Learning-based Resource Allocation for Backscatter-aided Vehicular Networks

Wali Ullah Khan, Tu N. Nguyen, Furqan Jameel, Muhammad Ali Jamshed,
Haris Pervaiz, Muhammad Awais Javed, Riku Jäntti

***Abstract*—Heterogeneous backscatter networks are emerging as a promising solution to address the proliferating coverage and capacity demands of next-generation vehicular networks. However, despite its rapid evolution and significance, the optimization aspect of such networks has been overlooked due to their complexity and scale. Motivated by this discrepancy in the literature, this work sheds light on a novel learning-based optimization framework for heterogeneous backscatter vehicular networks. More specifically, the article presents a resource allocation and user association scheme for large-scale heterogeneous backscatter vehicular networks by considering a collaboration centric spectrum sharing mechanism. In the considered network setup, multiple network service providers (NSPs) own the resources to serve several legacy and backscatter vehicular users in the network. For each NSP, the legacy vehicle user operates under the macro cell, whereas, the backscatter vehicle user operates under small private cells using leased spectrum resources. A joint power allocation, user association, and spectrum sharing problem has been formulated with an objective to maximize the utility of NSPs. In order to overcome challenges of high dimensionality and non-convexity, the problem is divided into two subproblems. Subsequently, a reinforcement learning and a supervised deep learning approach have been used to solve both subproblems in an efficient and effective manner. To evaluate the benefits of the proposed scheme, extensive simulation studies are conducted and a comparison is provided with benchmark techniques. The performance evaluation demonstrates the utility of the presented system architecture and learning-based optimization framework.**

***Index Terms*—Vehicular communications, Machine learning, Spectrum interoperability, Backscatter communication, Resource management**

I. Introduction

The tremendous growth in the number of smart and connected vehicles is going to consume the wireless spectrum and network resources in the coming years [1]. With the emergence of novel Intelligent Transportation System (ITS) applications such as autonomous driving, smart traffic management, and infotainment, there will be a growing demand for efficient use of spectrum and cost-effective network infrastructure [2]. To effectively meet the requirement of forthcoming vehicular networks, different solutions such as non-orthogonal multiple access, device-to-device (D2D) communications, machine learning, blockchain, wireless social networking and fog computing approaches have been proposed in the literature [3]–[14]. However, with ubiquitous connectivity and ultimate functionality, come several important challenges. Firstly, massive amounts of data need to be collected by and transferred across low-powered devices in ways that consume as little energy as possible [15]. Secondly, the spectrum needs to be shared appropriately, so as to ensure the user fairness in dense networks [16]. Thirdly, the interference needs to be kept at a reasonable level, in order to prevent random fluctuations in the throughput [17]. These limitations of low rates, poor communications reliability, and high interference have profound implications on the design of dense communication platforms such as vehicular networks [18].

Lately, one of the most emerging technologies has been backscatter communication. The novelty of this technique is to design ultra-low-power devices for the communication by using the existing nearby radio signals [19]. These devices can benefit from a simple cost and energy effective design as compared to the traditional communication systems. In backscatter communication, the communicating node also acts as an energy source, which raises some additional challenges as compared to the traditional systems. Backscatter communication is a derivative of radio-frequency identification (RFID)-based systems [20]. Although backscatter communication may act as a key enabler of massive IoT and vehicular networks, there is still room to make this technology more favorable in large-scale setup. In this regard, machine learning techniques have recently been proposed to incorporate much-needed intelligence in the wireless networks [21]–[23]. Therefore, many latest works from network optimization to hardware design take into account the learning aspect to make the wireless networks more fault-tolerant and smarter. The basic idea of machine learning techniques is to get a computer program to learn from experience and complete the task in the future with these experiences. These learning techniques are generally divided into three main categories, i.e., supervised learning, unsupervised learning, and reinforcement learning [24].

Another important aspect from the network operator's point-of-view is an efficient spectrum sharing mechanism [25],

Corresponding author: Tu N. Nguyen

Wali Ullah Khan is with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, 1855 Luxembourg City, Luxembourg (Emails: waliullah.khan@uni.lu, waliullahkhan30@gmail.com).

Tu N. Nguyen is with the Department of Computer Science, Kennesaw State University, Marietta, GA 30060, USA (Email: tu.nguyen@kennesaw.edu).

Furqan Jameel and Riku Jäntti are with the Department of Communications and Networking, Aalto University, 02150 Espoo, Finland (email: furqanjameel01@gmail.com and riku.jantti@aalto.fi).

Muhammad Ali Jamshed is with James Watt School of Engineering, University of Glasgow, UK. (email: muhammadali.jamshed@glasgow.ac.uk).

Haris Pervaiz is with the School of Computing and Communications, Lancaster University, UK (email: h.b.pervaiz@lancaster.ac.uk).

M. A. Javed is with the Department of Electrical and Computer Engineering, COMSATS University Islamabad, Islamabad 45550, Pakistan (email: awais.javed@comsats.edu.pk).

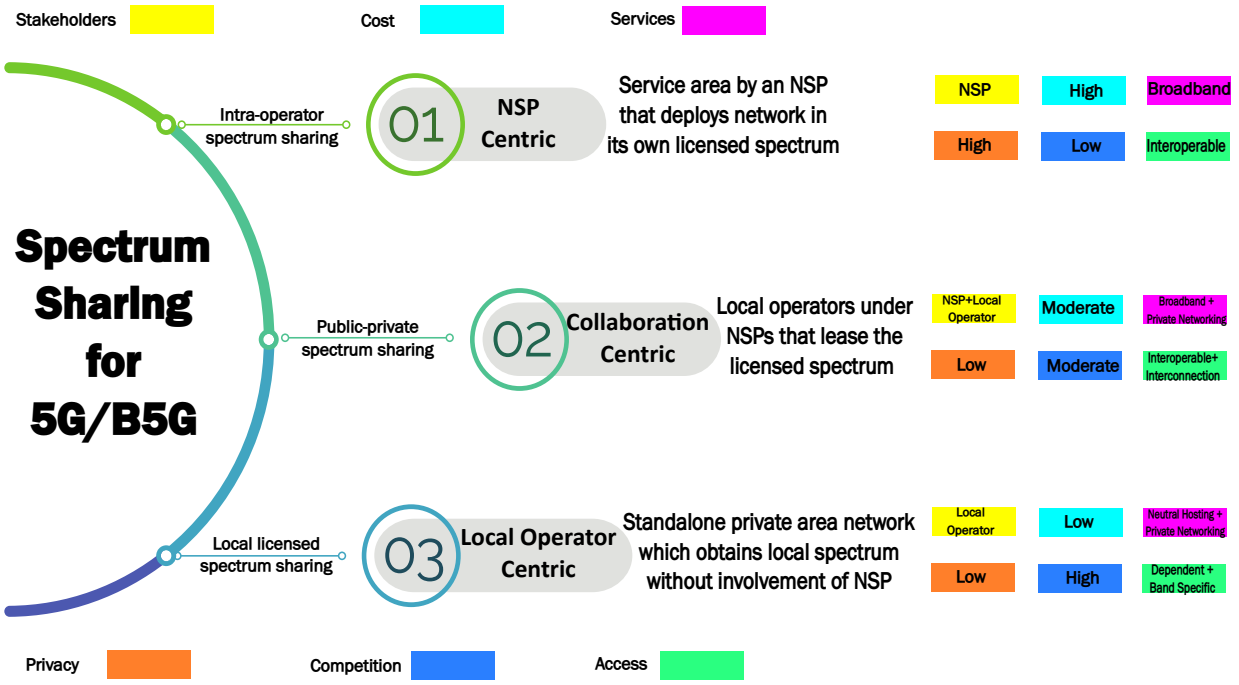


Figure 1. Spectrum sharing landscape in 5G and beyond 5G era.

[26]. As shown in Fig. 1, the spectrum sharing landscape is rapidly changing from a network service provider (NSP) centric to a more diverse and distributed local operator centric paradigm. Although all three approaches have proven their feasibility under certain conditions, it is difficult to identify a single best approach that meets all the requirements of future networks [27]. Thus, we focus our attention on a more hybrid spectrum sharing approach, i.e., collaboration centric spectrum sharing. In this paradigm, there exists a harmony among NSPs and local operators. More specifically, local operators (having standalone and private networks) can lease the spectrum from the NSPs resulting in a public-private sharing of spectrum. Due to having a moderate cost and enabling interoperability of bands, we anticipate that collaboration centric paradigm would play a critical role in most of the future heterogeneous wireless networks such as vehicular networks. With this intent, our work aims to advocate the utility of a collaboration centric spectrum sharing paradigm and employs supervised and reinforcement learning techniques to improve the performance of such heterogeneous vehicular networks.

II. Related Work and Motivation

Due to recent hype in machine learning, many studies use one of the machine learning techniques to enhance different metrics of the wireless networks. The concept of reinforcement learning has also been implemented on cognitive radio (CR) networks. In [28], authors proposed CR based networks for video surveillance applications. The results showed that the reinforcement learning-based approach is an effective way that enables SUs (honest users) to evade PUs (malicious users) so that the network performance is improved. The predictive learning model in CR based reinforcement learning was considered in [29] and the learning time, prediction accuracy and prediction errors have been observed. The concept of reinforcement learning has also been applied in 5G millimeter wave (mmWave) communication in a massive multiple-input multiple-output (MIMO) network in [30]. The joint optimization of beamwidth and transmit power was considered by taking the sensitivity of mmWave into account. Numerical results showed an increase in data rate from 7.5 Gbps to 9 Gbps in comparison to baseline schemes. In [31], the author studied the problem of providing buffer aware video streaming to wireless channel users. A joint optimization problem considering bandwidth allocation and buffer management for maximizing effective video was considered. Simulation results in the proposed deep reinforcement learning algorithm are effective for buffer aware video streaming. Vu *et al.* in [32] considered the optimal path selection and rate allocation in mmWave network and found the best path by means of the reinforcement learning algorithm. The proposed approach achieves 99.999% reliability and reduced latency. In [33], the authors considered the problem of rate allocation and optimal path selection in the self-backhaul mmWave network. Hence, a new scheme was proposed taking into consideration multiple antenna diversity, mm-wave bandwidth, and traffic splitting. The proposed solution ensures reliable communication and guaranteed probability.

The reinforcement learning-based artificial intelligence algorithms are applicable to wide areas of wireless communication such as D2D communication. The authors in [34] used a reinforcement learning-based power control algorithm in underlay D2D communication and compared a centralized Q-learning based algorithm with distributed Q-learning. It was shown that distributed Q-learning users are enabled to self-organize by learning independently, thus, reducing the overall complexity of the system. In [35], the problem of

Table I
COMMONLY USED NOTATIONS AND THEIR DEFINITIONS.

Not.	Definitions
M	Number of NSPs
J	Number of SBS
T_1	1st Switch
T_2	2nd Switch
W_m	Spectrum of m-th NSP
β	VUE association indicator
α	Spectrum sharing factor
B	Energy harvesting capacity of backscatter tag
P_t	Transmit power
h_u	Channel gain between MBS and legacy VUE
h_{st}	Channel gain between SBS and backscatter transmitter
h_{tr}	Channel gain between backscatter transmitter and receiver
h_{sr}	Channel gain between SBS and the backscatter receiver
μ	Backscatter reflection coefficient
η	Energy conversion efficiency
ζ_u	Profit per VUE per rate unit
$\psi_{m,1}$	Cost unit of MBS
$\psi_{m,2}$	Cost unit of SBS
ω_m	Backhaul price per bit
ϕ	Ratio of cost units of MBS and SBS
s	State of environment
a	Actions taken by the agent
r	Reward of the agent
Ω	Discount factor

vehicle-to-vehicle (V2V) transmission of the message was considered. Platooning is a key technology in smart cities for efficient V2V communication. The authors proposed a cooperative reinforcement learning (CRL) using long-term evolution (LTE) technology. The proposed scheme outperforms the other schemes in terms of delay in cooperative awareness message (CAM). Moreover, resource utilization was also efficient in the proposed scheme. Qiu *et al.* in [36] explored the idea of joint mode selection and power adaptation using D2D communication. The authors proposed a joint optimization of different transmission modes and power levels using Q-learning. The algorithm provides optimal energy efficiency, in comparison to the state-of-art. In [37], the achieved sum rate is maximized by using a more realistic channel formulation using finite-state Markov channels (FSMC). The complexity of the stated problem was high but has outperformed other schemes. Of late, Jameel *et al.* have also exploited reinforcement learning to investigate the problems of interference mitigation to improve the system performance of backscatter communication [38], [39].

There exist a few studies on backscatter communication that employ machine learning techniques [40], [41]. For instance, the authors of [42] used a supervised machine learning technique (support vector machine) to detect the signal from a backscatter tag by transforming the tag detection into a classification task. The learning algorithm divides a signal into two groups based on the energy features. It was shown that the proposed scheme outperforms the conventional random forest technique. In [43], the authors proposed to use machine learning for channel estimation in backscatter systems. They design a semi-blind channel estimator using a typical machine learning technique called expectation maximization. They also derived Cramer-Rao lower bounds of estimated parameters to verify the utility of the proposed technique and validated the results using simulations. In [44] the authors discussed the ambient backscatter communication that enables wireless devices to communicate without utilizing radio resources. The system is modeled by the Markov decision process and the optimal channel is obtained by the iterative algorithm. If the channel distribution is unknown, Q- learning method is implemented to find a suboptimal solution. The physical layer security of backscatter tags is another important issue [45]. To address this using machine learning techniques, the authors of [46] used a multiobjective genetic algorithm. More specifically, they reduced the antenna side lobes and obtained optimal Pareto fronts. The simulation result indicates that the proposed antenna design not only improves the physical layer security but also provides improved energy efficiency. In [47], Fan *et al.* provided a detailed summary of different machine learning techniques used for activity identification of backscatter tags while the authors of [48] proposed to use deep learning for intelligent user association. Beside that, there also exist several works on backscatter communication using non-learning approaches [49]–[59].

As evident from the aforementioned review, the studies incorporating the learning aspect in backscatter communications are very few in numbers. Moreover, the deployment of backscatter tags along with legacy users in a heterogeneous network has not been studied. In addition, it is important to allocate spectrum resources (both licensed and unlicensed) effectively to guarantee the quality of service (QoS) to all the devices. To our best knowledge, such types of studies on backscatter communications are still missing in the literature and need the utmost attention prior to deploying a massive number of backscatter devices. In simple terms, we make our best effort to settle the following questions:

- **Question 1**: How to optimize the performance of a heterogeneous network having legacy users and backscatter tags?
- **Question 2**: Compared to conventional greedy methods, can applying learning techniques in heterogeneous backscatter networks bring more performance gains?
- **Question 3**: What are the associated costs (in terms of overhead and computation) to train large-scale backscatter networks with multiple NSPs.

III. CONTRIBUTION AND ORGANIZATION

In order to address the aforementioned questions both from the perspective of the interoperability of backscatter and legacy users and from the view-point of the evolution of large-scale backscatter tags are nontrivial in nature. In this regard, we provide a detailed architecture of heterogeneous backscatter networks and the proposed learning-based optimization framework for solving the resource allocation and user association problem in detail. A list of commonly used notations throughout this paper is outlined in Table I. The main contribution of this work is summarized as follows:

1) A wireless heterogeneous backscatter network architecture is introduced that can be operated by multiple NSPs. Each NSP can own a macro BS (MBS), whereby, every MBS controls various small BSs (SBSs). The

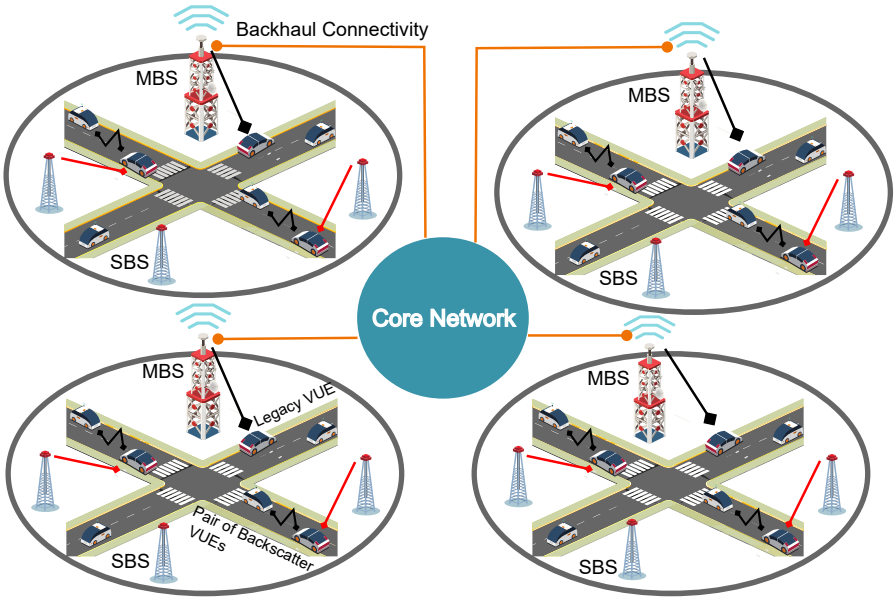


Figure 2. An illustration of system model having multiple MBS and SBS. The MBS and SBS are connected to the core network via backhaul links. Each MBS has multiple SBS and a legacy UE. Whereas, each SBS has a backscatter transmitter-receiver pair that reuse the spectrum resources.

spectrum is divided among MBSs and SBSs. The SBSs have energy harvesting backscatter tags and receivers under their coverage area and the spectrum is reused among different SBSs. These details on network setup are provided in Section IV.

2) To maximize the total utility of NSPs, resource allocation, and user association optimization problem is formulated. The proposed problem considers both the cost paid for using the backhaul network resources and revenue earned from the users. The formulated problem addresses the power allocation, user association, and spectrum sharing issues with explicit constraints on resource utilization. The problem formulation steps are presented in Section V.
3) The optimization problem is then divided into two subproblems. The problem of spectrum sharing and user association is addressed using a reinforcement learning technique. Specifically, the Q-learning approach is used where an agent learns the optimal policy for associating users and allocating spectrum resources. On the other hand, the power allocation problem is resolved by training and testing a deep neural network. The details of both solutions can be found in Section VI.
4) The effectiveness of the learning-based optimization framework is demonstrated by performing extensive simulations. By providing a reliable solution to the resource distribution and user association problem, it is shown in Section VII that both users and NSPs can be benefited. Some concluding remarks and future research directions are detailed in Section VIII.

IV. System Model

We consider a collaboration centric spectrum sharing paradigm in a geographical area having multiple NSPs with MBSs and SBSs operating at sub-6 GHz band, as illustrated in Fig. 2. It is considered that there are M NSPs, whereby each NSPs has one MBS and J SBSs. For each NSP, we consider an OFDM-based network for which the spectrum is divided into MBS and SBSs. Assuming the total spectrum bandwidth as W, the bandwidth of the m-th NSP then becomes $W_m = \frac{W}{|M|}$. The set of BSs of m-th NSP is denoted by $\mathcal{S}_m$ and $\mathcal{S}_m^j$ denotes the j-th SBS of m-th NSP, such that $\mathcal{S}_m^j \in \mathcal{S}_m$. In addition, $\mathcal{S}_m^0$ refers to the MBS belonging to the m-th NSP such that $\mathcal{S}_m^0 \cup \mathcal{S}_m^j = \mathcal{S}_m$. We also consider that the set of vehicle user equipments (VUEs) connecting to m-th NSP and j-th SBS are denoted as $\mathcal{U}_m^j$, whereas, the VUEs connected to m-th NSP belong to $\mathcal{U}_m$, such that $\mathcal{U}_m^j \in \mathcal{U}_m$. Likewise, VUEs connecting to m-th NSP and its MBS are denoted as $\mathcal{U}_m^0$ such that $\mathcal{U}_m^0 \cup \mathcal{U}_m^j = \mathcal{U}_m$. Without loss of generality, we consider that a VUE can connect to only one BS at a time and each MBS and SBS serves only one VUE such that the total number of VUEs are $M(J+1)$. Here, the VUEs are divided into two categories (i.e., pair of backscatter VUEs and legacy or cellular VUE) based on their connection to the BS. Specifically, the pair of backscatter VUEs are assumed to be communicating to SBS. On the other hand, the legacy VUEs are considered to be communicating with the MBS. Thus, the VUE association indicator for the u-th user connected to the j-th SBS of the m-th NSP at any l-th time slot can be given as

$$\beta_u^{m,j}(l) = \begin{cases} 1, & \text{if } u\text{-th VUE associates with } j\text{-th SBS} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

A. Backscatter Model

For backscatter communications in SBS, we consider that the pair of backscatter UE consists of a transmitter and a receiver that is equipped with a single antenna. Each backscatter VUE is assumed to use a reflection amplifier that

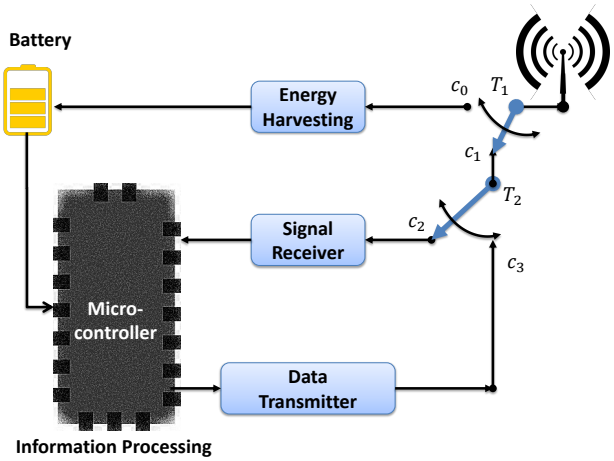


Figure 3. Block diagram of the backscatter transmitter. The backscatter transmitter consists of an antenna, a microcontroller and a recharge-able battery. Based on the configuration of the switches, the tag can either use the received energy for harvesting or for data transmission.

is characterized by the negative load impedance [60]. The channel coefficients between the SBS and backscatter transmitter, between SBS and backscatter receiver, and between the backscatter transmitter and receiver are denoted as h_{st}, h_{sr}, and h_{tr}. During each time slot, the backscatter transmitter has to decide whether to operate in energy harvesting mode or backscatter mode. The circuit diagram is given in Fig. 3. As illustrated in the figure, each backscatter transmitter is equipped with a rechargeable battery. By exploiting the harvested energy from the SBS, the backscatter transmitter can improve its life cycle while virtually operating in a battery-less manner. The backscatter transmitter can switch the connection point between T_1 and T_2. The switch T_1 determines whether the received RF signal needs to be used for energy harvesting (i.e., $T_1 = c_0$) or for information decoding/ transfer (i.e., $T_1 = c_1$). In the case of information decoding/ transfer, the second switch T_2 determines whether the information needs to be decoded (i.e., $T_2 = c_2$) or the data needs to be transferred (i.e., $T_2 = c_3$).

When the switch is $T_1 = c_0$, the backscatter transmitter of the u-th VUE connected to the j-th SBS of the m-th NSP uses the RF signal to harvest energy and operates in so-called energy harvesting mode. By converting the RF signal into direct current, the energy harvesting circuit recharges the battery. The collected energy can be used for charging the battery or transferring the data to the backscatter receiver. Thus, the harvested energy can be denoted as

$$E_h^{m,j} = \eta |h_{st}|^2 P_t^{m,j}, \tag{2}$$

where the energy harvesting efficiency is represented as η and P_t is the transmit power of the SBS. From the expression, we can note that the harvested energy can be zero if the received power is too low. Without loss of generality, the capacity of the battery is quantized into B units such that B is an integer. The backscatter transmitter is assumed to consume v units of energy during energy harvesting mode, while it consumes $v^{'}$ units of energy during the backscattering phase, where $1 \leq v < v^{'} < B$. To incorporate a realistic communication procedure, it is considered that before each time slot, the tag can operate in backscattering mode when $E_h^{m,j} \geq v^{'}$. When this condition is not satisfied, the backscatter transmitter must harvest energy.

When the backscatter transmitter switches to backscattering mode, the connection switches like $T_1 = c_1$ and $T_2 = c_3$. From the circuit diagram, it can be seen that for this case, the amount of harvested energy would be zero and the received power would be used to transfer the data to the receiver. Generally, the transmission distance from SBS to the backscatter transmitter is much larger than the distance between the backscatter transmitter and backscatter receiver. Due to this high dynamic range, the backscatter receiver can apply successive interference cancellation to obtain the interference-free signal from its own SBS [61].

B. Transmission Setup

Each NSP leases a specific portion of the spectrum to the SBSs and uses the remainder of the spectrum for its own MBS. Note that in the considered network setup, the NSPs cannot lease the network spectrum to each other. Thus, within a particular NSP, the spectrum shared among MBS and SBSs is, respectively, given as $f_m = \alpha_m W_m$ and $f_m^s = (1 - \alpha_m) W_m$. Here, $0 < \alpha_m < 1$ represents the spectrum allocation factor. Let us define $R_u^{m,j}$ which denotes the throughput of the u-th VUE connected to j-th SBS. The maximum achievable throughput of the u-th VUE connected to the MBS of the m-th NSP can be given as

$$R_u^{m,0}(l) = f_m \log_2 \left(1 + \frac{P_t^{m,0} |h_u^{m,0}|^2}{N_0}\right), \tag{3}$$

where $P_t^{m,0}$ denotes the transmit power MBS of m-th NSP, $h_u^{m,0}$ is the channel gain from the MBS to the legacy VUE. Moreover, N_0 denotes the power of the additive white Gaussian noise (AWGN).

Since the backscatter VUEs use reflection amplifiers, the peak power constraint can be relaxed and the maximum achievable throughput can be written as

$$R_u^{m,j}(l) = f_m^s \log_2 \left(1 + \frac{\mu P_t^{m,j} |h_{st}^{m,j}|^2 |h_{tr}^{m,j}|^2}{N_0 + \mathbb{I}_1 + \mathbb{I}_2}\right), \tag{4}$$

where $\mathbb{I}_1 = \sum_{k=1, k \neq j}^{J} \mu P_t^{m,k} |h_{st}^{m,k}|^2 |h_{tr}^{m,k}|^2$ represents the interference from other backscatter VUEs under the same NSP's SBSs, $\mathbb{I}_2 = \sum_{k=1, k \neq j}^{J} P_t^{m,k} |h_{sr}^{m,k}|^2$ denotes the interference from other SBSs of the same NSP and $P_t^{m,j}$ is the transmit power of j-th SBS of the m-th NSP. Moreover, $h_{st}^{m,j}$ denotes the channel gain from the SBS to the backscatter transmitter and $h_{tr}^{m,j}$ represents the channel gain between backscatter transmitter and backscatter receiver. Furthermore, μ represents the reflection coefficient which is kept constant for all backscatter VUEs. Also note that $P_t^{m,k}$ represents the interference power from other SBSs under the same NSP's MBS during the downlink transmission, whereas, the interfering channel gains are given as $h_{tr}^{m,k}$ and $h_{sr}^{m,k}$ from

other backscatter tags to desired receiver and from other SBSs of same NSP to the desired receiver, respectively. The total achievable throughput for the u-th VUE at l-th time slot can be given as

$$C_u^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}) = \beta_u^{m,0}(l)R_u^{m,0}(l) + \sum_{j=1}^{J}\beta_u^{m,j}(l)R_u^{m,j}(l) = \sum_{j\in\mathcal{S}_m}\beta_u^{m,j}(l)R_u^{m,j}(l), \tag{5}$$

where $\boldsymbol{\beta} = \{\beta_u^{m,j}(l)\}$, $\boldsymbol{\alpha} = \{\alpha_m(l)\}$, and $\boldsymbol{P} = \{P_t^{m,j}(l)\}$ represent the VUE association indicators, spectrum sharing policy, and power allocation policy, respectively.

V. Problem Formulation

This section provides the detailed information about the definition of utility function and pinpoints the specific constraints for the problem formulation. We first develop a utility function by taking into account the profit and loss for the NSP. Later on, based on the utility function, we formulate a utility maximization problem with specific set of constraints.

A. Utility function

The aforementioned details in the previous section reveal that each NSP can maximize its own profit by serving more VUEs (while improving the throughput) and by reducing the costs associated with the exchange of data within the network. In this regard, metrics like VUE association, spectrum sharing among MBS and SBSs, and power allocated to each BS come into play. Overlooking the importance of any of these resources may lead to inefficient network performance, thereby, resulting in reduced profit for certain cost of resources. Thus, the overall utility can be defined as a sum function of the profit obtained and the cost of serving the VUEs. Mathematically, we can define it as

$$\sum_{Uti}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}) = \sum_{i\in I}\sum_{m\in M}\sum_{u\in\mathcal{U}_m}\zeta_u U_u^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}) - \sum_{i\in I}\sum_{m\in M}\Upsilon_i^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}) - \sum_{i\in I}\sum_{m\in M}\Psi_i^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}), \tag{6}$$

where the first term in (6) indicate the benefit obtained by serving the VUEs and ζ_u denotes the profit per VUE per rate unit. The value of $U_u^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P})$ can be given as

$$U_u^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}) = \log(C_u^m). \tag{7}$$

where the logarithmic function has been used as a common choice of the utility function for maintaining user fairness. Note that linear utility functions result in a trivial solution for the case of throughput maximization, where providing more resources to VUEs with low rates is desirable. Hence, following the approach of [62], we opt to use logarithmic utility function in the remainder of this paper which is more close to the resource allocation philosophy of the practical networks.

In (6), Υ_i^m shows the cost of the radio and power resources. It can be written as

$$\Upsilon_i^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}) = \psi_{m,1}\sum_{u\in\mathcal{U}_m}\beta_u^{m,0}f_m p^{m,0} + \psi_{m,2} \times \sum_{u\in\mathcal{U}_m}\sum_{j\in\mathcal{S}_m}\beta_u^{m,j}f_m^s p^{m,j}, \tag{8}$$

where the product of bandwidth and power is used for quantifying the consumption of BS resources, whereas, the coefficients $\psi_{m,1}$ and $\psi_{m,2}$ indicate the cost unit of MBS and SBS, respectively. Note that due to increased cost of operating MBS, it is intuitive that $\psi_{m,1} \geq \psi_{m,2}$.

Also, Ψ_i^m in (6) refers to the cost of using backhaul resources during the communication. More specifically, this cost depends on many factors ranging from the amount of backhaul data and the type of backhaul technique. Mathematically, it is given as

$$\Psi_i^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}) = \omega_m\sum_{u\in\mathcal{U}_m}C_u^m(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}), \tag{9}$$

where the price per bit for every backhaul transmission is represented as ω_m.

B. Problem Formulation

The proposed work aims to jointly optimize the allocated power $\boldsymbol{P}$, VUE association $\boldsymbol{\beta}$, and spectrum sharing $\boldsymbol{\alpha}$. In light of the aforementioned analysis, the utility maximization problem can be formulated as

$$\mathbf{P1} \quad \max_{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}} \Sigma_{Uti}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{P}) \tag{10}$$

$$\text{s.t.} \quad \text{C1}: \ \beta_u^{m,j} \in \{0,1\}, \tag{10a}$$

$$\text{C2}: \ 0 \leq P_t^{m,j} \leq P_{\max}^j, \tag{10b}$$

$$\text{C3}: \ 0 \leq E_h^{m,0} \leq B, \tag{10c}$$

$$\text{C4}: \ 0 \leq \alpha_m \leq 1. \tag{10d}$$

where C1 is the constraint on the VUE association indicator such that VUE can be connected to one BS at a time. The second constraint C2 ensures that the maximum transmit power limit is not exceeded. The third constraint C3 is for limiting the total energy harvesting capacity of the tags due to the limitations of the energy reservoir. The C4 ensures that the spectrum resources are distributed accordingly in the MBS and SBS.

By observing the problem formulated in **P1**, we can observe that this problem is a non-convex combinatorial integer programming problem which is NP-hard and cannot be solved directly in a polynomial time [63]. Hence, the best possible strategy for solving it can be through a brute-force method which is capable of providing an optimal solution with an expensive computational complexity, infeasible for large-scale wireless systems. Moreover, the amount of real-time information required for these brute-force methods can be

overwhelming to collect due to the dynamic nature of wireless systems. Thus, in order to provide a near-optimal solution, we intend to utilize the learning-based optimization framework for solving the problem formulated in **P1**.

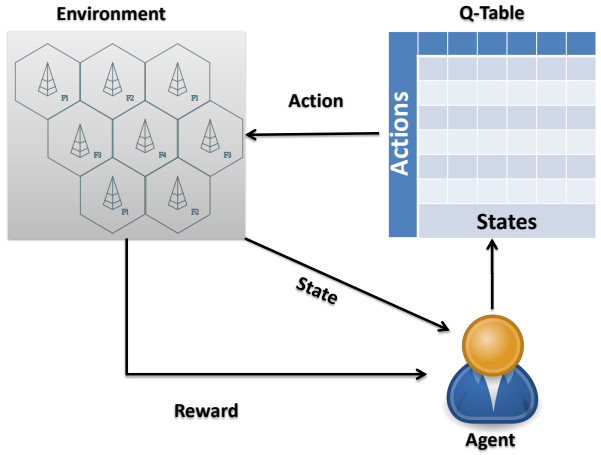


Figure 4. Illustration of general Q-learning technique. The technique consists of an agent and an environment on the principle of exploration and exploitation. The agent interacts with the environment based on a state-action table. As a result of its actions, the agent receives a reward.

VI. Proposed Solution

In this section, we first reformulate the problem **P1** to make it solvable by dividing the original problem into two subproblems. Firstly, the subproblem of VUE association and spectrum sharing is solved by keeping the transmission power fixed. Thus, the problem formulated in **P1** is converted into a first subproblem defined in **P2**. After utilizing the time-sharing relaxation, the reformulated problem **P2** is still non-convex in nature. The problem **P2** is then changed into **P3** by further transforming the constraint C4 into a linear constrained by utilizing an auxiliary continuous variable. The newly transformed problem **P3** is solved using Q-learning to find the user association metrics (β) and spectrum sharing partition (α) among the MBS and SBS within an NSP. After solving the problem defined in **P3**, we will replace the near-optimal user association metrics (β) and spectrum sharing partition (α) into the original problem defined in **P1** to solve the second subproblem. For the given near-optimal values of α and β obtained from the solution of **P3**, the original problem formulated in **P1** is rewritten as **P4** and will be solved to find the near-optimal power allocation vector P by training a deep neural network. In the following subsections, we provide solutions for both subproblems.

A. Proposed Q-learning Solution for VUE Association and Spectrum Sharing

Here, we first transform the subproblem to have a feasible solution. The problem **P1** can be written as

$$\mathbf{P2} \quad \max_{\boldsymbol{\beta},\boldsymbol{\alpha}} \Sigma_{Uti}(\boldsymbol{\beta},\boldsymbol{\alpha}) \tag{11}$$

$$\text{s.t.} \quad \text{C1, C4.} \tag{11a}$$

From (11), we can observe that **P2** has a binary variable $\beta_u^{m,j}$, due to which the problem is still non-convex. Following the detailed approach of [64], we can relax $\beta_u^{m,j}$ of C1 to have real values from 0 to 1. This means that $0 \leq \beta_u^{m,j} \leq 1$ can be viewed as a factor for sharing time such that u remains associated with a BS for that duration. Despite relaxing the problem, it is still non-convex due to the variables involved in the objective function. To address this situation, we introduce another auxiliary continuous variable $\widehat{\alpha_u^{m,j}} = \alpha_u^{m,j}\beta_u^{m,j}$. It is obvious that BS would not allocate spectrum resources to the u which is not associated with it. Correspondingly, the problem **P2** can be transformed into **P3** with the constraints given as

$$\mathbf{P3} \quad \max_{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}} \Sigma_{Uti}(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}}) \tag{12}$$

$$\text{s.t.} \quad \text{C1}: \ 0 \leq \beta_u^{m,j} \leq 1 \tag{12a}$$

$$\widehat{\text{C4}}: \ 0 \leq \widehat{\alpha_u^{m,j}} \leq \beta_u^{m,j}. \tag{12b}$$

From (12), one can observe that our objective function is the maximizing sum of concave functions, where, the constraints are linear. Therefore, the convexity of the objective function can be proved by showing the continuity of function and with the help of perspective operation of log [65], [66]. To provide a solution to **P3**, we use Q-learning to find the optimal policy for the given set of constraints.

Q-learning is one of the most popular reinforcement learning techniques. The reinforcement learning problem consists of an environment and either single or multiple agents. As shown in Fig. 4, an agent observes a current state and takes action according to a stochastic policy π. Generally, there are three elements of a Q-learning model, i.e., states, actions, rewards. During each time slot l, the agent selects an action by observing the state of the model and gets a reward (or no reward) in response to that action. After several interactions, the scheme aims to maximize the overall reward of the network and converges. The details of states, actions and rewards for our case are provided below:

- **State**: It is considered that the agent is centrally controlled and has the information regarding the transmit power, channel gains and energy harvesting capacity. Then, at any time slot l, we define state as

$$s_l = [P_t^1, h_u^1 E_h^1, \ldots, P_t^l, h_u^l E_h^l \ldots, P_t^N, h_u^N E_h^N]. \tag{13}$$

  Note that N is the total number of BS in a geographic area from all the NSPs. More specifically, N can be defined as $N = M(J+1)$.
- **Action**: During each episode, the agent tries to optimally allocate the resources by performing VUE association and spectrum sharing for all VUEs and BSs.

$$a_l = [\beta_u^1, \widehat{\alpha_u^1}, \ldots \beta_u^l, \widehat{\alpha_u^l}, \ldots, \beta_u^N, \widehat{\alpha_u^N}]. \tag{14}$$

- **Reward**: After performing some action during each epoch, the agent is going to receive a reward. In general the reward is associated to the objective function. For the case of our network setup, the main objective is

to maximize the payoff of the NSPs[1]. Therefore, the reward of agent is positively related to the amount of data exchanged between the VUEs in the network. We define the immediate reward $r_{l+1} \in R$ as the amount of information successfully transferred to the VUEs.

$$R(s_l, a_l) = \begin{cases} +1, \text{if } \Theta > \Xi \\ 0, \text{Otherwise}, \end{cases} \tag{15}$$

where $\Xi = \Upsilon_i^m(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) + \Psi_i^m(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}})$ and $\Theta = \sum_{u \in \mathcal{U}_m} \zeta_u U_u^m(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}})$.

For any l-th time slot, the Q-learning uses the value function $Q(s_l, a_l)$ for each state-action pair. This value is stored in a table called Q-table, which is regarded as the long-term reward of the agent. Given a certain policy π, it can be written as

$$Q(s, a) = \mathbb{E}\left\{ \sum_{g=0}^{G} \Omega^g r_{l+g+1} | s_l = s, a_l = a, \pi \right\}, \tag{16}$$

where G represents the length of one episode and $0 \leq \Omega \leq 1$ is discount factor. In this regard, it is worth mentioning that the agent has to make a tradeoff between exploration and exploitation. Thus, the agent would have to decide whether it needs to focus on the current immediate reward or further explore the environment for future rewards. This is done by the discount factor in the learning setup. If Ω tends toward 1 then it means that the agent focuses on the exploration, whereas, when Ω approaches 0 then it considers the immediate reward. As a result of this, the optimal Q-function which satisfies the Bellman optimality equation can be given as

$$Q^*(s, a) = \max_{\pi} Q^{\pi}(s, a). \tag{17}$$

One of the simplest ways to choosing the appropriate action by the agent is using the greedy method. This policy selects the largest Q-value in each state, such that

$$\pi(s) = arg \max_{a} Q^{\pi}(s, a). \tag{18}$$

This is called Q-learning process which is an iterative process having following procedure

$$Q_{l+1}(s_l, a_l) = Q_l(s_l, a_l) + \kappa_l \Big\{ r_{l+1} + \Omega \max_{a} Q_l(s_{l+1}, a) - Q_l(s_l, a_l) \Big\}, \tag{19}$$

where κ_l represents the learning rate at l-th time step. The main operations of the proposed Q-learning approach are provided in Algorithm 1. The algorithm begins by defining various environmental variables such as discount factor, decay rate, states (13) and actions (14). The environment simulation begins by deploying BSs and VUEs in the network and for each BS, there is a table that consists of all possible states as its rows and actions as its columns. For each time slot, the algorithm selects a rate of exploration and chooses a reward as per uniform distribution while comparing it with the exploration rate. If the value of r is greater than the exploration rate, then the Q-value maximizing action is selected. The action is performed and the reward is calculated (as per (15)) while updating the table entries according to (19). In the end, the state is updated and the loop continues. This process outputs the optimized sequence of VUE association and spectrum sharing factor for different actions on the network.

[1]Our network setup does not take into account the competition among different NSPs. Such game theoretic communication competitive network architecture of NSPs would be considered in the future works.

Algorithm 1 Q-learning for UE Association and Spectrum Sharing.

procedure *Q-Learning Process*
Input: Transmit power, channel gains and energy harvesting capacity.
Start:
  Define states and actions as per (13) and (14)
  Define the discount factor Ω.
  Define the threshold for minimum exploration Ω_{min}
  Define decay rate λ
  Initialize states, actions, and overall utility.
  for each episode **do**
  Start a heterogeneous network simulator.
    while each time slot **do**
    Select the rate of exploration $\Omega := \max(\Omega.\lambda, \omega_{\min})$
    Randomly select r using uniform distribution.
      if $r \leq \Omega$ **then**
        Randomly select an action a.
      else
        Choose the action for maximizing the reward, i.e., $a := \arg\max_a Q(s, a)$.
      end if
    Calculate the reward r_l as per (15) for action a_l.
    Evaluate the impact of a_l on s_{l+1}.
    Update the table entry as per (19).
    Switch to the next state, i.e., s_{l+1}.
    end while
  end for
Output: Optimized sequence of VUE association and spectrum sharing for different action on network.
end procedure

B. Proposed Deep Learning Solution for Power Allocation

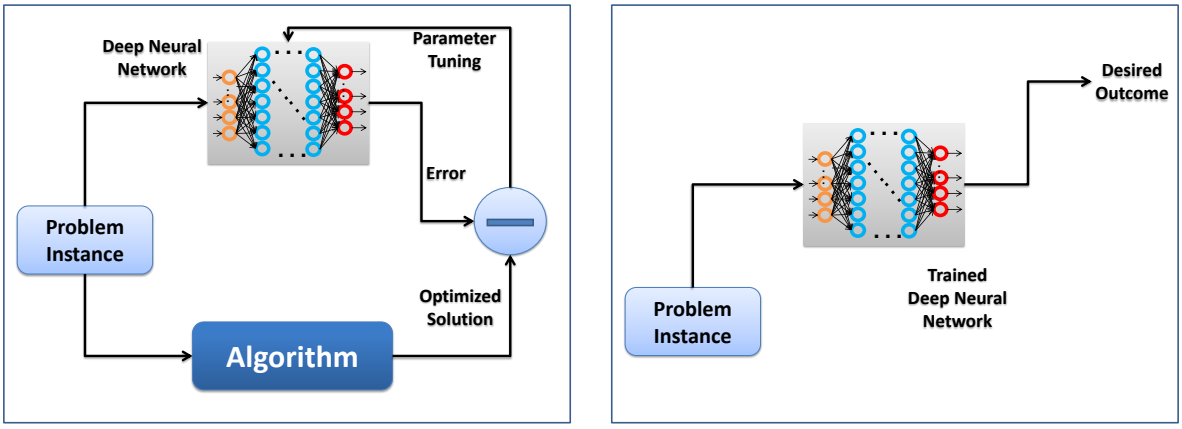
In this section, we provide a solution for optimal power allocation by fixing the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Thus, **P1** can now be written as

$$\textbf{P4} \quad \max \sum_{i \in I} \sum_{m \in M} \sum_{u \in \mathcal{U}_m} \zeta_u U_u^m(\boldsymbol{P}) - \sum_{i \in I} \sum_{m \in M} \Upsilon_i^m(\boldsymbol{P}) - \sum_{i \in I} \sum_{m \in M} \Psi_i^m(\boldsymbol{P}) \tag{20}$$

$$\text{s.t.} \quad \text{C2}: \ 0 \leq P_t^{m,j} \leq P_{\max}^j \tag{20a}$$

$$\text{C3}: \ 0 \leq E_h^{m,0} \leq B. \tag{20b}$$

Due to the reformulated problem, the solution to **P4** can be obtained by using classical approaches such as weighted minimum mean square error [67] or steepest descent method [68]. However, such approaches are non-scalable due to the high computational cost incurred by these algorithms. Moreover, due to the involvement of complex matrix inversion and



(a) (b)

Figure 5. Deep learning solution for power allocation (a) Training setup (b) Testing setup.

bisection during each iteration, the real-time implementation of these algorithms is a challenging task.

In order to maintain realistic resource management in wireless networks (due to the effect of fast fading and dynamic conditions), we treat the algorithms for power allocation as a 'black box' and aim to learn the complex interplay of system variables. Thus, we build a deep neural network for optimizing the allocation of power by approximating a weighted minimum mean square error algorithm. By obtaining the optimal power factors for the particular set of inputs from the said algorithm, the deep neural network can be trained to learn the relationship between inputs and outputs, as shown in Fig. 5 (a). Since the validation/ testing requires only simple operations like matrix multiplication, such a neural network can not only reduce the computation complexity but also help in minimizing the processing time. Moreover, with the recent developments in cloud computing, training these networks is fairly convenient. It is because of such a neural network can be trained offline in the cloud and the trained model can be utilized for real-time testing and validation.

1) Deep Learning Model: Our deep learning model uses a neural network approach to learn the complex interdependence of inputs and outputs. Specifically, the neural network is composed of a number of hidden layers along with having a single input and a single output layer. The input layer takes the channel coefficients and provides power allocation at the output layer. Besides this, for the hidden layers, we use rectifier linear unit (ReLU) as the activation function, whereas, to ensure the energy constraint, the output layer uses a special type of activation function such as

$$O = \min(B, \max(y, 0)), \tag{21}$$

where O denotes the activation function output. Without loss of generality, we consider that the maximum transmits power of all the BS is constant such that $P^{j}_{\max} = P_{\max}, \ \forall j \in J$. Also, it is worth pointing out that the expression in (21) does not contain the constraint on maximum power. It is because maximum power is always greater than the energy harvesting capabilities of the VUEs, such that $0 \leq E_h^{m,0} \leq B < P_{\max}$. As previously mentioned, we fix the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and for each tuple, i.e., channel realization, maximum transmit power and noise, we generate optimized power vectors from the weighted minimum mean square error algorithm. Subsequently, we obtain the training sample as the tuple of the channel and transmit power. Then, the process is repeated multiple times to obtain the entire training, validation and testing data set. The validation data is generated for cross-validation, early stoppage during the training, and appropriate model selection.

2) Training and Testing of Model: In order to optimize the weights of our neural network, we use the entire training set of channel realizations and transmit power. We have employed RMSProp as the optimization algorithm which is a reliable construction of the stochastic gradient descent method. By running the average of the recent gradients, the RMSProp divides the learning rate for that weight. Through exhaustive cross-validation, we select the decay rate of learning as 0.9 and select an appropriate learning rate. The truncated normal distribution has been used to initialize the weights of the neural network which also improves the performance of our neural network. More specifically, we generate the weights from a normal distribution, however, value is dropped and regenerated if the random number generated has an absolute value is greater than 2. Subsequently, to ensure the normalization of variance at the output, the weights of a neuron are divided by the square root of a total number of inputs. In this way, we normalize the variance of each output of the neuron.

For the sake of testing our deep learning model, we use the same distribution for generating channel realizations as used for the training phase. These generated channel realizations are passed through our deep neural network and the optimized powers are obtained at the output, as shown in Fig. 5 (b). Subsequently, the utility function in (20) is computed using the transmit power from the neural network. The results show that the obtained power allocation improves the performance of the network.

C. Computation Complexity

Our proposed optimization framework uses two learning techniques, i.e., Q-learning and supervised learning (i.e., deep

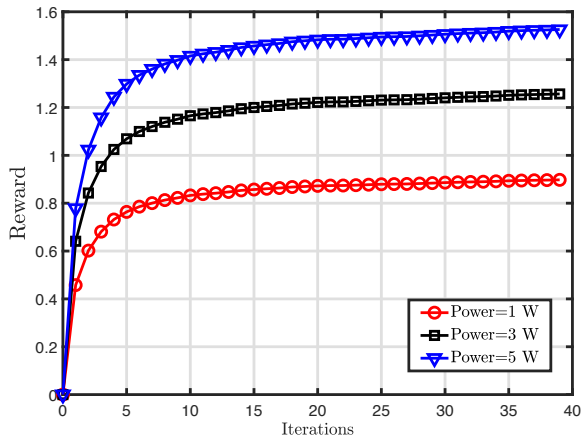


Figure 6. Reward as a function of the number of iterations.

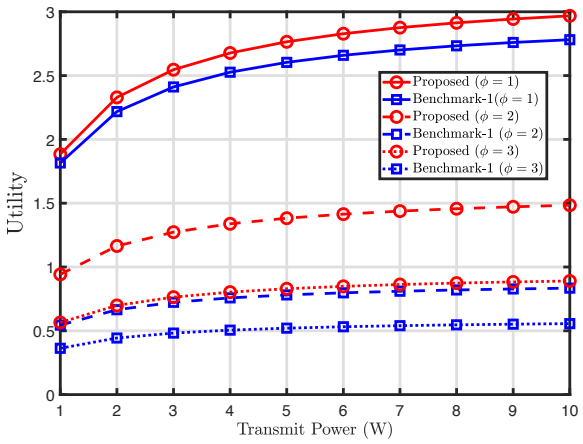


Figure 7. Utility of NSPs against BSs transmit power.

neural networks). In general, the complexity of both learning techniques varies significantly and strictly depends on the application. For instance, a larger amount of data may require more training time for the convergence of the learning model. In a similar manner, an increase in the number of hidden layers may result in consuming more time in order to reliably train the models. Thus, due to the inherent "black box" nature and hierarchical structure of these learning models, it is difficult to provide exact computational complexity during training. Additionally, since inappropriate training may lead to issues like under-fitting and over-fitting, the hyperparameter tuning and corresponding training duration may vary from one scenario to another.

However, despite these indefinite characteristics of learning techniques, many studies indicate a reasonable complexity under basic training conditions (see [69] and references therein). Moreover, since the training can be performed in an independent manner, the complexity of such learning techniques is less of an issue. The training of these models can be performed in the cloud and, then, the trained models can be used for testing in real-time. After a pre-specified period of time, these models can be completely or partially retrained in the cloud [70]. Since the resources in the cloud are less constrained, retraining can be performed frequently. The analysis of the frequency of retraining these models is beyond the scope of this work.

VII. Performance Evaluation

This section provides a detailed discussion of the simulation results as per the abovementioned analysis. The position of MBS is kept fixed, whereas, the SBSs are randomly deployed in a coverage area of MBS. Then, fixing the location of the SBS, the locations of VUEs are changed during each iteration. For performing Monte-Carlo simulations, the total number of iterations is 10,000, whereby, to reduce the effect of randomness, average values are taken in the simulations. During each simulation, consider that there are two NSPs and each NSP owns one MBS. In addition, two SBS associated with the MBS of a particular NSP.

Table II
Simulation Parameters and their Value.

Simulation Parameters	Values
Transmit Power P_t	1 W
Fading Channel	Rayleigh fading
Realizations	10^4
No. of NSPs	2
No. of SBS per MBS	2
Cost of MBS $\psi_{m,1}$	100 units/MHz
Cost of SBS $\psi_{m,2}$	80 units/MHz
Price of Backhaul Resources ω_m	1 unit/ Mbps
Learning rate	0.2
Bandwidth W_m	20 MHz
Reflection Coefficient μ	0.5
Hidden layers	5
Energy Conversion Efficiency η	0.8
Decay rate	0.9
Epochs	1000

For a fair comparison, we have provided two benchmark schemes in the results. "Benchmark-1" refers to the fixed resource allocation and VUE association scheme (also known as hard slicing [71]). In that, each BS is allocated a fixed share of resources no matter if they use it or not. The second benchmark scheme called "Benchmark-2" refers to the traditional maximum power allocation strategy, such that a BS always transmits at its maximum power [72]. This scheme disregards any instantaneous QoS requirements of the VUEs and traffic conditions in the network. Unless mentioned otherwise, the simulation parameters and their values are provided in Table II.

In Fig. 6, we have illustrated the results for Q-learning reward as a function of the number of iterations. In general, it can be seen that the reward increases with an increase in the number of iterations. This shows that if sufficient time is spent by the agent in exploring the environment, then it can maximize its long-term reward. Moreover, it can also be observed that with an increase in transmit power, the overall reward of the agent increases. This increase can be attributed to the fact that an increase in transmit power improves the overall throughput of the network. However, one can also note that with an increase in the value of transmit power, it

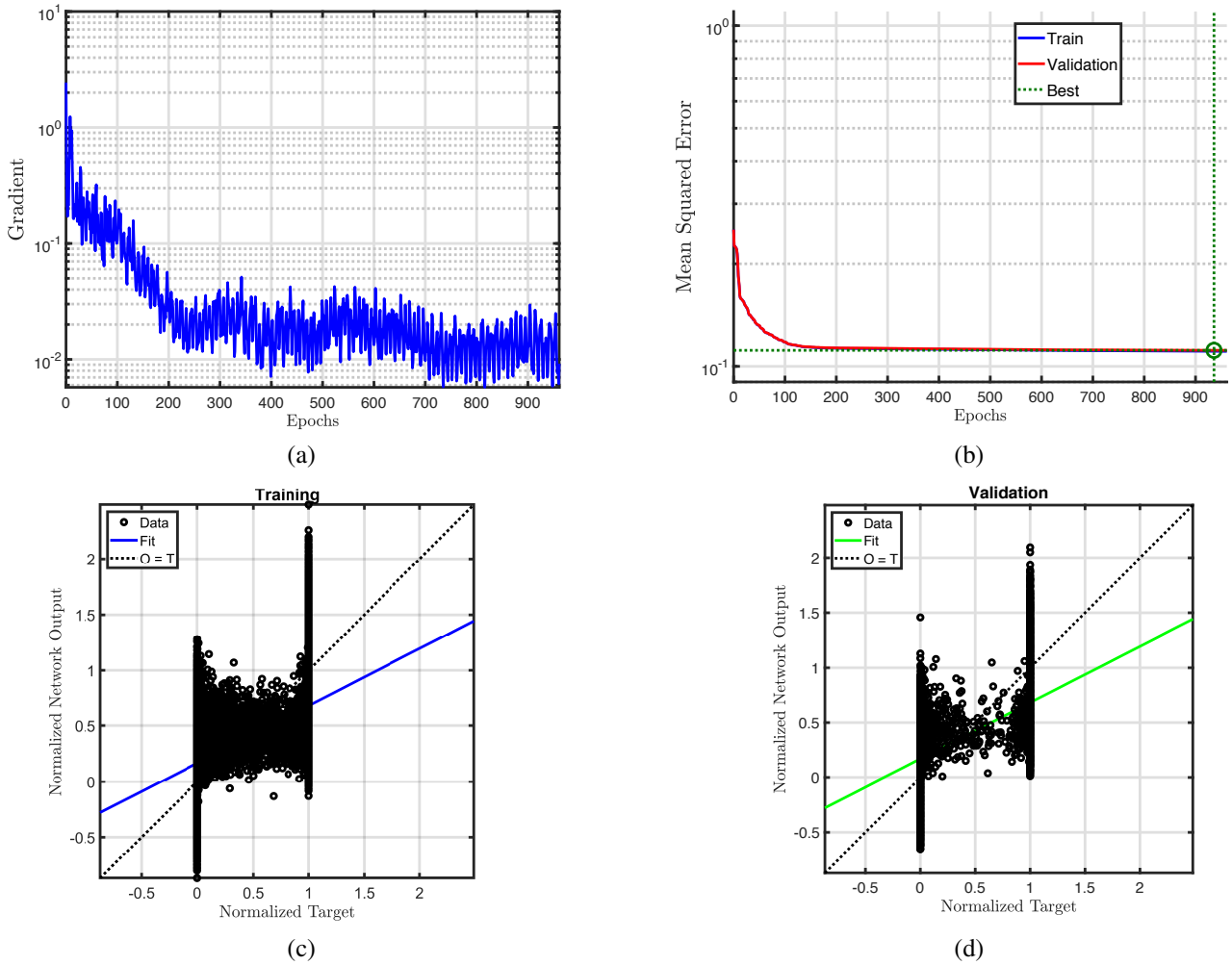


Figure 8. Traing and validation plots. (a) Gradient versus the number of epochs (b) Mean square error against the number of epochs (c) Regression plot for training data (d) Regression plot for validation data.

becomes difficult for the agent to reach a point of stability due to increased interference in the network. Thus, it takes a larger number of iterations to converge when the transmit power is increased from 1 W to 5 W.

Fig. 7 compares our proposed Q-learning technique with the benchmark greedy policy by plotting utility against the increasing values of transmit power of BSs, where all BSs under NSPs are assumed to transmit with the same power. Note that the overall utility generally increases with an increase in the transmit power. Moreover, it can be seen that the proposed learning framework outperforms the conventional benchmark technique. For providing more in-depth insights, we have proved curves for different values of the ratio of costs of MBS and SBS, i.e., ϕ. It can be seen that when $\phi = 1$, the curves of proposed and benchmark techniques are very close. This indicates that the greedy policy performs better when the ratio is small. However, as the value of ϕ increases both proposed and benchmark techniques suffer to maintain the utility at the same level, thereby, resulting in loss of utility. Yet, the proposed technique greatly outperforms the benchmark technique. It is also evident that the gap between the curves of proposed and benchmark increases. This shows the inability of the benchmark technique to cope with changes and dynamicity in the model.

Fig. 8 generally demonstrates the training and validation process of the proposed deep neural network approach for improving the overall utility. In particular, Fig. 8 (a) shows the reduction in the value of gradient over a different number of epochs. It is worth noting that we ran 1000 epochs for training our neural network. From the figure, it can be observed that the value of the gradient first decreases rapidly and becomes almost stable after crossing 700 epochs. As a result of this descent in the value of gradient, the mean square error at the output of the neural network drops. This phenomenon is illustrated in Fig 8 (b). It can also be observed from the figure that the validation error generally decreases with an increase in the amount of training. Also, for the specified number of epochs, the lowest value of mean square error is reached at around 950-th epoch. Fig 8 (c) & (d) demonstrate the regression plots for both training and validation data. It can be seen that the training and validation curves fit the data. This fitting helps the neural network to predict the response when data is varied. Note that for the generated data, the fit is reasonably good.

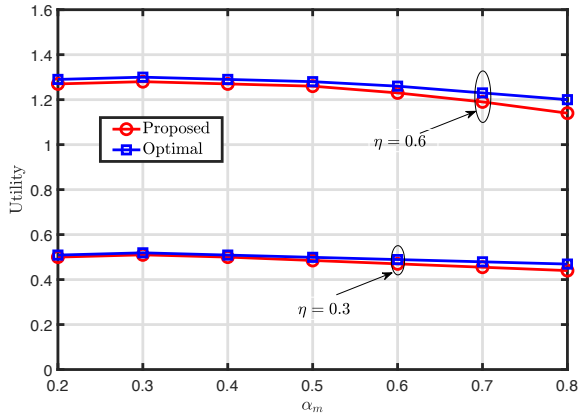


Figure 9. Utility of NSPs against the spectrum allocation factor.

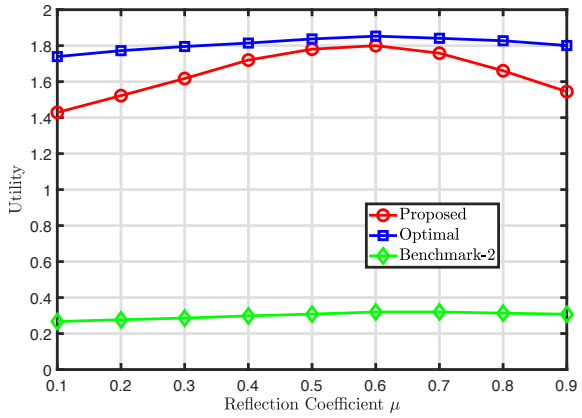


Figure 10. Utility of NSPs against the reflection coefficient of backscatter VUEs.

Fig. 9 shows a comparison of achievable utility for the optimal and proposed learning-based framework. In general, we note that the utility for both optimal and proposed techniques first increases and then gradually decreases. However, the separation between proposed and optimal increases for larger values of α_m. To understand the impact of energy conversion efficiency, we demonstrate the optimal and proposed curves for different values of η. It can be seen that a higher value of energy conversion efficiency ensures higher utility. By contrast, for $\eta = 0.3$, the overall utility decreases for both optimal and proposed techniques. We also note that the impact of energy conversion efficiency significantly diminishes at lower values. Thus, indicating that for a poor energy harvesting hardware, the performance of optimal and proposed techniques becomes almost identical. This results in reducing the gap between the utility curves of optimal and proposed techniques.

In Fig. 10, we provide a comparison between the optimal, proposed and benchmark techniques. In general, we note that the overall utility generally increases with an increase in the reflection coefficient and then decreases. This is because an increase in the value of the reflection coefficient indicates that the more part of the incident signal is reflected to the receiver. However, as this value continues to increase, the interference level from the other backscatter tags also grows which contributes to a decline in the overall utility. Despite this, we can observe that the proposed neural network-based approach closely follows the optimal technique. However, the proposed learning-based technique is fairly simple as it requires only linear multiplications once the model is fully trained. In contrast, the iterative optimization method is rigorous and complex, which may not be desirable for large-scale heterogeneous networks.

VIII. Conclusions and Future Work

The collaboration centric and heterogeneous backscatter communications will enable connected and smart vehicular networks. In this regard, this article has provided some key insights on optimizing such networks. More specifically, this work improves the performance of heterogeneous backscatter vehicular networks with the help of learning techniques, including reinforcement learning and supervised deep learning. The proposed optimization framework uses Q-learning for VUE association and spectrum sharing, whereas, power allocation was performed with the help of deep neural networks. The results clearly indicate that the utility of NSPs can be more enhanced using the learning-based framework as compared to conventional benchmark techniques. It was also demonstrated that the proposed learning-based optimization framework entails reasonable complexity which is suitable for large-scale heterogeneous backscatter vehicular networks. We anticipate that the results provided here would significantly contribute to the widespread deployment of heterogeneous backscatter vehicular networks.

Although the results provided by our work show considerable promise, many extensions can be derived from this study. For instance, this work considers single antennas at the backscatter tags. However, with the emergence of energy-efficient schemes, backscatter tags having multiple antennas are also becoming popular. As an extension of this work, the performance of a heterogeneous vehicular network consisting of multi-antenna backscatter tags can be optimized. Another potential work can be done from the perspective of multiple access schemes. In this work, we consider orthogonal multiple access, however, we expect that using non-orthogonal multiple access (NOMA) can significantly improve the performance of such networks. This interesting work is also left for future studies.

Acknowledgment

This work was supported in part by the Academy of Finland under Project No. 319003 and EPSRC GCRF DARE: Distributed Autonomous and Resilient Emergency Management System under Project No. EP/P028764/1.

References

[1] W. U. Khan, F. Jameel, T. Ristaniemi, S. Khan, G. A. S. Sidhu, and J. Liu, "Joint Spectral and Energy Efficiency Optimization for Downlink NOMA Networks," *IEEE Trans. Cogn. Commun. Netw.*, pp. 1–1, 2019.

[2] F. Jameel *et al.*, "Efficient power-splitting and resource allocation for cellular V2X communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3547–3556, June 2021.

[3] W. U. Khan *et al.*, "NOMA-enabled optimization framework for next-generation small-cell IoV networks under imperfect SIC decoding," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2021.

[4] M. A. Javed and S. Zeadally, "Ai-empowered content caching in vehicular edge computing: Opportunities and challenges," *IEEE Network*, vol. 35, no. 3, pp. 109–115, 2021.

[5] Z. Ali *et al.*, "Artificial intelligence techniques for rate maximization in interference channels," *Physical Communication*, vol. 47, p. 101294, 2021.

[6] F. Jameel *et al.*, "Reinforcement learning in blockchain-enabled IIoT networks: A survey of recent advances and open challenges," *Sustainability*, vol. 12, no. 12, p. 5161, 2020.

[7] U. M. Malik, M. A. Javed, S. Zeadally, and S. u. Islam, "Energy efficient fog computing for 6G enabled massive IoT: Recent trends and future opportunities," *IEEE Internet of Things Journal*, to be published, DOI: 10.1109/JIOT.2021.3068056.

[8] M. Z. Khan, M. Rahim, M. A. Javed, F. Ghabban, O. Ameerbakhsh, and I. Alfadli, "A d2d assisted multi-hop data dissemination protocol for inter-uav communication," *International Journal of Communication Systems*, vol. 34, no. 11, p. e4857.

[9] I. Elgendy, W. Zhang, and H. H. et al., "Joint computation offloading and task caching for multi-user and multi-task mec systems: reinforcement learning-based algorithms," *Wireless Networks*, vol. 27, pp. 2023–2038, 2021.

[10] M. S. Ebrahimi Shahabadi, H. Tabrizchi, M. Kuchaki Rafsanjani, B. Gupta, and F. Palmieri, "A combination of clustering-based under-sampling with ensemble methods for solving imbalanced class problem in intelligent systems," *Technological Forecasting and Social Change*, vol. 169, p. 120796, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040162521002286

[11] U. Farooq, M. W. Shabir, M. A. Javed, and M. Imran, "Intelligent energy prediction techniques for fog computing networks," *Applied Soft Computing*, vol. 111, p. 107682, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494621006037

[12] M. Rahim, S. Ali, A. N. Alvi, M. A. Javed, M. Imran, M. A. Azad, and D. Chen, "An intelligent content caching protocol for connected vehicles," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 4, p. e4231, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.4231

[13] M. Rahim, M. A. Javed, A. N. Alvi, and M. Imran, "An efficient caching policy for content retrieval in autonomous connected vehicles," *Transportation Research Part A: Policy and Practice*, vol. 140, pp. 142–152, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0965856420306893

[14] M. A. Naeem, T. N. Nguyen, R. Ali, K. Cengiz, Y. Meng, and T. Khurshaid, "Hybrid cache management in iot-based named data networking," *IEEE Internet of Things Journal*, pp. 1–1, 2021.

[15] F. Jameel *et al.*, "NOMA-enabled backscatter communications: Toward battery-free IoT networks," *IEEE Internet of Things Magazine*, vol. 3, no. 4, pp. 95–101, Dec. 2020.

[16] Z. Ali *et al.*, "Fair power allocation in cooperative cognitive systems under NOMA transmission for future IoT networks," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 575–583, 2022.

[17] W. U. Khan *et al.*, "Energy efficiency maximization for beyond 5G NOMA-enabled heterogeneous networks," *Peer-to-Peer Networking and Applications*, pp. 1–15, 2021.

[18] S. Zeadally, M. A. Javed, and E. B. Hamida, "Vehicular communications for its: Standardization and challenges," *IEEE Communications Standards Magazine*, vol. 4, no. 1, pp. 11–17, 2020.

[19] R. Long, H. Guo, L. Zhang, and Y.-C. Liang, "Full-duplex backscatter communications in symbiotic radio systems," *IEEE Access*, vol. 7, pp. 21 597–21 608, 2019.

[20] F. Jameel, T. Ristaniemi, I. Khan, and B. M. Lee, "Simultaneous harvest-and-transmit ambient backscatter communications under Rayleigh fading," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 166, 2019.

[21] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, 2019.

[22] P. Do, T. H. V. Phan, and B. B. Gupta, "Developing a vietnamese tourism question answering system using knowledge graph and deep learning," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 20, no. 5, Jun. 2021. [Online]. Available: https://doi.org/10.1145/3453651

[23] M. Hammad, M. H. Alkinani, and B. G. et al., "Myocardial infarction detection based on deep neural network on imbalanced data," *Multimedia Systems*, pp. 1–13, 2021.

[24] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, 2016.

[25] R. I. Ansari, H. Pervaiz, S. A. Hassan, C. Chrysostomou, M. A. Imran, S. Mumtaz, and R. Tafazolli, "A new dimension to spectrum management in iot empowered 5g networks," *IEEE Network*, 2019.

[26] W. U. Khan *et al.*, "Spectral efficiency optimization for next generation NOMA-enabled IoT networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 284–15 297, Dec. 2020.

[27] M. Matinmikko-Blue, S. Yrjölä, V. Seppänen, P. Ahokangas, H. Hämmäinen, and M. Latva-aho, "Analysis of spectrum valuation approaches: The viewpoint of local 5g networks in shared spectrum bands," in *IEEE DySPAN*. IEEE, 2018, pp. 1–9.

[28] A. Syed, K. Yau, H. Mohamad, N. Ramli, and W. Hashim, "Channel selection in multi-hop cognitive radio network using reinforcement learning: An experimental study," 2014.

[29] S. Tubachi, M. Venkatesan, and A. Kulkarni, "Predictive learning model in cognitive radio using reinforcement learning," in *IEEE ICPCSI*. IEEE, 2017, pp. 564–567.

[30] T. K. Vu, M. Bennis, M. Debbah, M. Latva-Aho, and C. S. Hong, "Ultra-reliable communication in 5G mmWave networks: A risk-sensitive approach," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 708–711, 2018.

[31] Y. Guo, R. Yu, J. An, K. Yang, Y. He, and V. C. Leung, "Buffer-Aware Streaming in Small Scale Wireless Networks: A Deep Reinforcement Learning Approach," *IEEE Trans. Veh. Technol.*, 2019.

[32] T. K. Vu, C.-F. Liu, M. Bennis, M. Debbah, and M. Latva-Aho, "Path selection and rate allocation in self-backhauled mmwave networks," in *IEEE WCNC*. IEEE, 2018, pp. 1–6.

[33] T. K. Vu, M. Bennis, M. Debbah, and M. Latva-Aho, "Joint Path Selection and Rate Allocation Framework for 5G Self-Backhauled mm-wave Networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2431–2445, 2019.

[34] S. Nie, Z. Fan, M. Zhao, X. Gu, and L. Zhang, "Q-learning based power control algorithm for D2D communication," in *IEEE PIMRC*. IEEE, 2016, pp. 1–6.

[35] S. Sharma and B. Singh, "Cooperative Reinforcement Learning Based Adaptive Resource Allocation in V2V Communication," in *IEEE SPIN*. IEEE, 2019, pp. 489–494.

[36] Y. Qiu, Z. Ji, Y. Zhu, G. Meng, and G. Xie, "Joint mode selection and power adaptation for D2D communication with reinforcement learning," in *IEEE ISWCS*. IEEE, 2018, pp. 1–6.

[37] A. Moussaid, W. Jaafar, W. Ajib, and H. Elbiaze, "Deep Reinforcement Learning-based Data Transmission for D2D Communications," in *IEEE WiMob*. IEEE, 2018, pp. 1–7.

[38] F. Jameel *et al.*, "Reinforcement learning for scalable and reliable power allocation in SDN-based backscatter heterogeneous network," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2020, pp. 1069–1074.

[39] ——, "Towards intelligent IoT networks: Reinforcement learning for reliable backscatter communications," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[40] T. T. Anh, N. C. Luong, D. Niyato, Y.-C. Liang, and D. I. Kim, "Deep Reinforcement Learning for Time Scheduling in RF-Powered Backscatter Cognitive Radio Networks," *arXiv preprint arXiv:1810.04520*, 2018.

[41] A. Rahmati and H. Dai, "Reinforcement Learning for Interference Avoidance Game in RF-Powered Backscatter Communications," *arXiv preprint arXiv:1903.03600*, 2019.

[42] Y. Hu, P. Wang, Z. Lin, M. Ding, and Y.-C. Liang, "Machine Learning Based Signal Detection for Ambient Backscatter Communications," in *IEEE ICC*. IEEE, 2019, pp. 1–6.

[43] S. Ma, Y. Zhu, G. Wang, and R. He, "Machine Learning Aided Channel Estimation for Ambient Backscatter Communication Systems," in *IEEE ICCS*. IEEE, 2018, pp. 67–71.

[44] X. Wen, S. Bi, X. Lin, L. Yuan, and J. Wang, "Throughput Maximization for Ambient Backscatter Communication: A Reinforcement Learning Approach," in *IEEE ITNEC*. IEEE, 2019, pp. 997–1003.

[45] W. U. Khan *et al.*, "Secure backscatter communications in multi-cell NOMA networks: Enabling link security for massive IoT networks," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2020, pp. 213–218.

[46] T. Hong, C. Liu, and M. Kadoch, "Machine learning based antenna design for physical layer security in ambient backscatter communications," *Wireless Communications and Mobile Computing*, vol. 2019, 2019.

[47] X. Fan, F. Wang, F. Wang, W. Gong, and J. Liu, "When RFID Meets Deep Learning: Exploring Cognitive Intelligence for Activity Identification," *IEEE Wireless Commun.*, p. 2, 2019.

[48] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Intelligent User Association for Symbiotic Radio Networks using Deep Reinforcement Learning," *arXiv preprint arXiv:1905.04041*, 2019.

[49] W. U. Khan, E. Lagunas, A. Mahmood, S. Chatzinotas, and B. Ottersten, "Integration of backscatter communication with multi-cell NOMA: A spectral efficiency optimization under imperfect SIC," *arXiv preprint arXiv:2109.11509*, 2021.

[50] M. Ahmed *et al.*, "Backscatter sensors communication for 6G low-powered NOMA-enabled IoT networks under imperfect SIC," *arXiv preprint arXiv:2109.12711*, 2021.

[51] W. U. Khan *et al.*, "Energy-efficient resource allocation for 6G backscatter-enabled NOMA IoV networks," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2021.

[52] X. Li *et al.*, "Physical layer security of cognitive ambient backscatter communications for green Internet-of-things," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1066–1076, Sept. 2021.

[53] W. U. Khan, X. Li, M. Zeng, and O. A. Dobre, "Backscatter-enabled NOMA for future 6G systems: A new optimization framework under imperfect SIC," *IEEE Communications Letters*, vol. 25, no. 5, pp. 1669–1672, May 2021.

[54] A. Ihsan *et al.*, "Energy-efficient backscatter aided uplink NOMA roadside sensor communications under channel estimation errors," *arXiv preprint arXiv:2109.05341*, 2021.

[55] W. U. Khan, F. Jameel, N. Kumar, R. Jäntti, and M. Guizani, "Backscatter-enabled efficient V2X communication with non-orthogonal multiple access," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1724–1735, Feb. 2021.

[56] F. Jameel *et al.*, "Multi-tone carrier backscatter communications for massive iot networks," in *Wireless-powered backscatter communications for Internet of things*. Springer, 2021, pp. 39–50.

[57] W. U. Khan, N. Imtiaz, and I. Ullah, "Joint optimization of NOMA-enabled backscatter communications for beyond 5G IoT networks," *Internet Technology Letters*, vol. 4, no. 2, p. e265, 2021.

[58] F. Jameel *et al.*, "Time slot management in backscatter systems for large-scale IoT networks," in *Wireless-powered backscatter communications for Internet of things*.

[59] W. U. Khan *et al.*, "Secure backscatter-enabled NOMA system design in 6G era," *Internet Technology Letters*, p. e307.

[60] F. Amato, C. W. Peterson, B. P. Degnan, and G. D. Durgin, "Tunneling RFID tags for long-range and low-power microwave applications," *IEEE J. Radio Freq. Identif.*, vol. 2, no. 2, pp. 93–103, 2018.

[61] R. Duan, E. Menta, H. Yiğitler, and R. Jäntti, "Hybrid Beamformer Design for High Dynamic Range Ambient Backscatter Receivers," *arXiv preprint arXiv:1901.05323*, 2019.

[62] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, 2013.

[63] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[64] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, 2006.

[65] S. Gortzen and A. Schmeink, "Optimality of dual methods for discrete multiuser multicarrier resource allocation problems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3810–3817, 2012.

[66] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[67] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.

[68] N. Wang, C. He, T. A. Gulliver, and V. K. Bhargava, "Generalized queue-aware resource management and scheduling for wireless communications," *IEEE Access*, vol. 3, pp. 1298–1312, 2015.

[69] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PloS one*, vol. 13, no. 3, p. e0194889, 2018.

[70] X. Peng, K. Ota, and M. Dong, "Edge computing based traffic analysis system using broad learning," in *IEEE AICON*. Springer, 2019, pp. 238–251.

[71] N. Huin, J. Leguay, S. Martin, P. Medagliani, and S. Cai, "Routing and Slot Allocation in 5G Hard Slicing," 2019.

[72] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Trans. Veh. Tech.*, vol. 65, no. 3, pp. 1652–1661, 2015.