in which case two channels (one downlink and one uplink) will be assigned to the D2D pair. Although the performance of a D2D pair working in cellular mode is the same as that of a regular cellular user, we assume that any other D2D pair will not reuse its channel.

The uplink signal-to-noise ratio (SINR) K of D2D pair can be expressed as:

$$\xi_{k,up}^{(1)} = \frac{p_k^{(1)} h_{k,B}^D}{\sigma_N^2} \tag{5}$$

$p_k^{(1)}$represents the transmitting power K of D2D pair in cellular mode, and σ_N^2 is the power of the additive white Gaussian noise (AWGN).

Similarly, the downlink SINR can be expressed as:

$$R_i^{D_{(0)}} = log(1 + \xi_{k,up}^{(1)}) \tag{6}$$

Similar to [20], we use the probability of interruption as a reliability metric. Under terminal threshold γ_0 and allowable terminal probability P_0, the reliability requirement of D2D pair($k \in K$) is expressed as:

$$P\{r_i^{D(0)} \leq \gamma_0\} \leq P_0 \tag{7}$$

According to [20], the reliability constraint (7) can be transformed into (8) under Rayleigh fading:

$$r_i^{D(0)} \leq reff = \frac{\gamma_0}{\ln(\frac{1}{1-p_0})} \tag{8}$$

reff is the effective shutdown threshold.

Assuming that packet size, maximum tolerable delay and the probability of tolerable interrupt are the same for all D2D pairs, and uplink SINR should be greater than the given threshold ξ_{min} in order to ensure QoS of DU.

$$\min\{\xi_{k,up}^{(1)}\} \geq \xi_{\min} \tag{9}$$

B. DEDICATED MODE

A dedicated mode is considered when two users are nearby, and a CU has an empty channel that is not currently in use. In the private mode, SINR of D2D pair can be expressed as:

$$R_i^{D(1)} = \xi_k^{(2)} = \frac{p_k^{(2)} h_k^D}{\sigma_N^2} \tag{10}$$

$p_k^{(2)}$ is the transmitting power of D2D pair k in dedicated mode.

C. REUSE MODE

In this mode, the two DUs communicate directly by reusing the existing CU channel, which further improves the spectral efficiency. However, this will cause interference between D2D and its channel CU. In reuse mode, when the uplink channel of CUm is reused, the SINR of D2D pair k can be expressed as:

$$\xi_{k,m}^{(3)} = \frac{p_{k,m}^{(3)} h_k^D}{p_{k,m}^c h_{k,m} + \sigma_N^2} \tag{11}$$

Where $p_{k,m}^{(3)}$ and $p_{k,m}^c$ are respectively the transmitting power of D2D pair k and CUm when D2D pair reuses CU. At the same time, channel reuse will also cause interference to the same channel CU, and the SINR of CUm with the interference of D2D pair k can be expressed as:

$$\xi_{k,m}^c = \frac{p_{k,m}^c h_{m,B}^C}{p_{k,m}^{(3)} h_{k,B}^D + \sigma_N^2} \tag{12}$$

$$R_i^{D(2)} = \log_2(1 + \xi_{k,m}^c) \tag{13}$$

If the spectrum of the j_{th} CU is not reused by DU, then its rate $R_J^{C,u}$ is:

$$R_J^{C.u} = \log_2(1 + \frac{p_{k,m}^c h_{m,B}^C}{\sigma_N^2}) \tag{14}$$

And if the spectrum of the j th CU is reused by DU, the rate is:

$$R_J^{C,r} = \log_2(1 + \frac{p_{k,m}^c h_{m,B}^C}{p_{k,m}^{(3)} h_{k,B}^D + \sigma_N^2}) \tag{15}$$

Since CU has a higher priority, its QoS should be ensured, and DU is allowed to reuse channels only if $\xi_{k,m}^c \geq \xi_{min}$. In other words, only when the SINR requirements of both the D2D pair and the interfering CU are satisfied, can a D2D pair share the channel with the CU.

IV. D2D PROBLEM MODEL

We will maximize the throughput of the system with the joint mode selection, channel allocation and power control while ensuring the SINR of CU and DU. $x = x_{(1)}, x_{(2)}, x_{(3)}$ is expressed as the mode selection and channel allocation matrix, $x_{(1)}$ and $x_{(2)}$ are k-dimension indicator vectors of cellular mode and dedicated mode, where, if the D2D pair works in (dedicated) cellular mode, then $x_k^{(1)} = 1 (x_k^{(2)} = 1)$. $x_{(3)}$ is the $k * m$ channel allocation indicator matrix in the multiplexing mode. If D2D is reused for K, $x_{k,m}^{(3)} = 1$; otherwise, $x_{k,m}^{(3)} = 0$.

$P = P^{(1)}, P^{(2)}, P^{(3)}, P^c$ is a power matrix, $P^{(1)}$, $P^{(2)}$ and $P^{(3)}$ represent the transmitting power at the time of mode selection, and their values are the same as $x_{(1)}$, $x_{(2)}$ and $x_{(3)}$ respectively.

And then, the joint mode selection, channel allocation, and power control problems can be modeled as:

$$\begin{aligned}(p^*, x^*) = \arg\max_{p,x} \{ & \sum_{k=1}^{K} x_k^{(1)} \log(1 + \frac{p_k^{(1)} h_{k,B}^D}{\sigma_N^2}) \\ & + \sum_{k=1}^{K} x_k^{(2)} \log(1 + \frac{p_k^{(2)} h_k^D}{\sigma_N^2}) \\ & + \sum_{k=1}^{K} \sum_{m=1}^{M} x_{k,m}^{(3)} \log(1 + \frac{p_{k,m}^{(3)} h_k^D}{p_{k,m}^c h_{k,m} + \sigma_N^2}) \\ & + \sum_{k=1}^{K} \sum_{m=1}^{M} x_{k,m}^{(3)} \log(1 + \frac{p_{k,m}^{(c)} h_{k,B}^C}{p_{k,m}^{(3)} h_{k,B}^{(D)} + \sigma_N^2}) \\ & + \sum_{m=1}^{M} (1 - \sum_{k=1}^{K} x_{k,m}^{(3)}) \log(1 + \frac{p_m^c h_{m,B}^C}{\sigma_N^2}) \end{aligned} \tag{16}$$

$$x_k^{(1)}, x_k^{(2)}, x_k^{(3)} \in \{0, 1\}, \forall k, m, \tag{16a}$$

$$\sum_{k=1}^{K} x_k^{(1)} \leq \min\{N_U, N_D\}, \tag{16b}$$

$$2 \sum_{k=1}^{K} x_k^{(1)} + \sum_{k=1}^{K} x_k^{(2)} \leq N_U + N_D, \tag{16c}$$

$$x_k^{(1)} + x_k^{(2)} + \sum_{k=1}^{M} x_{k,m}^{(3)} \leq 1, \forall m, \tag{16d}$$

$$\sum_{k=1}^{K} x_{k,m}^{(3)} \leq 1, \forall m, \tag{16e}$$

$$x_k^{(1)} \frac{p_k^{(1)} h_{k,B}^D}{\sigma_N^2} + x_k^{(2)} \frac{p_k^{(2)} h_k^D}{\sigma_N^2}, \tag{16f}$$

$$(1 - \sum_{k=1}^{K} x_{k,m}^{(3)}) p_m^c + \sum_{k=1}^{K} x_{k,m}^{(3)} p_{k,m}^c \leq P_{\max}^C, \forall k, \tag{16g}$$

$$\begin{aligned} & x_k^{(1)} \frac{p_k^{(1)} h_{k,B}^D}{\sigma_N^2} + x_k^{(2)} \frac{p_k^{(2)} h_{k,B}^D}{\sigma_N^2} \\ & + \sum_{m=1}^{M} x_{k,m}^{(3)} \frac{p_{k,m}^{(3)} h_K^D}{p_{k,m}^{(3)} h_{k,B}^D + \sigma_N^2} \end{aligned} \tag{16h}$$

$$\begin{aligned} & \sum_{k=1}^{K} x_{k,m}^{(3)} \frac{p_{k,m}^c h_{m,B}^C}{p_{k,m}^{(3)} h_{k,B}^D + \sigma_N^2} \\ & + (1 - \sum_{k=1}^{K} x_{k,m}^{(3)}) \frac{p_m^c h_{m,B}^C}{\sigma_N^2} \geq \xi_{\min}, \forall m. \end{aligned} \tag{16i}$$

P_{max}^D and P_{max}^C respectively represent the maximum transmitting power of DUs and CUs, ξ_k^d, and ξ_m^c are the SINR of D2D pair and CU. The constraint (16b) is that the number of D2D pairs in cellular mode should be not greater than the number of unused uplink channels. (16c) means that the number of channels used for D2D communication in cellular and dedicated mode should not exceed the total number of

unused channels. (16d) means that any D2D pair will choose up to one of three patterns. (16e) indicates that a CU can only be reused by at most one D2D pair. (16f) Moreover, (16g) indicates that the transmitting power of DUs and CUs cannot exceed the maximum. Furthermore, (16h) and (16i) show that the SINR of DUs and CUs should have a minimum.

V. MODE SELECTION,Channel ALLOCATION AND POWER CONTROL ALGORITHM

In this section, we will address the optimization problem presented in (16). Suppose (16) has second order continuous partial derivative on region D, denoted as $A = f''_{p^*p^*}(p^*, x^*)$, $B = f''_{p^*x^*}(p^*, x^*)$,$C = f''_{x^*x^*}(p^*, x^*)$, and $A > 0$ is always on D, and $AC - B^2 \geq 0$ [23]. Then (16) contains p^* and x^* binary variables, so this problem is non-concave and cannot be solved directly. We overcome the problem down into three subproblems and solve them separately. The original optimization problem is rewritten as:

$$
\begin{aligned}
(p^*, x^*) &= \arg\max\{\sum_{k=1}^{K} x_k^{(1)} Q_1 + \sum_{k=1}^{K} x_k^{(2)} Q_2 \\
&\quad + \sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m}^{(3)} Q_3 + \sum_{m=1}^{M}(1 - \sum_{k=1}^{K} k_{k,m}^{(3)}) Q_4\}, \\
Q_1 &= \arg\max\{\log(1 + \frac{p_k^{(1)} h_{k,B}^D}{\sigma_N^2})\}, \\
Q_2 &= \arg\max\{\log(1 + \frac{p_k^{(2)} h_k^D}{\sigma_N^2})\}, \\
Q_3 &= \arg\max\{\log(1 + \frac{p_{k,m}^{(3)} h_k^D}{p_{k,m}^{(c)} h_{k,m} + \sigma_N^2}) \\
&\quad + \log(1 + \frac{p_{k,m}^{(c)} h_{m,B}^C}{p_{k,m}^{(3)} h_{k,B}^D + \sigma_N^2})\}, \\
Q_4 &= \arg\max\{\log(1 + \frac{p_m^c h_{m,B}^C}{\sigma_N^2})\}. \qquad (17)
\end{aligned}
$$

It can be seen from (17) that the optimization problem is composed of two layers. Internal control is the power control represented by Q_1, Q_2, Q_3, and Q_4 to determine the optimal transmission power of DUs and CUs in each mode. The other is the decision-making process of communication mode and channel allocation. So we optimize the inner layer and the outer layer to obtain the optimal solution.

A. ALGORITHM1 (PATTERN SELECTION ALGORITHM BASED ON HEURISTIC ALGORITHM)

To solve the problem of light load under the cellular mode and dedicated mode selection, $x_{k,m}^{(3)} = 0, \forall k, m$, (17) can be simplified to:

$$
x^* = \arg\max\{\sum_{k=1}^{K} x_k^{(1)} \gamma_k + \sum_{k=1}^{K} x_k^{(2)} \theta_k\}, \qquad (18)
$$

A sufficient number of empty channels can eliminate the constraints (16b) and (16c). The best mode selection for each D2D pair can be found by comparing the capacity of the cellular mode and the dedicated mode. In other words, if the capacity of the dedicated mode is greater than that of the cellular mode, the D2D pair will choose the dedicated mode and vice versa.

If there are insufficient empty channels, pattern selection has to be done jointly, making the problem even more complicated. Since the abundance of channels consumed by the cellular mode is twice that of the dedicated mode, we compare the capacity of the cellular mode with the capacity of the dedicated mode. Only when the capacity of the cellular mode is twice that of the dedicated mode, D2D will choose the cellular mode; otherwise, the dedicated mode will be selected, and the D2D pairs must use the dedicated mode when the uplink channel is not available.

The detailed algorithm is shown in Algorithm 1. Its main computational complexity is the sorting of k-dimension vector T, the complexity is $O(Klog_2^K)$.

Algorithm 1 Heuristic Algorithm for the Light Load Scenario

1: Calculate the capacity of each DU working in the cellular mode and the dedicated mode, as: $T^{(1)} = (T_1^{(1)}, \cdots, T_k^{(1)}, \cdots, T_K^{(1)})$, $T^{(2)} = (T_1^{(2)}, \cdots, T_k^{(2)}, \cdots, T_K^{(2)})$, where $T_k^{(1)} = \log(1 + \frac{p_{\max}^D h_{k,B}^D}{\sigma_N^2})$ and $T_k^{(2)} = \log(1 + \frac{p_{\max}^D h_k^D}{\sigma_N^2})$
2: Construct a vector T, whose element is the higher of $T^{(1)}$ and $2T^{(2)}$, $T = (\max\{T_1^{(1)}, 2T_1^{(2)}\}, \cdots, \max\{T_K^{(1)}, 2T_K^{(2)}\})$.
3: Initialize the allocated uplink channel counter $n_1 = 0$, the allocated downlink channel counter $n_2 = 0$.
4: **while** $n_1 < N_U$ and $n_2 < N_D$ **do**
5: select $k^* = arg\max T_k$.
6: **if** $T_{k^*} = T_{k^*}^{(1)}$ **then**
7: set $n_1 = n_1 + 1$, $n_2 = n_2 + 1$, and $x_{k^*}^{(1)} = 1$.
8: **else**
9: **if** $N_U - n_1 > N_D - n_2$ **then**
10: set $n_1 = n_1 + 1$ and $x_{k^*}^{(2)}$.
11: **else**
12: set $n_2 = n_2 + 1$ and $x_{k^*}^{(2)}$.
13: **end if**
14: **end if**
15: set $T_{k^*}^{(2)} = T_{k^*}^{(1)} = 0$
16: **end while**
17: **for** $i = 1$ to $maxN_U - n_1, N_D - n_2$ **do**
18: select $k^* = arg\max T_k^{(2)}$, and set $x_{k^*}^{(2)} = 1$
19: **end for**

B. ALGORITHM2 (CHANNEL ALLOCATION ALGORITHM BASED ON MATCHING THEORY)

We simplify by disregarding the cellular mode. There are for two reasons for this. First, for adjacent DU, the channel

gain between D2D pairs is usually more significant than that between DU and eNB so that D2D communications will improve system capacity and energy efficiency. Second, in a dedicated mode, each D2D pair occupies only one channel, which also improves bandwidth utilization.

Let $x_k^{(1)} = 0, \forall k$, then simplify the problem in (17) to:

$$x^* = \arg\max\{\sum_{k=1}^{K} x_k^{(2)}\theta_k + \sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m}^{(3)}\eta_{k,m} + \sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m}^{(3)}\lambda_{k,m} + \sum_{m=1}^{M}(1 - \sum_{k=1}^{K} x_{k,m}^{(3)})\tau_m\} \quad (19)$$

$$x_k^{(2)}, x_{k,m}^{(3)} \in \{0, 1\}, \forall k, m, \quad (19a)$$

$$\sum_{k=1}^{K} x_k^{(2)} \leq N_U + N_D, \quad (19b)$$

$$x_k^{(2)} + \sum_{m=1}^{M} x_{k,m}^{(3)} \leq 1, \forall k, \quad (19c)$$

$$\sum_{k=1}^{K} x_{k,m}^{(3)} \leq 1, \forall m. \quad (19d)$$

Furthermore, the optimization problem can be reformulated by inserting a constraint into the objective function.

$$x^* = \arg\max \sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m}^{(3)}(\eta_{k,m} + \lambda_{k,m} - \theta_k - \tau_m) \quad (20)$$

$$x_{k,m}^{(3)} \in \{0, 1\}, \forall k, m, \quad (20a)$$

$$\sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m}^{(3)} \geq K - N_U - N_D, \quad (20b)$$

$$\sum_{k=1}^{K} x_{k,m}^{(3)} \leq 1, \forall m, \quad (20c)$$

$$\sum_{m=1}^{M} x_{k,m}^{(3)} \leq 1, \forall k. \quad (20d)$$

We can easily infer $\eta_{k,m} + \lambda_{k,m} - \theta_k - \tau_m < 0, \forall k, m$, so the constraint (20b) can be forced to be equal to $\sum_{k=1}^{K}\sum_{m=1}^{M} x_{k,m}^{(3)} = K - N_U - N_D$.

We define $\rho_{k,m} = \eta_{k,m} + \lambda_{k,m} - \theta_k - \tau_m$, and the utility matrix is:

$$\Theta = \begin{bmatrix} \rho_{1,1} & \cdots & \rho_{1,m} & \cdots & \rho_{1,M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k,1} & \cdots & \rho_{k,m} & \cdots & \rho_{k,M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{K,1} & \cdots & \rho_{K,m} & \cdots & \rho_{K,M} \end{bmatrix} \quad (21)$$

The optimization problem is now required to select $K - N_U - N_D$ elements from the matrix in such a way that the resulting rows and columns are selected with at most one element. If the total number of elements is equivalent to K, i.e., $N_U = N_D = 0$, then the problem will be an allocation problem. Therefore, we propose an optimal solution to the assignment problem using the maximum weight binary matching algorithm with polynomial time complexity. To allocate at most one resource block to a D2D pair, let us take a look at a bipartite graph in which the vertex set U is a set of D2D pairs and the vertex set V is a set of resource blocks. The edge weight of resource block K in D2D pair and subframe N is determined by $\lambda_d^k[n] = r_d^k[n]/R_d[n-1]$.

To allocate a maximum of T resource blocks to D2D pairs, we create a new bipartite graph $G' = (U', V, E')$, where vertex set U' is a set of D2D pairs that repeat T times, and vertex set V is a set of resource blocks. Therefore, as showed in Figure 2, in the case of $T = 2$, the edge weight is $\lambda_d^k[n]$. Once the graph G' is formed, we can apply the maximum weight binary matching algorithm on G' to get the best resource allocation result. In algorithm 2, we explain the proposed resource allocation scheme for D2D users.

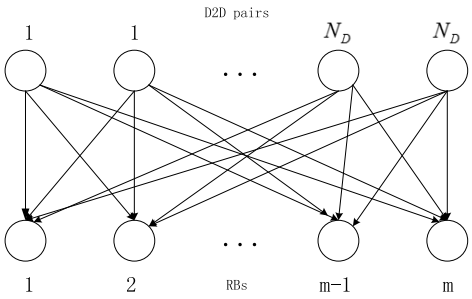

FIGURE 2. Bipartite graph for allocating multiple shared resource block (T=2).

Now we discuss the complexity of the proposed maximum weight binary matching algorithm. For the general case, when a D2D pair can share at most T resource blocks, the complexity of calculating edge weight is $O(N_D \times T_2)$. To determine the maximum weight bipartite matching algorithm of the bipartite graph, we use the original dual method [15] to realize the Blossom algorithm. The computational complexity of this algorithm is $O(n_3)$, where n is the total number of nodes in the graph given by $n = (T + 1)N_D$. Therefore, the overall complexity of this algorithm amounts to $O(N_D \times T_2 + n_3)$.

C. ALGORITHM3 (POWER CONTROL ALGORITHM BASED ON DQN)

The maximum throughput of a D2D user pair multiplexed over a cellular user channel can be obtained by mode selection and channel allocation. This section focuses on how to allocate the appropriate power for each D2D user pair. In many applications, like video games, the current strategy has a long-term impact on cumulative returns [24], from which the DQN algorithm can achieve significant results, and power control, however, the discount factor is set to 0. So the DQN algorithm is designed to maximize the Q function. That is:

$$\max Q = \max E_\pi[r^t | s^t = s, a^t = a]. \quad (22)$$

Algorithm 2 Heuristic Algorithm for the Medium Load Scenario

1: Utilize the Hungarian algorithm for K D2D pairs and M CUs with cost matrix Θ
2: Let $\omega = (\omega_1, \cdots, \omega_k, \cdots, \omega_K)(\omega_k \in \{1, 2, \cdots, M\}, \forall k)$ denote the resulting channel assignment vector by the Hungarian algorithm and $T = (T_1, \cdots, T_k, \cdots, T_K)$, where $T_k = \rho_{k,\omega_k}$
3: Denote π as the permutation on $1, 2, \cdots, K$ which rearranges T into a non-ascending order.
4: **for** $k = 1 : K - N_D - N_U$ **do**
5: set $x^{(3)}_{\pi k,\omega_{\pi k}} = 1$.
6: **end for**
7: **for** $k = K - N_D - N_U + 2 : K$ **do**
8: set $x^{(2)}_{\pi k} = 1$.
9: **end for**

For the power control, $s = h^t$, $a = p^t$. Then let $r^t = R^i$ to get:

$$\max Q = \max E_\pi[C^t | h^t, p^t]. \tag{23}$$

The policy is determined during the execution period, so (23) can be rewritten as:

$$\max Q = \max C^t(h^t, p^t), \tag{24}$$

As shown in Figure 3, the optimal solution p^{t^*} of (3) is determined only by the current CSI h^t, and (h^t, p^t) is used to calculate the rate R^t. The optimal power p^{t^*} can be obtained by taking h^t as the input of DQN theoretically, but the DQN of this design is non-convex, and it is difficult to find the best so that poor performance. Therefore, we proposed two auxiliary functions: C^{t-1} and P^{t-1}. Since the channel can be modeled as a first-order Markov process, the solution of the preceding period can help DQN approach the optimal. Rewrite (24) as:

$$\max Q = \max C^t(h^t, p^t, C^{t-1}, p^{t-1}), \tag{25}$$

Once $\gamma = 0$ and $\gamma_t = C_t$, the replay memory is also reduced to (s_t, a_t, r_t). As an estimator, DQN can predict the current consumption rate of response power level under certain CSI.

In our proposed model-free two-step training framework, DQN uses a deep reinforcement learning algorithm for offline pre-training for the first time in a simulated wireless communication system. Due to the high demand for a data-driven algorithm, this process narrows the pressure of online training. And then with the help of transfer learning, the trained DQN need to be further fine-tuned in the actual scene.

In a certain cellular network, each BS-user link is regarded as an agent and thus a multi-agent system is studied. However, for a large amount of learning data, training time, and appropriate DNN parameters are required for multi-agent training, leading to the problematic training consequence. Therefore, centralized training is consideration using the experience playback memory of all agents to train only one agent, and

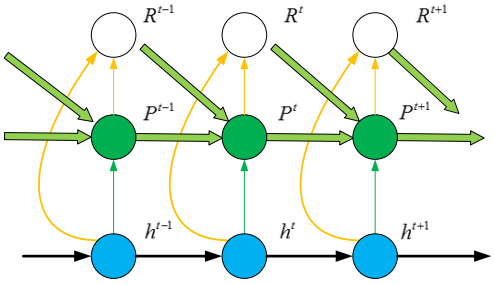

FIGURE 3. The solution of DQN is determined by CSI h^t, along with downlink rate C^{t-1} and transmitting power p^{t-1}.

then the agent's learning strategy during distributed execution. For the designed DQN designed, the composition of replay memory is as following:

1) Statement: The statement design of an agent (n, k) is very important, for the entire environment information is redundant and irrelevant elements must be removed. We assume that the environment is in a logarithmic complete set and define the interference factors as:

$$\tau^t_{n,k} = \{\underbrace{1, \cdots, 1}_{k-1}, \{\log_2(1 + \frac{p^c_{k,m} h^C_{m,B}}{p^{(3)}_{k,m} h^D_{k,B} + \sigma^2_N})\}\} \tag{26}$$

The channel amplitude of the interference sources are normalized by the amplitude of the required links, and since the channel amplitudes usually vary by orders of magnitude, it is preferable to express them logarithmically. Base $\Gamma^t_{n,k}$ is represented as $(|D_n| + 1)K - 1$, and in order to further reduce the input dimension and the computational complexity, the elements in radix $\Gamma^t_{n,k}$ are sequence successively, and only the first C elements are retained. The rate $C^{t-1}_{n,k}$ and transmitting power $p^{t-1}_{n,k}$ of the corresponding uplink of the link act as two additional parts of the DQN input at the last time slot. Therefore, a statement is composed of three characteristics:$s^t_{n,k} = \Gamma^t_{n,k}, C^{t-1}_{n,k}, p^{t-1}_{n,k}$.

2) Action: In (3), the upstream power is a continuous variable constrained by the maximum power. Considering limited action space of DQN, the transmitting power is quantized as $|A|$, and the settings are as follows, where P_min is non-zero minimum transmitting power:

$$A = \{0, P_{\min}, P_{\min}(\frac{P_{\max}}{P_{\min}})^{\frac{1}{|A|-2}}, \cdots, P_{\max}\}, \tag{27}$$

3) Rewards: In order to increase the transmission rate of agents and reduce the incidence of interference, other studies have carefully designed reward functions, most of which however are actually suboptimal methods to solve the target function. In this paper, the system sum-rate is directly shared by all agents as a reward functions instead. The feasibility of this method is demonstrated in the training simulation of small and medium-sized cellular networks.

The selection of right action is based upon accurate estimation, and thus DQN is aimed to search for optimal

parameter θ_q^* to minimize the loss:

$$\theta_q^* = \arg\min \frac{1}{2}(Q(s, a; \theta_q) - r_s^a)^2. \quad (28)$$

The gradient concerning θ_q is given as

$$\nabla\theta_q = (Q(s, a; \theta_q) - r_s^a)\nabla_{\theta_q} Q(s, a; \theta_q) \quad (29)$$

The optimal action a^* is selected to maximize the Q value, and it is given by

$$a^* = \arg\max Q(s, a; \theta_q). \quad (30)$$

During training, a dynamic ε-greedy policy is adopted to control the exploration probability, and ε_k is defined as

$$\varepsilon_k = \varepsilon_1 + \frac{k-1}{N_e - 1}(\varepsilon_{N_e} - \varepsilon_1), k = 1, \cdots, N_e \quad (31)$$

where N_e denotes the episode times, ε and ε_{N_e} are initial and final exploration probabilities, respectively.

In terms of computational complexity, because there are too many variables describing the time complexity of deep learning, it cannot be accurately described. For example, the time complexity of neural network training is $O(E*D/B*T)$, where E is epochs, D is the data set size, B is Batch Size, and T is the time complexity of a single iter. Here T can continue to be decomposed into $O(T) = O(L*n)$, L is the average time complexity of each layer, and n is the number of layers. Among them, L can continue to be decomposed into $O(L) = O(M*N*K*K*H*W)$. It is assumed that each layer has several calculations in conv2d, where M and N are the number of input and output channels, respectively. K is the size of the convolution kernel, H and W are the spatial sizes of the output feature map respectively. In summary, the time complexity of the training process is related to at least 10 free variables, so we do not specifically analyze the computational complexity of DQN.

A detailed description of our DQL algorithm is presented in Algorithm 3.

VI. SIMULATION

According to the hexagon element model [11], as shown in Figure 4, we summarizes the simulation parameters that measure the performance of the proposed algorithm in Table 2. The hexagon element model is constituted of CUs and D2D transmitters are uniformly distributed. The range of D2D communication is defined as the distance between the D2D receiver and its corresponding transmitter. It is assumed that the D2D receiver is uniformly distributed around the D2D transmitter, and the range is RD2D. In order to examine the system performance of RD2D with different value ranges, we change it from $10m$ to $100m$, and the step length is $10m$. All other parameters related to the DQN are given in Table 3. Note that the listed parameters are selected from multiple simulation tests to balance complexity and performance of DRL algorithm.

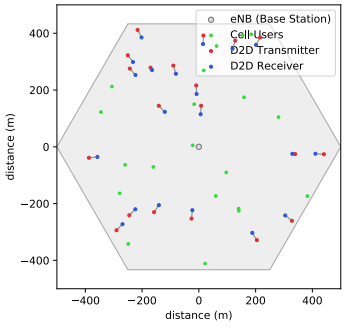

FIGURE 4. Hexagonal cell model.

Algorithm 3 DQN Algorithm

input : Episode times N_e,exploration times T, learning rate η_q, initial and final exploration probability ε_1, ε_{N_e}
Initialization : Initialize DQN $Q(s, a; \theta_q)$ with random parameter θ_q.
for $k = 1$ to N_c **do**
  Update ε_k by (31)
  Receive initial state s.
  for $t = 1$ to T **do**
    if $rand() < \varepsilon_k$ **then**
      Randomly select action $a^t \in A$ with uniform probability.
    else
      Select action a^t by (30)
    end if
    Execute action a^t, achieve reward r^t and observe new state s^{t+1}.
    Calculate gradient $\nabla\theta_q$ by(29), and update parameter along negative gradient direction: $\theta_q \leftarrow \theta_q - \eta_q\nabla\theta_q$
    $s^t \leftarrow s^{t+1}$
  end for
end for
Output : Learned DQN $Q(s, a; \theta_q)$

TABLE 2. List of notations.

Parameter	Values
Cell layout	Single hexagonal cell
ISD	500 m
Spectrum allocation (uplink)	5 MHz
Available resource blocks	25
Max D2D transmit power	250 mW
Max CU transmit power	250 mW
User distribution	Uniform
Number of CUs(N_C)	20
Number if D2D pairs(N_D)	10%,20%,...,100% of active CUs
Path loss	PL=128.1+37.6log(d)
Shadowing	Log-normal distribution
Fast fading	Rayleigh fading
UE noise figure	5 dB
UE thermal noise density	-174 dBm

TABLE 3. Simulation Parameters for DRL.

Parameter	Values
Learning rate	0.001
Discount factor	0.70
Initial exploration	1
Final exploration	0.01
Total exploration steps	1000
Replay memory size	3000
Minibatch size	8
Network update frequency	2
Target network update frequency	30
Number of steps in each epoch	10
Weights in reward function	0.1,0.9,1,1

Figure 5 shows the relationship between the total throughput of the different scenarios and the number of D2D pairs N. It can be seen that the scheme combining the three algorithms improve the total throughput of the system compared with other schemes.

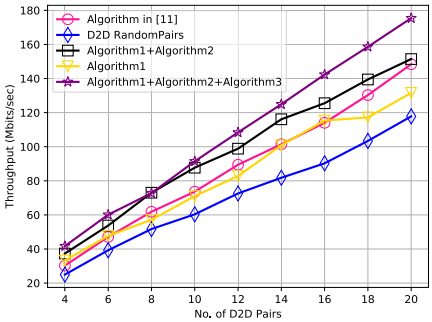

FIGURE 5. Throughput versus different CU numbers (r=5,P^D_{max}=250mW,P^C_{max}=250mW,B=1).

We can also observe that the mode selection of the heuristic algorithm alone has similar performance to the algorithm based on the bipartite graph proposed in [11], because DU and CU have no interference between the same channel in the cellular mode or the dedicated mode. Therefore, the maximum throughput can be guaranteed when DU and CU have the maximum power, which is similar to the resource allocation scheme proposed in [11].

Figure 6 shows the impact of the link length of the D2D pair on the system throughput of different algorithms. As can be observed in the figure, the system throughput of all algorithms decreases with the increase of the maximum transmission distance, which is that the gain of each algorithm decreases as the distance increases. Besides, when the maximum transmission distance increases, the transmitting power of the D2D transmitter has to be increased to guarantee the transmission quality but reduce the throughput. Nonetheless

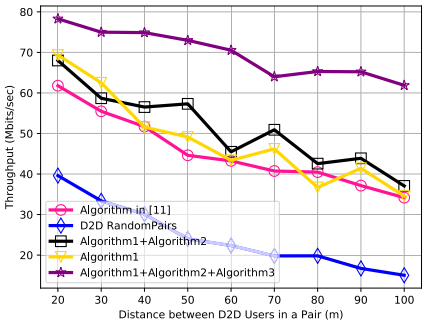

FIGURE 6. Overall system throughput for different D2D distances(NU=20,ND=20P^D_{max}=250mW,P^C_{max}=250mW,B=1).

on the whole, the combined algorithm is superior to the other four comparison algorithms.

Figure 7 displays the change in throughput after increasing the resource block while keeping the number of D2D pairs constant. As showed from the figure, the system will achieve higher throughput as the resource blocks increase. Moreover, the algorithm combining the three modes can make the system throughput is increasingly rapidly.

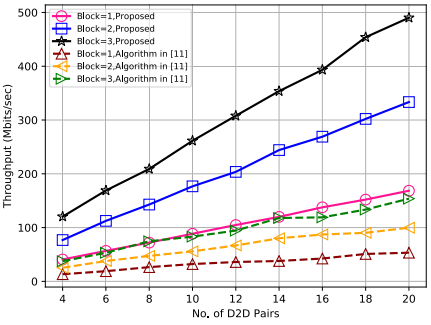

FIGURE 7. Overall system throughput for different Block(r=5, NU=20, ND=20, P^D_{max}=250mW, P^C_{max}=250mW, B=1).

The key technology for DQN is the point of experience playback. The correlation between training samples was broken through random and uniform sampling in experience playback. Simultaneously, multiple samples in the past were used for averaging, which also smoothed the distribution of training samples and alleviated the problem of sampling distribution changes. Figure 8 shows the smoothing Training step with different learning rates to evaluate the Algorithm3. As shown from the graph, the Smoothing Training Step index is the highest at a learning rate of 0.001. A high learning rate may lead to a local optimal rather than a globally optimal, so considering the actual real-time execution of the algorithm. The learning rate is selected as 0.001.

To assess the fairness of the average user data rate of our proposed scheduling algorithm, we measure it using the Fairness index of Jain's fairness Index (JFI). Let $\bar{R}_d$ be the average data rate of D2D user D on subframe N. Since the total number of D2D users is ND, and the JFI equity index is

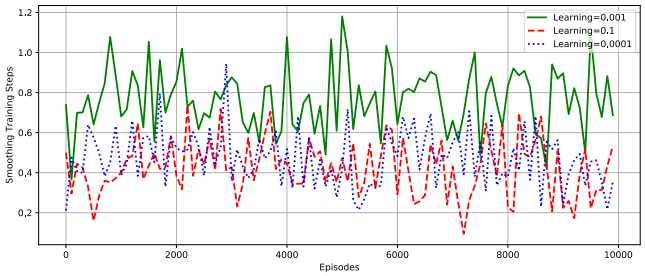

FIGURE 8. With different Learning rate, the smoothing training steps during training period.

defined as:

$$\eta = (\sum_{d} \bar{R}_d / (N_D \sum_{d} \bar{R}_d^2)) \tag{32}$$

The range of JFI values is $[1/N, 1]$, where $JFI = 1$ when all users have the exact same rate. Therefore, the closer JFI is 1, the better the fairness between users will be proved. As shown in Figure 9, the JFI of the proposed scheme is about 0.9, which is much higher than that of the pre-optimization scheme. Moreover, with the given number of cellular users, this advantage will be any more evident as the number of D2D users increases.

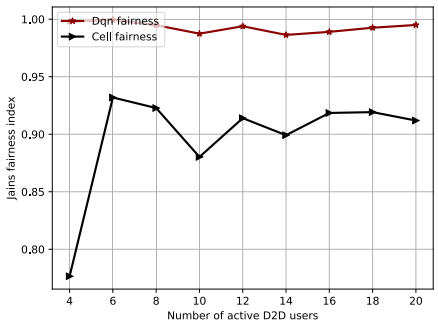

FIGURE 9. Comparison of jain's fairness index between PF and MR scheduling with increasing number of D2D pairs for r=5, NU=20, ND=20, P^{D}_{max}=250mW, P^{C}_{max}=250mW, B=1.

In the pre-optimization scheme, the base station tends to allocate the resource block to the DU, bringing the maximum rate gain. When the number of CU channels is fixed, the increase of DU will make it more and more challenging to meet the rate requirements of part DU, and the performance of users with poor performance will be worse and worse. Further, in our proposed scheme, the system always prioritizes those users with the worst performance no matter how many D2D users there are, so the performance gap between users will not increase further.

Figure 10 indicates the D2D communication pairs' cumulative distribution function curve of effective outage threshold. It can be observed in the figure that the scheme combined with the three algorithms is superior to the other four schemes. Due to the effective outage threshold increase, D2D pairs tend to choose an immense transmission power to

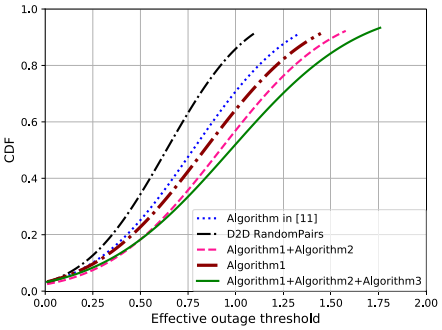

FIGURE 10. Cumulative distribution function curve of effective outage threshold.

ensure the reliability requirements, causing more severe interference to the adjacent D2D pairs. Furthermore, the proposed scheme of combining the three algorithms can effectively suppress the interference by the adaptive selection of the best transmission mode.

Figure 11 shows the training process of Power Control. The reward value is the average value of the numerical simulation obtained over 10 training sessions. It can be seen from the experimental results that the average reward of iteration increases with the increase of the number of interactions between user agents. This shows that the proposed method combined with the three schemes can be a successful and effective learning strategy, and the algorithm can converge faster.

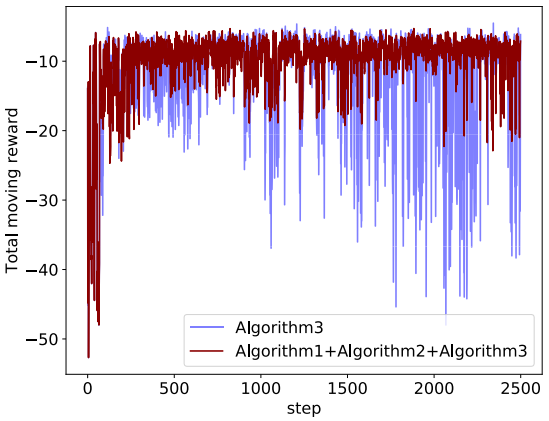

FIGURE 11. Learning process of combining three algorithms.

VII. CONCLUSION

In this paper, we consider joint mode selection, channel allocation and power control in D2D communication in cellular networks. The throughput of whole system is optimized by combining the three proposed algorithms, and the SINR of cellular and D2D links is guaranteed. The optimization problem is decomposed into three subproblems: Transmission power control, joint mode selection and channel allocation for each D2D pair. And we find that the proposed scheme combining three algorithms can effectively improve system performance through numerical simulation. In the future, our research will be extended to the scene of ultra-dense network,

considering the interference between cells, and designing an effective distribution scheme reasonably.

REFERENCES

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[2] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," 2020, *arXiv:2002.03491*. [Online]. Available: https://arxiv.org/abs/2002.03491

[3] N. Cheng, H. Zhou, L. Lei, N. Zhang, Y. Zhou, X. Shen, and F. Bai, "Performance analysis of vehicular device-to-device underlay communication," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5409–5421, Jun. 2017.

[4] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3541–3551, Aug. 2013.

[5] C.-H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2752–2763, Aug. 2011.

[6] P. Janis, V. Koivunen, C. Ribeiro, J. Korhonen, K. Doppler, and K. Hugl, "Interference-aware resource allocation for device-to-device radio underlaying cellular networks," in *Proc. IEEE 69th Veh. Technol. Conf. (VTC Spring)*, Apr. 2009, pp. 1–5.

[7] B. Kaufman, J. Lilleberg, and B. Aazhang, "Spectrum sharing scheme between cellular users and ad-hoc device-to-device users," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1038–1049, Mar. 2013.

[8] Y. Jiang, Q. Liu, F. Zheng, X. Gao, and X. You, "Energy-efficient joint resource allocation and power control for D2D communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6119–6127, Aug. 2016.

[9] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 136–144, Jun. 2014.

[10] X. Diao, J. Zheng, Y. Wu, and Y. Cai, "Joint computing resource, power, and channel allocations for D2D-assisted and NOMA-based mobile edge computing," *IEEE Access*, vol. 7, pp. 9243–9257, 2019.

[11] I. Mondal, A. Neogi, P. Chaporkar, and A. Karandikar, "Bipartite graph based proportional fair resource allocation for D2D communication," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, Mar. 2017, pp. 1–6.

[12] Y. Liu, Y. Wang, R. Sun, and Z. Miao, "Distributed resource allocation for D2D-assisted small cell networks with heterogeneous spectrum," *IEEE Access*, vol. 7, pp. 83900–83914, 2019.

[13] K. Doppler, C.-H. Yu, C. B. Ribeiro, and P. Janis, "Mode selection for device-to-device communication underlaying an LTE-advanced network," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2010, pp. 1–6.

[14] S. Hakola, T. Chen, J. Lehtomaki, and T. Koskela, "Device-to-device (D2D) communication in cellular network-performance analysis of optimum and practical communication mode selection," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2010, pp. 1–6.

[15] M.-H. Han, B.-G. Kim, and J.-W. Lee, "Subchannel and transmission mode scheduling for D2D communication in OFDMA networks," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2012, pp. 1–5.

[16] Z. Qin, H. Ye, G. Y. Li, and B. F. Juang, "Deep learning in physical layer communications," *CoRR*, vol. abs/1807.11713, pp. 1–14, Mar. 2018.

[17] Z. Wei, D. W. K. Ng, and J. Yuan, "Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3944–3961, Sep. 2017.

[18] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference-aware graph based resource sharing for device-to-device communications underlaying cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 140–145.

[19] H. Yang, X. Xie, and M. Kadoch, "Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4157–4169, May 2019.

[20] L. Liang, S. Xie, G. Y. Li, Z. Ding, and X. Yu, "Graph-based resource sharing in vehicular communication," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4579–4592, Jul. 2018.

[21] J. Chen, H. Yin, L. Cottatellucci, and D. Gesbert, "Feedback mechanisms for FDD massive MIMO with D2D-based limited CSI sharing," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5162–5175, Aug. 2017.

[22] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, vol. 159. Norwell, MA, USA: Kluwer, 1992.

[23] W. Zhao and S. Wang, "Resource sharing scheme for device-to-device communication underlaying cellular networks," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4838–4848, Dec. 2015.

[24] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, and G. Ostrovski, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

[25] J. Chen, H. Yin, L. Cottatellucci, and D. Gesbert, "Dual-regularized feedback and precoding for D2D-assisted MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6854–6867, Oct. 2017.

YONGWEN DU was born in Lanzhou, Gansu, China, in 1974. He received the M.S. and Ph.D. degrees from Northwestern Polytechnical University. His main research interests include mobile edge computing, game theory, and intrusion detection.

WENXIAN ZHANG was born in Benxi, Liaoning, China, in 1995. He received the B.S. degree from the University of Science and Technology Liaoning, China. He is currently pursuing the master's degree with the School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu, China. His research interests include mobile edge computing and machine learning.

SHAN WANG is currently pursuing the master's degree in computer application technology with Lanzhou Jiaotong University. Her main research interest includes blockchain technology.

JINZONG XIA was born in Jinan, Shandong, China, in 1996. He received the B.S. degree from Shandong Jiaotong University, Jinan. He is currently pursuing the master's degree with the School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu, China. His research interests include information security and wireless sensor networks.

HYTHEIM ALHAG MOHAMMAD received the bachelor's degree (Hons.) in computer systems and networks from the Sudan University of Science and Technology, in 2009, Graduate Project: Real Time Intrusion Detection System. He is currently pursuing the master's degree with Lanzhou Jiaotong University under the supervision of Prof. Yang Jun. He worked with Arabian Computer Company, Sudan. He worked with AZ Technology Company Ltd., Sudan. His main research interest includes 3D shape correspondence.