telecom

MDPI

Article

A Semi-Distributed Scheme for Mode Selection and Resource Allocation in Device-to-Device-Enabled Cellular Networks Using Matching Game and Reinforcement Learning

Ibrahim Sami Attar [1], Nor Muzlifah Mahyuddin [1,*] and M. H. D. Nour Hindia [2]

[1] School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Nibong Tebal 14300, Malaysia; ibrahim_sami@student.usm.my
[2] Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia; nourhindia@um.edu.my
* Correspondence: eemnmuzlifah@usm.my

Abstract: Device-to-Device (D2D) communication is a promising technological innovation that is significantly considered to have a substantial impact on the next generation of wireless communication systems. Modern wireless networks of the fifth generation (5G) and beyond (B5G) handle an increasing number of connected devices that require greater data rates while utilizing relatively low power consumption. In this study, we present joint mode selection, channel assignment, and power allocation issues in a semi-distributed D2D scheme (SD-scheme) that underlays cellular networks. The objective of this study is to enhance the data rate, Spectrum Efficiency (SE), and Energy Efficiency (EE) of the network while maintaining the performance of cellular users (CUs) by creating a threshold of data rate for each CU in the network. Practically, we propose a centralized approach to address the mode selection and channel assignment problems, employing greedy and matching algorithms, respectively. Moreover, we employed a State-Action-Reward-State-Action (SARSA)-based reinforcement learning (RL) algorithm for a distributed power allocation scheme. Furthermore, we suggest that the sub-channel of the CU is shared among several D2D pairs, and the optimum power is determined for each D2D pair sharing the same sub-channel, taking into consideration all types of interferences in the network. The simulation findings illustrate the enhancement in the performance of the proposed scheme in comparison to the benchmark schemes in terms of data rate, SE, and EE.

Keywords: device-to-device (D2D); resource allocation; greedy algorithm; matching theory; power allocation; reinforcement learning (RL); SARSA

Academic Editor: Mario E. Rivero-Angeles

Received: 23 December 2024
Revised: 7 February 2025
Accepted: 11 February 2025
Published: 13 February 2025

Citation: Attar, I.S.; Mahyuddin, N.M.; Hindia, M.H.D.N. A Semi-Distributed Scheme for Mode Selection and Resource Allocation in Device-to-Device-Enabled Cellular Networks Using Matching Game and Reinforcement Learning. *Telecom* **2025**, *6*, 12. https://doi.org/10.3390/telecom6010012

1. Introduction

The growing demand for improved broadband services for mobile devices and the rapid growth of different fields of application, such as vehicle-to-vehicle communication, automating factories, cellular healthcare services, and augmented and virtual reality solutions, require the development of new (B5G) network architectures. These architectures need to be capable of supporting lower energy consumption and higher area capacity compared to current networks [1]. Considering these demands, the problem of limited network capacity is an important challenge for the evolution of emerging wireless networks. In addition to the limited availability of the spectrum, the development of the cellular network is leading to concerning levels of energy consumption [2]. The importance of EE has been growing as a consequence of economic, operational, and environmental considerations [3].

Furthermore, the rapid increase in data rate demands and power-intensive mobile systems and applications, in combination with the scarcity of available spectrum resources, necessitates the exploration of innovative networking solutions.

D2D communication is a promising technology among the emerging B5G solutions, providing potential solutions for the aforementioned demands. D2D communication enables direct transmission of data traffic instantaneously via the D2D transmitter toward a D2D receiver without passing via a cellular gNB [4]. Moreover, it requires low power transmission that enhances EE and allows for the spectrum sharing of the available resources, ultimately improving spectral efficiency. D2D communications utilizing the resources of wireless cellular networks will be crucial in enhancing the capacity of incoming B5G systems [5]. The advantages of implementing D2D communication over wireless cellular networks are becoming widely recognized, particularly for data offloading, content sharing, EE, coverage expansion, and enhanced utilization of the spectrum. Furthermore, an additional demand in B5G networks is the capacity to deal with extensive communications resulting from the rapid increase in connected devices in conventional cellular networks [6].

Mode selection and resource allocation are crucial issues for creating and maintaining direct connections between D2D users within cellular networks. Furthermore, the distribution of network resources among D2D pairs and cellular users can be achieved efficiently, which could improve SE and EE by controlling interferences in the network.

Mode selection is an essential challenge in D2D communications that enhances EE and average data rate and minimizes interference in the network. It decides whether D2D users can perform in direct mode or cellular mode. The mode selection procedure is flexible, leading to decreased latency and increased spectrum resource utilization. The gNB can assign three D2D communication modes to each D2D pair, including direct D2D mode (DM), relay-assisted D2D mode (RM), and local route D2D mode (LM) [7].

D2D communication can utilize either a licensed or unlicensed spectrum for channel assignment to establish direct connections, assigned as in-band and out-band D2D communications, respectively [8]. According to out-band D2D communication, users communicate via direct communication utilizing an unlicensed spectrum, separated from that utilized by cellular users in the network. On the other hand, in-band D2D communication may be categorized into two classifications: underlay and overlay. According to the framework of underlaying in-band communication, the spectrum is shared and assigned to D2D pairs and CUs. Moreover, in-band communication overlay identifies specific parts of the whole spectrum for D2D communications and the other portion for CUs. This study investigates the idea of underlaying in-band D2D communication, focusing on improving the system's performance. However, D2D pairs experience interference when implemented with the underlay in-band architecture, caused by the utilization of shared sub-channels [9]. Consequently, the control of interference becomes a significant research concern within the field of D2D communication networks.

Despite the improvements in cellular networks enabled by D2D communication, many concerns should be addressed. Conventional cellular networks are significantly affected by the interferences imposed by D2D communications due to the sharing of cellular user resources. These concerns have a substantial effect on the performance of cellular user communications and could impact the future development of D2D technology. Therefore, it is essential to figure out efficient strategies for the purpose of allocating resources to D2D communication pairs without maximizing the complexity of the network's topology. Specifically, effective allocation of resources for D2D communications is an important challenge in the cellular network. Following the mode selection procedure, optimal sub-channel resource allocation is necessary to achieve the goal of D2D technology [10].

Power control and interference mitigation should be considered through the resource allocation process. Additionally, to enhance EE and guarantee QoS for both D2D pairs and CUs, appropriate power allocation approaches need to be designed. Blindly implementing power control to D2D communications in a cellular network could reduce system efficiency. D2D transmitters should optimize their transmission power to enhance EE, SINR demands, and system performance.

Machine Learning (ML) has been demonstrated to be highly beneficial in multiple areas due to its ability to precisely predict future scenarios and address complex problems with huge datasets [11]. Artificial intelligence approaches have been indicated to be an excellent technique for tackling complicated non-convex optimization problems in communication networks. In the field of wireless communication networks, reinforcement learning (RL) based on ML techniques has been utilized to solve the issue of power allocation [12].

2. Related Work

Various current investigations have tackled the problem of mode selection, channel assignment, and power allocation for D2D communication underlay cellular networks. For instance, in [13], the resource allocation problem has been formulated based on the investigation of the system model in many-to-many matching D2D communication underlying cellular networks. An overlapping coalition game method that utilizes a candidate sequence is introduced to improve the D2D transmission data rate. The simulation findings showed that the efficiency of the proposed method, which incorporates an optimization technique for candidate sequences, outperforms that of the overlapping approach with random optimization.

In [14], a genetic algorithm-based joined power and channel allocation was investigated in D2D-underlayed cellular networks. Efficient allocation of transmitted power and shared channels for every user is achieved by a genetic algorithm optimization, that assists in minimizing interference. The simulation outcomes illustrated that the proposed approach outperformed the fixed, random, and particle swarm optimization approaches in terms of maximizing the total resource utilization. In [15], the authors addressed the issues of mode selection, spectrum usage, and power management in D2D underlay cellular communications by the use of a hierarchical game approach. For mode selection and channel assignment issues, the optimal solution is achieved by employing the hedonic coalition game. Moreover, the non-cooperative game is employed to tackle the issue of power management in the proposed scheme. The algorithm effectively addressed the issue of fair allocation of resources among users, while simultaneously improving the system throughput.

In [16], the interference management issue was investigated in order to effectively enable the coexistence of two technologies: a massive MIMO and a D2D communication system that shares uplink network resources under cellular networks. To achieve a distributed solution, they formulated the problem via a matching theory and presented a resource optimization technique utilizing the principles of many-to-many matching to enhance the performance of the system. The numerical outcomes illustrated that the suggested method effectively improved the network performance by leveraging the diversification benefits of massive MIMO and matching users based on their preferences. Moreover, in [17], the paper presented a dynamic resource allocation strategy for D2D communications under cellular networks. The suggested resource allocation technique incorporates a Q-learning-based power management algorithm to allocate optimum powers for D2D pairs, with the goal of optimizing the throughput. The introduced algorithm employs a Q-learning technique to enhance the transmission power for the D2D pairs that utilize the same resource block at the establishment of a unique D2D pair. The model findings showed that the suggested

technique can achieve quick convergence and outperform the random power allocation algorithm in terms of overall throughput. In [18], the authors examined techniques for allocating resources in D2D communications underlaying cellular networks. A unique DC programming method for the difference of convex functions is provided to successfully address this complex resource allocation problem. The simulation findings showed that the introduced system reaches the maximum weighted sum-data rate in comparison to the benchmark methods, while simultaneously guaranteeing the satisfaction of QoS demands for each D2D pair and CUs.

Furthermore, in [19], the quality of cell-edge user coverage was enhanced with the development of a D2D-relay communication system for underlay cellular networks. Subsequently, the Lagrange dual approach-based power allocation technique was developed to effectively allocate power levels to the users in the network. The suggested technique converges in time, and an optimal closed-form solution is derived. The simulation findings demonstrated that the suggested strategy highly improved the network coverage, the data rates of users, and the SE of the system. In [20], a multi-agent Q-learning approach is presented to enhance the throughput of D2D communications underlying cellular networks. First, the multi-agent Q-learning technique-based channel resource allocation is implemented. Furthermore, to tackle the issue of slow convergence in the Q-learning algorithm for the D2D communications system, the authors included a Fuzzy C-Means algorithm into a multi-agent Q-learning framework. This integration aimed to enhance the utilization of power management by employing a Q-learning approach. The results showed that the implementation of the multi-agent Q-learning approach enhances the system's throughput. In [21], the authors proposed an auction technique based on a D2D relay selection algorithm underlaying cellular networks. The cooperative willingness of user relay devices was evaluated from a social level. As social relationships become stronger, the willingness to cooperate increases and the transmission power of the relay also increases. Subsequently, they determined a relationship between the outage probability and the transmission power. Next, an auction approach was employed to stimulate relays to enhance the transmitted power via monetary incentives. The results demonstrated the technique's superiority over previous relay methods, as it not only enhanced the system's data rate but also decreased the probability of a communications outage.

Additionally, the researchers in [22] suggested a graph coloring approach to address interference issues in D2D communications-based cellular networks. The primary objective is to exploit the weighted prioritization of spectrum resources, allowing several D2D pairs to use the same resources as cellular users inside the network. Once the spectrum allocation has been achieved, the power management process is employed to reduce the transmission power of D2D pairs which leads to minimizing the interferences and enhancing the energy consumption of the entire cell. The simulation outcomes illustrated that the suggested method successfully mitigated co-channel interference, enhanced system throughput, and minimized power consumption in comparison to the traditional techniques. The authors in [23] suggested a method based on the D2D multicast clusters approach along with a Q-Learning-aided approach to tackle the issue of joint sub-channel assignment and power management for D2D communication under cellular networks. An agglomerative hierarchical clustering approach using unsupervised ML was suggested to establish clusters of D2D pairs, considering user preferences and ensuring reliable D2D multicast communications across D2D users. The numerical simulation outcomes indicated that the suggested approach provided substantial benefits for the throughput and EE of the introduced system compared to existing techniques. In [24], the paper examined the issues of joint channel assignment and power management in D2D communications underlaying cellular networks based on the concept of the NOMA technique. The suggested approach divided

the issue into two sub-issues: channel assignment and power management. The assignment of channels to the pairs is shown using a matching game theory, providing a stable solution while considering the interferences imposed by users in the network. The power allocation problem is addressed in both phases utilizing difference-of-convex programming that is optimized iteratively by the Frank–Wolfe method. The performance investigations have shown that the suggested method enhanced SE, fairness, and network connectivity. In [25], the authors examined the issue of resource allocation for D2D communications in cellular networks. A one-to-many matching game was implemented to enhance spectrum utilization and tackle frequency interference problems. They introduced a resource allocation algorithm based on the relationship between the distance of the users and the interference levels to assign channel resources efficiently. This technique enables D2D pairs to efficiently reuse channel resources utilized by CUs in close proximity while minimizing interference between D2D pairs and CUs. Moreover, the particle swarm optimization technique was employed to tackle the optimum power allocation problem, aiming to obtain the highest transmission rate of the network. The simulation findings demonstrated that this approach enhanced system data rate and performance, while simultaneously minimizing the computational complexity.

In [26], a resource allocation technique is suggested, based on the QL algorithm, for D2D communication in the unlicensed spectrum. This approach allows for dynamic allocation of transmitted power to D2D users based on updating network traffic conditions, leading to improved performance of the coexisting system. The agent's states are determined by different factors such as fairness, SNR, and data rate of CUs. The agent's actions can be determined by the different duty cycles and transmission power levels. The agent can learn the duty cycles and optimum power transmission via iterative interaction responses to the network. The comprehensive simulation findings indicated that the proposed technique outperforms the compared approaches in resource allocation fairness and throughput. The authors in [27] tackled a joint optimization issue of resource allocation, optimum power management, and relay-selection in a two-way relaying approach underlaying cellular networks. The power management problem is solved using particle swarm optimization, while the relay selection problem is addressed via the one-to-one stable matching technique. The numerical outcomes indicated that employing stable matching in the relay selection problem significantly increased the performance of the network, D2D data rate, and EE. In [28], the authors investigated the enhancement of system fairness and throughput in an underlying D2D communications scheme by employing joint channel allocation and power management methodologies. They suggested an iterative resource allocation approach based on RL by considering the channel parameters. The authors proposed an enhanced reinforcement learning-based SARSA algorithm. The simulation findings illustrated the enhancement of the system's throughput, EE, and SE. In [29], the authors examined EE optimization for D2D communications underlaying cellular networks. The primary objective of the study is to improve the performance of the scheme as well as reduce the power conception of the equipment while maintaining the QoS for every user. The Q-learning algorithm is applied to achieve optimum communication between the users and gNB in the proposed network. The simulation findings demonstrated that the proposed system enhanced the transmission performance as compared with the traditional systems.

In [30], the authors proposed a joint channel allocation and power allocation issue for D2D connections within cellular networks. The main goal of this research is to optimize the EE of the proposed scheme while satisfying the QoS of CUs and D2D communications. The suggested problem is NP-hard and complex to solve; therefore, an iterative solution is introduced. The authors introduced Dinkelbach's method to address the problem and

obtain the optimal solution. The simulation findings showed the advantage of the suggested technique relative to conventional schemes.

3. Research Contributions

This paper introduces joint mode selection, channel assignment, and power allocation issues of D2D communications underlaying cellular networks in an uplink scenario, where a single sub-channel can be shared via a CU and several D2D pairs. The main goal of this study is to enhance the data rate, SE, and EE of the proposed network, while simultaneously satisfying the minimal QoS demands for CUs and D2D pairs. The suggested approach is a semi-distributed architecture, in which mode selection and channel assignment issues are centralized, while the power management issue is solved with a distributed technique. The complexity of this technique is lower in comparison to centralized solutions. Firstly, the optimal D2D mode can be achieved by utilizing a greedy algorithm that chooses the best mode among DM, RM, and LM based on maximum SINR. Furthermore, the provided channel assignment method is based on a matching algorithm that takes into consideration the priorities of both the D2D pairs and the CUs, which differs from most of the existing research that mainly investigates the preferences of D2D pairs only. The channel assignment technique based on two-sided preference achieves stable matching with minimal complexity. Moreover, the channel assignment method enables the reuse of a single CU sub-channel throughout several D2D pairs, resulting in higher SE. A higher number of D2D pairs may be served with limited spectrum resources through the implementation of this type of resource-sharing scheme.

In addition, power management is achieved for each D2D pair by applying the SARSA-based RL algorithm. This low-complexity distributed RL algorithm has the ability to calculate the optimum power for each D2D pair that enhances the EE of the network. While several studies have attempted to address the issue of resource allocation throughout D2D pairs, they have either insufficiently accounted for the potential that D2D users might interfere with each other or have assigned resources based on the assumption that D2D pairs have a constant power transmission. The proposed approach differs from previous studies that either assumed a fixed D2D power or neglected the interferences between the D2D pairs. The main contributions of the introduced scheme can be illustrated as described below:

1. The problem of joint mode selection, channel assignment, and power allocation is formulated for D2D communications underlaying cellular networks by utilizing the uplink resources of CUs. The optimization problem is formulated to enhance the data rate, SE, and EE of the network while considering QoS characteristics related to D2D pairs and CUs simultaneously.
2. By employing a greedy algorithm, a mode selection technique is introduced to choose the optimum mode throughout DM, RM, and LM across every D2D pair in the network. The computational representation is formulated based on the highest SINR.
3. channel assignment method is introduced, using a matching algorithm to assign the optimal sub-channel to the D2D pairs in the network. The channel assignment approach based on a two-sided preferences list provides stable matching with low complexity. The first preference list consists of the data rates of D2D pairs arranged in descending order according to their highest value. The second preference list consists of the interference effect of CUs on D2D pairs when they share the same sub-channel, arranged in ascending order according to their lowest impact value.
4. low-complexity distributed SARSA-based RL algorithm is implemented to address the issue of power control and allocate the optimum power level for each D2D pair to enhance the EE of the network.

5. The effectiveness of the suggested method has been shown by simulations, particularly in terms of the data rate, SE, and EE of the network in comparison to conventional systems.

The remainder of this paper is organized as follows: Section 4 provides a comprehensive clarification about the system model. Section 5 defines the mode selection, channel assignment, and power allocation problem. Section 6 discusses the challenge and illustrates the methodologies utilized for joint mode selection, channel assignment, and the power allocation scheme. Section 7 presents and discusses the simulation parameters as well as the simulation results. Section 8 concludes the paper.

4. System Model

In this study, we explore the principle of the D2D communication scheme underlaying cellular networks, which involves spectrum sharing across multiple D2D users. For every D2D pair, we analyze three different modes, which include DM, RM, and LM, as illustrated in Figure 1. The DM enables direct transmission of data from the transmitter to the receiver of each D2D pair. Moreover, the second D2D mode is RM which aims to establish relay communication between two distant devices. An idle user is used as a relay node in this mode to facilitate the creation of a connection between the transmitter and receiver of the D2D pairs. Regarding the LM, it is suggested that auxiliary antennas be installed on the gNB to enhance the connections of the LM [31]. The data are transferred via the gNB instead of entering the core, which means that the gNB acts as a node to support the connections of faraway users to establish a D2D pair.

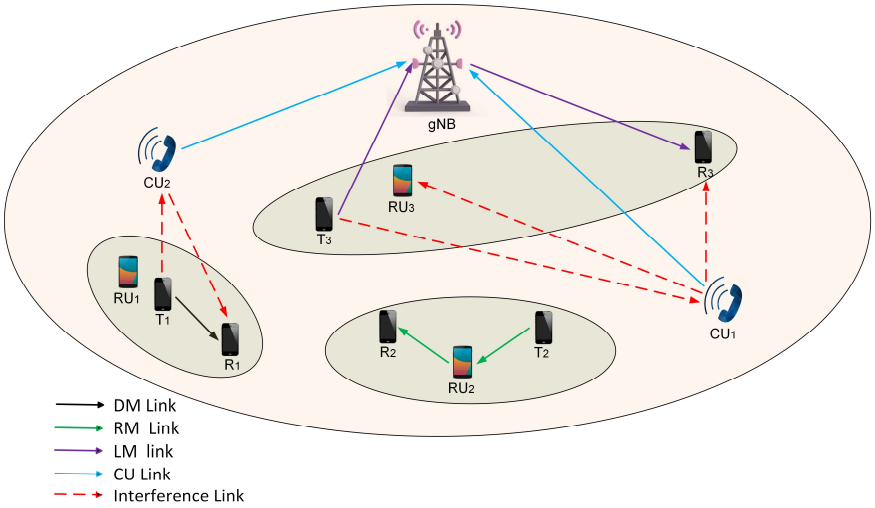

Figure 1. System model.

In this study, we examine an environment where N represents the D2D pairs' number. Furthermore, T_n, R_n, and RU_n represent the D2D transmitter, receiver, and relay of the pair n, respectively. Let us suppose W is the total bandwidth in the network which is partitioned to a number of sub-channels indicated as K. Table 1 illustrates the scheme symbols.

Table 1. Scheme symbols.

Symbol	Description
N	D2D pairs' number
M	CUs number
T_n	D2D transmitter
R_n	D2D receiver

Table 1. *Cont.*

Symbol	Description
RU_n	Relay user equipment
W	System Bandwidth
K	Sub-channels' number
G_n, G_{RUm}	Channel gain for *nth* D2D pair, R_n to *mth* CU, respectively
G_{nm}, G_{mn}	Channel gain across *nth* D2D pair and *mth* CU, *mth* CU and *nth* D2D pair, respectively
P_n, P_C, P_Z	Power transmission of the D2D pair, CU, and gNB, respectively
G_{TR}, G_{TRU}, G_{TZ}	Channel gain between T_n to R_n, RU_n, and gNB, respectively
G_{RUR}, G_{ZR}	Channel gain across RU_n, and gNB to R_n
σ	Noise power
$\Upsilon_{n,k}$, $\Upsilon_{m,k}$	Received SINR for D2D pair n and cellular user m utilizing sub-sub-channel k, respectively
$\mathcal{I}_D$, $\mathcal{I}_R$, $\mathcal{I}_L$	Interference impact of CUs, other D2D transmitters, and D2D relays on the *nth* DM, RM, and LM, respectively
$G_{T'R}$, $G_{T'RU}$, $G_{T\prime Z}$	Channel gain between other D2D transmitters to R_n, RU_n, and gNB, respectively
$G_{RU\prime R}$, $G_{RU'RU}$, $G_{RU\prime Z}$	Channel gain across other D2D relays to R_n, RU_n, and gNB, respectively
$\Pi^1_{n,k}$, $\Pi^2_{n,k}$, $\Pi^3_{n,k}$	Binary mode sub-channel assignment indicator of DM, RM, and LM respectively
$R_{n,k}$, $R_{m,k}$	Data rate for *nth* D2D pair and *mth* CU, respectively
$\mathcal{M}_D$, $\mathcal{M}_R$, $\mathcal{M}_L$	Mode selection sets of DM, RM, and LM respectively
$\mathcal{Q}^D$, $\mathcal{Q}^R$, $\mathcal{Q}^L$	DM, RM, and LM, each has respective quotas
ℓ^p, ℓ^s	Primary and secondary preference lists, respectively
$\mathcal{s}_t$, $\mathcal{a}_t$, $\mathcal{R}_t$	State, action, and reword of the SARSA algorithm
$\mathcal{s}_{t+1}$, $\mathcal{a}_{t+1}$	Next state and next action of the SARSA algorithm
α	The learning rate
$\mathcal{Y}$	The discount factor
$\mathcal{V}^*(\mathcal{s})$	The value function

Let $P_n = \{p_1, p_2, \ldots, p_n, \ldots, p_N\}$, P_C, P_Z denote the transmission power of the T_n, CU, gNB, respectively. The transmitted power for every D2D pair P_n is determined from the available power levels set, ranging from p_{min} to p_{max}, while the transmission power of CUs is supposed to be fixed. In the proposed scheme, Π is the binary mode sub-channel assignment indicator matrix with $\Pi \in \{0,1\}$, where $\Pi^1_{n,k}$, $\Pi^2_{n,k}$, and $\Pi^3_{n,k}$ represent the binary mode sub-channel assignment indicators for DM, RM, and LM, respectively. If the *nth* D2D pair is utilizing the sub-channel k of the CU m, then $\Pi = 1$, otherwise, $\Pi = 0$. The SINR of the *nth* D2D pair utilizing *kth* shared sub-channel of the *mth* CU at time slot t is given as follows:

$$\Upsilon_{n,k}(t) = \frac{P_n G_{TR}(t)\Pi^1_{n,k} + (P_n G_{TRU}(t) + P_n(t)G_{RUR}(t))\Pi^2_{n,k} + (P_n G_{TZ}(t) + P_Z G_{ZR}(t))\Pi^3_{n.k}}{\mathcal{I}_{D2D} + P_C(t)G_{mn} + \sigma} \quad (1)$$

where G_{TR}, G_{TRU}, G_{TZ}, represent the channel gain between T_n to R_n, RU_n, and gNB, respectively. G_{RUR} and G_{ZR} refer to the channel gain between RU_n and gNB to R_n. Furthermore, G_{mn} refers to the channel gain of the D2D pair *nth* and the *mth* CU. σ denotes the noise power. The expression $\mathcal{I}_{D2D}$ can be illustrated as follows:

$$\mathcal{I}_{D2D} = \mathcal{I}_D + \mathcal{I}_R + \mathcal{I}_L, \quad (2)$$

where $\mathcal{I}_D$, $\mathcal{I}_R$, and $\mathcal{I}_L$ demonstrate the impact of the interferences on the *nth* D2D pair in DM, RM, and LM caused by gNB, other D2D T_n, and other RU_n. The mathematical representation of $\mathcal{I}_D$, $\mathcal{I}_R$, and $\mathcal{I}_L$ can be shown as follows:

$$\mathcal{I}_D = P_Z G_{ZR}(t) \Pi^3_{n,k} + \sum_{n=1}^{N} P_n G_{T'R} \Pi^1_{n,k} + \sum_{n=1}^{N} P_n G_{RUR}(t) \Pi^2_{n,k}, \tag{3}$$

$$\begin{aligned} \mathcal{I}_R = (P_Z G_{ZRU}(t) + P_Z G_{ZR}(t))\, \Pi^3_{n,k} + \sum_{n=1}^{N} (P_n G_{T'RU}(t) + P_n G_{T'R}(t))\, \Pi^1_{n,k} \\ + \sum_{n=1}^{N} (P_n G_{RU'RU}(t) + P_n G_{RU'R}(t)) \Pi^2_{n,k}, \end{aligned} \tag{4}$$

$$\mathcal{I}_L = \sum_{n=1}^{N} (P_n G_{T'Z}(t) + P_n G_{T'R}(t)) \Pi^1_{n,k} + \sum_{n=1}^{N} (P_n G_{RU'Z}(t) + P_n G_{RU'R}(t)) \Pi^2_{n,k}, \tag{5}$$

where $G_{T'R}$, $G_{T'RU}$, and $G_{T'Z}$ are the channel gain between other D2D transmitters to R_n, RU_n, and gNB, respectively. Moreover, $G_{RU'R}$, $G_{RU'RU}$, and $G_{RU'Z}$ represent the channel gain between other D2D relays to R_n, RU_n, and gNB, respectively. The SINR of the *mth* CU utilizing the *kth* sub-channel at time slot t may be expressed as follows:

$$\Upsilon_{m,k}(t) = \frac{P_C(t) G_{mn}}{\sum_{n=1}^{N} p_n(t) G_n + \sum_{n=1}^{N} p_n(t) G_{RUm} + \sigma}, \tag{6}$$

The data rate of D2D pair n utilizing the uplink *kth* sub-channel can be determined at time slot t as follows:

$$R_{n,k}(t) = W\, log_2(1 + \Upsilon_{n,k}(t)), \tag{7}$$

Furthermore, the data rate of CU m utilizing the *kth* channel can be determined at time slot t as follows:

$$R_{m,k}(t) = W\, log_2(1 + \Upsilon_{m,k}(t)), \tag{8}$$

SE shows the effectiveness of using the available spectrum in terms of the data rate obtained regarding a given bandwidth. Thus, the SE for the *nth* D2D communication pair can be given as follows:

$$SE_{n,k}(t) = \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} R_{n,k}(t)}{W}, \tag{9}$$

Based on the obtained data rate and energy consumption, the EE of the D2D communications scheme at time slot t is given as follows:

$$EE_{n,k}(t) = \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} R_{n,k}(t)}{\sum_{n=1}^{N} p_n(t) + p_{cir}}, \tag{10}$$

where p_{cir} denotes the D2D pair circuit power consumption.

5. Problem Formulation

In this study, an optimization problem in D2D networks is investigated, specifically concentrating on joint mode selection, channel assignment, and power allocation optimization issues. In our proposed system, D2D users can choose among the available three D2D modes, including DM, RM, or LM based on maximum SINR. Moreover, a network that has been completely loaded is regarded as having no dedicated channels for D2D pairs to utilize. Moreover, the optimum power level can be obtained from the range P_{min} to P_{max}. This paper aims to optimize the sum data rate, SE, and EE of the proposed D2D

communications scheme while guaranteeing the QoS demands for both D2D pairs and CUs. The following is the formulation of the optimization problem:

$$\max_{\mathscr{M},\Pi,P} \sum_{n=1}^{N} \sum_{k=1}^{K} R_{n,k}(t), SE_{n,k}(t), \text{ and } EE_{n,k}(t), \tag{11}$$

s.t.

$$\sum_{n=1}^{N} \sum_{k=1}^{K} R_{n,k}(t) \geq R_{min}^{th}(t), \qquad \forall k \in 1 \ldots K, \tag{11a}$$

$$\sum_{n} \mathscr{M} = 1 \qquad n \in N, \tag{11b}$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} \Pi_{n,k}^{1}, \Pi_{n,k}^{2}, \text{ or } \Pi_{n,k}^{3} \leq K \qquad \forall k \in 1 \ldots K, \tag{11c}$$

$$\sum_{m} \Pi_{m} = 1 \qquad m \in UEs, \tag{11d}$$

$$\sum_{n=1}^{N} \Pi_{k,n} \leq 3, \qquad \forall n \in N, \tag{11e}$$

$$P_{min} \leq P_{n} \leq P_{max}, \qquad \forall N, \quad RU, \tag{11f}$$

$$P_{C}, P_{Z} = P_{max}, \tag{11g}$$

Constraint (11a) specifies the minimal data rate for the *nth* D2D pair in shared sub-channel k. Constraint (11b) denotes that each D2D pair n chooses one mode among the D2D modes including DM, RM, or LM. Constraint (11c) indicates that the binary mode sub-channel indicator matrix for each D2D including DM, RM, and LM is equivalent to or less than the total number of sub-channels k. Moreover, the constraint (11d) indicates that every cellular user m utilizes a distinct sub-channel k. The constraint (11e) indicates that each sub-channel may be utilized a maximum of three times. The constraints (11f) and (11g) denote that T_n, CUs, and gNB utilize specific transmission power.

To sufficiently address the optimization problem expressed in (11), it should be divided into two sub-issues: joint mode selection and channel assignment, as well as power management. Since it is an MINLP problem, the optimization problem is NP-hard and involves computational difficulties.

6. Proposed Joint Mode Selection and Resource Allocation Scheme (SD-Scheme)

In the present part, we introduce an SD-scheme underlaying cellular networks. Joint mode selection, channel assignment, and power allocation are considered with the aim of optimizing the sum data rate, SE, and EE. First, the mode selection issue is tackled by employing a greedy algorithm based on maximum SINR to select the optimum mode among DM, RM, and LM for every D2D pair. After that, the matching algorithm is implemented to tackle the problem of sub-channel assignment by exploiting the two-sided preference lists to optimize the utilization of spectrum resources. Finally, the power allocation issue is solved by introducing SARSA-based RL to obtain the optimum power for each D2D pair in the proposed scheme.

Several important factors inspired the decision to choose the SARSA algorithm for this study. Firstly, the modeling of the network in the complicated and proposed scenario of resource allocation for D2D communication is simply unpracticable. Therefore, the model-free feature of this system is particularly advantageous in this particular scenario. Moreover, the SARSA algorithm is highly applicable to decentralized decision-making and

enables agents to self-sufficiently learn the most optimal policies, which aligns effectively with the architecture of the proposed network. The proposed approach utilizes an SARSA algorithm to train and update the power level states by considering the environment feedback information.

6.1. Mode Selection and Channel Assignment Scheme (C-Scheme)

Firstly, this study examines a mode selection technique for D2D communication by employing a greedy algorithm. The method focuses on direct, relay, and local route D2D modes to enhance the performance of the D2D scheme. The mode selection issue is tackled based on the maximum SINR to calculate the best mode for each D2D pair, while considering the distance between the transmitter and the receiver. This will guarantee the chosen mode optimizes signal quality while taking into account the physical relationship of the communication devices, thereby maximizing overall performance. However, a threshold level is considered for the distance between the transmitter and the receiver for each D2D communication in the proposed network.

Let us suppose that $\mathscr{M}$ is a set of 0 and 1 elements that are applied to represent which mode is chosen. The following formula is applied to select the best mode:

$$\mathscr{M}_{D2Dn} = \mathscr{M}_D + \mathscr{M}_R + \mathscr{M}_L, \tag{12}$$

where $\mathscr{M}_D$, $\mathscr{M}_R$, and $\mathscr{M}_L$ represent the mode selection sets of DM, RM, and LM, respectively. If the data rate in a DM is greater than in an RM and an LM for each given n D2D communication pair, then $\mathscr{M}_D = 1$ and $\mathscr{M}_R = \mathscr{M}_L = 0$, and similarly for other cases. While a value in the set of the mode selection $\mathscr{M}_{D2Dn}$ is 1, the associated mode is selected ($\mathscr{M}_D$, $\mathscr{M}_R$, or $\mathscr{M}_L$), subsequently adding that particular data rate of that D2D pair n to the sum data rate of the network. Conversely, when a value in the mode selection set $\mathscr{M}_{D2Dn}$ is valued as 0, it leads to a missing contribution to the sum data rate of that particular D2D pair n.

Once the optimal D2D mode is determined in a scenario including DM, RM, and LM, the gNB employs the matching method to assign optimum reused sub-channels for the D2D pairs, which increases the spectrum utilization in the proposed network. This part introduces the model of a channel assignment issue to optimize the sum data rate and SE accordingly. We define the channel assignment formula in which D2D pair n shares the sub-channel k with CU m at time slot t as follows:

$$\Pi = \left(\Pi^1_{n,k}, \Pi^2_{n,k}, \Pi^3_{n,k}\right)_{N \times 3K}, \tag{13}$$

The vectors $\Pi^1_{n,k}$, $\Pi^2_{n,k}$, and $\Pi^3_{n,k}$ indicate the possibility that the D2D pair is assigned to DM, RM, or LM, using the shared sub-channel k with CUs. For each D2D mode in the proposed scheme, the quota ($\mathbb{Q}$) can be determined. $\mathbb{Q}$ represents the threshold of mode-channel assignment for every D2D pair within the framework. The $\mathbb{Q}$ features can be illustrated as follows:

- $\sum_{k=1}^{K} \sum_{n=1}^{N} \Pi^{1}_{n,k} \leq \mathbb{Q}^D, \quad \forall n \in \mathscr{M}^D$
- $\sum_{k=1}^{K} \sum_{n=1}^{N} \Pi^{2}_{n,k} \leq \mathbb{Q}^R, \quad \forall n \in \mathscr{M}^R$
- $\sum_{k=1}^{K} \sum_{n=1}^{N} \Pi^{3}_{n,k} \leq \mathbb{Q}^L, \quad \forall n \in \mathscr{M}^L$

Based on the above criteria, it is crucial that the overall set of D2D pairs across the modes, that are assigned to sub-channel k, should not exceed $\mathbb{Q}$. In the matching game, the establishment of the two-sided preference lists can be represented as follows:

$$\ell^p = \left\{R^D_{n,k}(t), R^R_{n,k}(t), R^L_{n,k}(t)\right\}, \tag{14}$$

$$\ell^s = \{P_C(t)G_{nm}\}, \tag{15}$$

where ℓ^p and ℓ^s denote preference lists that consist of D2D pairs arranged in descending sequence according to their highest data rate and CUs organized in ascending sequence according to their minimal interference effect, respectively.

While performing the matching game theory, every D2D pair within the cell is proposed to earn the sub-channels with its higher priority. Consequently, the gNB admits the D2D pairs with the highest priority while refusing the remaining pairs. For further clarification, if the D2D pair n gives a proposal to pick sub-channel k based on its greatest utility in ℓ^p, then sub-channel k is subsequently assigned to the exact D2D pair n according to the least interference impact utility function in ℓ^s. Moreover, the matching process continues till all devices in the network are paired to enhance the system performance.

To provide a more detailed explanation of the resource allocation process, we now describe the matching game-based sub-channel assignment in greater depth.

Matching Process Execution:

- Each D2D pair initially proposes to the sub-channel that provides the highest data rate based on its preference list.
- The gNB evaluates all proposals and initially assigns sub-channels to D2D pairs while ensuring that the total number of assigned pairs does not exceed the predefined threshold $\mathcal{Q}$.
- If a sub-channel receives multiple proposals, the gNB selects the D2D pairs that maximize SE and rejects lower priority requests.
- Rejected D2D pairs then propose to their next preferred sub-channel, and this process iterates until a stable matching is achieved, meaning no further changes can improve the overall network performance.

This iterative matching ensures an efficient and interference-aware sub-channel allocation strategy that enhances both spectral efficiency and system stability. The complete mode selection and channel assignment approach is detailed in Algorithm 1.

Algorithm 1. C-scheme algorithm

Input: M, N, K
Output: $\Pi_{n,k}$
Initialization $\Pi_{n,k} = \text{zeros}(N, 3\text{K})$
for 1 to M
determine $\Upsilon_{m,k}(t)$
determine $R_{m,k}(t)$
end for
for 1 to N
calculate $\Upsilon_{n,k}(t)$
calculate $R_{n,k}(t)$, $SE_{n,k}(t)$, $EE_{n,k}(t)$
end for
find $\mathcal{M}_{D2Dn}$ based on maximum $\Upsilon_{n,k}(t)$
find Q from the matrix $\mathcal{M}_{D2Dn}$
for 1 to N
for 1 to K
calculate $R^D_{n,k}(t)$, $R^R_{n,k}(t)$, $R^L_{n,k}(t)$ with regards to $\mathcal{M}_{D2Dn}$
calculate the interference impact of CUs
end for

Algorithm 1. *Cont.*

17: end for
18: sort ℓ^p in descending order
19: sort ℓ^s in ascending order
20 the most preferred sub-channel k is matched by D2D pair n based on ℓ^p and ℓ^s
21: if $R_{n,k} \geq R_{th}$
22: set $\Pi_{n,k} = 1$
23: else
24: set $\Pi_{n,k} = 0$
25: end if

6.2. Proposed RL-Based Power Allocation Scheme

The power allocation optimization can be achieved using a dynamic distribution scheme. A wireless network with real-time communications demands immediate training and learning of the D2D pairs to provide distributed power allocation without imposing a significant load over the gNB. Thus, ML provides a potential solution with a wide range of applications in the execution of dynamic resource allocation and tackles many challenges associated with prospective communications networks. The transmitter of the D2D pair, performing as the intelligent agent in this scenario, has the ability to learn and make the most suitable decision to enhance the network performance.

One of the most advanced ML techniques is RL. RL utilizes an approach based on trial and error to determine the best resource allocation decisions. Moreover, RL works efficiently without any previous information about the system environment, in contrast to the traditional techniques. RL may enhance performance by facilitating rapid detection of optimum solutions or decisions in comparison with conventional centralized techniques.

SARSA is a reinforcement learning approach utilized to identify the best action in a dynamic resource allocation system. This research presents an SARSA-based approach to solving the issue of power distribution, including the following elements:

Agent: the agent is the D2D transmitter and serves as a crucial element in the power allocation issue.

State: The state of the SARSA algorithm includes essential network information, such as interference levels, channel conditions, and user location. These factors describe the present state of the environment, operating as inputs to an agent's decision-making in power distribution. In this case, the agent indicates the connection of D2D pairs.

Action: The action is an activity performed via the agent. The power distribution levels established by the D2D pairs constitute the action, which comprises a range of powers from P_{min} to P_{max}.

Reward: The reward function in the SARSA-based power management approach is defined as the EE of each D2D pair within the system. The agent's interactions with the environment in SARSA are illustrated in Figure 2.

In the specific state $\mathcal{s}_t$, the action $\mathcal{a}_t$ is selected, and the reward $\mathcal{R}_t$ is allocated to the D2D pair (agent) for each action performed. The agent subsequently transitions to the newly created state $\mathcal{s}_{t+1}$ and executes another action, $\mathcal{a}_{t+1}$ for its present state $\mathcal{s}_{t+1}$. Moreover, the pattern $\mathcal{s}_t$-$\mathcal{a}_t$-$\mathcal{R}_t$-$\mathcal{s}_{t+1}$-$\mathcal{a}_{t+1}$ defines the sequence of procedures for the suggested SARSA algorithm. The Q-value is firstly set to zero value, then the proposed algorithm modifies the Q-table in accordance with the current policy.

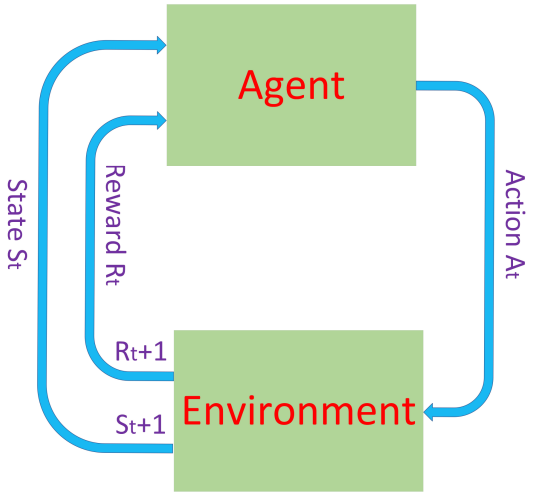

Figure 2. Agent–environment interactions in SARSA.

The algorithm shows the agent inside the SARSA framework has to accomplish several episodes. At each time step t, the agent in the state $\mathcal{s}_t$ chooses an action $\mathcal{a}_t$ based on greedy strategy. Afterwards, the agent receives the reward $\mathcal{R}_t$ and proceeds to the following state $\mathcal{s}_{t+1}$, where they choose the action $\mathcal{a}_{t+1}$ in dependence on the Q-table. The state–action equation can be represented as:

$$\mathcal{Q}(\mathcal{s}_t, \mathcal{a}) = (1 - \mathcal{a})\mathcal{Q}(\mathcal{s}_t, \mathcal{a}) + \alpha[\mathcal{R}_{t+1} + \mathcal{Y}\mathcal{Q}(\mathcal{s}_{t+1}, \mathcal{a}_{t+1})], \tag{16}$$

Here, α represents the learning rate of the agent, $\mathcal{R}_{t+1}$ denotes the reward function of the next state, and $\mathcal{Y}$ signifies the discount factor. In the framework of RL, the agent aims to optimize the reward by adopting an optimal policy. The optimum policy can be calculated through the Bellman formula:

$$\mathcal{V}^*(\mathcal{s}) = \max_{\mathcal{a} \in \mathcal{A}} \mathcal{Q}^*(\mathcal{s}, \mathcal{a}), \tag{17}$$

Subsequently, the value function is identified by the following equation:

$$\mathcal{V}(\mathcal{s}) = \max_{\mathcal{a} \in \mathcal{A}} \mathcal{Q}(\mathcal{s}, \mathcal{a}), \tag{18}$$

The given equation is utilized to identify the optimal action value in order to optimize $\mathcal{Q}(\mathcal{s}_t, \mathcal{a}_t)$ for each state involved.

$$\mathcal{a} = \arg\max_{\mathcal{a} \in \mathcal{A}} \mathcal{Q}(\mathcal{s}, \mathcal{a}), \tag{19}$$

To decide what action $\mathcal{a}$ is going to be chosen during a particular time t, the Exploration and Exploitation Policy (EEP) feature is employed as follows:

$$a_t = \begin{cases} \arg\max_{\mathcal{a} \in \mathcal{A}} \mathcal{Q}(\mathcal{s}, \mathcal{a}) & exploitation \\ rand_{\mathcal{a} \in \mathcal{A}}(\mathcal{a}) & exploration \end{cases}, \tag{20}$$

In Equation (20), the 'ε − greedy' strategy is used while performing EEP, indicating the fact that the probabilities of exploitation and exploration are ε and $1 - \varepsilon$, accordingly. Furthermore, a Markov Decision Process model of D2D communications underlaying cellular networks is provided to allocate power for each D2D pair using SARSA-based RL. Moreover, the actions of the agents, which are represented as $\mathcal{a}$, include a set of transmission powers, are indicated by P_n, and are allocated to the D2D transmitters T_n.

The reward function is formulated based on EE for each D2D pair n employing sub-channel k as follows:

$$R_n = \begin{cases} EE_{n,k}(\sqrt{R_{n,k}(t) - R_{min}^{th}}) & if\ R_{n,k}(t) \geq R_{min}^{th} \\ EE_{n,k}(-e^{(R_{n,k}(t) - R_{min}^{th})}) & if\ R_{n,k}(t) < R_{min}^{th} \end{cases}, \quad (21)$$

Exploration involves a comprehensive examination of the network, gathering information, and randomly selecting actions to evaluate their efficiency. However, exploitation exploits the advantage of previous decisions according to the Q-table. Exploration and exploitation have trade-off features in power distribution techniques that utilize RL.

SARSA is more appropriate for our case than Q-learning because of its on-policy nature, which updates action values according to the current policy, hence offering a more adaptive response to network changes. This functionality is especially beneficial in D2D communication systems, where interference levels and network topology may vary. The Q-learning method, while efficient in static situations, fails to update according to real-time rules, potentially limiting its flexibility in dynamic contexts. Our investigation demonstrates that SARSA surpasses Q-learning in EE and power consumption, particularly under changing interference levels, highlighting its enhanced capability to handle real-time interference in high-density networks. Algorithm 2 defines the extensive structure of the SARSA strategy.

Algorithm 2. SARSA algorithm for power allocation issue

Initialize: N, M, P_{max}, P_{min}, $\Pi_{n,k}$, $\mathcal{Q}(\mathcal{s}, \mathcal{a})$ table, $\mathcal{Y}$, ε, α
For episode $\in \{1, \ldots, EP\}$ do
Reset $\mathcal{s}$, $t = 0$
Select the level of power between P_{max} and P_{min}, utilizing policy derived Q (ε-greedy)
For $t \in \{0, \ldots, T-1\}$ do
Every agent performs an action $\mathcal{a} \in A$ as well as observes $\mathcal{R}$ and $\mathcal{s}_{t+1}$
Check (11a), (11f), and (11g)
If the conditions are satisfied, then
Establish action $\mathcal{a} \in A$, $\mathcal{R}$, $\mathcal{s}_{t+1}$
End if
Each agent takes an action $\mathcal{a}_{t+1} \in A$ and observes $\mathcal{R}$ and $\mathcal{s}_{t+1}$
Update the Q-table
$S \leftarrow \mathcal{s}_{t+1}$, $A \leftarrow \mathcal{a}_{t+1}$
End until all D2D pairs connect or the total iteration numbers is reached
End for
Output: optimum power for each D2D pair

7. Simulation Results

This part demonstrates the performance evaluation of the proposed SD-scheme regarding sum data rate, EE, SE, outage probability, and power saving. The SD-scheme that effectively joins the centralized mode selection and channel assignment scheme (C-scheme) is introduced based on greedy and matching algorithms, respectively, with the distributed power allocation scheme based on the SARSA algorithm. The suggested approach is compared with multiple traditional schemes, including a channel allocation scheme based on matching theory with the goal of optimizing EE [24], a channel allocation scheme based on the greedy algorithm [30], and power allocation schemes based on Q-learning in [17,23,29]. Table 2 contains the parameters employed in the simulation.

Table 2. Simulation parameters.

Parameter	Value
Entry 1	data
Cell radius	500 m
CUs number	20
D2D pairs number	2–38
Transmitter to receiver distance	10–100 m
Total Bandwidth	5 MHz
Noise power	−174 dBm//Hz
Sub-channel bandwidth	240 KHz
D2D transmitter power	0–23 dBm
gNB transmitting power	35 dBm
SINR threshold	3 dBm
Shadowing standard deviation	8 dB
Circuit power	0.1 watt
CU transmitting power	23 dBm
Discount factor	0.9
Learning rate	0.2
Epsilon	0.2

Figure 3 shows the EE comparison of the proposed SD-scheme with the suggested C-scheme, a channel allocation scheme based on matching theory with the goal of optimizing EE [24], a traditional channel allocation scheme based on the greedy algorithm [30], and power allocation schemes based on Q-learning-RL in [17,23,29] versus the number of D2D pairs. The EE rises with the incremental in the number of D2D pairs. As illustrated in Figure 3, the introduced approach demonstrates performance superiority and outperforms the benchmark schemes. The substantial boost in EE of the suggested technique demonstrates its efficacy in managing resources with the increase in D2D pairs, resulting in allocating optimal transmission power to each pair based on its requirements. Conversely, conventional systems provide limited enhancements, stabilizing at lower EE values as the number of D2D pairs escalates. The papers [24,30] confirm the worst EE since these schemes utilize conventional centralized approaches which lead to high control overhead, delayed or inflexible decisions, and insufficient optimum power allocation. In contrast, the proposed SARSA-based RL outperformed the Q-learning algorithm schemes [17,23,29], especially, where the pair number equals 10 and higher. The reason is that Q-learning often exhibits excessive exploration in some scenarios, strongly seeking optimum behaviors that enhance throughput, perhaps resulting in increased energy consumption. This conduct may adversely affect EE, particularly in D2D networks where sustaining low consumption of energy is essential. Furthermore, SARSA offers a more appropriate balance between exploration (engaging in new actions) and exploitation (utilizing the currently optimal action). This balanced approach in EE guarantees that power consumption is low during long exploratory stages since the SARSA strategy is continually adjusted depending on actual performance.

Figure 4 demonstrates the effectiveness of the proposed SD-scheme on total power saving versus the D2D pair numbers in the network. It is obvious that the power saving of the suggested scheme and the benchmark schemes increase as the number of D2D pairs increases. When the number of D2D pairs increases, the total power saving obtained by the suggested technique shows a constant and substantial enhancement, especially at the highest D2D pairs number. The introduced scheme outperforms the conventional systems as shown in the figure. The reason is that the SARSA algorithm is on-policy learning that responds to varying network situations instantaneously. Moreover, the introduced SARSA algorithm optimized the power transmission for every D2D pair continually in response to

the channel conditions and interferences established by the CUs or other D2D pairs inside the cell.

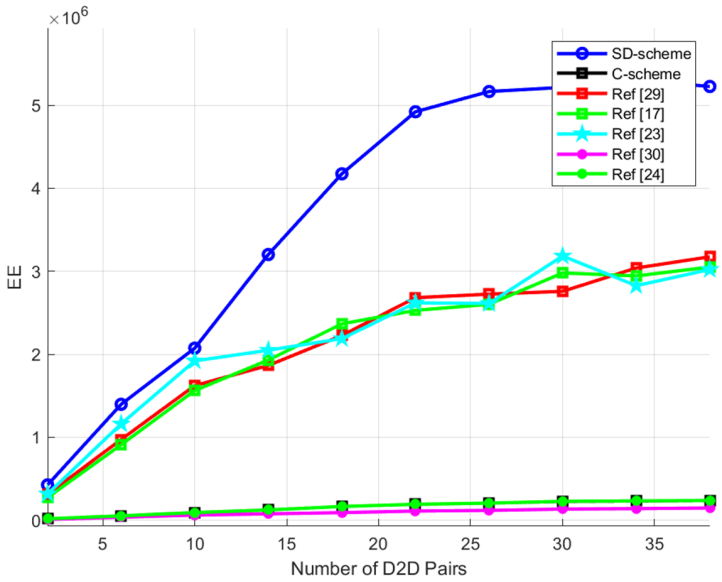

Figure 3. EE versus number of D2D pairs [17,23,24,29,30].

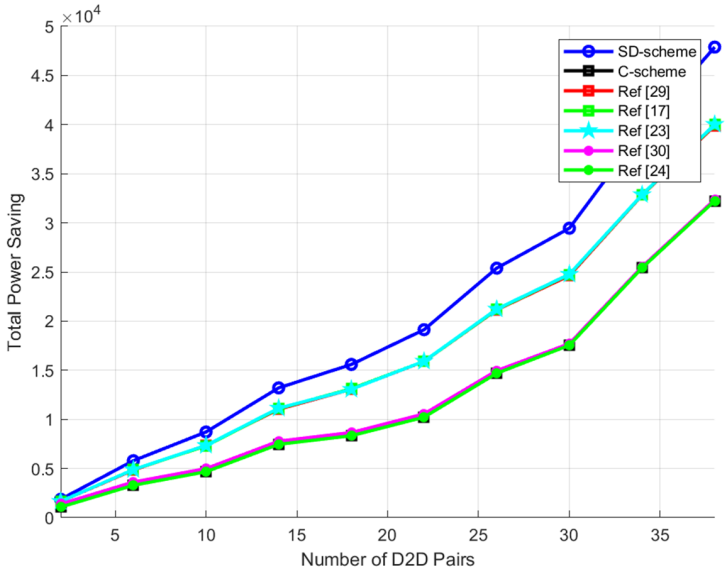

Figure 4. Total power saving versus number of D2D pairs [17,23,24,29,30].

Figure 5 compares the sum data rate of the SD-scheme to benchmark algorithms with various numbers of D2D pairs. The increment in the sum data rate is proportional to the D2D pair increase. It is clear that the [24,30] approaches outperform the proposed scheme in terms of sum data rate since these approaches employ maximum transmission power which results in reducing EE in the system. Moreover, in comparison to the benchmark schemes, the SD-scheme provides a higher sum data rate. The reason is that optimizing mode selection and channel assignment mitigates the effects of co-tier interference resulting from spectrum sharing between CUs and other D2D pairs. Consequently, the sum data rate of the network is enhanced accordingly. Clearly, the sum data rates increase slightly when the number of D2D pairs is between 25 and 40, due to the second reuse of the CUs' spectrum that increases the interference in the network. Furthermore, the suggested strategy enhances network performance by effectively balancing the utilization of resources and interference mitigation.

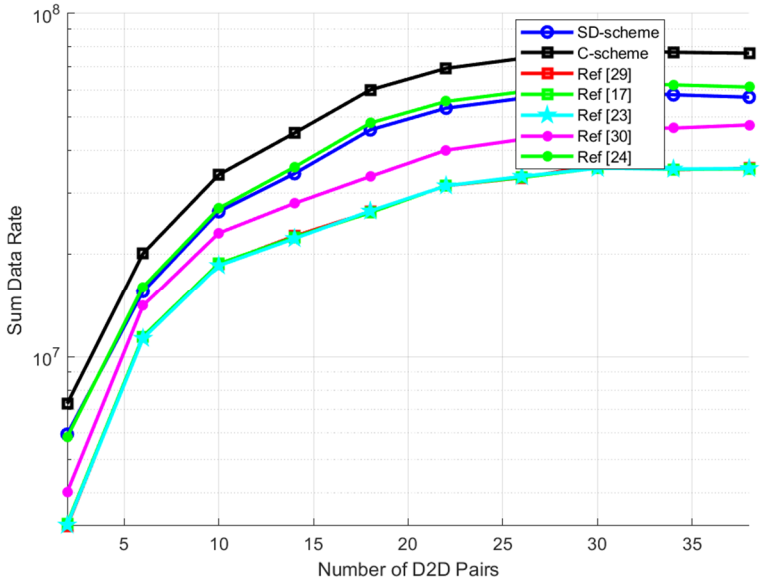

Figure 5. Sum data rate versus the number of D2D pairs [17,23,24,29,30].

Table 3 provides an overview of performance values, particularly EE, total power savings, and sum data rate, to elucidate the benefits of the SD-scheme over the C-scheme. The findings clearly indicate that the SD-scheme achieves better performance based on elevated EE, total power savings, and sum data rate, demonstrating its efficacy in resource allocation for D2D communications within cellular networks.

Table 3. Performance values.

No. of Pairs	EE (Bit/Sec/Watt)		Total Power Saving (Watt)		Sum Data Rate (Bit/Sec)	
	SD-Scheme	C-Scheme	SD-Scheme	C-Scheme	SD-Scheme	C-Scheme
5	1.12	0.71	0.49	0.27	1.3×10^7	2.0×10^7
10	2.06	0.95	0.91	0.49	2.6×10^7	3.4×10^7
15	3.45	1.54	1.40	0.77	3.8×10^7	4.9×10^7
20	4.52	2.12	1.74	0.93	4.9×10^7	6.4×10^7
25	5.14	2.93	2.43	1.37	5.5×10^7	7.2×10^7
30	5.21	3.15	2.93	1.74	5.9×10^7	7.8×10^7
35	5.25	3.40	4.22	2.72	5.8×10^7	7.8×10^7

Figure 6 shows the SE evaluation of the SD-scheme with the benchmark schemes in relation to the D2D pair numbers. The term SE rises incrementally with the number of D2D pairs since the sub-channel reuse indicator improves correspondingly with the increase in D2D pairs. Consequently, the co-tier interferences among shared channels of CUs increase. The SD-scheme demonstrates superior efficiency relative to other approaches. This approach demonstrated better SE when the number of D2D pairs ranged from 2 to 25, related to the minimal interference across the shared channels during the single reuse of the CUs' uplink channel. Moreover, the SE decreases in cases where there are 25–40 D2D pairs due to multiple times reusing the uplink spectrum of CU resources. When the transmitted power of the D2D pair is set to the maximum value, an increased data rate and SE are obtained due to a strong signal which results in increasing the interference and decreasing EE accordingly. The significant efficacy of the proposed strategy is attributed to its adaptive resource allocation mechanism, which responds flexibly to changing network situations. This flexibility ensures potential improvements in SE, which is especially important in heavily loaded D2D environments where interference management is critical.

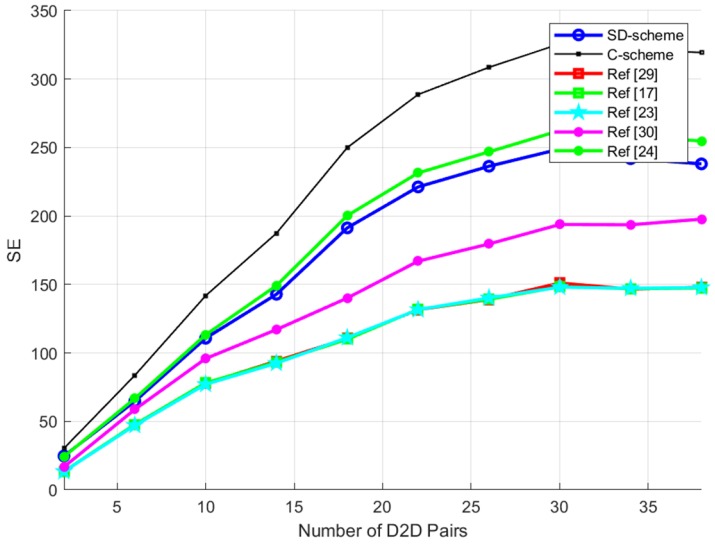

Figure 6. SE versus the number of D2D pairs [17,23,24,29,30].

Figure 7 shows the comparison of the outage probability with various numbers of D2D pairs. The outage probability of the SD-scheme increases in accordance with the number of D2D pairs because of the increased network interferences. Nonetheless, the suggested method and other channel-optimized algorithms reduce the probability of outages by ensuring the most efficient utilization of the resources in order to meet the minimal QoS demands for each D2D pair. By optimizing reused channels, the suggested system efficiently allows multiple D2D pairs using one particular CU's sub-channel. Despite the number of D2D pairs growing, this approach maintains low outage probability and reduced interference. Because of the unmanaged interferences of the resources that are shared among D2D pairs and between D2D pairs and CUs from the other side in benchmark methods, the outage probability significantly increases.

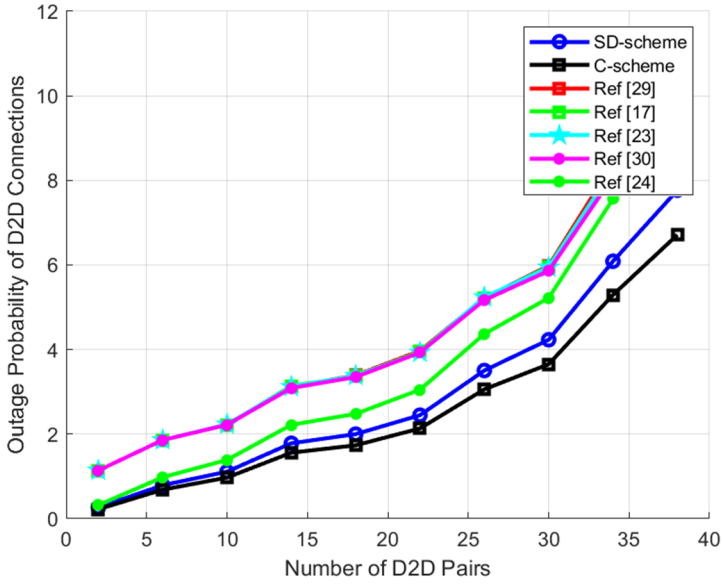

Figure 7. Outage probability versus the number of D2D pairs [17,23,24,29,30].

The EE analysis for varying transmission power of CUs is illustrated in Figure 8. It is clear that there are slight decreases in EE with the incremental increase in CUs transmitting power due to high interference created by CUs and the higher transmission power of D2D pairs required to satisfy QoS demands. The EE of the suggested SD-scheme demonstrates

better performance in comparison to traditional schemes. The reason is that a lack of scalability and flexibility with a higher power level can be observed in traditional systems like [24,30], which show better establishing EE but rapidly decrease as CU power rises. The suggested approach, on the other hand, demonstrates its better energy management and adaptation to varied network conditions by maintaining better EE despite changes in power levels.

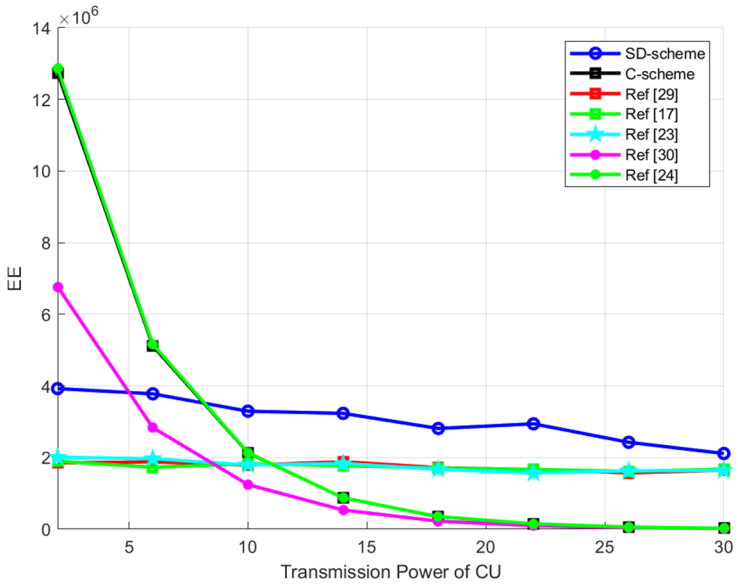

Figure 8. EE versus transmission power of CUs [17,23,24,29,30].

Figure 9 shows the sum data rate comparison of the suggested SD-scheme with traditional schemes versus the transmission power levels of CUs. The sum data rate demonstrates a slight decrease with the increase in CU transmission power. The first reason is due to the increased interference of D2D pairs that share the same spectrum of CUs. Secondly, the D2D pairs experience higher competition for limited resources in the network. The conventional approaches fail to efficiently allocate resources which results in low SINR and low sum data rate accordingly.

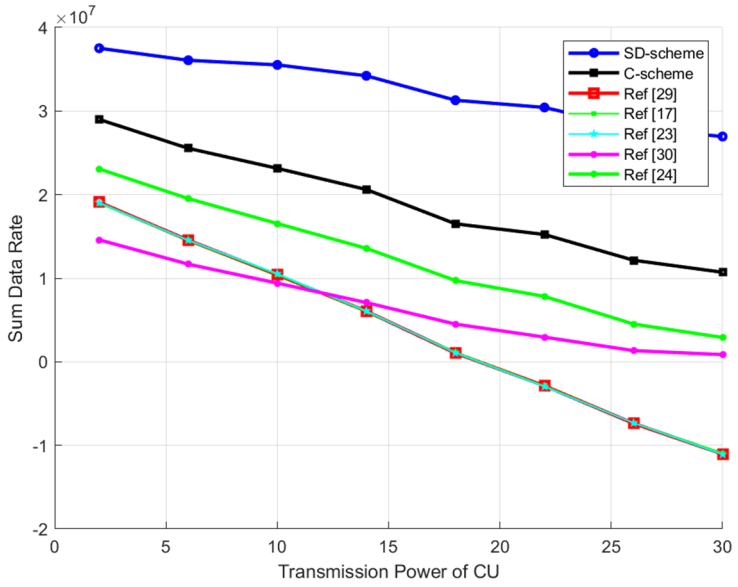

Figure 9. Sum data rate versus transmission power of CUs [17,23,24,29,30].

Figure 10 demonstrates the total power saving of the proposed SD-scheme over varying transmission power levels of CUs. It is obvious that the total power saving remains stable up to 15 dBm, after that, the total power saving increases sharply due to several reasons. Firstly, the pairs can barely perform with the same power level without any requirement to enhance the transmission power strategies. Furthermore, the interference from CUs substantially affects the D2D links; consequently, the D2D pairs tend to overcome the interference threshold. As illustrated in Figure 10, the proposed SD-scheme outperforms the benchmark approaches. The reason is that the proposed SARSA-based RL allocates optimum power for each D2D pair according to the immediate demands of users and ensures that the D2D pairs are utilizing the lowest possible power to prevent co-tier interference and guarantee communication between the transmitter and receiver for the D2D link.

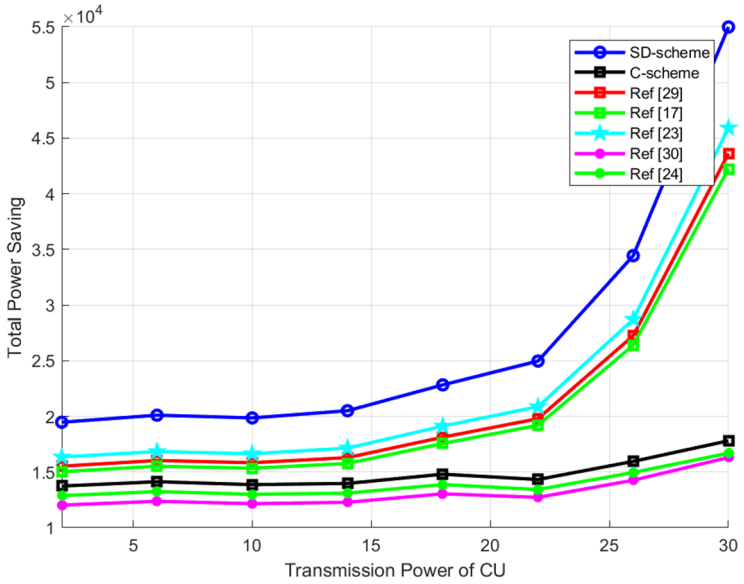

Figure 10. Power saving versus transmission power of CUs [17,23,24,29,30].

Figure 11 illustrates the SE different levels in relation to different available techniques versus the CU transmission power. When the transmission power of CUs escalates, the SE decreases to specific thresholds, except for the [24,30] methods, since these approaches employ maximum transmission power. The reason behind this decrease is the conflict of sub-channel reuse since the transmission power of CUs increases. Hence, the D2D pairs experience higher interference due to shared resources between them.

Figure 12 demonstrates the probability of connections for D2D modes including DM, RM, and LM versus the maximum distance of relays to D2D pairs. Regarding DM, the figure indicates a consistent probability with the variety of relay distances since the distance between relays and D2D pairs has no impact on direct D2D mode. Furthermore, the probability of connection of RM decreases with the incremental distance, highlighting the dependence of D2D pairs on the proximity of relays to establish efficient communication links. When the distance of relays increases, the D2D pairs tend to communicate with each other through LM. Consequently, LM increases with the increase in the distance between relays to D2D pairs.

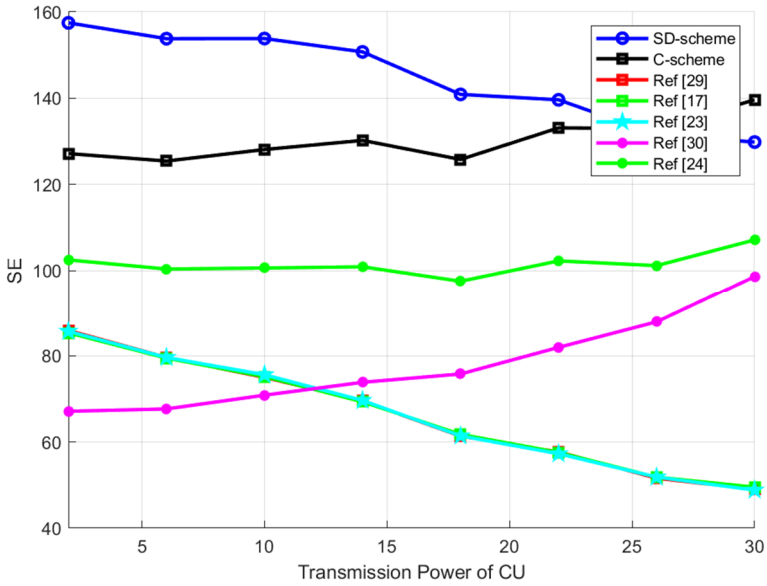

Figure 11. SE versus transmission power of CUs [17,23,24,29,30].

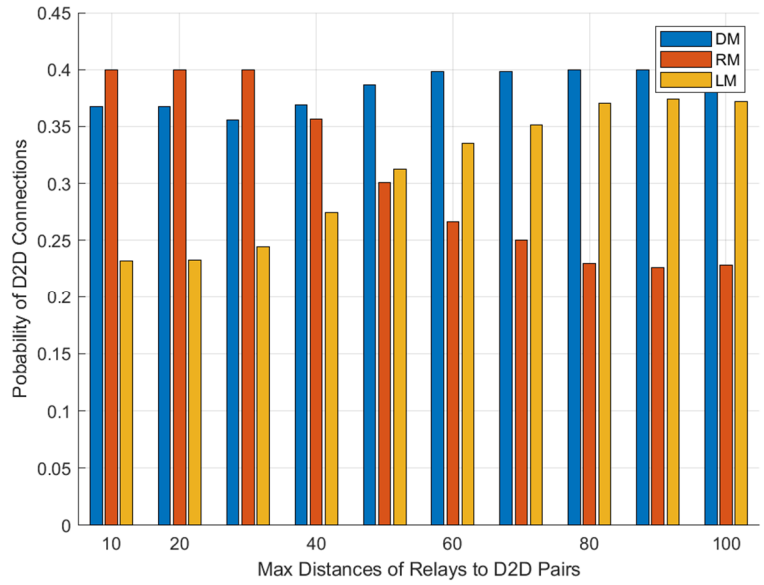

Figure 12. Probability of D2D connections versus maximum distance of relays to D2D pairs (m).

Figure 13 demonstrates the probability of connections of D2D pairs including DM, RM, and LM in comparison to the highest distance between the D2D pair. It is obvious that the DM reaches the highest probability of connections when the maximum distance across the D2D pair is limited. Furthermore, the probability of connections for DM decreases with the increase in distance since the DM mode depends on the proximity between transmitter and receiver, which allows the D2D pair to operate efficiently. On the other hand, the probability of connections for both RM and LM increases with the increase in the maximum distance between the D2D pair. The reason is that when the distance is medium, the pair tend to choose RM to maintain the QoS threshold. Moreover, when the distance is high, the pair prefers to choose the LM for the same previous reason since the power of the gNB is greater than that of the D2D transmitter. Consequently, the D2D pair chooses either DM, RM, or LM depending on the maximum SINR which is impacted by the distance.

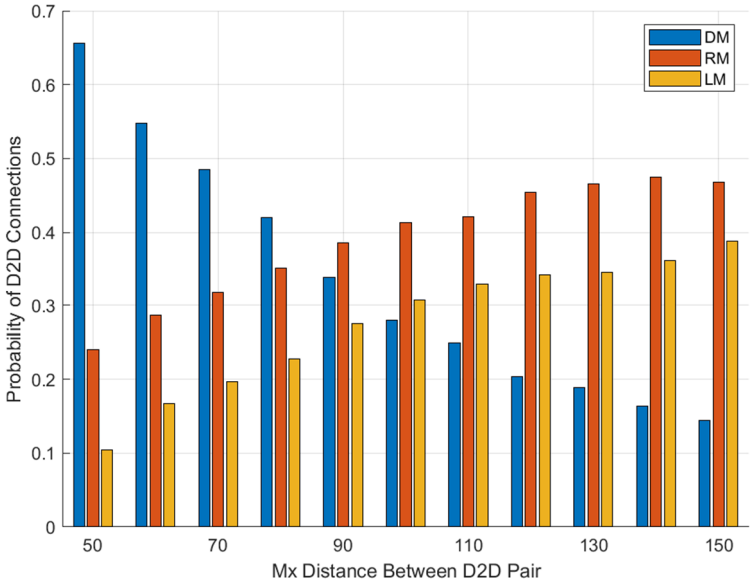

Figure 13. Probability of D2D connections versus max distance between D2D pairs.

Figure 14 demonstrates the number of disconnected pairs versus the number of D2D pairs in the network with the goal of comparing the performance of the proposed SD-scheme with the C-scheme. When the number of D2D pairs increases, the proposed schemes increase in terms of the number of disconnected D2D pairs because of the increased interference between D2D pairs and CUs. It is obvious that the proposed SD-scheme achieves slightly better performance than the C-scheme because of its better ability to mitigate interference and utilize resources effectively. The performance divergence across both schemes increases with the number of D2D pairs, indicating that the SD-scheme exhibits more adaptability and stability under higher network loads. This enhancement is due to the SD-scheme's optimized resource allocation technique, which reduces connection failures and guarantees superior connectivity for D2D users.

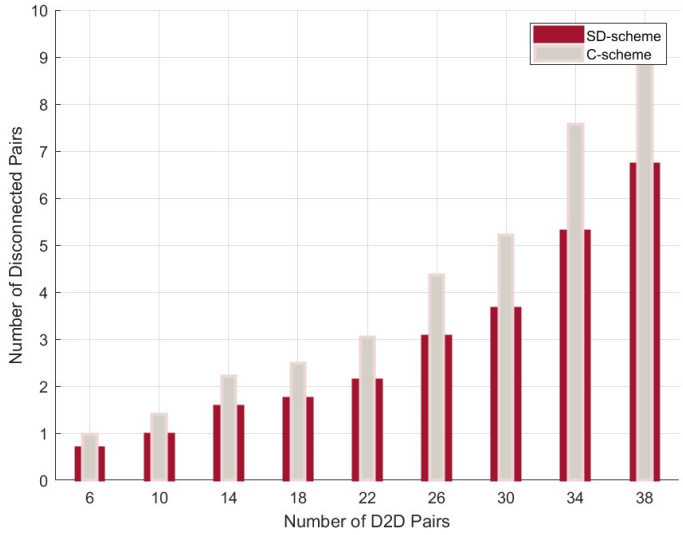

Figure 14. Number of disconnected pairs versus number of D2D pairs.

8. Conclusions

This paper presents an SD-scheme for mode selection, channel assignment, and power allocation in D2D communications underlaying cellular networks. The primary goal of this research is to enhance D2D performance while maintaining the QoS demands of CUs.

Furthermore, the sub-channel of each single CU can be reused across several D2D pairs, and the detrimental effects of the interference that occurs across D2D pairs are considered during the process of resource allocation. The joint problems of mode selection, channel assignment, and power allocation are MINLP and NP and difficult to resolve. Therefore, we designed a hybrid scheme: a centralized mode selection and channel assignment approach, followed by a distributed power management approach. The initial process involves the concept of a centralized greedy-based mode selection and two-sided preference lists channel assignment approach, followed by the implementation of a distributed power control approach in the subsequent phase. Moreover, an SARSA-based RL power control method has been proposed to iteratively update the transmission power for each D2D pair utilizing the same assigned sub-channel of individual CUs with the goal of improving the EE of the D2D communications in the network. The simulation findings demonstrated that the introduced scheme yields better performance with low complexity and outperforms traditional and Q-learning schemes in terms of data rate, SE, and EE. Future research may include examining the influence of other networking factors such as users' mobility as well as including modern equipment like unmanned aerial vehicles and satellites to the network.

Author Contributions: Conceptualization, N.M.M.; Data curation, N.M.M.; Formal analysis, I.S.A.; Investigation, I.S.A.; Methodology, I.S.A. and M.H.D.N.H.; Project administration, N.M.M.; Software, I.S.A.; Supervision, N.M.M.; Validation, M.H.D.N.H.; Writing—original draft, I.S.A.; Writing—review and editing, M.H.D.N.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created in this study. All relevant data are included within the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Siddiqui, M.U.A.; Qamar, F.; Tayyab, M.; Hindia, M.N.; Nguyen, Q.N.; Hassan, R. Mobility Management Issues and Solutions in 5G-and-Beyond Networks: A Comprehensive Review. *Electronics* **2022**, *11*, 1366. [CrossRef]
2. Alhashimi, H.F.; Hindia, M.N.; Dimyati, K.; Hanafi, E.B.; Safie, N.; Qamar, F.; Azrin, K.; Nguyen, Q.N. A Survey on Resource Management for 6G Heterogeneous Networks: Current Research, Future Trends, and Challenges. *Electronics* **2023**, *12*, 647. [CrossRef]
3. Gismalla, M.S.M.; Azmi, A.I.; Bin Salim, M.R.; Abdullah, M.F.L.; Iqbal, F.; Mabrouk, W.A.; Othman, M.B.; Ashyap, A.Y.I.; Supa'At, A.S.M. Survey on Device to Device (D2D) Communication for 5GB/6G Networks: Concept, Applications, Challenges, and Future Directions. *IEEE Access* **2022**, *10*, 30792–30821. [CrossRef]
4. Alibraheemi, A.M.H.; Hindia, M.N.; Dimyati, K.; Izam, T.F.T.M.N.; Yahaya, J.; Qamar, F.; Abdullah, Z.H. A Survey of Resource Management in D2D Communication for B5G Networks. *IEEE Access* **2023**, *11*, 7892–7923. [CrossRef]
5. Alibraheemi, A.M.H.; Hindia, M.N.; Izam, T.F.T.M.N.; Dimyati, K. Spectrum Efficient Mode Selection and Resource Allocation Optimization for D2D Communication in HetNet: A Multi-Agent Q-Learning Approach. *IEEE Access* **2024**, *12*, 131217–131229. [CrossRef]
6. Alhashimi, H.F.; Hindia, M.N.; Dimyati, K.; Hanafi, E.B.; Izam, T.F.T.M.N. Joint Optimization Scheme of User Association and Channel Allocation in 6G HetNets. *Symmetry* **2023**, *15*, 1673. [CrossRef]
7. Chen, C.-Y.; Sung, C.-A.; Chen, H.-H. Capacity maximization based on optimal mode selection in multi-mode and multi-pair D2D communications. *IEEE Trans. Veh. Technol.* **2019**, *68*, 6524–6534. [CrossRef]
8. Attar, I.S.; Mahyuddin, N.M.; Hindia, M.H.D.N. Joint mode selection and resource allocation for underlaying D2D communications: Matching theory. *Telecommun. Syst.* **2024**, *87*, 663–678. [CrossRef]
9. Jayakumar, S.; Nandakumar, S. A review on resource allocation techniques in D2D communication for 5G and B5G technology. *Peer-to-Peer Netw. Appl.* **2021**, *14*, 243–269. [CrossRef]
10. Gu, W.; Zhu, Q. Social-aware-based resource allocation for NOMA-Enhanced D2D communications. *Appl. Sci.* **2020**, *10*, 2446. [CrossRef]
11. Hassan, A.N.; Al-Chlaihawi, S.; Khekan, A.R. Artificial intelligence techniques over the fifth generation mobile networks. *Indones. J. Electr. Eng. Comput. Sci.* **2021**, *24*, 317–328. [CrossRef]

12. Alhashimi, H.F.; Hindia, M.N.; Dimyati, K.; Hanafi, E.B.; Tengku Mohmed Noor Izam, T.F. Reinforcement Learning Based Power Allocation for 6G Heterogenous Networks. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin, Germany, 2024; pp. 128–141. [CrossRef]
13. Sheng, J.; Liu, S.; Huang, T.; Wu, Y. Overlapping Coalition Game for Resource Allocation in Many-to-Many D2D Communication. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 1738530. [CrossRef]
14. Rosas, A.A.; Shokair, M.; Dessouky, M.I. Genetic Based Approach for Optimal Power and Channel Allocation to Enhance D2D Underlaied Cellular Network Capacity in 5G. *Comput. Mater. Contin.* **2022**, *72*, 3751–3762. [CrossRef]
15. Sun, Y.; Miao, M.; Wang, Z.; Liu, Z. Resource Allocation Based on Hierarchical Game for D2D Underlaying Communication Cellular Networks. *Wirel. Pers. Commun.* **2021**, *117*, 281–291. [CrossRef]
16. Dejen, A.A.; Wondie, Y.; Forster, A. Distributed Throughput and Energy Efficient Resource Optimization When D2D and Massive MIMO Coexist. *J. Commun. Inf. Netw.* **2022**, *7*, 278–295. [CrossRef]
17. Jiang, S.; Zheng, J. A Q-learning based Dynamic Power Control Algorithm for D2D Communication Underlaying Cellular Networks. In Proceedings of the 13th International Conference on Wireless Communications and Signal Processing, WCSP 2021, Changsha, China, 20–21 October 2021; IEEE: Piscataway, NJ, USA, 2021. [CrossRef]
18. Chang, H.-H.; Liu, L.; Bai, J.; Pidwerbetsky, A.; Berlinsky, A.; Huang, J.; Ashdown, J.D.; Turck, K.; Yi, Y. Resource Allocation for D2D Cellular Networks with QoS Constraints: A DC Programming-Based Approach. *IEEE Access* **2022**, *10*, 16424–16438. [CrossRef]
19. Xing, X.; Cao, J.; Zhou, H. Improving Quality of Service for Cell-Edge Users in D2D-Relay Networks. *Wirel. Pers. Commun.* **2022**, *126*, 1789–1804. [CrossRef]
20. Wei, Y.; Qu, Y.; Zhao, M.; Zhang, L.; Yu, F.R. Resource allocation and power control policy for device-to-device communication using multi-agent reinforcement learning. *Comput. Mater. Contin.* **2020**, *63*, 1515–1532. [CrossRef]
21. Wang, H.; Wang, Y.; Tang, L.; Xia, Y. D2D Social Selection Relay Algorithm Combined with Auction Principle. *Sensors* **2022**, *22*, 9265. [CrossRef]
22. Hamid, A.K.; Al-Wesabi, F.N.; Nemri, N.; Zahary, A.; Khan, I. An optimized algorithm for resource allocation for D2D in heterogeneous networks. *Comput. Mater. Contin.* **2022**, *70*, 2923–2936. [CrossRef]
23. Jiang, F.; Zhang, L.; Sun, C.; Yuan, Z. Clustering and resource allocation strategy for D2D multicast networks with machine learning approaches. *China Commun.* **2021**, *18*, 196–211. [CrossRef]
24. Awad, M.K.; Baidas, M.W.; El-Amine, A.A.; Al-Mubarak, N. A matching-theoretic approach to resource allocation in D2D-enabled downlink NOMA cellular networks. *Phys. Commun.* **2022**, *54*, 101837. [CrossRef]
25. Gao, J.; Meng, X.; Yang, C.; Zhang, B.; Yi, X. Resource Allocation for D2D Communication Underlaying Cellular Networks: A Distance-Based Grouping Strategy. *Wirel. Commun. Mob. Comput.* **2023**, *2023*, 8594323. [CrossRef]
26. Pei, E.; Zhu, B.; Li, Y. A Q-learning based Resource Allocation Algorithm for D2D-Unlicensed communications. In Proceedings of the IEEE Vehicular Technology Conference, Helsinki, Finland, 25–28 April 2021. [CrossRef]
27. El-Nakhla, O.M.; Obayya, M.I.; Kishk, S.E. Stable Matching Relay Selection (SMRS) for TWR D2D Network With RF/RE EH Capabilities. *IEEE Access* **2022**, *10*, 22381–22391. [CrossRef]
28. Jayakumar, S.; Nandakumar, S. Reinforcement learning based distributed resource allocation technique in device-to-device (D2D) communication. *Wirel. Netw.* **2023**, *29*, 1843–1858. [CrossRef]
29. Lee, S.-H.; Shi, X.-P.; Tan, T.-H.; Lee, Y.-L.; Huang, Y.-F. Performance of Q-learning based resource allocation for D2D communications in heterogeneous networks. *ICT Express* **2023**, *9*, 1032–1039. [CrossRef]
30. Gour, R.; Tyagi, A. Joint uplink–downlink resource allocation for energy efficient D2D underlaying cellular networks with many-to-one matching. *Phys. Commun.* **2023**, *58*, 102016. [CrossRef]
31. Wang, H.; Xiao, P.; Li, X. Channel Parameter Estimation of mmWave MIMO System in Urban Traffic Scene: A Training Channel-Based Method. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 754–762. [CrossRef]