Coordinated Machine Learning for Channel Reuse and Transmission Power Allocation for D2D Communication

Ishtiaq Ahmad, Zdenek Becvar, Pavel Mach
Faculty of Electrical Engineering, *Czech Technical University in Prague*, Prague, Czech Republic
{ahmadish, zdenek.becvar, machp2}@fel.cvut.cz

***Abstract*—Mutual reuse of communication channels among device-to-device (D2D) pairs enhances the spectral efficiency of the mobile networks. However, the interference among D2D pairs mutually reusing the same channels imposes a significant challenge. In combination with allocation of the transmission power of each pair for the reused channels, the problem of joint D2D channel reuse and transmission power allocation becomes NP-hard. Thus, we employ deep deterministic policy gradient (DDPG) to decide how the D2D channels should be reused by the D2D pairs. Then, for the reused channels, we allocate the transmission power of the D2D pairs sharing the channels using deep neural network (DNN). However, combining the DDPG-based channel reuse with the DNN-based transmission power allocation leads to an accumulation of errors introduced by DDPG and DNN. The accumulated errors degrade the overall communication capacity. Thus, we also introduce a coordination between DNN and DDPG to suppress the effect of the error accumulation. Simulation results demonstrate that the proposed DDPG-based channel reuse even without coordination increases the sum capacity by 15% compared to state-of-the-art works. On top of this gain, the coordination of both DDPG and DDN adds another 12% in the sum capacity.**

***Index Terms*—Channel reuse, D2D communication, transmission power allocation, machine learning, coordination.**

I. Introduction

A direct communication between two devices in proximity, known as Device-to-Device (D2D) communication, is a promising trend enabling to increase the data rates and spectral efficiency of mobile networks [1]. This is due to the fact that two devices, namely a D2D transmitter and a D2D receiver, form a D2D pair facilitating a direct transmission of data without sending data through a base station (BS) [2].

The D2D communication operates in two modes: *1)* a *shared* mode, where the D2D devices utilize resources allocated to cellular devices communicating with the base station, and *2)* a *dedicated* mode, where the D2D devices use separated resources not assigned to the cellular devices [3]. The shared mode generally provides higher spectral efficiency than the dedicated one. However, achieving such efficiency often necessitates complex solutions for resource allocation and management, which should be able to suppress mutual interference between cellular and D2D devices. Consequently, ensuring communication reliability and overall quality of service (QoS) is challenging in the shared mode [4]. Since the shared mode cannot guarantee reliability due to varying and unpredictable interference, a dedicated mode is preferred by D2D devices with stringent QoS requirements demanding a high reliability, such as vehicular or public safety applications.

In the D2D dedicated mode, an efficient spectrum utilization can be facilitated by the reuse of available resources originally allocated to each D2D pair [5]. The authors in [6] propose the channel reuse of D2D pairs aiming to minimize interference by reusing the channel at least once while exploiting the minimum number of channels. In [5], [6], however, the sum capacity is limited, as only a restricted number of channels are reused to ensure a low interference and a low complexity.

The reuse of all channels in order to maximize the sum capacity of D2D pairs in the dedicated mode is assumed in [7]. However, there is no guarantee of any minimal capacity for the D2D pairs leading often to the case, where some D2D pairs have no channels at all. In this regard, the authors in [8] propose a game theory-based channel reuse and the Lagrangian method for power allocation to guarantee a minimal capacity while maximizing the sum capacity of the users. Further, the authors in [9], present an algorithm where all D2D pairs reuse all the access channels concurrently to maximize spectral efficiency. Then, in [10], the authors propose a two-stage graph coloring problem to reduce the interference at the reused channels. Unfortunately, if the number of D2D devices reusing channels increases, as expected in future generations of networks, the above-proposed solutions [8]–[10] become too complex and unable to solve the problem in real-time.

Recently, machine learning has gained attention in addressing complex resource allocation problems in wireless communication. For example, deep neural networks (DNNs) are used for transmission power allocation [11] or deep reinforcement learning is employed to manage radio resources [12], [13]. Moreover, the authors in [14], introduce a pointer neural network to optimize the channel and power allocation to maximize the total throughput of D2D and cellular devices while adhering to interference threshold constraints. In [12]–[14], the authors consider a shared mode for D2D communication and do not ensure any minimal capacity to the devices making the solutions not suitable for scenarios, where a reliable communication is required.

In this paper, we focus on the problem of D2D channel

This work was supported by the project No.23-05646S funded by the Czech Science Foundation.

reuse in dedicated mode to maximize the sum capacity of D2D devices. To this end, we employ deep deterministic policy gradient (DDPG), which determines the channels to be reused and selects the D2D pairs to reuse these channels. Unlike most of the related works, we set limits neither to the number of reused channels nor to the number of channels used by each D2D pair. Such generalization of the reuse is possible due to the low complexity of developed DDPG compared to solutions adopted in related works.

To further cope with interference due to reuse, we also allocate the transmission power of D2D pairs to the reused channels via DNN, as in [15]. The D2D transmission power for individual channels determined by the DNN is fed into the DDPG for channel reuse. Unfortunately, simply concatenated machine learning-based solutions for the D2D transmission power allocation and the decision on D2D channel reuse lead to a relatively high cumulative learning error degrading the final sum capacity of the D2D pairs, as we show in the paper. One could consider to simply merge the D2D power allocation and the D2D channel reuse into a single DNN; however, as shown in [16], such merging results either in a very large and hard-to-train neural network or in a poor performance due to dependencies among inputs and outputs of the DNN (power allocated to individual channels would depend on the channel reuse and vice versa). Therefore, on top of the proposed DDPG for the D2D channel reuse, another contribution of our work consists in proposed coordination between the machine learning for the power allocation and for the channel reuse while maximizing the sum capacity of D2D devices with a constraint on the the minimum capacity of the D2D pairs.

To summarize, the *key novelty* of our work resides in the introduction of *DDPG for the prediction of D2D channel reuse* and in the *mutual coordination* of the DDPG for the D2D channel reuse with the DNN for the D2D power allocation via feedback from the environment (network) to mitigate the accumulated errors introduced by individual learned models. We show that the proposed predicted D2D channel reuse improves the sum capacity of the system compared to the solution based on machine learning algorithms without coordination as well as compared to the recent related works.

The rest of the paper is structured as follows. In Section II, we introduce the system model and formulate the problem. Sections III and IV describe the proposed channel reuse based on DDPG and the coordination of DNN and DDPG, respectively. Section V presents simulation results and Section VI concludes this paper.

II. System Model and Problem Formulation

In this section, we present the system model followed by the description of machine learning adopted for the power allocation. After that, we formulate the targeted problem.

A. Communication model

As illustrated in Fig. 1, we consider a general urban area comprising M D2D devices forming $N = \lfloor M/2 \rfloor$ D2D pairs operating in dedicated D2D mode [9] while exploiting K channels. Then, communication capacity of the n-th D2D pair at the k-th channel with reuse is:

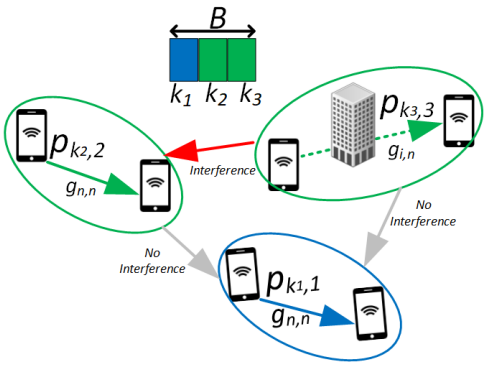

Fig. 1. System model with devices communicating directly using D2D; bandwidth originally allocated to individual pairs can be reused, as in case of bandwidth chunks k_2 and k_3 allocated originally for the second and third pairs, respectively, in this example.

$$c_{k,n} = \frac{B}{N} \log_2 \left(1 + \frac{p_{k,n} g_{n,n}}{\sigma_0 + I_{k,n}} \right) \quad (1)$$

where B is the total bandwidth (we assume B is initially split equally among D2D pairs, as in [21], but later on each D2D pair can use wider bandwidth due to reuse), $p_{k,n}$ represents the transmission power of the transmitter in the n-th D2D pair, $g_{n,n}$ denotes the channel quality between the transmitter and the receiver of the n-th D2D pair, σ_0 stands for the noise level, and $I_{k,n}$ is the sum of interference caused by all pairs reusing the same k-th channel to the receiver of the n-th D2D pair, expressed as:

$$I_{k,n} = \sum\nolimits_{i=1, i \neq n}^{i=N} \alpha_{k,i} p_{k,i} g_{i,n} \quad (2)$$

where $p_{k,i}$ is the transmission power of the i-th interfering device using the k-th communication channel, $g_{i,n}$ presents the channel quality between the i-th interfering D2D pair to the receiver of the n-th D2D pair, and $\alpha_{k,i} = 1$ indicates the k-th communication channel is reused by the i-th D2D pair while $\alpha_{k,i} = 0$ otherwise.

B. Architecture of DNN for Power Allocation

This section describes a general architecture of DNN conventionally employed for power allocation (DNN-PA), as, e.g., in [11], [15], [17]. The DNN-PA is comprised of an input layer, H hidden layers, and an output layer [17]. The inputs to the DNN-PA are represented by D2D channel qualities $g_{n,n}$ between the transmitter and the receiver of the n-th D2D pair. The D2D channel qualities undergo processing across the hidden layers. The h-th hidden layer is composed of V_h neurons and the neurons in the h-th layer are with weights $[w_h^1, \ldots, w_h^{V_h}]$. The hidden layers are fully connected and are activated using the sigmoid function. The sigmoid function guarantees that the output power allocation value remains constrained within the desired range. The output layer is activated by a regression function, which predicts the continuous transmission power $\hat{p}_{k,n}$.

C. Problem Formulation

The objective of this paper is to maximize the sum capacity of D2D devices while ensuring the minimum required capacity c_{min} of the D2D pairs. The maximum sum capacity is achieved via combined power allocation and smart reuse of the D2D communication channels allowing multiple D2D pairs to reuse channels. Then, the targeted sum capacity maximization problem is expressed as:

$$\begin{aligned} \boldsymbol{p}^*, \boldsymbol{\alpha}^* = & \underset{\substack{0 \leq \sum_k p_{k,n} \leq p_{max}, \\ \alpha_{k,n} \in \{0,1\}, \forall_n, \forall_k}}{\arg\max} \sum_{n=1}^{N} \sum_{k=1}^{K} c_{k,n} \\ & a) \quad \alpha_{k,n} \in \{0,1\} \quad \forall_n \in \{1,..,N\}, \forall_k \in \{1,..,K\} \\ & b) \quad 0 < \sum\nolimits_k p_{k,n} \leq p_{max} \quad \forall_n \in \{1,..,N\} \\ & c) \quad c_{k,n} \geq c_{\text{min}} \quad \forall_n \in \{1,..,N\} \end{aligned} \tag{3}$$

where $\boldsymbol{p}^*$ and $\boldsymbol{\alpha}^*$ are the optimal p and α, respectively. the constraint (3a) limits the range of the reuse indicator variable, (3b) ensures that the power assigned to the D2D pair over all channels is restricted to the range between 0 and the maximum allowed transmission power p_{max}, and (3c) ensures that $c_{k,n}$ is greater than minimum required capacity c_{min}.

The problem defined in (3) is a mixed integer non-linear programming (MINLP) problem and is NP-hard. Therefore, we employ DDPG as a promising approach for addressing the challenges associated with channel reuse. To address the non-convex nature of the transmission power allocation, we exploit the DNN-based power allocation proposed in [11]. Note that we do not introduce any novelty in the power allocation itself, since this area is well-investigated. Instead we focus on integration of power allocation with channel reuse via coordination of machine learning tools.

In the following sections, we first present the proposed adaptive channel reuse using DDPG-CR for D2D devices and then describe the proposed coordination to minimize the prediction errors through coordination of DNNs and DDPG during the exploitation of trained machine learning tools.

III. Proposed Adaptive Channel Reuse using DDPG-CR for D2D Devices

In this section, we first present the details of the incorporation of the channel reuse problem into the MDP framework. Then, we describe the detailed architectural description of DDPG employed for channel reuse.

A. Incorporate channel reuse into MDP framework

We break down a challenging channel reuse problem into a framework based on the principles of Markov Decision Processes (MDP). The MDP is defined by a tuple (S, A, P, R), where S represents the state space, A is the action space, the transition probability P defines the probability of the state s transitioning to state s' after the execution of the action a, and R represents the immediate reward following the execution of the action a. At each state, the MDP selects the action that maximizes expected sum of discounted future rewards. The individual components of the MDP are elaborated as follows.

State Space: In this work, the state space of DDPG-CR comprises two elements: 1) the D2D channel qualities (e.g., predicted according to [20]); and 2) predicted power allocation over reused channels. These two aspects provide key and sufficient information about interference, thus, influencing channel reuse. Then, the state space for DDPG-CR is defined as $S(t) = \{g_{n,n}, p_{k,n}\}$

Action Space: The agent controls the channel reuse via the action a. The action space is defined as a determination is the k-the channel should be reused by the n-th D2D pair, i.e., determining $\alpha_{k,n}$, hence, the action of DDPG-CR is defined as $A(t) = \{\alpha_{k,n}\}$

Reward: The system computes the immediate reward based on the action taken and, then, updates the system state. Our objective is to maximize the overall sum capacity; hence, the immediate reward $R(t)$ in the time slot t is formulated as $R_{(s,a)}(t) = \max \mathbb{E}[c_{k,n}]$.

The maximization of the sum capacity is achieved via maximizing the cumulative reward, which is defined as the profit received by the system over a long-term period of T time slots. Hence, the cumulative reward is defined as:

$$R = \max \mathbb{E}[\sum_{t=1}^{T} R_{(s,a)}(t)] \tag{4}$$

where $\mathbb{E}[.]$ represents the maximization of the expected cumulative reward. The cumulative reward refers to the actual reward obtained by a decision-making policy in a particular instance t. The expected cumulative reward is the average sum of rewards received over T. Before the reuse of channels by D2D pairs, each pair is initially assigned with a dedicated channel B/N for all D2D pairs [8]. These channels are then available for reuse by other pairs using DDPG-CR.

B. Architecture description of DDPG for channel reuse

The DDPG-CR comprises an actor-critic reinforcement learning-based architecture that involves two DNNs: the actor DNN and the critic DNN. The actor DNN input states are composed of factors related to the D2D communication, i.e., the D2D channel qualities and the predicted D2D power allocation. The output is the probability distribution of the action, i.e., the channels reuse value $\alpha_{k,n}$ for the n-th D2D pair and k-th channel. The states undergo processing through H^a_{CR} hidden layers that are fully interconnected and are subsequently activated using the rectified linear unit (ReLU) activation function. The ReLU function is computationally efficient and introduces non-linearity to the model, which is crucial in the learning process for complex D2D environmental patterns imposed via interference caused by channel reuse.

The critic DNN evaluates the channel reuse actions chosen by the actor, providing feedback on the effectiveness of the chosen actions for a given state. In addition, critic DNN estimates the expected cumulative reward from a given state-action pair. The critic DNN consists of states related to the D2D dedicated environment as inputs and H^c_{CR} hidden layers

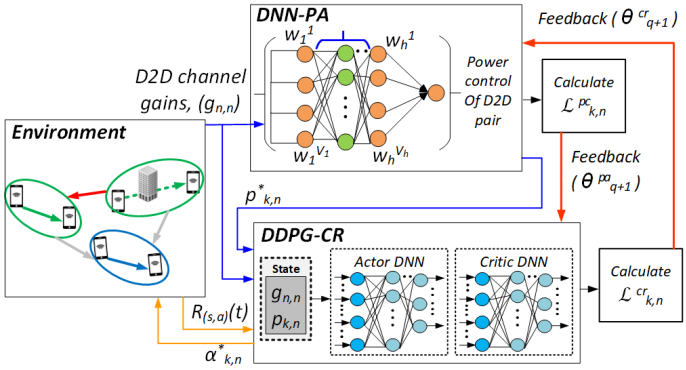

Fig. 2. Coordinated architecture of DNN-PA and DDPG-CR.

that capture the complex relationships between input state and expected cumulative reward. The output layer of critic DNN gives estimated values of the channel reuse value $\alpha_{k,n}$. The actor DNN maximizes the expected cumulative reward (i.e., the sum capacity), and the critic DNN minimizes difference between the estimated and actual cumulative rewards.

In the proposed algorithm, the DDPG-CR makes decisions on reusing allocated channels in the following way. First, the actor DNN defines a parameterized policy and chooses the channel reuse for all D2D pairs according to the channel qualities, and predicted power allocation. Then, the critic DNN evaluates the current policy by processing the rewards received from the environment and calculating the loss function. Both the critic and actor DNNs are updated based on the loss function, which represents the difference between the estimated value and the true value of the state-value function $V(s(t))$. The state-value function $V(s(t))$, represents the reward function starting from specific state $s(t)$ following the policy π, which determines the actions in each state. The state-value function $V(s(t))$ is represented as:

$$V(s(t)) = \mathbb{E}[R_{(s,a)}(t)|s(t), \pi] \tag{5}$$

The critic DNN estimates the expected reward using a state-value function. Whereas, the actor DNN determines the true reward by selecting the action for a given state. The critic DNN learns to improve its predictions over time by comparing the expected reward with the actual rewards.

The loss function for updating the actor and critic DNNs $\mathcal{L}_{\text{ac}}$ is defined as $\mathcal{L}^{\text{ac}}(t) = V(s(t+1)) - V(s(t))$. The critic DNN updates the weights of its DNN according to the loss function as $w_{t+1} \leftarrow w_t + \beta_c \delta_w (\mathcal{L}^{\text{ac}}(t))^2$, where β_c is the learning rate of critic DNN and the squared loss function helps to amplify discrepancies, aiding efficient gradient-based optimization, and ensures non-negativity and smoothness in convergence (better generalization). Now, the actor DNN takes the channel to reuse decision maximizing the expected cumulative rewards according to the current state. The weight θ^{ac} of the actor DNN is updated using the loss function and the policy gradient as:

$$\theta^{\text{ac}}_{t+1} \leftarrow \theta^{\text{ac}}_t + \beta_a \delta_{\theta^{\text{ac}}} [\log_{\pi_{\theta^{\text{ac}}}}(a(t)|s(t))] \mathcal{L}^{\text{ac}}(t)] \tag{6}$$

where β_a is learning rate of actor DNN and $\pi_{\theta^{\text{ac}}}(a(t)|s(t))$ is output probability for each action computed by actor DNN.

IV. Proposed Coordination of DNN and DDPG during Exploitation of Trained Networks

The coordination between the DNN-PA and DDPG-CR is connected to the exploitation (inference) stage of the trained DNN-PA and DDPG-CR as illustrated in Fig. 2. The coordination of DNN-PA with DDPG-CR during D2D communication stems from the necessity to fulfill the constraints inherent in the targeted sum capacity maximization problem. We implement the coordination from DDPG-CR to DNN-PA using a loss function that gives feedback to DNN-PA to update its internal weights. The feedback indicates that there is a potential error in the predictions of power allocation. In the loss function, we consider a logarithmic component that continuously encourages a high sum capacity and penalizes a low sum capacity. We also incorporate the penalty in case the constraints are not fulfilled, i.e., if the differences $[c_{\text{min}} - c_{k,n}]$ and $[I_{k,n} - I_{\text{max}}]$ are positive. Unlike the traditional representation of typical loss functions based on mean square error, in our case, the penalties are added to satisfy the constraints in (3). Hence, the loss function used as the feedback from DDPG-CR to DNN-PA is defined as:

$$\begin{aligned}\mathcal{L}^{\text{cr}}_{k,n} = \mathbb{E}[&-\log(c_{k,n}) + \max(0, [c_{\text{min}} - c_{k,n}]) \\ &+ \max(0, [I^r_{k,n} - I_{\text{max}}])\end{aligned} \tag{7}$$

The loss function $\mathcal{L}^{\text{cr}}_{k,n}$ is calculated every iteration and is continuously fed back to DNN-PA to update its weights. The weights are updated based on the objective function's gradients during the training so that:

$$\theta^{\text{cr}}_{q+1} \leftarrow \theta^{\text{cr}}_q + \epsilon \nabla_{\theta^{\text{cr}}} \mathcal{L}^{\text{cr}}_{k,n}(\theta^{\text{cr}}_q) \tag{8}$$

where q is the iteration of the update, $0 < \epsilon \ll 1$ is the initial learning rate of both DNNs (note that both DNNs are initialized with the same initial learning rate to maintain synchronization and a similar learning pace of both), $\nabla_{\theta^{\text{cr}}} \mathcal{L}^{\text{cr}}_{k,n}(\theta_q)$ is the gradient of the loss function $\mathcal{L}^{\text{cr}}_{k,n}$ for the parameters θ at the iteration q, and $\nabla_{\theta^{\text{cr}}}$ is the gradient element, which is computed concerning the loss function and is used to adjust the parameters to minimize the loss. The value of the gradient is determined by the partial derivatives of the loss function $\nabla_{\theta^{\text{cr}}} \mathcal{L}^{\text{cr}}_{k,n} = (\frac{\partial \mathcal{L}^{\text{cr}}_{k,n}}{\partial \theta^{\text{cr}}_1}, \frac{\partial \mathcal{L}^{\text{cr}}_{k,n}}{\partial \theta^{\text{cr}}_2}, ..., \frac{\partial \mathcal{L}^{\text{cr}}_{k,n}}{\partial \theta^{\text{cr}}_\eta})$ concerning each layer's parameters, where η represents the number of parameters.

Similar to the feedback from DDPG-CR to DNN-PA, the loss function as feedback from DNN-PA to DDPG-CR is defined as:

$$\begin{aligned}\mathcal{L}^{\text{pa}}_{k,n} = \mathbb{E}[&-\log(c_{k,n}) + \max(0, [c_{\text{min}} - c_{k,n}]) \\ &+ \max(0, [p_{k,n} - p_{max}])\end{aligned} \tag{9}$$

The weights of the actor and critic DNNs are updated based on the objective function's gradients during the training, similar to (8), i.e.,

$$\theta^{\text{pa}}_{q+1} \leftarrow \theta^{\text{pa}}_q + \epsilon \nabla_{\theta^{\text{pa}}} \mathcal{L}^{\text{pa}}_{k,n}(\theta^{\text{pa}}_q) \tag{10}$$

The proposed channel reuse and coordination during the exploitation of the trained DNN and DDPG are summarized

in the following steps: *i)* the D2D channel qualities $g_{n,n}$ are fed into DNN-PA to predict the transmission power $p_{k,n}$ for the n-th D2D pair at the k-th channel; *ii)* using D2D channel qualities $g_{n,n}$ $p^*_{k,n}$, the channel reuse $\alpha^*_{k,n}$ is predicted using DDPG-CR; *iii)* $\mathcal{L}^{\text{cr}}_{k,n}$ s determined using (7) and $\mathcal{L}^{\text{PA}}_{k,n}$ is determined using (9) to maximize the sum capacity of all D2D users over all channels; *iv)* the weights of DNN-PA and DDPG-CR are updated via feedback using (8) and (10).

Algorithm 1 Coordination of DNN and DDPG via feedback.

1: **for** iteration q **do**
2: Predict $p_{k,n}$ using DNN-PA
3: Calculate $\mathcal{L}^{pa}_{k,n}$ using (9) to ensure $c_{k,n} > c_{\text{min}}$ & $p_{min} < \sum_{k=1}^{K} p_{k,n} \leq p_{max}$
4: Use $p_{k,n}$ and $g_{n,n}$ to predict $\alpha_{k,n}$ using DDPG-CR
5: Determine $\mathcal{L}^{cr}_{k,n}$ using (7) to ensure $c_{k,n} > c_{\text{min}}$ & $I^r_{k,n} < I_{\text{max}}$
6: Update θ^{pa}_{q+1} using (10) and θ^{cr}_{q+1} using (8)
7: **end for**

V. Performance Evaluation

This section describes a simulation scenario and settings and, then, provides a discussion of the simulation results.

A. Simulation scenario, models and competitive algorithms

We consider four BSs and 16-96 devices (composing 8-48 D2D pairs) randomly distributed so that the maximum distance between the D2D transmitter and the D2D receiver is 50 meters. The minimum and maximum transmission power for D2D transmitters is 0 dBm and 23 dBm, respectively [17]. The carrier frequency is set to 2 GHz and bandwidth is 20 MHz. A common level of thermal noise of -110 dBm is assumed. We consider a mixed LoS/NLoS scenario. In the case of LoS, the path loss is modeled in line with the 3GPP outdoor-to-outdoor environment defined in [19]. The NLoS channel interrupted by one or more buildings is subject to an additional attenuation of 10 dB per wall [20].

DNN-PA is composed of three hidden layers with 60, 30, and 20 neurons in their respective layers. The batch size is set to 32, and the learning rate is 0.01. In DDPG-CR, the actor DNN comprises three hidden layers with 32, 16, and 8 neurons, and the critic DNN consists of two hidden layers with 50 and 20 neurons in each layer. These settings are determined by a trial-and-error approach. The experience buffer length is 10^6, the discount factor is 0.99, the mini-batch size is 64, the actor learning rate is 0.001, and the critic learning rate is 0.01. The code of implemented proposal is available at GitLab [1].

We compare the performance of our proposed concept of the coordination of DNN-PA and DDPG-CR with the following competitive works:

- *DDPG-CR and DNN-PA w/o coordination:* The state-of-the-art solution for transmission power allocation [11] based on DNN is implemented in a cascade/sequential way *without any coordination* with the proposed DDPG-CR based channel reuse.
- *Minimum interference-based channel reuse (MI-CR) and DNN-PA:* The state-of-the-art work for transmission power allocation according to [11] and the channel reuse minimizing interference proposed in [6] are integrated. The D2D pairs follow two conditions for channel reuse: 1) D2D pairs utilize the sub-band of at least one other D2D pair; 2) exploit the minimum number of channels for reuse to reduce control overhead and complexity.
- *DNN-PA without reuse:* Only state-of-the-art transmission power allocation [11] based on DNN is implemented *without channel reuse*.

[1]Code of the proposal in Matlab: https://gitlab.fel.cvut.cz/mobile-and-wireless/codes/publications/Coordinated-Machine-Learning-for-Channel-Reuse-and-Transmission-Power Allocation-for-D2D-Communication

B. Simulation results

In Fig. 3, we investigate the sum capacity of the proposal and competitive state-of-the-art works for varying numbers of D2D devices. The sum capacity increases with the number of D2D devices since the channels are exploited more efficiently.

Comparing the sum capacity of the proposal, where DNN-PA with DDPG-CR are coordinated to the case without coordination, DNN-PA with MI-CR, and DNN-PA with no reuse reveals significant gain in the sum capacity introduced by the proposal of up to 12.9%, 27%, and 35.5%, respectively. More specifically, there is almost a 15.2% gain by employing DDPG-based channel reuse in D2D communication compared to MI-CR and no reuse case. Moreover, the additional 11.8% gain is obtained due to the coordination of DNN-PA and DDPG-CR. In addition, we show that if we increase c_{min} from 2 Mbps to 10 Mbps, the sum capacity is decrease by 8.9% in the proposed algorihtm and 14%, 39.7%, and 55% in DNN-PA with DDPG-CR without coordination, DNN-PA with MI-CR, and DNN-PA with no reuse, respectively.

Fig. 4 depicts the impact of the c_{min} on the ratio of satisfied D2D devices, referring to devices that achieve the capacity of at least c_{min}. As the minimum required capacity increases, the satisfaction ratio decreases across all algorithms due to limited communication resources within the system. Comparatively, the proposed approach with coordination between DDPG-CR and DNN-PA surpasses the DDPG-CR and DNN-PA without coordination, DNN-PA and MI-CR, and DNN-PA without reuse by around 10.1%, 18.3%, and 25.5%, respectively. The gain is around 8.2% by employing DDPG-based channel reuse while attaining the satisfaction of D2D devices compared to the MI-CR case. Moreover, the additional 10.1% gain is attained due to the coordination of DNN-PA and DDPG-CR.

In Fig. 5, we delve into the convergence of the proposed coordinated DNN-PA and DDPG-CR and compare it with the convergence of DNN-PA and DDPG-CR without coordination. Notably, the coordinated DNN-PA and DDPG-CR of varying numbers of D2D devices, i.e., 16 and 48 D2D devices demonstrate an enhanced sum capacity convergence, with improvements of around 12.6%, 26.5%, and 36.1% compared to the DDPG-CR and DNN-PA without coordination, DNN-PA and MI-CR, and DNN-PA without reuse. The gain is

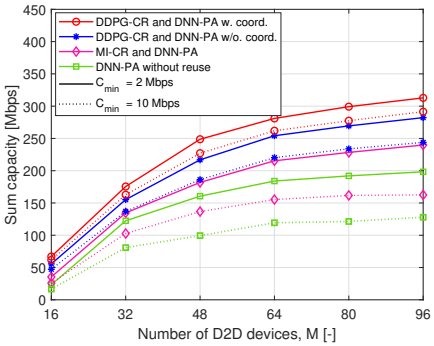

Fig. 3. Sum capacity for various numbers of D2D devices (full solid line is for $c_{\min} = 2$ Mbps, whereas dotted line is for $c_{\min} = 10$ Mbps).

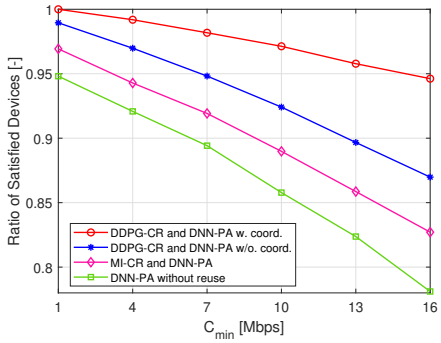

Fig. 4. Ratio of satisfied devices for various c_{min} (M = 48, $c^r_{k,n} > c_{min}$).

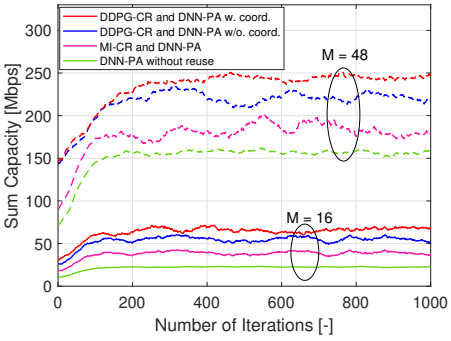

Fig. 5. Convergence of proposed DNN-PA and DDPG-CR with coordination for 16 and 48 D2D devices.

around 14.8% by employing DDPG-CR and 12.1% gain due to the coordination of DNN-PA and DDPG-CR.

VI. Conclusion

We address the challenge of efficiently reusing D2D communication channels in mobile networks to enhance the sum capacity of the D2D devices using deep reinforcement learning-based DDPG for D2D pairs. Additionally, we utilize DNNs to predict D2D transmission power for each device at each channel to cope efficiently with interference due to reuse. Since both machine learning solutions (for reuse and power allocation) naturally impose an error in their decision, we also propose coordination between DNN and DDPG to reduce prediction errors. The simulation results demonstrate the effectiveness of the proposed scheme, showcasing an enhancement in sum capacity by up to 15% for employing DDPG-CR, and an increase of around 12% due to the coordination of DNN-PA and DDPG-CR.

References

[1] M. S. M. Gismalla, *et al.*, "Survey on Device to Device (D2D) Communication for 5GB/6G Networks: Concept, Applications, Challenges, and Future Directions," *IEEE Access*, vol. 10, 2022.

[2] M. H. Khoshafa, *et al.*, "On the Physical Layer Security of Underlay Relay-Aided Device-to-Device Communications," *IEEE Trans. on Veh. Techn.*, vol. 69, no. 7, July 2020.

[3] Q. Wang, *et al.*, "Mode selection for D2D communication underlaying a cellular network with shared relays," in *Conf. on Wire. Commun. and Signal Proces. (WCSP)*, Hefei, China, 2014.

[4] J. Dai, et al., "Analytical Modeling of Resource Allocation in D2D Overlaying Multihop Multichannel Uplink Cellular Networks," *IEEE Trans. on Veh. Techn.*, vol. 66, no. 8, Aug. 2017.

[5] Y. Zhang *et al.*, "Incentive Compatible Overlay D2D System: A Group-Based Framework without CQI Feedback," *IEEE Trans. on Mob. Compu.*, vol. 17, no. 9, 2018.

[6] H. Saini *et al.*, "Churn Rate Aware Interference Management for Device-to-Device Communications in SDNs," in *IEEE Conf. on Commun. Workshops (ICC Workshops)*, Rome, Italy, 2023.

[7] A. Abrardo, *et al.*, "Distributed Power Allocation for D2D Communications Underlaying/Overlaying OFDMA Cellular Networks," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 3, March 2017.

[8] M. Najla *et al.*, "Reuse of Multiple Channels by Multiple D2D Pairs in Dedicated Mode: A Game Theoretic Approach," *IEEE Trans. on Wireless Commun.*, vol. 20, no. 7, July 2021.

[9] L. Eslami, *et al.*, "Joint Mode Selection and Resource Allocation for D2D and Femtocell Users in Dense Heterogeneous Networks With Full Frequency Reuse," *IEEE Trans. Veh. Tech.*, vol. 72, no. 11, Nov. 2023.

[10] C. Sun *et al.*, "Joint mode selection and resource allocation based on many-to-many reuse in D2D-aided IoT cellular networks," *Internet of Things*, vol. 25, 2024.

[11] W. Lee and R. Schober, "Deep Learning-Based Resource Allocation for Device-to-Device Communication," *IEEE Trans. on Wireless Commun.*, vol. 21, no. 7, July 2022.

[12] H. Yang, *et al.*, "Distributed Deep Reinforcement Learning-Based Spectrum and Power Allocation for Heterogeneous Networks," *IEEE Trans. on Wireless Commun.*, vol. 21, no. 9, Sept. 2022.

[13] V. Vishnoi, *et al.*, "A Deep Reinforcement Learning Scheme for Sum Rate and Fairness Maximization Among D2D Pairs Underlaying Cellular Network With NOMA," *IEEE Trans. on Veh. Techn.*, vol. 72, no. 10, Oct. 2023.

[14] L. Zhu, *et al.*, "Machine Learning-Based Resource Optimization for D2D Communication Underlaying Networks," in *IEEE Veh. Techn. Conf. (VTC2020-Fall)*, Victoria, BC, Canada, 2020.

[15] D. Ron *et al.*, "DNN-Based Dynamic Transmit Power Control for V2V Communication Underlaid Cellular Uplink," *IEEE Trans. on Veh. Techn.*, vol. 71, no. 11, Nov. 2022.

[16] I. Ahmad, Z. Becvar, P. Mach and D. Gesbert, "Coordinated Machine Learning for Energy Efficient D2D Communication," *IEEE Wireless Commun. Lett.*, 2024.

[17] M. Najla et al., "Machine Learning for Power Control in D2D Communication based on Cellular Channel Gains," in *IEEE Globecom Workshop, Waikoloa, HI, USA*, 2019.

[18] R. Li *et al.*, "Energy-Efficient Resource Allocation for High-Rate Underlay D2D Communications With Statistical CSI: A One-to-Many Strategy," *IEEE Trans. on Veh. Techn.*, vol. 69, no. 4, 2020.

[19] 3GPP TS 36.814, "Further advancements for E-UTRA physical layer aspects," *3GPP Technical Specification #36.814*, vol. 9.2.0, 2017.

[20] M. Najla *et al.*, "Predicting Device-to-Device Channels from Cellular Channel Measurements: A Learning Approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, 2020.

[21] Y. Huang *et al.*, "Mode Selection, Resource Allocation, and Power Control for D2D-Enabled Two-Tier Cellular Network," *IEEE Trans. on Commun.*, vol. 64, no. 8, Aug. 2016.