Coordinated Spatial Reuse With Shared AP Index Optimization for IEEE 802.11be WLAN Enhancement

Jaewook Jung, Gangwoo Lee, and Jong-Moon Chung, *Fellow, IEEE*

***Abstract*—As Internet of Things (IoT) technology rapidly develops, the use of IEEE 802.11 wireless local area networks (WLANs) has significantly increased. To meet further growing IoT demands, the IEEE 802.11be standards (Wi-Fi 7) include coordinated spatial reuse (CSR) technology to achieve a higher area sum throughput in dense WLAN environments. The IEEE 802.11be CSR scheme adjusts the transmission power considering interference through coordination between access points (APs). However, in addition to adjusting the transmission power, selecting a shared AP to perform concurrent transmission is also important in improving the area sum throughput performance. For this purpose, in this paper, SAIOC scheme is proposed, which aims to find optimal transmission power and shared AP index to maximize expected sum throughput derived based on a semi-Markov model developed. The proposed SAIOC system uses multi-agent deep reinforcement learning (MA-DRL) to efficiently maintain an optimized performance in dynamically changing conditions. The simulation results show that the proposed scheme can provide an improved performance compared to other related benchmarked schemes.**

***Index Terms*—IEEE 802.11be, coordinated spatial reuse, spatial reuse, deep reinforcement learning, multi-agent, Wi-Fi 7.**

I. Introduction

IEEE 802.11be task group (TG) TGbe began standardization in May of 2019 to further improve the IEEE 802.11ax standards (which were published in May of 2021), where the D7.00 version of the IEEE 802.11be standards has been released and the official publication will be released in May of 2025. The IEEE 802.11be standards' objective is to support extremely high throughput (EHT) to satisfy the data rate requirements of upcoming applications, such as 4 K/8 K video, augmented reality (AR), extended reality (XR), and online game services [1], [2]. The key components of the IEEE 802.11be standards (Wi-Fi 7) include multi-access point (AP) coordination, which aims to mitigate co-channel interference through sharing necessary information, such as channel state information (CSI) between adjacent APs, and multi-link operation (MLO), which utilizes multiple links through multiple radio interfaces simultaneously for transmission [3]. Along with these developments, IEEE 802.11 wireless local area networks (WLANs) support Internet of Things (IoT) services with significant advantages in cost effectiveness, simple deployment, backward compatibility, and pervasive connectivity with various devices. As a result, the global IEEE 802.11 WLAN IoT market is expected to grow to an estimated $3.2 billion by 2030 [4]. To deal with the significantly increasing number of WLAN devices, the IEEE 802.11 standards introduce new features to enhance the spectral efficiency in dense deployment environments. The overlapping basic service set (OBSS) packet detection (PD) based spatial reuse (SR) scheme is included in the IEEE 802.11ax (Wi-Fi 6) high efficiency WLAN (HEW) standards [5]. The OBSS PD based SR scheme can enhance the spectral efficiency by enabling adjustment of parameters for channel contention. The OBSS PD based SR scheme adjusts the clear channel assessment (CCA) threshold along with the transmission power taking into account the interference from concurrent transmissions. However, because the IEEE 802.11ax standards only specify the boundary values of the parameters used to control OBSS PD based SR without specifying how to control these parameters, OBSS PD based SR systems commonly cannot provide an optimal performance.

The IEEE 802.11be standards improve the area sum throughput using the coordinated SR (CSR) scheme. CSR improves the area sum throughput through concurrent transmission by sharing transmission opportunities (TXOPs) (which is acquired through the backoff process) with other APs. In CSR, an AP that has acquired a TXOP and shares it is referred to as a 'sharing AP,' and an AP that has received a TXOP shared by a sharing AP is referred to as a 'shared AP.' In addition to sharing a TXOP, the CSR scheme adjusts the transmission power using link information from coordination between the APs, which can include received signal strength indication (RSSI) information for interference management [6]. Fig. 1 shows an example of adjusting the transmission power to mitigate the interference range between concurrent transmission so as not to affect each other's transmission. Since the CSR scheme adjusts its transmission power considering the influence of interference from concurrent transmission, the CSR scheme can provide an improved performance when compared to the OBSS PD based SR scheme [7]. Since CSR concurrent transmission induces

Received 16 November 2024; revised 8 May 2025; accepted 4 July 2025. Date of publication 14 July 2025; date of current version 21 November 2025. This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) funded by the Ministry of Land, Infrastructure and Transport under Grant RS-2022-00143782 of the Republic of Korea. Recommended for acceptance by H. Elsawy. *(Corresponding author: Jong-Moon Chung.)*

Jaewook Jung and Jong-Moon Chung are with the School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea (e-mail: qazaq9669@yonsei.ac.kr; jmc@yonsei.ac.kr).

Gangwoo Lee was with the School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea. He is now with Korea Telecom (KT) Seongnam-si 13606, South Korea (e-mail: gang5541@yonsei.ac.kr).

Digital Object Identifier 10.1109/TNSE.2025.3588679

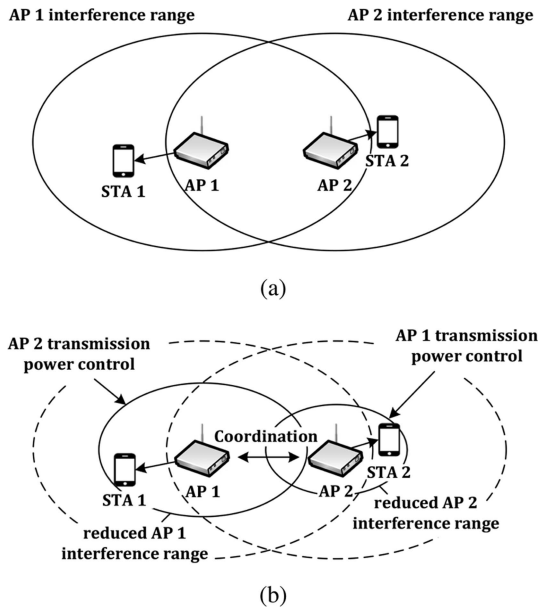


Fig. 1. Characteristics of the CSR operation (a) prior to transmission power control (where the nodes are subject to each other's interference range) and (b) after transmission power control (in which mutual interference is mitigated).

interference to other nodes, the overall network performance also depends on the selection of the shared AP by the sharing AP, which may not be maximized despite the adjustment of transmission power. Therefore, the sharing AP should optimally determine the shared AP index for concurrent transmission to enhance the area sum throughput performance.

However, even the process of deriving the optimal transmission power alone (excluding the shared AP index) that can maximize the area sum throughput is a strong NP-hard problem, since it is a sum-rate maximization problem that is equivalent to a reduced combinatorial optimization problem on a graph [8]. As there is no algorithm that can solve NP-hard problems in polynomial time, it is difficult to derive an optimal solution of a NP-hard problem analytically, and even harder to approximate due to its complexity. Even though there are approximation algorithms for NP-hard problems, only a worst-case solution can be derived in which this process will have low scalability [9]. For this reason, heuristic algorithms that search for near optimal solutions through trial and error are commonly used to solve NP-hard problems. However, proper heuristic algorithms need to be selected among numerous heuristic algorithms, based on considerations of the domain specific properties of the problem, and extensive tailoring and knowledge are required to find near optimal solutions. In addition, heuristic methods should be re-executed to solve the problem whenever the environment changes. For this reason, applying heuristic algorithms to dynamically changing network environments makes it very challenging to find the optimal solution adaptively. To overcome these drawbacks, reinforcement learning (RL) is considered as a promising technology to effectively deal with NP-hard problems [10], [11]. RL systems search for the optimal action that can maximize the reward among selectable actions for each state through interaction with the environment. Since RL does not rely on pre-existing knowledge of the environment and performs learning based on interaction with the environment, it is an effective tool to use in solving NP-hard problems and facilitates adaptive decision-making in dynamically changing network conditions [12].

This motivated the design of the shared AP index optimized CSR (SAIOC) scheme, which can find both the optimal transmission power and the shared AP index that can maximize the area sum throughput. In this paper, the saturated transmission probability of CSR is derived based on the developed semi-Markov chain model to consider the backoff process and transmission duration. The optimization problem with respect to the transmission powers and shared AP indexes is formulated using the derived saturated transmission probability, which aims to maximize the expected area sum throughput based on the constraints. The proposed SAIOC scheme applies multi-agent deep RL (MA-DRL) to enable parallel processing, which maximizes common rewards through cooperation between agents. This objective is consistent with the goals of the IEEE 802.11be CSR scheme, which improves the area sum throughput performance by utilizing information from coordination between APs. As transmission only occurs when the backoff counter expires or the AP is designated as a shared AP, action asynchrony arises when applying RL to derive the optimal transmission power and shared AP index. For this purpose, the MA-DRL based SAIOC optimization framework is constructed to find the optimal transmission power and shared AP index, in which the state space, action space, and reward function are configured to address asynchrony of action. The contributions of this paper are summarized as follows.

1) This paper proposes the SAIOC scheme that maximizes the area sum throughput in IEEE 802.11be WLANs. The proposed scheme can find the optimal transmission power and shared AP index that maximizes the expected sum throughput using MA-DRL based on collaborative rewards.
2) The saturated transmission probability of CSR is derived based on a semi-Markov chain model developed in this paper. Based on the derived saturated transmission probability, an optimization problem to maximize the expected sum throughput is formulated with respect to the transmission power and shared AP index.
3) The SAIOC optimization framework uses MA-DRL to maintain the optimal transmission power and shared AP index, where the state space, action space, and reward function are configured considering the asynchrony of actions due to the backoff process.
4) The performance of the SAIOC scheme is evaluated based on Python simulation. The simulation results demonstrate that the proposed SAIOC scheme can outperform the benchmark schemes in terms of various metrics.

The subsequent sections of this paper are organized as follows. Section II presents the related work of the SR schemes. In Section III, the system model is described. In Section IV, the process of constructing the optimization problem is described and the SAIOC framework that can derive the solution of the

optimization problem is described. In Section V, the performance of the proposed SAIOC scheme is analyzed by evaluating and comparing with other benchmark schemes. In Section VI, summary and discussion of the proposed SAIOC scheme are provided. Finally, Section VII concludes the paper.

II. Related Works

Due to the potential of enhancing the spectral efficiency in dense deployment environments, much research on SR has been conducted, focusing on techniques such as transmission power control (TPC) and dynamic sensitivity control (DSC) (which controls the CCA threshold) to overcome the disadvantages of the IEEE 802.11 WLAN distributed coordination function (DCF). In [13], the authors analyze the relationship between the throughput, packet loss rate (PLR), and the physical carrier sensing (PCS) threshold, and propose a QoS-aware heuristic algorithm to adjust the PCS threshold based on a derived analytical model. In [14], a power controlled multiple access (PCMA) scheme which considers the acceptable boundary of the interference to adjust the transmission power is proposed. In [15], an incremental-power carrier-sensing (IPCS) scheme, which enhances physical carrier-sensing to improve the transmission performance through periodic interference monitoring and PCS threshold control is proposed. In [16], the authors propose a power and rate control algorithm (PRC) that controls the transmission power and data rate in a decentralized manner based on the signal interference level. In [17], the authors propose a CCA threshold optimization method that considers the capture threshold to increase the packet transmission success rate. In [18], a TPC and DSC scheme in which an artificial neural network (ANN) based controller configures all users to improve the fairness performance is proposed. In [19], the authors propose a dynamic CCA threshold (DCCA) that sets the CCA threshold using a heuristic mechanism, depending on whether the transmission is successful or not. In [20], a location-enhanced DCF (LED) algorithm that performs interference prediction and blocking assessment using the location information in the frame header is proposed. In [21], the authors propose a CCA threshold optimization scheme by comparing the interference range and carrier sense range in a mesh network environment. In [22], a CCA threshold control method that performs transmission without frame loss, according to the polar coordinates of STAs detected by the AP through a set of cameras is proposed. In [23], the authors perform simulations while adjusting the CCA threshold under an IEEE 802.11n network environment and present an optimal CCA threshold adjustment method based on the results. In [24], the authors derive the network capacity in terms of transmit power and CCA threshold and show that SR is affected by the ratio of the transmission power to the CCA threshold, and propose a distributed power and rate control algorithm that allows each node to adjust its transmission power and data rate. In [25], a fine-grained adaptation of the carrier sense threshold (FACT) algorithm is proposed, which calculates the desired signal-to-interference ratio (SIR) based on radio frequency parameters, and adjusts the CCA threshold and transmission power based on the SIR. In [26], the authors propose the federated reinforcement multiple access (FRMA) algorithm, which performs transmit and wait configurations based on distributed DRL, and applies federated learning techniques to improve the transmission performance while preserving channel access fairness. In [27], the authors propose a dynamic CCA threshold adjustment scheme based on the transmission loss rate. In [28], the authors model a spatial channel selection game based on the distributed channel selection problem with user location, and propose a technique to select the optimal channel based on the location profile, transmission range, and interference graph. In [29], the authors propose a SRDCF that adjusts the transmission power based on location information and the required SIR, considering the signal capture phenomena. In [30], an estimation-based adaptive physical carrier sensing (APCS) scheme that dynamically adjusts the CCA threshold by estimating the interference range and carrier sense range using the minimum received power in a mesh network environment is proposed. In [31], the authors propose a method to formulate the SR problem regarding the CCA threshold and transmission power control as a multi-armed bandit problem and perform optimization through Bayesian optimization based on a Gaussian process.As the OBSS PD based SR scheme was officially incorporated into the IEEE 802.11ax (Wi-Fi 6) standards, research on the OBSS PD based SR scheme of IEEE 802.11 WLANs has been more active. In [32], the authors propose a control OBSS PD sensitivity threshold (COST) scheme, which can address the drawbacks of the existing dynamic sensitivity control (DSC) scheme by adjusting the OBSS_PD threshold adaptively, using the measured interference level and RSSI level. In [33], a RSSI to OBSS threshold (RTOT) scheme is proposed, which utilizes both a carrier sense threshold (CST) control scheme and a transmission power control (TPC) scheme. The RTOT scheme derives the OBSS_PD threshold based on the beacon RSSI and then derives the transmission power using the derived OBSS_PD threshold. In [34], the authors present an analytical model of the IEEE 802.11ax OBSS PD based SR scheme and derive the optimal OBSS_PD level to maximize the defined SR gain metric. In [35], the authors formulate optimization problems relating to the number of SR opportunities and packet error rate (PER) and propose a decentralized dynamic OBSS_PD level selection algorithm, which determines the optimal OBSS_PD threshold using the locally computed PER. In [36], the authors propose the optimized transmission power controlled OBSS PD based SR (OTOP) scheme, which derives the generalized transmission success probability based on stochastic geometry analysis and computes the optimal transmission power to maximize the transmission success probability of transmissions. In [37], the authors propose the rate-adaptive inter-BSS carrier elimination-based OBSS_PD threshold (RACEBOT) algorithm, which records each RSSI with their OBSS frame count (OFC) over a specific period and adjusts the transmission power and OBSS_PD threshold based on the OFC. In [38], the authors formulate the problem of selecting OBSS_PD levels for multiple users as a multi-armed bandit problem and propose a method to calculate the optimal OBSS level and transmission power based on Thompson sampling. In [39], the authors derive the optimal transmission power and CCA threshold which are used in the

proposed model-agnostic meta-learning (MAML) algorithm. In [40], the authors propose the link-aware SR (LSR) scheme, which performs simultaneous transmission by sharing the TXOP through distributed coordination among BSSs based on control information in the header. In [41], the authors propose a method to perform simultaneous transmission considering interference, which sets up the modulation and coding scheme (MCS) through repeated update Q-learning. In [42], the authors prove that the problem of setting the optimal transmission power and transmission channel is an exact potential game and propose a method for setting the optimal transmission power and transmission channel based on game theory. In [43], the authors propose the distributed Bayesian optimization for improving SR (INSPIRE) scheme, which derives the optimal transmission power and OBSS_PD level through Bayesian optimization based on a Gaussian process. In [44], the contextual bandit OBSS PD (CB-OBSS/PD) algorithm is proposed, which learns whether to perform or postpone transmission based on the state using contextual multi-armed bandits. In [45], the authors propose the interference-based dynamic channel algorithm (IB-DCA), which distinguishes between SR links and non-SR links and determines transmission based on the SR flag information in the header. In [46], the authors propose algorithms in which each agent performs learning based on a deep neural network (DNN) and uses federated learning to calculate the optimal transmission power. In [47], a method to derive the optimal OBSS_PD level of the master AP and slave AP based on the packet loss rate and the number of associated STAs using deep deterministic policy gradient (DDPG) is presented. In [48], the authors propose an algorithm to derive the transmission power and OBSS_PD level through Bayesian optimization based on a Gaussian process that maximizes the expected improvement (EI). In [49], the authors propose a deep Q network (DQN) based SR scheme in which each agent considers other agents as a part of the environment without cooperation and derives the CCA threshold and transmission power based on DQN.

As multi-AP coordination is mentioned as one of the main candidate features in the IEEE 802.11be Project Authorization Request (PAR), research is also being conducted to enhance the SR performance through multi-AP coordination. In [50], the authors develop an analytical throughput model and propose a coordinated time division multiple access/SR (c-TDMA/SR) scheduler that can identify the best subset of concurrent transmissions for TXOP sharing under a TDMA protocol. In [51], the optimal transmission powers for involved APs are derived to maximize the area sum throughput based on the path loss equation for each AP. In [52], the authors propose a scheduling technique that calculates the parameterized SR (PSR) transmission power based on an acceptable receiver interference level and allocates resources alternately between PSR-favorable and PSR-unfavorable transmissions. In [53], the authors propose an algorithm for CSR grouping and traffic scheduling for simultaneous transmission through multi-AP coordination. In [54], a DRL channel access (DRLCA) algorithm is proposed, which derives the optimal CCA threshold and contention window based on DQN, utilizing the knowledge contained in the beacon frame of each AP. In [55], the authors propose a DRL channel access (DLCA) algorithm, in which the centralized AP controller (APC) sets the channel configuration of the APs based on a greedy algorithm, while each AP learns whether to transmit or not on the assigned channel based on its own DQN model, and first-order model-agnostic meta-learning (FOMAML) is performed through the APC for fast convergence. In [56], a method to reduce the amount of information shared by deciding whether to learn information based on the Q-value criterion is introduced. The scheme determines whether to transmit and the MCS level using Q-learning. In [57], the authors propose a hierarchical multi-armed bandits (MAB) based CSR group selection scheme, where the first-level MAB agent learns whether APs will transmit in the current TXOP, and the second-level MAB agent decides the AP-station (STA) pairs. In [58], a bidirectional CSR algorithm that adjusts the transmission power and CCA thresholds by considering the performance of both sharing and shared APs, based on RSSI and transmission power information is proposed. In [59], the authors introduce a centralized CSR algorithm in which the centralized network controller calculates the transmission power and MCS level of concurrently transmitting APs based on the RSSI measurement report from the beacon frame. In [60], the authors propose a coordinated MAB solution for the CCA threshold and transmission power configuration, which derives the optimal SR parameters based on a greedy algorithm and Thompson sampling with shared rewards between agents. In [61], a deep MA-CMAB-based transmission power and CCA threshold adjustment algorithm is introduced, which applies DRL transfer learning techniques to adapt to dynamic environmental changes. In [62], the authors propose a SR grouping algorithm based on the information exchanged between APs and the interference model, in which simultaneous transmission is successfully performed.

In Table I, the differences between the proposed SAIOC and existing studies are summarized. Previous studies on SR schemes have primarily focused on optimizing SR parameters using rule-based algorithms (RBA), deep supervised learning (DSL), RL, and single-agent DRL (SA-DRL). Research applying RBA to SR techniques typically derives SR parameters based on predefined rules. However, such methods often assume deterministic network topologies, rely on heuristic approaches, or exhaustively search through all combinations to determine SR parameters. Consequently, these methods face challenges when attempting to apply them in dynamically changing network environments or general scenarios. In addition, there have been studies that apply artificial intelligence (AI) algorithms such as DSL, RL, and SA-DRL to derive SR parameters. These approaches typically operate under a centralized scheme, where a main controller computes the SR parameters for all APs. This can lead to the curse of dimensionality as the number of APs increases. Even in distributed settings where each AP computes its own SR parameters, single-agent settings are often used, which fail to account for the learning processes of other agents. As a result, these approaches may encounter stability problems, as each agent performs learning to maximize each reward independently. Furthermore, using simple metrics such as throughput or fairness as the reward in these learning frameworks can limit the ability to enhance the overall performance.

TABLE I
A SUMMARY OF DIFFERENCES BETWEEN THE PROPOSED SAIOC SCHEME AND EXISTING STUDIES

Reference	TPC	DSC	TXOP Sharing	Optimization Method	Control Paradigm	Multi-AP Coordination	General Analytic Model
[25]	✗	✓	✗	RBA	Centalized	✗	✗
[59]	✓	✗	✗	RBA	Centalized	✓	✗
[14]	✓	✗	✗	RBA	Distributed	✗	✗
[42]	✓	✗	✗	RBA	Distributed	✓	✓
[13], [15], [17], [20], [24], [28], [30]	✗	✓	✗	RBA	Distributed	✗	✗
[26]	✗	✓	✗	RBA	Distributed	✗	✓
[27]	✗	✓	✗	RBA	Distributed	✓	✓
[19]	✗	✗	✓	RBA	Distributed	✗	✓
[53], [57], [62]	✗	✗	✓	RBA	Distributed	✓	✗
[22], [32], [33], [34], [35], [37]	✓	✓	✗	RBA	Distributed	✗	✗
[23], [31], [36]	✓	✓	✗	RBA	Distributed	✗	✓
[16], [38], [43], [48], [60]	✓	✓	✗	RBA	Distributed	✓	✗
[58]	✓	✓	✗	RBA	Distributed	✓	✓
[44], [52]	✓	✗	✓	RBA	Distributed	✗	✗
[18], [40], [45] [50], [51]	✓	✗	✓	RBA	Distributed	✓	✗
[56]	✗	✗	✓	RL	Distributed	✓	✗
[29]	✓	✓	✗	DSL	Centalized	✓	✗
[61]	✓	✓	✗	DSL	Distributed	✓	✗
[39], [46]	✓	✓	✗	DSL	Federated	✓	✗
[47]	✗	✗	✓	SA-DRL	Centalized	✗	✓
[41]	✓	✗	✓	SA-DRL	Distributed	✗	✗
[54]	✗	✓	✗	SA-DRL	Distributed	✓	✗
[55]	✗	✗	✓	SA-DRL	Distributed	✓	✓
[21]	✗	✗	✓	SA-DRL	Federated	✓	✓
[49]	✓	✓	✗	SA-DRL	Distributed	✗	✗
SAIOC Scheme	✓	✗	✓	MA-DRL	Distributed	✓	✓

- TPC: transmit power control
- DSC: dynamic sensitivity control
- RBA: rule-based algorithm
- RL: reinforcement learning
- DSL: deep supervised learning
- SA-DRL: single-agent deep RL
- MA-DRL: multi-agent deep RL

The proposed SAIOC scheme derives the transmission probability of nodes (to obtain the optimal transmission power and shared AP index) based on a semi-Markov chain of the backoff algorithm. The proposed SAIOC scheme applies a MA-DRL algorithm to derive the optimal SR parameters, facilitating adaptive decision-making in dynamically changing network conditions. By leveraging the expected sum throughput based on the derived saturated transmission probabilities as a common reward through a multi-agent setting, the proposed scheme can achieve faster convergence and enhance the sum throughput performance.

III. SYSTEM MODEL

The system model used in this paper is introduced in this section. There are K APs and each AP is indexed by $\mathcal{K} = \{1, 2, \ldots, K\}$, which transmits frames using transmission power $\mathcal{P} = \{P_1, P_2, \ldots, P_K\}$, respectively. The STAs are associated with the AP that has the strongest reception power and let Y_k be the set of STAs associated with AP $k \in \mathcal{K}$. The STA $y_k^t \in Y_k$ is the current receiver of AP $k \in \mathcal{K}$ for time slot t.

A. CSR Protocol Model

The CSR protocol consists of the CSR preparation phase, backoff phase, CSR trigger phase, and CSR transmission phase as in Fig. 2. In the CSR preparation phase, CSR-capable APs share their capabilities with other APs and collect the necessary information for CSR operations. CSR-capable APs inform their CSR capability and intention to neighboring APs by including the CSR capability information in the beacon frame or management frame. To obtain RSSI information from other APs, APs send measurement report requests to the associated STAs. The STAs perform the measurement based on the beacon frames and send measurement reports (which include the RSSI information of neighboring APs) to the associated AP. Let $r_{u,v}$ be the RSSI of

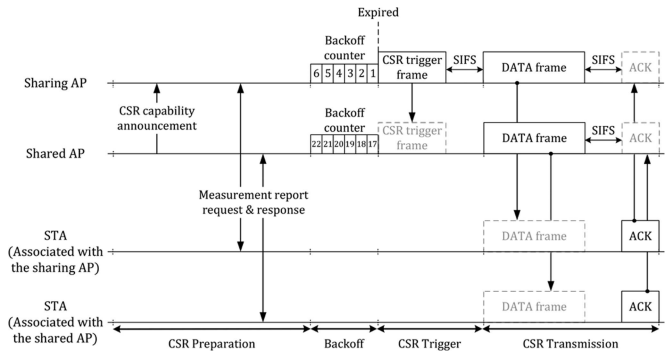


Fig. 2. The procedures of the CSR scheme.

CSR AP index	AP index	AP k BSS associated STA index				
	k	1	⋯	Y_k-2	Y_k-1	Y_k
1	$r_{1,k}$	$r_{1,1}$	⋯	r_{1,Y_k-2}	r_{1,Y_k-1}	r_{1,Y_k}
2	$r_{2,k}$	$r_{2,1}$	⋯	r_{2,Y_k-2}	r_{2,Y_k-1}	r_{2,Y_k}
3	$r_{3,k}$	$r_{3,1}$	⋯	r_{3,Y_k-2}	r_{3,Y_k-1}	r_{3,Y_k}
⋮	⋮	⋮	⋯	⋮	⋮	⋮
K	$r_{K,k}$	$r_{K,1}$	⋯	r_{K,Y_k-2}	r_{K,Y_k-1}	r_{K,Y_k}

Fig. 3. Management information base of CSR.

node v from node u. Based on the RSSI measurement report received from the STA and AP's measured RSSI of other CSR AP, each AP creates a CSR management information base (MIB), as shown in Fig. 3, and uses it when controlling the CSR parameters [63]. Afterward, each AP performs a channel contention process based on the backoff algorithm, and the AP that occupies the channel takes on the role of the sharing AP, which shares its TXOP with other APs. It is assumed that the maximum number of shared APs is configurable for each CSR group as M, and the maximum number of shared APs is identical among all APs within the same CSR group. Since the sharing AP can share its acquired TXOP with multiple shared APs, each AP $k \in \mathcal{K}$ has a set of shared AP indexes to share the acquired TXOPs, which are indexed by $\mathcal{I} = \{I_1, I_2, \ldots, I_K\}$. The sharing AP designates the shared AP to perform concurrent transmissions and transmits a CSR trigger frame to the shared APs to initiate CSR concurrent transmission. The CSR trigger frame includes the shared AP index information, transmission power information of both the sharing AP and shared AP and the transmission duration information. When the shared AP receives the CSR trigger frame, it immediately starts data frame transmission regardless of the channel contention process. The transmissions of the sharing AP and the shared AP are synchronized by the CSR trigger frame so concurrent transmissions will end at the same time as in Fig. 2.

B. Network Model

Based on transmission power information in the CSR trigger frame, the received power of node v from node u can be estimated based on CSR MIB as (in dBm scale)

$$P_{u,v}^{dBm} = P_u^{dBm} + r_{u,v} - P_{ref}^{dBm} \tag{1}$$

where P_{ref}^{dBm} is the reference transmission power for transmitting a beacon frame and management frame (in dBm scale). The signal-to-interference-plus-noise ratio (SINR) of STA y_k from AP k can be defined as

$$\tau_{k,y_k} = \frac{P_{k,y_k}}{\sum_{u\in\mathcal{K}\setminus\{k\}} P_{u,y_k} + N_0} \tag{2}$$

where P_{k,y_k} represents the received power of STA y_k from AP k (converted into the linear scale based on the expression $10^{(P_{u,v}^{dBm}-30)/10}$) and N_0 is the noise spectral density. Since the wireless channel model includes a small-scale Rayleigh fading component along with a large-scale path loss model, the bit error rate (BER) for phase shifting keying (PSK) and quadrature amplitude modulation (QAM) of STA y_k from AP k over Rayleigh fading can be expressed as in (3), shown at the bottom of this page [64]. Given the BER resulting from the modulation scheme as in (3), the corresponding BER for irregular low-density parity-check (LDPC) codes decoding can be estimated as follows [65], [66]

$$\epsilon_{b,k}^{LDPC} = \sum_{n=2}^{\psi_2 N} \frac{n}{N} \frac{(\lambda_2 \rho'(1))^n}{2n} \epsilon_{b,k}^{MOD} \tag{4}$$

where N is the length of the code, $\lambda(x) = \sum_{i=2}^{M_b} \lambda_i x^{i-1}$ is the bit node degree distibution of the code with the maximum bit node degree M_b, $\rho(x) = \sum_{i=2}^{M_c} \rho_i x^{i-1}$ is the check node degree distribution of the code with the maximum check node degree M_c, and $\psi_i = \frac{\lambda_i / i}{\sum_j \lambda_j / j}$ is the fraction of degree-i bit nodes. Then, the frame error rate (FER) can be expressed as follows using (4)

$$p_{f,k} = 1 - \left(1 - \epsilon_{b,k}^{LDPC}\right)^{N_f} \tag{5}$$

where N_f denotes the number of bits per frame.

IV. Proposed SAIOC Scheme

In this section, details of the proposed SAIOC scheme are presented. The process of deriving the saturated transmission probability of CSR and how to formulate the optimization problem is described. Thereafter, the SAIOC optimization framework is described and details of the proposed SAIOC scheme's implementation are described.

A. Saturated Transmission Probability Analysis

The saturated transmission probability of CSR is derived based on a semi-Markov chain. It is assumed that the exponential backoff algorithm is adopted. The state of a semi-Markov chain model consists of a backoff counter and backoff stage, as shown in Fig. 4. Let $b(t)$ be the stochastic process of the backoff counter and let $s(t)$ be the stochastic process of the backoff stage for each time slot. Let W_n be the contention window of backoff stage n. The backoff counter is determined between $(0, W_n)$ with random equal probability for each backoff stage.

$$\epsilon_{b,k}^{MOD} = \begin{cases} \frac{1}{2}\left(1 - \sqrt{\frac{\tau_{k,y_k}/2}{1+\tau_{k,y_k}/2}}\right), & M \in \{2,4\} \text{ for } M\text{-PSK} \\ \left(\frac{\sqrt{M}-1}{\sqrt{M}}\right) \frac{1}{\log_2 M} \sum_{i=1}^{\sqrt{M}/2} \left(1 - \sqrt{\frac{1.5\left((2i-1)^2 \tau_{k,y_k} \log_2 M\right)}{M-1+1.5\left((2i-1)^2 \tau_{k,y_k} \log_2 M\right)}}\right), & M \in \{16,64,256,1024,4096\} \text{ for } M\text{-QAM} \end{cases} \tag{3}$$

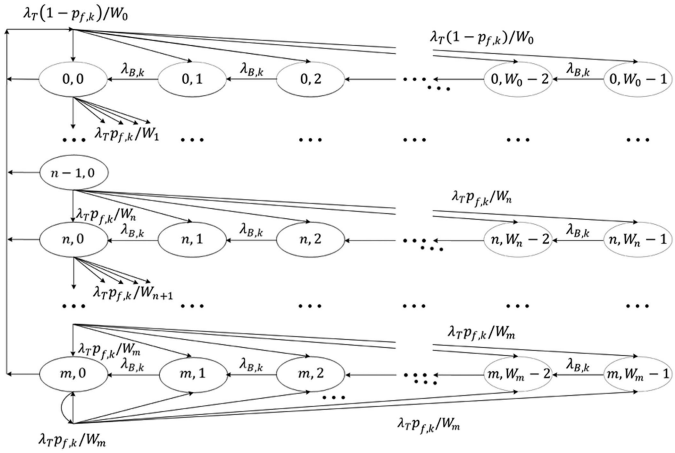


Fig. 4. Semi-Markov chain model of CSR.

The backoff counter decrements if the channel is in idle state during the DCF inter-frame space (DIFS) interval and freezes if the channel is in busy state. When the backoff counter reaches zero, the node proceeds to transmit its frames. Let σ_D be the DIFS time and let σ_S be the slot time. Let N_D be the number of slots for the DIFS duration, in which N_D can be approximated as $N_D \approx \left\lceil \frac{\sigma_D}{\sigma_S} \right\rceil$. If the transmission is a failure, transition to the next backoff stage and the contention window increases as a binary exponential pattern. Let W_0 be the minimum contention window size, then the contention window of backoff stage n can be expressed as $W_n = 2^n W_0$. Let m be the maximum backoff stage in which the contention window does not increase above. Let $q_{n,l}$ be the stationary distribution of the state, which can be expressed as $q_{n,l} = \lim_{t\to\infty} p\{s(t) = n, b(t) = l\}$ for $n \in [0, m]$ and $l \in [0,\ W_n - 1]$. Since the shared AP performs transmission immediately after being designated by the sharing AP and the corresponding backoff counter is maintained, the backoff counter state has sojourn time, which is affected by the DIFS waiting time and frame transmission time. The probability that the node is designated as a shared AP can be expressed as $p_{c,k} = G_1(\sum_{u\in\mathcal{K}\setminus\{k\}} g(C_u = k))$, where $g(\cdot)$ and $G(\cdot)$ are respectively the probability distribution function (PDF) and cumulative distribution function (CDF) of the degenerate distribution. The degenerate distribution returns a value of one if the condition is met and a value of zero otherwise. Let T be the transmission duration time and let N_T be the number of slots of the transmission duration, where N_T can be approximated as $N_T \approx \left\lceil \frac{T}{\sigma_S} \right\rceil$. The steady state should be analyzed based on a semi-Markov chain that simultaneously considers the state sojourn time and state transition probability. To consider the transition rate of the semi-Markov chain, the channel idle probability and backoff counter state sojourn time are respectively derived as in *Lemma 1* and *Lemma 2*.

Lemma 1: The channel idle probability of node k, denoted by $p_{i,k}$, is defined as $\frac{1}{\Gamma(\eta)}\gamma\left(\eta, P_{th}\sum_{v\in\mathcal{K}\setminus\{k\}} P_{v,k}\right)$ where $\Gamma(\cdot)$ is the gamma function, η is the number of the transmitter and $\gamma(\cdot)$ is the lower incomplete gamma function. □

Proof: Since the Rayleigh fading coefficient follows an exponential distribution, the received power of node k, $P_{received,k}$, is the sum of the exponential distribution variables and therefore follows a gamma distribution, as shown below

$$P_{received,k} \sim \text{Gamma}\,(\eta,\ \theta) \tag{6}$$

where η is the shape parameter (which is the number of the transmitters) and θ is the scale parameter (i.e., $\theta = 1/\sum_{v\in\mathcal{K}\setminus\{k\}} P_{v,k}$). Since the channel idle probability is the probability that the received power does not exceed the CCA threshold P_{th} such as $p_{i,k} = \mathbb{P}(P_{received,k} \le P_{th})$, the channel idle probability can be derived based on CSR MIB using the CDF of the gamma distribution as follows

$$p_{i,k} = F\left(P_{th}; \eta, \theta\right) = \frac{1}{\Gamma(\eta)}\gamma\left(\eta, P_{th}\sum_{v\in\mathcal{K}\setminus\{k\}} P_{v,k}\right) \tag{7}$$

where $F(x;\eta,\theta)$ is the CDF of the gamma distribution. ■

Lemma 2: The mean backoff counter sojourn time of node k is defined as $\frac{(1-(p_{i,k})^{N_D})}{(1-p_{i,k})(p_{i,k})^{N_D}} + p_{c,k}N_T$. □

Proof: During the backoff process, the backoff counter is decreased after detecting whether the channel is idle during the DIFS interval. Therefore, the waiting time for the channel idle state during the DIFS interval should be considered in the backoff counter state transition. Let N_W be the number of waiting slots for the channel idle state during the DIFS interval. Based on the geometric distribution, N_W is derived as follows

$$N_W = \frac{\sum_{n=1}^{N_D}(p_{i,k})^{n-1}}{(p_{i,k})^{N_D}} = \frac{\left(1-(p_{i,k})^{N_D}\right)}{(1-p_{i,k})(p_{i,k})^{N_D}}. \tag{8}$$

In addition, if the node is designated as a shared AP by the sharing AP, transmission is immediately performed and after transmission is completed, the backoff process is performed again with the existing backoff counter value, so the probability of being designated as a shared AP should be considered. Let $N_{B,k}$ be the mean of the number of backoff counter state sojourn slots, which can be derived as follows

$$N_{B,k} = (1-p_{c,k})N_W + p_{c,k}(N_T + N_W) = N_W + p_{c,k}N_T \tag{9}$$

which takes into account both the DIFS waiting time and the concurrent transmission duration. ■

Based on the derived mean number of backoff counter state sojourn slots, the transition rate of the backoff counter state can be derived as $\lambda_{B,k} = 1/N_{B,k}$ based on (9). Since transmission is performed when the backoff expires, the sojourn slot time of the corresponding state is derived as $\lambda_T = 1/N_T$. The state transition rate of the semi-Markov chain can be expressed as follows.

$$\begin{cases} \lambda\left\{n, l | n, l+1\right\} = \lambda_{B,k}, & n \in [0,m], l \in (0, W_n - 1] \\ \lambda\left\{0, l | n, 0\right\} = \lambda_T, & n \in [0,m], l \in (0, W_n - 1] \\ \lambda\left\{n, l | n-1, 0\right\} = \lambda_T, & n \in [0,m], l \in (0, W_n - 1] \\ \lambda\left\{m, l | m, 0\right\} = \lambda_T, & l \in [0, W_m - 1] \end{cases} \tag{10}$$

The only non-null one-step state transition probability of the semi-Markov chain can be expressed as follows.

$$\begin{cases} p\{n,l|n,l+1\}=1, & n\in[0,m], l\in[0,W_n-1) \\ p\{0,l|n,0\}=\frac{1-p_{f,k}}{W_0}, & n\in[0,m], l\in[0,W_n-1] \\ p\{n,l|n-1,0\}=\frac{p_{f,k}}{W_n}, & n\in(0,m], l\in[0,W_n-1] \\ p\{m,l|m,0\}=\frac{p_{f,k}}{W_m}, & l\in[0,W_m-1] \end{cases} \tag{11}$$

Based on the state transition probabilities and state transition rates defined previously, the saturated transmission probability of CSR is derived in *Proposition 1*.

Proposition 1: The saturated transmission probability of CSR node k, denoted by $p_{s,k}$, is defined as (12), shown at the bottom of this page. □

Proof: Since a node performs transmission when the backoff counter reaches zero, or the node is designated as a shared AP, the transmission probability of node k can be expressed as follows.

$$p_{s,k}=\sum_{n=0}^{m}\sum_{l=1}^{W_n-1} q_{n,l}p_{c,k}+\sum_{n=0}^{m} q_{n,0} \tag{13}$$

Since the semi-Markov chain requires a balance between incoming and outgoing probabilities in each state, the following relationship is established

$$q_{n,l}=\frac{W_n-l}{W_n}q_{n,0}\frac{\lambda_T}{\lambda_{B,k}} \tag{14}$$

where $n\in[0,m]$ and $l\in[0,W_n-1]$. Since $q_{n,0}=q_{0,0}(p_{f,k})^n$ and $q_{m,0}=q_{0,0}\frac{(p_{f,k})^m}{1-p_{f,k}}$, the relation is established as in (15), shown at the bottom of this page. Then, the following is obtained.

$$q_{0,0}=\frac{2\lambda_{B,k}\left(1-p_{f,k}\right)}{\left(\lambda_T\left(W_0-1\right)+2\lambda_{B,k}\right)+\lambda_T W_0 p_{f,k}\frac{\left(1-(2p_{f,k})^m\right)}{(1-2p_{f,k})}} \tag{16}$$

Since $\sum_{n=0}^{m} q_{n,0}=\frac{q_{0,0}}{(1-p_{f,k})}$ from [67], the following relationship is established.

$$\sum_{n=0}^{m} q_{n,0}=\frac{2\lambda_{B,k}}{\left(\lambda_T\left(W_0-1\right)+2\lambda_{B,k}\right)+\lambda_T W_0 p_{f,k}\frac{\left(1-(2p_{f,k})^m\right)}{(1-2p_{f,k})}} \tag{17}$$

Based on (14), the following relation can be derived.

$$\sum_{n=0}^{m}\sum_{l=1}^{W_n-1} q_{n,l}p_{c,k}=\sum_{n=0}^{m}\frac{W_n-1}{2}q_{n,0}\frac{\lambda_T}{\lambda_{B,k}}p_{c,k} \tag{18}$$

Since $q_{n,0}=q_{0,0}(p_{f,k})^n$ and $q_{m,0}=q_{0,0}\frac{(p_{f,k})^m}{1-p_{f,k}}$, (18) can be expressed as in (19), shown at the bottom of this page. Then, the saturated transmission probability of node k can be derived as (12) based on (16), (17), and (19). ■

B. Problem Formulation

Using (12), the objective function (which is based on the expected sum throughput) can be defined as follows

$$\mathbb{E}[S]=\sum_{k\in\mathcal{K}} p_{s,k}B_k\log_2\left(1+\frac{P_{k,y_k}}{\sum_{u\in\mathcal{K}\setminus\{k\}}P_{u,y_k}+N_0}\right) \tag{20}$$

where $p_{s,k}$ is saturated transmission probability of node k derived in (12), B_k is channel bandwidth of node k, $P_{u,v}$ is the received power derived from MIB as in (1) (in a linear scale), and N_0 is the noise spectral density. Then, the optimization problem with respect to the transmission power and shared AP index to maximize the expected sum throughput can be defined as follows.

$$\begin{aligned} &\underset{\boldsymbol{P},\boldsymbol{I}}{\text{maximize}}\ \mathbb{E}[S] \\ &\text{s.t.}\quad C1: P_k\in[0,P_{ref}],\forall k\in\mathcal{K} \\ &\qquad C2: I_k\subset\{1,2,\ldots,K\},\forall k\in\mathcal{K} \\ &\qquad C3: |I_k|\le M,\forall k\in\mathcal{K} \end{aligned} \tag{21}$$

In (21), P_k is the transmission power of node k (in linear scale), I_k is the set of shared AP indexes of node k, $\boldsymbol{P}$ is the vector with

$$p_{s,k}=\frac{2\lambda_{B,k}\left(1-2p_{f,k}\right)+\left(\lambda_T\left(\left(W_0-1\right)\left(1-2p_{f,k}\right)+W_0p_{f,k}\left(1-(2p_{f,k})^m\right)\right)\right)p_{c,k}}{\left(\lambda_T\left(W_0-1\right)+2\lambda_{B,k}\right)\left(1-2p_{f,k}\right)+\lambda_T W_0 p_{f,k}\left(1-(2p_{f,k})^m\right)} \tag{12}$$

$$1=\sum_{n=0}^{m}\sum_{l=1}^{W_n-1} q_{n,l}=\frac{q_{0,0}}{2}\left[\frac{\lambda_T}{\lambda_{B,k}}\left(W_0\left(\sum_{n=0}^{m-1}(2p_{f,k})^n+\frac{(2p_{f,k})^m}{1-p_{f,k}}\right)-\frac{1}{1-p_{f,k}}\right)+\frac{2}{1-p_{f,k}}\right] \tag{15}$$

$$\begin{aligned} \sum_{n=0}^{m}\frac{W_n-1}{2}q_{n,0}\frac{\lambda_T}{\lambda_{B,k}}p_{c,k} &= \frac{\lambda_T}{\lambda_{B,k}}p_{c,k}r_{0,0}\left\{\sum_{n=0}^{m-1}\frac{2^nW_0-1}{2}\left(p_{f,k}\right)^n+\frac{2^mW_0-1}{2}\frac{(p_{f,k})^m}{1-p_{f,k}}\right\} \\ &= \frac{\lambda_T r_{0,0}}{2\lambda_{B,k}}\left(\frac{\left(W_0-1\right)\left(1-2p_{f,k}\right)+W_0p_{f,k}\left(1-(2p_{f,k})^m\right)}{\left(1-2p_{f,k}\right)\left(1-p_{f,k}\right)}\right)p_{c,k} \end{aligned} \tag{19}$$

components P_k for $\forall k \in \mathcal{K}$, and $\boldsymbol{I}$ is the vector with components I_k for $\forall k \in \mathcal{K}$. The transmission power is constrained to the IEEE 802.11 reference transmission power P_{ref} (in linear scale). In addition, the number of shared APs that a sharing AP can designate is restricted to the configured maximum number of shared APs M as aforementioned.

Proposition 2: The optimization problem in (21) is a strong NP-hard problem. □

Proof: If it is proven that a specific case of a problem is NP-hard, then it follows that the general problem is also NP-hard. Therefore, a specific case of the optimization problem in (21) is examined for NP-hardness. When the saturated transmission probability is one, the optimization problem in (21) becomes equivalent to the sum rate maximization problem as follows [8]

$$\underset{\boldsymbol{P}}{\text{maximize}} \sum_{k\in\mathcal{K}} B_k \log_2\left(1+\frac{P_{k,y_k}}{\sum_{u\in\mathcal{K}\setminus\{k\}} P_{u,y_k}+N_0}\right)$$
$$\text{s.t.} \quad P_k \in [0, P_{ref}], \forall k \in \mathcal{K} \tag{22}$$

where B_k is the channel bandwidth of node k, $P_{u,v}$ is the received power derived from MIB as in (1) (in linear scale), P_k is the transmission power of node k (in linear scale), P_{ref} is the IEEE 802.11 reference transmission power (in linear scale), and N_0 is the noise spectral density. Since the objective function in (22) is convex in relation to each component of $\boldsymbol{P}$ and has a maximum value when all agents have either $P_k = 0$ or $P_k = P_{\text{ref}}$, the optimization problem in (22) is equivalent to finding the largest mutually non-interfering subset, which is a maximum independent set problem. The maximum independent set problem (which determines the maximum set of independent vertices in a graph) is well known to be a strong NP-hard problem in combinatorial optimization [8]. Therefore, since the specific case of the optimization problem (which eliminates influence of the saturated transmission probability from the general optimization problem) is strongly NP-hard, the general optimization problem formulated in (23) is strongly NP-hard. ■

C. SAIOC Optimization Framework

Since the formulated optimization problem in (21) is a NP-hard problem, finding the optimal solution in polynomial time analytically is difficult, particularly in dynamically changing network environments [68]. Since a RL system can effectively learn the optimal actions to take for a high-dimensional complicated task through interaction with the environment, RL is adopted in the proposed SAIOC scheme to derive the optimal solution of the formulated optimization problem. The APs are assigned to be the agents in the RL model and each agent observes the state, takes action, and receives a reward. Since the sharing APs that acquire the TXOP through the backoff process are the only ones that calculate the CSR parameters, asynchrony of action occurs when applying RL to derive the optimal transmission power and shared AP index. To deal with this, the state space, action space, and reward function of the SAIOC framework are defined as follows.

- *State Space:* An AP measures the received power level periodically for every timestep t to obtain the current channel state. Let s_k^t be the state space of AP k at timestep t. The state space can be expressed as

$$s_k^t = [I_k^t, R_k^t, L_k^t, X_k^t, P_k^t] \tag{23}$$

where I_k^t represents the shared AP indexes of AP k, R_k^t represents the measured power level of AP k, L_k^t represents the backoff counter of AP k, X_k^t represents the transmission index, which has a value of one if AP k performs transmission and zero otherwise, and P_k^t represents the transmission power of AP k.

- *Action Space:* Let a_k^t be the action space of AP k at timestep t. The SAIOC agents aim to derive the optimal transmission powers of the sharing AP and the shared AP, as well as the shared AP index, to maximize the expected sum throughput. As mentioned earlier, the backoff process leads to the occurrence of action asynchrony when applying RL. To address the asynchrony of action caused by the backoff process, an action masking technique is applied to filter out feasible actions based on the current state. Four cases can be configured based on the backoff counter information and transmission index information in the state space as follows.
  1) *Case 1. Immediately after acquiring the TXOP:* Suppose that $L_k^t = 0$ and $X_k^t = 0$. In this case, the backoff counter has expired but a transmission is not performed. Since AP k performs the role of a sharing AP immediately after acquiring TXOP through the backoff process, the feasible action in this case is to derive the shared AP index, as well as the transmission powers of the sharing AP and the shared AP.
  2) *Case 2. Performing transmission as a sharing AP:* Suppose that $L_k^t = 0$ and $X_k^t = 1$. In this case, the backoff counter has expired and a transmission is in progress. Since AP k has obtained a TXOP and is performing concurrent transmission with the shared AP specified in Case 1, the feasible action in this case is to maintain the same transmission powers and shared AP index as in the previous time slot.
  3) *Case 3. Performing transmission as a shared AP:* Suppose that $L_k^t > 0$ and $X_k^t = 1$. In this case, the backoff counter has not expired, but a transmission is being performed. Since AP k is designated as a shared AP by the sharing AP and is performing concurrent transmission with the sharing AP, the feasible action in this case is to update neither the transmission power nor the shared AP index.
  4) *Case 4. During the backoff process:* Suppose that $L_k^t > 0$ and $X_k^t = 0$. In this case, the backoff counter has not expired and a transmission is not performed. Since AP k has not been designated as a shared AP and the backoff process is being performed, the feasible action in this case is to update neither the transmission power nor the shared AP index.

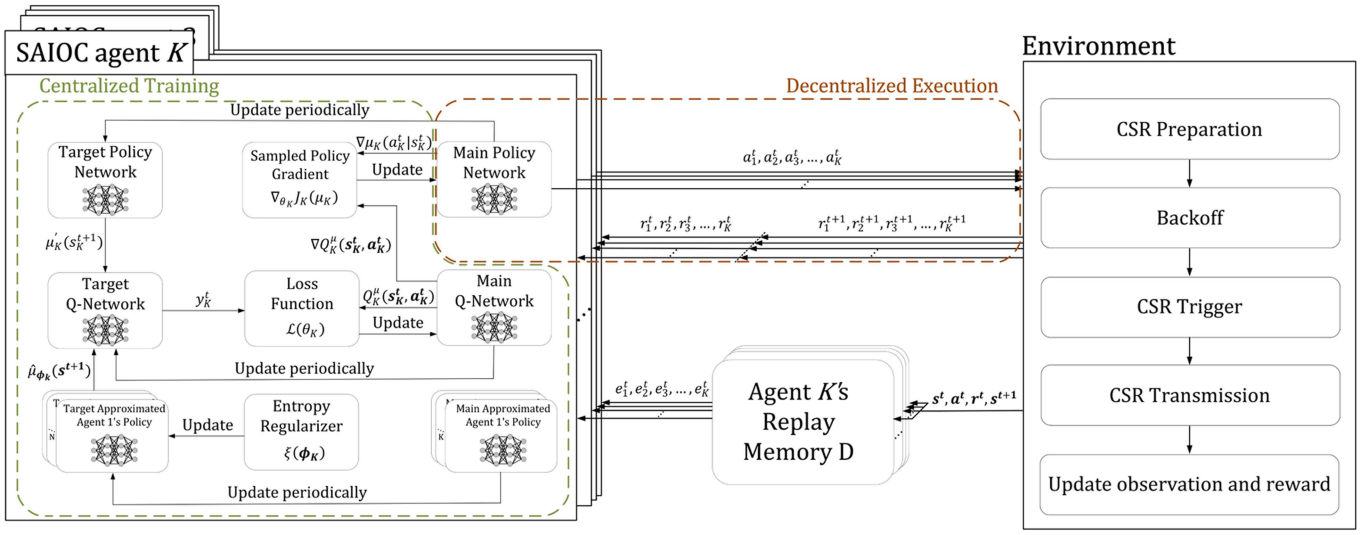


Fig. 5. The proposed SAIOC optimization framework.

Therefore, the action space can be expressed for each case as

$$a_k^t = \begin{cases} [a_p^t, \boldsymbol{a_c^t}, \boldsymbol{a_s^t}], & \text{if } L_k^t = 0, X_k^t = 0 \\ [a_p^{t-1}, \boldsymbol{a_c^{t-1}}, \boldsymbol{a_s^{t-1}}], & \text{if } L_k^t = 0, X_k^t = 1 \\ [0, 0, 0], & \text{if } L_k^t > 0 \end{cases} \quad (24)$$

where a_p^t represents the transmission power of the sharing AP at timestep t, $\boldsymbol{a_c^t}$ represents set of the shared AP indexes at timestep t (which is projected onto the shared AP index constraint through the floor function and the hyperbolic tangent function) and $\boldsymbol{a_s^t}$ represents the set of the transmission power of shared APs at timestep t (which are projected onto the transmission power constraint through a hyperbolic tangent function).

- *Reward Function:* Since the formulated optimization problem aims to maximize the expected sum throughput of all agents, all agents should carry out learning to maximize expected sum throughput using the same collaborative reward information. The collaborative reward function of timestep t can be expressed as follows

$$r_k^t = \sum_{k \in \mathcal{K}} p_{s,k} B_k \log_2 \left(1 + \frac{P_{k,y_k}}{\sum_{v \in \mathcal{K} \setminus \{k\}} P_{v,y_k} + N_0} \right) \quad (25)$$

which is the objective function of the formulated optimization problem, as in (21).

In Fig. 5, an overview of the SAIOC scheme is provided, which represents the interaction between the agents and the environment. The proposed SAIOC scheme aims to find the optimal shared AP index and the transmission power to maximize the expected sum throughput based on a multi-agent DDPG (MA-DDPG) DRL model. The goal of the MA-DDPG DRL model is to obtain an optimal policy which maximizes the expected long-term common reward through collaboration between the agents. The MA-DDPG DRL model originated from the DDPG DRL model. The MA-DDPG DRL model learns the deterministic policy, and outputs actions deterministically like DDPG, but for multi-agents. The MA-DDPG DRL model is a model-free off-policy algorithm, where the generated trajectories are stored in the experience replay buffer and the model is trained based on randomly selected samples from an experience replay buffer to alleviate the correlation of sequential sampling. The proposed SAIOC scheme utilizes information of other APs attained from coordination between the APs for learning.

There are two main reasons why the proposed SAIOC scheme applies the policy-based MA-DDPG DRL model [69] instead of the value-based multi-agent DQN (MA-DQN) DRL model [70]. First, the SAIOC system needs to address a continuous action domain. The value-based learning models have limitations in learning continuous action domain tasks because there are countless actions in the continuous action space and it is impossible to assign a value to each action for each state. However, the policy-based RL models learn about the action policy itself through policy parameter adjustments without needing to have values assigned to all actions. This is why the policy-based MA-DDPG DRL model can deal with continuous action domain tasks more effectively and why the proposed SAIOC scheme applies the MA-DDPG DRL model. Second, the SAIOC system needs to address the multi-agent environment. Since all agents need to update their transmission power and shared AP index in the SAIOC system, if a single-agent RL model is applied, a non-stationary environment problem can occur due to each agent's action to maximize their individual rewards. Compared to this, in a multi-agent RL model, all agents aim to maximize the common reward through collaboration, which is appropriate for the SAIOC scheme's objective.

The proposed SAIOC optimization framework has an actor-critic architecture, which is composed of an actor network and critic network. The actor network determines actions for the state based on policy and the critic network estimates the value of an action based on the action-value function. If both

networks used for learning estimate the value of action for their own learning, an action overestimation problem may occur. Therefore, the proposed SAIOC optimization framework divides into two networks where one is a training network (that generates samples and participates in the learning) and the other is a target network (that estimates the value of the action to reduce instability occurring in the learning process). The target network is not trained and updated as the training network periodically with soft update rate τ (which is delayed update). The proposed SAIOC optimization framework uses centralized training with a decentralized execution architecture, where the information of other agents is used in the training, but only local information of a corresponding agent is used in the execution. The information of the other agents can be obtained from coordination between the APs. Let μ_k be the deterministic policy of agent k with parameter set θ_k^{μ} and μ_k' be the target deterministic policy of agent k with parameter set $\theta_k^{\mu'}$. Let $\boldsymbol{\mu}$ be the vector with components μ_k for $\forall k \in \mathcal{K}$ and $\boldsymbol{\theta^{\mu}}$ be the vector with components θ_k^{μ} for $\forall k \in \mathcal{K}$. Let $\boldsymbol{\mu'}$ be the vector with components μ_k' for $\forall k \in \mathcal{K}$ and $\boldsymbol{\theta^{\mu'}}$ be the vector with components $\theta_k^{\mu'}$ for $\forall k \in \mathcal{K}$. Let $\boldsymbol{s_t}$ be the vector with components s_k^t for $\forall k \in \mathcal{K}$, $\boldsymbol{a_t}$ be the vector with component a_k^t for $\forall k \in \mathcal{K}$, and $Q_k^{\boldsymbol{\mu}}$ be the centralized action-value function of agent k with parameter set θ_k^Q, which takes the actions and the states of all agents as input. Let $\boldsymbol{Q}$ be the vector with components $Q_k^{\boldsymbol{\mu}}$ for $\forall k \in \mathcal{K}$ and $\boldsymbol{\theta^{Q}}$ be the vector with components θ_k^Q for $\forall k \in \mathcal{K}$. Let $Q_k'^{\boldsymbol{\mu'}}$ be the target action-value function of agent k with parameter set $\theta_k^{Q'}$ and target policies. Let $\boldsymbol{Q'}$ be the vector with components $Q_k'^{\boldsymbol{\mu'}}$ for $\forall k \in \mathcal{K}$ and $\boldsymbol{\theta^{Q'}}$ be the vector with components $\theta_k^{Q'}$ for $\forall k \in \mathcal{K}$. In addition, $\mathcal{D}$ represents the experience replay buffer which contains the tuple $(\boldsymbol{s_t}, \boldsymbol{a_t}, r_k^t, \boldsymbol{s_{t+1}})$ and $J_k(\mu_k)$ is the expected return of agent k. The gradient of the expected return of agent k can be expressed as in (26)

$$\nabla_{\theta_k^{\mu}} J_k\left(\mu_k\right) = \mathbb{E}_{\boldsymbol{s_t}, \boldsymbol{a_t} \sim \mathcal{D}}\left[\nabla_{\theta_k^{\mu}} \mu_k\left(s_k^t | \theta_k^{\mu}\right) \nabla_{a_k^t} Q_k^{\boldsymbol{\mu}}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right)\right] \tag{26}$$

where $a_k^t = \mu_k(s_k^t|\theta_k^{\mu})$. The policy parameters are updated in the direction of the gradient of the expected return in (26), which maximizes the expected return of the policy. Let $\mathcal{L}(\theta_k^Q)$ be the loss function of the centralized action-value function of agent k. The loss function of the centralized action-value function can be expressed based on the Bellman equation as follows

$$y = r_k^t + \gamma Q_k'^{\boldsymbol{\mu'}}\left(\boldsymbol{s_{t+1}}, \boldsymbol{a_{t+1}}\right) \tag{27}$$

$$\mathcal{L}\left(\theta_k^Q\right) = \mathbb{E}_{\boldsymbol{s_t}, a_k^t, r_k^t, \boldsymbol{s_{t+1}}}\left[\left(Q_k^{\boldsymbol{\mu}}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right) - y\right)^2\right] \tag{28}$$

where the centralized action-value function is updated to minimize the temporal difference (TD) error of the target value with discount factor γ, as in (27) and (28). Since the critic network is trained based on the actions of other agents in the next timestep, as in (28), the policies of the other agents are needed for learning. However, since the policies of the other agents are not shared and unknown, MA-DDPG attempts to approximate the policies of the other agents via learning. Let $\hat{\mu}_{\phi_k^v}$ be the approximated policy of agent v from agent k where ϕ_k^v is the parameter for approximation. Let $\hat{\boldsymbol{\mu}}_{\boldsymbol{\phi_k}}$ be the vector with components $\hat{\mu}_{\phi_k^v}$ for $\forall k \in \mathcal{K},\ \forall v \in \mathcal{K}\backslash\{k\}$ and $\boldsymbol{\phi_k}$ be the vector with components ϕ_k^v for $\forall k \in \mathcal{K},\ \forall v \in \mathcal{K}\backslash\{k\}$. Let $\xi(\phi_k^v)$ be the loss function for the policy approximation. The loss function of the policy approximation can be expressed using the entropy regularizer based on

$$\xi\left(\phi_k^v\right) = -\mathbb{E}_{s_v^t, a_v^t}\left[\log \hat{\mu}_{\phi_k^v}\left(a_v^t | s_v^t\right) + \lambda H\left(\hat{\mu}_{\phi_k^v}\right)\right] \tag{29}$$

Algorithm 1: SAIOC Scheme.

Initiate experience replay buffer $\mathcal{D}$
Initiate actor network $\boldsymbol{\mu}$ with random weight $\boldsymbol{\theta^{\mu}}$
Initiate critic network $\boldsymbol{Q}$ with random weight $\boldsymbol{\theta^{Q}}$
Initiate target actor network $\boldsymbol{\mu'}$ as $\boldsymbol{\theta^{\mu'}} = \boldsymbol{\theta^{\mu}}$ and target critic network $\boldsymbol{Q'}$ as $\boldsymbol{\theta^{Q'}} = \boldsymbol{\theta^{Q}}$
Initiate approximated policies of other agents $\hat{\boldsymbol{\mu}}_{\boldsymbol{\phi_k}}$ with random weight $\boldsymbol{\phi_k}$
for each episode i **do**
  for each agent k **do**
    for each timestep t **do**
      if receive CSR capability announcement **then**
        Send measurement report request
      Get action $a_k^t = \mu_k(s_k^t|\theta_k^{\mu})$ from actor network based on (24)
      if transmission duration **then**
        Send data frame using adjusted transmission power
      else
        if backoff counter expired **then**
          Send CSR trigger frame including shared AP index, transmission duration, and transmission power information
          Send data frame using adjusted transmission power
        else
          Update backoff counter
      Get next state s^{t+1} and reward r^t
      Store the transition tuple (s^t, a^t, r^t, s^{t+1}) in experience replay buffer $\mathcal{D}$
      Update policy approximation function based on (29) using latest sample
      Sample a random minibatch of transition tuple from experience replay buffer $\mathcal{D}$
      Compute target value y based on (30)
      Update actor network μ_k based on (26)
      Update critic network $Q_k^{\boldsymbol{\mu}}$ based on (28)
      for every T timesteps **do**
        Update target network parameter as $\theta_k^{Q'} = \tau\theta_k^{Q} + (1-\tau)\,\theta_k^{Q'}$ and $\theta_k^{\mu'} = \tau\theta_k^{\mu} + (1-\tau)\,\theta_k^{\mu'}$

where H is the entropy of the policy distribution and λ is entropy regularization coefficient. The approximated policy is updated to maximize the log probability of the action of agent v. Then, (27) can be expressed as follows using the approximated policies of the other agents.

$$\hat{y} = r_k^t + \gamma Q_k^{'\hat{\boldsymbol{\mu}}_{\boldsymbol{\phi_k}}} \left(\mathbf{s_{t+1}}, \hat{\mu}_{\phi_k^1}(s_1^t), \ldots, \hat{\mu}_{\phi_k^K}(s_K^t) \right) \tag{30}$$

D. Implementation of SAIOC Scheme

The pseudo code of the SAIOC scheme is presented in Algorithm 1. First, the experience replay buffer and actor network and critic network are initialized (line 1–3). In addition, the target networks $\boldsymbol{Q'}$ and $\boldsymbol{\mu'}$ as well as the approximated policies of the other agents are initialized (line 4–5). The learning process is iterated for each episode i and each agent k and each timestep t (line 6–8). For each timestep t, if the AP receives the CSR capability announcement from other APs, the AP sends a measurement report request to the associated STAs (line 9–10). The action is derived from the actor network based on (24) (line 11). If the AP is in the transmission duration, it performs data frame transmission using the adjusted transmission power (lines 12–13). Otherwise, if the AP's backoff counter expires, the CSR trigger frame is sent, which includes the shared AP index, transmission duration, and transmission power information. If neither condition is met, the AP updates its backoff counter (lines 14–19). Afterwards, the next state and reward are computed and stored in the experience replay buffer in the form of a transition tuple (line 20–21). The policy approximation function is updated based on (29) using the latest sample (line 22). To train the actor network and critic network, a random minibatch of transition tuples are sampled from the experience replay buffer and the target value is computed based on (30) and the target networks (line 23–24). The actor network is updated using (26) and the critic network is updated using the computed target value based on (28) (line 25–26). Finally, the target networks are updated through the training network every T timesteps (lines 27–28).

E. Complexity Analysis of SAIOC Scheme

The computation complexity of the proposed SAIOC scheme is derived based on floating point operations (FLOP) per second (FLOPS). To calculate FLOPS, addition, subtraction, multiplication, etc. are each counted as a single FLOP. Since each SAIOC agent consists of L fully connected layers of the actor network and J fully connected layers of the critic network, the computation complexity of offline training and online execution is determined by the number of matrix multiplications. Let $u_{a,l}$ be the number of neurons in the lth layer of the actor network, where $l \in \{1, \ldots, L\}$ and let $u_{c,j}$ be the number of neurons in the jth layer of the critic network where $j \in \{1, \ldots, J\}$. The FLOPS of matrix multiplications between the lth layer and $(l+1)$th layer of the actor network is $(u_{a,l} + u_{a,l} - 1)u_{a,l+1}$. In addition, the FLOPS of matrix multiplications between the jth layer and $(j+1)$th layer of the critic network is $(u_{c,j} + u_{c,j} - 1)u_{c,j+1}$. Therefore, the computation complexity of the training of the SAIOC agent is $O(\sum_{l=0}^{L-1} u_{a,l}u_{a,l+1} + \sum_{j=0}^{J-1} u_{c,j}u_{c,j+1})$ [71], [72]. In addition, since only the actor network is used for online execution, the computation complexity of execution of the SAIOC agent is reduced to $O(\sum_{l=0}^{L-1} u_{a,l}u_{a,l+1})$.

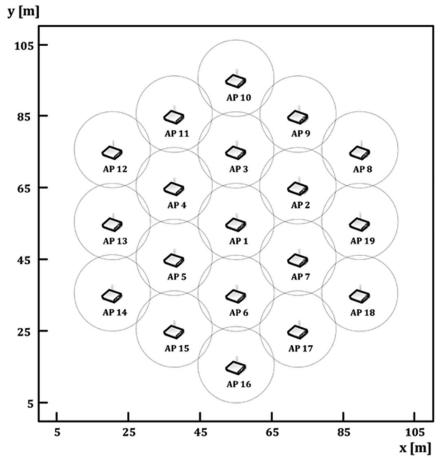


Fig. 6. Simulation environment.

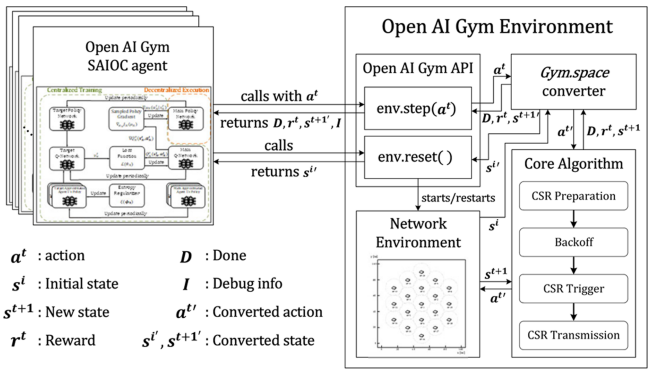


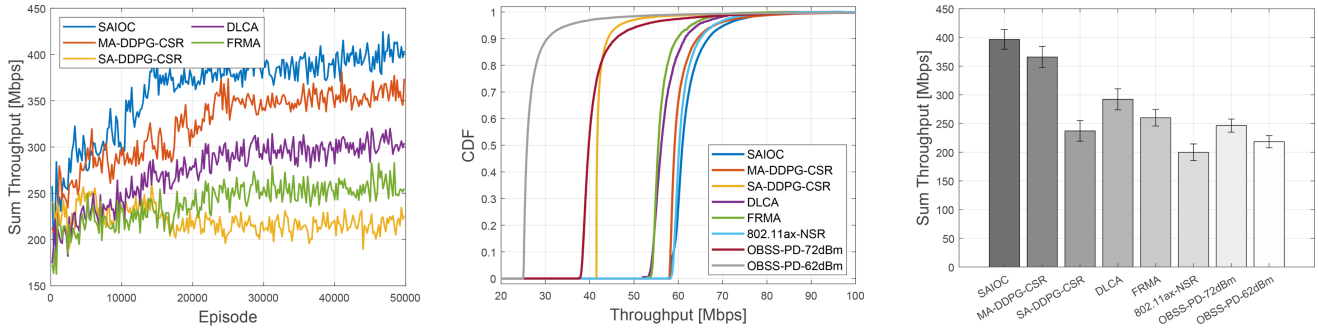
Fig. 7. Interaction diagram of the OpenAI Gym agents and environment for the packet-level simulation to evaluate the performance under the constructed simulation scenario.

V. Performance Evaluation

In this section, the simulation experiment environments applied in the performance evaluations are described, followed by the performance analysis of the proposed SAIOC scheme and other benchmark schemes.

A. Experiment Environments

The simulation was implemented based on the Python OpenAI Gym which is a Python toolkit for executing RL agents that operate on given environments. As previously mentioned, each AP participates in learning as an agent. The performance evaluation was conducted based on the indoor small BSS scenario of the TGax simulation scenarios where the hexagonal multi-cell layout is constructed as in Fig. 6 [73]. In the simulation scenario, 19 APs are deployed with a regular radius of 10 m and 50 STAs are deployed randomly in the simulation area. In Fig. 7, the interaction diagram of the OpenAI Gym agents and environment for the packet-level simulation under the constructed



(a) (b) (c)

Fig. 8. Throughput comparison based on the simulation scenario: (a) sum throughput according to learning episodes, (b) cumulative distribution function of each AP's throughput, and (c) sum throughput with standard deviation.

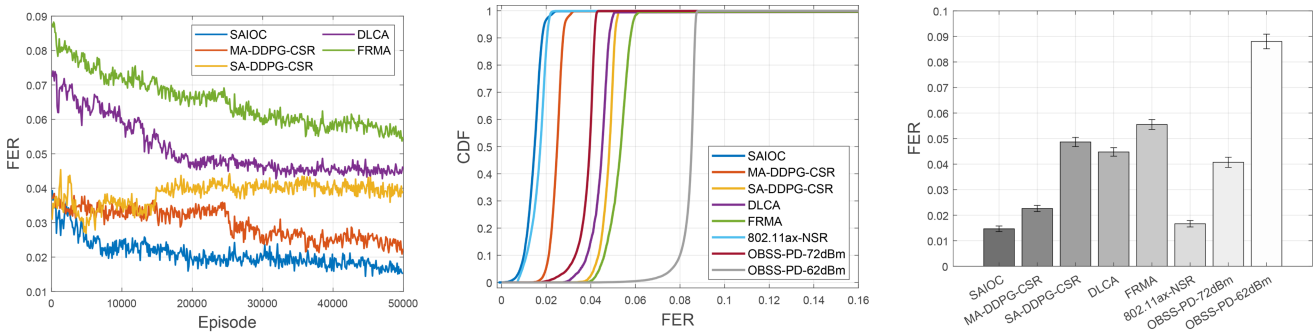
simulation scenarios is presented. The SAIOC agents interact with the OpenAI Gym environment using two methods defined by the Gym interface `env.step()` and `env.reset()`. To initiate a new episode, the SAIOC agents call `env.reset()` and return an initial state from the OpenAI Gym environment. In this process, since the OpenAI Gym SAIOC agents utilize observations in the form of the OpenAI Gym Spaces class objects, the OpenAI Gym environment converts the state into the form of an OpenAI Gym Spaces class object through the `gym.space` converter and returns it to the SAIOC agents. After that, the OpenAI Gym SAIOC agents calculate an action based on the returned state and call `env.step()`. Since the calculated action is also an OpenAI Gym Spaces class object, it is converted through the `gym.space` converter to be utilized in the OpenAI Gym environment. Based on the converted action, the core algorithm (which consists of a CSR preparation phase, backoff phase, CSR trigger phase, and CSR transmission phase) is applied under the constructed network environment, and reward information, observation information (converted to OpenAI Gym Spaces class objects), done information, and debug information are returned to the SAIOC agents. Then, the SAIOC agents calculate their action based on returned information and call `env.step()`. This process is repeated, and when the episode is done, `env.reset()` is called again and a new episode is initiated, thereby performing learning and packet-level simulation. Learning was conducted over 50,000 episodes and the simulation experiments were conducted over 10,000 episodes using the trained model. The performance of the proposed SAIOC scheme is compared with multi-agent DDPG CSR (MA-DDPG-CSR), single-agent DDPG CSR (SA-DDPG-CSR), DLCA [55], FRMA [26], IEEE 802.11ax OBSS PD based SR (OBSS-PD-SR) with OBSS_PD level of -62 dBm (OBSS-PD-62 dBm) and -72 dBm (OBSS-PD-72 dBm), and the IEEE 802.11ax with no SR (802.11ax-NSR) in terms of throughput, FER, delay and fairness performance [34]. Since IEEE 802.11ax OBSS PD based SR adjusts its transmission power along with its OBSS_ PD level to address interference from concurrent transmissions, the following relationship is established as $P_{PD} = P_{ref} - (\tau_{PD_{min}} - \tau_{PD})$, where P_{PD} is the adjusted transmission power, P_{ref} is the IEEE 802.11 reference transmission power, $\tau_{PD_{min}}$ is the minimum OBSS_PD level (e.g., –82 dBm), and τ_{PD} is the currently applied OBSS_ PD level. Based on this relationship, the transmission power corresponding to the OBSS_ PD levels of –62 dBm and –72 dBm are 1 dBm and 11 dBm, respectively. MA-DDPG-CSR only finds the optimal transmission power using the MA-DDPG based MA-DRL model, which simply uses the sum throughput as its reward function. The MA-DDPG-CSR scheme is compared with the proposed SAIOC scheme to confirm the level of performance improvements that come from the optimal shared AP index selection and SAIOC optimization framework enhancement. The parameters used in the simulation are summarized in Table II.

TABLE II
SIMULATION PARAMETERS

Parameter	Value
Channel Bandwidth	20 MHz
Packet Size	1500 Bytes
Number of APs	19
Number of STAs	50
Traffic Direction	Omni-Directional
Receiver Gain	0
Reference Transmission Power	21 dBm
Energy Detection Threshold	-82 dBm
Rx Noise Figure	7 dB

B. Throughput

In Fig. 8, the throughput performance of SAIOC is compared with MA-DDPG-CSR, SA-DDPG-CSR, DLCA, FRMA, OBSS-PD-62 dBm, OBSS-PD-72 dBm, and 802.11ax-NSR. Fig. 8(a) presents how the sum throughput performance changes in reference to the learning episodes. Since OBSS PD based SR operates based on the configured OBSS_PD level and transmission power, its performance does not change as the episodes progress. Therefore, in Fig. 8(a), the episode-wise sum throughput performance variations of the learning-based SR schemes, SAIOC, MA-DDPG-CSR, SA-DDPG, DLCA, and FRMA, are presented. As the proposed SAIOC scheme performs learning using the expected sum throughput as a reward, which considers the backoff process of the other APs, the results show that the proposed SAIOC is effective in finding the optimal transmission power levels and shared AP indexes as the learning episodes



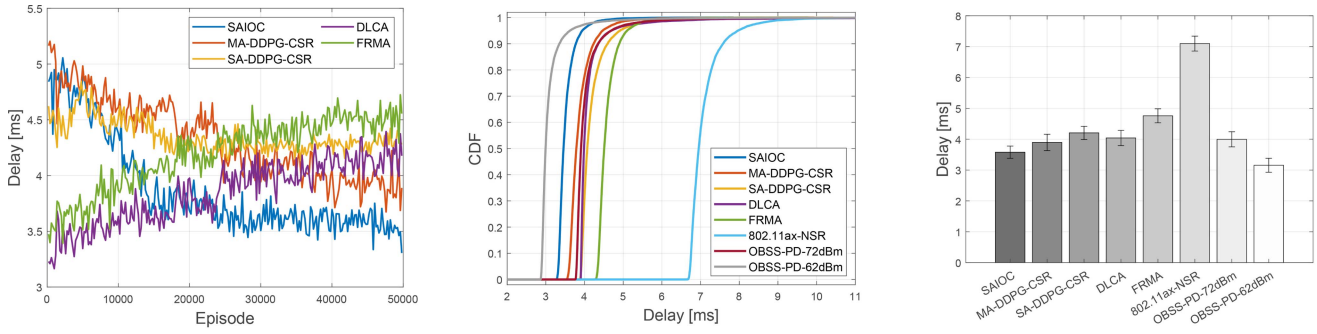
(a) (b) (c)

Fig. 9. FER comparison based on the simulation scenario: (a) average FER according to learning episodes, (b) cumulative distribution function of FER, and (c) average FER with standard deviation.

progress in the environment where the agents share and use the limited channel resources through contention. In addition, since the proposed SAIOC scheme performs transmission using both the optimal transmission power and the shared AP index, the proposed SAIOC shows the best sum throughput performance. MA-DDPG-CSR performs learning as the learning episodes progress, but shows a lower sum throughput performance than SAIOC because MA-DDPG-CSR does not derive the optimal shared index and simply uses the sum throughput as a reward without considering the backoff process of the other APs. On the other hand, since each agent of SA-DDPG-CSR uses individual reward functions for learning without consideration of the other agents, SA-DDPG-CSR will commonly fail to find the optimal transmission power level and shared AP index in environments where limited channel resources have to be shared with other agents. Since both DLCA and FRMA determine whether to access the channel using a DQN-based algorithm without performing a conventional backoff process, both schemes achieve a high sum throughput performance by eliminating the backoff time. Furthermore, DLCA exhibits a better sum throughput than FRMA, as it allocates the primary channel based on a greedy algorithm that considers proportional fairness (PF) at a centralized AP controller, thereby reducing the likelihood of collisions during transmission. Fig. 8(b) presents the CDF of each AP's throughput during the test episodes and Fig. 8(c) presents the sum throughput performance with an error bar indicating the one standard deviation point during the test episodes. While the proposed SAIOC scheme exhibits a similar distribution of per-AP throughput compared to 802.11ax-NSR, it achieves a superior sum throughput performance, as shown in Fig. 8(c). This result shows that the proposed SAIOC scheme enhances the sum throughput by utilizing an optimization framework to derive the optimal shared AP index and transmission power, thereby increasing the number of transmissions through TXOP sharing without degrading the performance due to the concurrent transmission interference. Similarly, MA-DDPG-CSR also can improve the sum throughput through concurrent transmission enabled by TXOP sharing, however, its performance is lower than that of the proposed SAIOC scheme, as it relies on a suboptimal shared AP index selection for simultaneous transmissions. Although both DLCA and FRMA exhibit a lower per-AP throughput compared to the 802.11ax-NSR scheme, both schemes achieve a higher sum throughput performance, as shown in Fig. 8(c). This is because both schemes eliminate the conventional backoff process and initiate transmissions immediately, thereby reducing the backoff delay and increasing the number of transmissions, which ultimately enhances the overall sum throughput. As shown in Fig. 8(c), both OBSS-PD-72 dBm and OBSS-PD-62 dBm achieve a higher sum throughput performance compared to 802.11ax-NSR. However, per-AP throughput distributions of both schemes exhibit a lower performance than 802.11-NSR. This is because transmissions are conducted with reduced transmit power (e.g., 1 dBm, 11 dBm) while adjusting the OBSS_PD level, which results in a lower SINR and subsequently reduced per-AP throughput. Nevertheless, the adjustment of the OBSS_PD level enables a greater number of transmission attempts, ultimately leading to an improved sum throughput.

C. FER

In Fig. 9, the FER performance of SAIOC is compared with the other benchmark schemes. Fig. 9(a) presents the average FER of the learning-based SR schemes according to the learning episodes. Since the proposed SAIOC scheme performs learning to maximize the expected sum throughput, which is derived based on the FER, the results show that the FER performance of SAIOC improves as the learning progresses. Even if the FER performance of MA-DDPG-CSR improves as the learning episodes progress, the learning speed and FER performance are lower than the proposed SAIOC scheme due to not being able to find the appropriate shared AP indexes. In contrast, SA-DDPG-CSR cannot properly learn due to the individual rewards in the multi-agent environment, so its performance remains at an almost constant level. Since both DLCA and FRMA perform channel contention and transmission without a conventional backoff process, both schemes exhibit a higher FER due to increased collisions and the FER performance improves as training progresses. In addition, since DLCA employs a greedy algorithm that incorporates proportional fairness (PF) to allocate primary channels to each AP, DLCA demonstrates a lower FER compared to FRMA.



(a) (b) (c)

Fig. 10. Delay comparison based on the simulation scenario: (a) average delay according to learning episodes, (b) cumulative distribution function of delay, and (c) average delay with standard deviation.

Fig. 9(b) presents the CDF of the APs' FER and Fig. 9(c) presents the average FER performance with the error bar representing the one standard deviation point. The proposed SAIOC shows the best FER performance and MA-DDPG-CSR shows the second-best FER performance compared to the other SR schemes. SA-DDPG-CSR has a lower FER performance than 802.11ax-NSR, as presented in Fig. 9(a), but a similar sum throughput performance is obtained from 802.11ax-NSR. This is because SA-DDPG-CSR has more TXOPs by utilizing the shared AP, but uses a suboptimal transmission power and shared AP index due to failing to learn the optimized way to make transmission power adaptations. In addition, both FRMA and DLCA exhibit a degraded FER performance due to collisions caused by concurrent transmissions. The FER performance of OBSS-PD-SR is lower than 802.11ax-NSR due to the reduced transmission power (e.g., 1 dBm, 11 dBm), but the throughput performance of OBSS-PD-SR is higher than 802.11ax-NSR because it uses increased TXOPs due to the adjusted OBSS_PD level, as presented in Fig. 9(a). Therefore, the throughput gains of OBSS-PD-SR is due to an increase in the TXOPs rather than an improvement in the transmission performance itself.

D. Delay

In Fig. 10, the delay performance of SAIOC is compared with the other benchmark schemes. To validate the delay performance of the SR schemes, the medium access control (MAC) layer delay is measured, which includes the channel access delay and retransmission delay due to transmission failure. Fig. 10(a) presents the average delay of the learning-based SR schemes according to the learning episodes. Since the proposed SAIOC scheme can mitigate the interference from concurrent transmissions using an optimal shared AP index and improve FER performance as learning proceeds, the results show that the delay performance of SAIOC improves as the learning progresses. Although the proposed SAIOC scheme causes additional delay due to the CSR preparation phase and CSR trigger phase of the CSR protocol as shown in Fig. 3, the proposed SAIOC scheme has the best delay performance among learning-based SR schemes by reducing the channel access delay due to the backoff process by sharing the TXOP between APs and reducing the retransmission delay by improving the FER performance (through learning) as in Fig. 10(a). The MA-DDPG-CSR scheme can improve the delay performance due to TXOP sharing and improve the FER performance by learning, as presented in Fig. 10(a), but the delay performance of MA-DDPG-CSR is lower than that of the proposed SAIOC scheme due to the lower FER performance based on the suboptimal shared APs. The SA-DDPG-CSR scheme reduces the channel access delay using the CSR operation, but shows a longer delay performance because the FER performance is not improved due to individual learning as in Fig. 10(a). Since DLCA and FRMA perform transmissions without a conventional backoff process, both schemes tend to defer transmissions as training progresses (in order to mitigate collisions caused by concurrent transmissions) and the transmission delay increases over the training duration.

Fig. 10(b) presents the CDF of the APs' delay and Fig. 10(c) presents the average delay performance with the error bar, representing the one standard deviation point. The proposed SAIOC shows a superior delay performance due to the TXOP sharing between APs through the CSR operation and the improved FER performance by using optimal SR parameters. The proposed scheme exhibits a lower delay performance compared to OBSS-PD-62 dBm. Although TXOP sharing reduces a portion of the backoff delay, the scheme still performs backoff with a CCA threshold of –82 dBm in order to share the TXOP as a sharing AP, which contributes to the overall delay. Both DLCA and FRMA exhibit a lower delay compared to 802.11ax-NSR, primarily because they initiate transmissions without performing the conventional backoff process, thereby reducing the backoff delay. OBSS-PD-72 dBm and OBSS-PD-62 dBm also demonstrate a superior delay performance by lowering the channel access delay through OBSS_PD level adjustments. As mentioned earlier, these results indicate that the SR schemes improve the sum throughput performance by increasing TXOPs, rather than by enhancing the transmission performance itself.

E. Fairness

In Fig. 11, the fairness performance of SAIOC is compared with the other benchmarked schemes. The fairness performance is evaluated based on the Jain's fairness index. The Jain's fairness index F is defined as $F = (\Sigma_{k=1}^{K} x_k)^2/(K\Sigma_{k=1}^{K} x_k^2)$, where K is the number of APs and x_k is the average throughput of AP k.

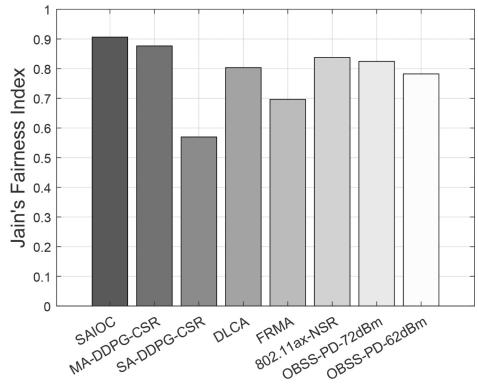


Fig. 11. Fairness comparison.

TABLE III
COMPARISON OF RUNNING TIME PER ITERATION

Scheme	Average Running Time (ms)	
	Training	Test
SAIOC	0.98284	0.45411
MA-DDPG-CSR	0.87569	0.4307
SA-DDPG-CSR	1.15956	0.53519
DLCA	2.66473	1.16329
FRMA	2.02316	1.04448
802.11ax-NSR	-	0.31989
OBSS-PD-72dBm	-	0.33874
OBSS-PD-62dBm	-	0.35986

TABLE IV
COMPARISON OF THE AVERAGE SAIOC PERFORMANCE FOR DIFFERENT SHARED AP NUMBERS

	Sum throughput (Mbps)	FER	Delay (ms)	SINR (dB)
SAIOC-1	393.753	0.015	3.588	24.848
SAIOC-2	407.624	0.0405	3.077	19.637
SAIOC-3	341.398	0.0443	2.752	15.587

The proposed SAIOC shows the highest fairness performance compared to the other schemes. This is because the proposed SAIOC scheme effectively performs learning of the expected sum throughput considering the other AP's backoff process and the common rewards. This is why the agents are not biased to one side. MA-DDPG-CSR shows the second highest fairness performance because it does not optimize the shared AP index to maximize common rewards. SA-DDPG-CSR has a low fairness performance since it performs learning to maximize each reward. Both DLCA and FRMA show a mid-level fairness performance, as all APs share learning parameters through meta-learning at the centralized AP controller, thereby preventing biased learning. In particular, the DLCA scheme further improves fairness by performing channel allocation based on PF considerations. The 802.11ax-NSR has a lower fairness performance because the APs cannot obtain fair TXOPs during the contention process in a dense network due to the conservative parameters. OBSS-PD-SR shows a low fairness performance because it uses different OBSS_PD levels and the transmission power depends on whether it is an inter-BSS or an intra-BSS.

VI. SUMMARY AND DISCUSSION

The performance of the proposed SAIOC scheme was compared with other SR schemes in various aspects. Based on the simulation results, it is confirmed that the proposed SAIOC scheme outperforms other SR schemes in terms of throughput, FER, delay, and fairness. The SAIOC scheme achieves a 7.78, 68.4, 34.35, 50.29, 99.14, 59.54, and 81.66 percent gain in sum throughput performance, a 60.94, 69.57, 66.74, 73.07, 10.71, 62.31, and 83.64 percent gain in average FER performance, a 7.14, 15.06, 12.36, 20.78, 50.37, 9.53, and −13.4 percent gain in average delay performance, and a 3.4, 59.06, 12.81, 30.16, 8.22, 9.94 and 15.84 percent gain in fairness performance when respectively compared to DLCA, FRMA, MA-DDPG-CSR, SA-DDPG-CSR, 802.11-NSR, OBSS-PD-62 dBm, and OBSS-PD-72 dBm. Table III presents the running time per iteration for each SR scheme. The results indicate that the proposed SAIOC scheme shows a higher running time per each iteration compared to DLCA, FRMA, MA-DDPG-CSR, SA-DDPG-CSR, 802.11-NSR, OBSS-PD-62 dBm, and OBSS-PD-72 dBm during the training phase, primarily due to the computational complexity of the MA-DRL optimization framework, as analyzed in Section IV-E. However, during the testing phase, the running time of the SAIOC scheme is comparable to that of the other schemes, owing to the CTDE structure. Although there is trade-off between performance improvement and computational overhead when applying the proposed SAIOC scheme, after the training phase, the computational overhead is significantly reduced by the CTDE structure, resulting that the performance gains outweigh the computational overhead. Table IV presents the performance of proposed SAIOC for different shared AP numbers. SAIOC-1, SAIOC-2, and SAIOC-3 represent the proposed SAIOC schemes that utilize one, two, and three shared APs, respectively. The results indicate that as the number of simultaneously utilized shared APs increases, the resulting interference from concurrent transmissions leads to a reduced SINR, which subsequently causes a higher FER and lower per-AP throughput. Notably, SAIOC-2 achieves the highest sum throughput among the three schemes. This is attributed to the increased number of transmission opportunities enabled by sharing TXOPs with more shared APs, which is further supported by its delay performance. In contrast, SAIOC-3 exhibits the lowest sum throughput despite the highest number of transmissions, due to excessive concurrent transmissions that significantly degrade the SINR, ultimately resulting in severely reduced per-AP throughput.

VII. CONCLUSION

In this paper, a SAIOC scheme that can maximize the sum throughput in dense IEEE 802.11 WLAN networks is proposed. The saturated transmission probability was developed based on the developed semi-Markov chain model and an optimization process that can maximize the expected sum throughput is formulated using the derived saturated transmission probability. Since the formulated optimization problem is a strong NP-hard problem, MA-DRL was used to control the SAIOC algorithm that was constructed by deriving the optimal solution of the

formulated problem. The simulation results demonstrate that the proposed SAIOC scheme can provide a superior performance compared to benchmark schemes in terms of throughput, FER, and fairness. In addition, further performance gains can be attained by appropriately configuring the number of shared APs according to the network environment within the proposed SAIOC framework.

REFERENCES

[1] D. Lopez-Perez, A. Garcia-Rodriguez, L. Galati-Giordano, M. Kasslin, and K. Doppler, "IEEE 802.11be extremely high throughput: The next generation of Wi-Fi technology beyond 802.11ax," *IEEE Commun. Mag.*, vol. 57, no. 9, pp. 113–119, Sep. 2019.

[2] J.-M. Chung, *Emerging Metaverse XR and Video Multimedia Technologies*. Berkeley, CA, USA: Springer Nature, 2023.

[3] C. Deng et al., "IEEE 802.11be Wi-Fi 7: New challenges opportunities," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2136–2166, Oct.–Dec. 2020.

[4] ResearchAndMarkets, "WiFi IoT Market Report: Trends, Forecast and Competitive Analysis to 2030," Dublin, Ireland, Jan. 2024.

[5] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on IEEE 802.11ax high efficiency WLANs," *IEEE Commun. Surveys. Tuts.*, vol. 21, no. 1, pp. 197–216, Jan.–Mar. 2019.

[6] S. Verma, T. K. Rodrigues, Y. Kawamoto, M. M. Fouda, and N. Kato, "A survey on Multi-AP coordination approaches over emerging WLANs: Future directions and open challenges," *IEEE Commun. Surv. Tutor.*, vol. 26, no. 2, pp. 858–889, 2nd Quart. 2024.

[7] K. Aio, *Coordinated Spatial Reuse Perform. Anal.*, Document IEEE 802.11, Piscataway, NJ, USA, Sep. 2019. [Online]. Available: https://mentor.ieee.org/802.11/documents?is_dcn=1534is_group=0TGbe

[8] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[9] A. Herzel, S. Ruzika, and C. Thielen, "Approximation methods for multiobjective optimization problems: A survey," *INFORMS J. Comput.*, vol. 33, no. 4, pp. 1284–1299, Feb. 2021.

[10] J.-M. Chung, *Emerging Secure Networks, Blockchains & Smart Contract Technologies*. Cham, Switzerland: Springer, 2024.

[11] T. Barrett, W. Clements, J. Foerster, and A. Lvovsky, "Exploratory combinatorial optimization with reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, New York, NY, USA, Apr. 2020, pp. 3243–3250.

[12] L. Ardon, "Reinforcement learning to solve NP-hard problems: An application to the CVRP," 2022, *arXiv:2201.05393*.

[13] S. Chen, Y. Zhu, Q. Zhang, Z. Niu, and J. Zhu, "On optimal QoS-aware physical carrier sensing for IEEE 802.11 based WLANs: Theoretical analysis and protocol design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1369–1378, Apr. 2008.

[14] J. P. Monks, V. Bhargavan, and W.-M. Hwu, "A power controlled multiple access protocol for wireless packet networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, Anchorage, AK, USA, Apr. 2001, pp. 219–228.

[15] S. P. Bingulac, L. Fu, S. C. Liew, and J. Huang, "Effective carrier sensing in CSMA networks under cumulative interference," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.

[16] T.-S. Kim, H. Lim, and J. C. Hou, "Understanding and improving the spatial reuse in multihop wireless networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 10, pp. 1200–1212, Oct. 2008.

[17] J. Deng, B. Liang, and P. K. Varshney, "Tuning the carrier sensing range of IEEE 802.11 MAC," in *Proc. 47th Annu. IEEE Glob. Commun. Conf. (GLOBECOM)*, Dalas, Texas, USA, 2004, pp. 2987–2991.

[18] I. Jamil, L. Cariou, and J.-F. Hélard, "Novel learning-based spatial reuse optimization in dense WLAN deployments," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, pp. 1–19, Aug. 2016.

[19] S. Merlin and S. Abraham, "Methods for improving medium reuse in IEEE 802.11 networks," in *Proc. 6th Annu. IEEE Consum. Commun. Netw. Conf.*, Las Vegas, NV, USA, 2009, pp. 1–5.

[20] T. Nadeem and L. Ji, "Location-aware IEEE 802.11 for spatial reuse enhancement," *IEEE Trans. Mobile Comput.*, vol. 6, no. 10, pp. 1171–1184, Oct. 2007.

[21] J. Zhu, X. Guo, L. L. Yang, and W. S. Conner, "Leveraging spatial reuse in 802.11 mesh networks with enhanced physical carrier sensing," in *Proc. IEEE Inte. Conf. Commun. (ICC)*, Paris, France, 2004, pp. 4004–4011.

[22] K. Murakami, T. Ito, and S. Ishihara, "Improving the spatial reuse of IEEE 802.11 WLAN by adaptive carrier sense threshold of access points based on node positions," in *Proc. 8th Int. Conf. Mobile Comput. Ubiquitous Netw.*, Hakodate, Japan, 2015, pp. 132–137.

[23] I. Jamil, L. Cariou, and J.-F. Helard, "Improving the capacity of future IEEE 802.11 high efficiency WLANs," in *Proc. 21st Int. Conf. Telecommun.*, Lisbon, Portugal, 2014, pp. 303–307.

[24] T.-S. Kim, J. C. Hou, and H. Lim, "Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks," in *Proc. 12th Annu. Int. Conf. Mobile Comput. Netw.*, Los Angeles, CA, USA, 2006, pp. 366–377.

[25] S. Kim, S. Yoo, J. Yi, Y. Son, and S. Choi, "FACT: Fine-grained adaptation of carrier sense threshold in IEEE 802.11 WLANs," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1886–1891, Feb. 2017.

[26] L. Zhang, H. Yin, Z. Zhou, S. Roy, and Y. Sun, "Enhancing WiFi multiple access performance with federated deep reinforcement learning," in *Proc. 92nd IEEE Veh. Technol. Conf. (VTC-Fall)*, Victoria, BC, Canada, 2020, pp. 1–6.

[27] P. Kulkarni and F. Cao, "Dynamic sensitivity control to improve spatial reuse in dense wireless LANs," in *Proc. 19th ACM Int. Conf. Model. Anal. Simul. Wireless Mobile Syst.*, 2016, pp. 323–329, [Online]. Available: http://dl.acm.org/citation.cfm?doid=2988287.2989138

[28] X. Chen and J. Huang, "Distributed spectrum access with spatial reuse," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 593–603, Mar. 2013.

[29] S. Kim, J. Cha, and J. Ma, "Design and theoretical analysis of throughput enhanced spatial reuse distributed coordination function for IEEE 802.11," *IET Commun.*, vol. 3, no. 12, pp. 1934–1947, Dec. 2009.

[30] J. Zhu et al., "Adapting physical carrier sensing to maximize spatial reuse in 802.11 mesh networks," *Wiley Online J. Wireless Commun. Mobile Comput.*, vol. 4, no. 8, pp. 933–946, Nov. 2004.

[31] J. Liu, Y. Liu, J. Zhang, X. Ge, A. Xu, and M. Zhao, "A bayesian optimization algorithm to improve the spatial reuse in the next-generation WLANs," in *Proc. 20th Int. Wireless Commun. Mobile Comput.*, Ayia Napa, Cyprus, 2024, pp. 1048–1053.

[32] I. Selinis, K. Katsaros, S. Vahid, and R. Tafazolli, "Control OBSS/PD sensitivity threshold for IEEE 802.11ax BSS color," in *Proc. IEEE 29th Annun. Int. Symp*. Pers. Indoor Mobile Radio Commun., Bologna, Italy, Sep. 2018, pp. 1–7.

[33] T. Ropitault, "Evaluation of RTOT algorithm: A first implementation of OBSS PD-based SR method for IEEE 802.11ax," in *Proc. 15th IEEE Annu. Consum. Commun. Netw. Conf.*, Las Vegas, Nevada, USA, Jan. 2018, pp. 1–7.

[34] L. Lanante and S. Roy, "Performance analysis of the IEEE 802.11ax OBSS PD-Based spatial reuse," *IEEE/ACM Trans. Netw.*, vol. 30, no. 2, pp. 616–628, Apr. 2022.

[35] S. Joshi, R. Roy, R. V. Bhat, P. Hathi, and N. Akhtar, "Dynamic distributed threshold control for spatial reuse in IEEE 802.11ax," in *Proc. Nat. Conf. Commun.*, 2022, pp. 373–378. [Online]. Available: https://ieeexplore.ieee.org/document/9806744

[36] J. Jung, J. Baik, Y. Kim, H.-S. Park, and J.-M. Chung, "OTOP: Optimized transmission power controlled OBSS PD based spatial reuse for high throughput in IEEE 802.11 be WLANs," *IEEE Internet Things J.*, vol. 10, no. 19, pp. 17110–17123, Oct. 2023.

[37] A. Karakoç, H. B. Yilmaz, and M. Ş. Kuran, "More WiFi for everyone: Increasing spectral efficiency in WiFi6 networks using a distributed OBSS/PD mechanism," *Turk. J. Electr. Eng. Comput. Sci.*, vol. 31, no. 3, pp. 660–677, 2023.

[38] A. Bardou, T. Begin, and A. Busson, "Mitigating starvation in dense WLANs: A multi armed bandit solution," *Ad Hoc Netw.*, vol. 138, pp. 1–33, Jan. 2023.

[39] P. E. Iturria-Rivera, M. Chenier, B. Herscovici, B. Kantarci, and M. ErolKantarci, "Meta bandit: Spatial reuse adaptation via meta-learning in distributed Wi-Fi 802.11ax," *IEEE Netw. Lett.*, vol. 5, no. 4, pp. 179–183, Dec. 2023.

[40] H. Lee, H.-S. Kim, and S. Bahk, "LSR: Link-aware spatial reuse in IEEE 802.11ax WLANs," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Nanjing, China, 2021, pp. 1–6.

[41] B. Yin, K. Yamamoto, T. Nishio, M. Morikura, and H. Abeysekera, "Learning-based spatial reuse for WLANs with early identification of interfering transmitters," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 1, pp. 151–164, Mar. 2020.

[42] H. Shimizu et al., "Joint channel selection and spatial reuse for starvation mitigation in IEEE 802.11ax WLANs," in *Proc. IEEE 90th Veh. Technol. Conf.*, Honolulu, Hawaii, USA, 2019, pp. 1–6.

[43] A. Bardou and T. Begin, "INSPIRE: Distributed bayesian optimization for ImproviNg SPatIal REuse in dense WLANs," in *Proc. 25th Int. ACM Conf. Model. Anal. Simul. Wireless Mob. Syst.*, Montreal, QC, Canada, 2022, pp. 133–142.
[44] H. Kim, G. Na, H. Im, and J. So, "Improving spatial reuse of wireless LANs using contextual bandits," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 6735–6749, Jul. 2024.
[45] A. Valkanis, A. Iossifides, P. Chatzimisios, M. Angelopoulos, and V. Katos, "IEEE 802.11ax spatial reuse improvement: An interference-based channel-access algorithm," *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 78–84, Jun. 2019.
[46] F. Wilhelmi et al., "Federated spatial reuse optimization in next-generation decentralized IEEE 802.11 WLANs," *ITU J. Future Evol. Technol.*, vol. 3, no. 2, pp. 118–133, Sep. 2022.
[47] H. Zhang, R. He, X. Fang, and L. Zhou, "DDPG-based Multi-AP cooperative access control in dense Wi-Fi networks," in *Proc. IEEE Veh. Technol. Conf.*, Hong Kong, 2023, pp. 1–6.
[48] A. Bardou and T. Begin, "Analysis of a decentralized Bayesian optimization algorithm for improving spatial reuse in dense WLANs," *Comput. Commun.*, vol. 208, pp. 158–170, Aug. 2023.
[49] Y. Huang and K. Chin, "A deep Q-network approach to optimize spatial reuse in WiFi networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6636–6646, Jun. 2022.
[50] D. Nunez, F. Wilhelmi, S. Avallone, M. Smith, and B. Bellalta, "TXOP sharing with coordinated spatial reuse in multi-AP cooperative IEEE 802.11be WLANs," in *Proc. IEEE 19th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, Jan. 2022, pp. 864–870.
[51] D. Zhu, L. Wang, G. Pan, and S. Luan, "Two enhanced schemes for coordinated spatial reuse in IEEE 802.11be: Adaptive and distributed approaches," *Comput. Netw.*, vol. 258, pp. 1–11, Feb. 2025.
[52] K. Chemrov, D. Bankov, E. Khorov, and A. Lyakhov, "Support of realtime applications in Wi-Fi 8 with multi-AP coordinated parameterized spatial reuse," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw.*, Istanbul, Turky, 2023, pp. 226–231.
[53] D. Nunez, M. Smith, and B. Bellalta, "Multi-AP coordinated spatial reuse for Wi-Fi 8: Group creation and scheduling," in *Proc. 21st IEEE Mediterr. Commun. Comput. Netw. Conf.*, Lazio, Italy, 2023, pp. 203–208.
[54] R. Yan, Z. Guo, P. Liu, Q. Lan, X-P. Zhang, and Y. Dong, "Multi-agent reinforcement learning based channel access optimization for IEEE 802.11bn," *IEEE Trans. Green Commun. Netw.*, vol. 9, no. 3, pp. 1429–1441, Sep. 2025.
[55] L. Zhang, H. Yin, S. Roy, and L. Cao, "Multiaccess point coordination for next-gen Wi-Fi networks aided by deep reinforcement learning," *IEEE Syst. J.*, vol. 17, no. 1, pp. 904–915, Mar. 2023.
[56] Y. Kihira, K. Yamamoto, A. Taya, T. Nishio, Y. Koda, and K. Yano, "Interference-free AP identification and shared information reduction for tabular q-learning-based WLAN coordinated spatial reuse," *IEICE Commun. Exp.*, vol. 11, no. 7, pp. 392–397, Jul. 2022.
[57] M. Wojnar et al., "IEEE 802.11bn Multi-AP coordinated spatial reuse with hierarchical multi-armed bandits," *IEEE Commun. Lett.*, vol. 29, no. 3, pp. 428–432, Mar. 2025.
[58] M. Talukder and J. Xie, "Enhanced coordinated spatial reuse: Bidirectional multiple AP coordination for IEEE 802.11be," in *Proc. IEEE Int. Conf. Commun.*, Rome, Italy, 2023, pp. 660–665.
[59] J. Haxhibeqiri et al., "Coordinated spatial reuse for WiFi networks: A centralized approach," in *Proc. 20th IEEE Int. Conf. Factory Commun. Syst.*, Toulouse, France, 2024, pp. 1–8.
[60] F. Wilhelmi, B. Bellalta, S. Szott, K. Kosek-Szott, and S. Barrachina-Muñoz, "Coordinated multi-armed bandits for improved spatial reuse in Wi-Fi," 2022, *arXiv:2412.03076*.
[61] P. E. Iturria-Rivera, M. Chenier, B. Herscovici, B. Kantarci, and M. Erol-Kantarci, "Cooperate or not cooperate: Transfer learning with multi-armed bandit for spatial reuse in Wi-Fi," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 2, pp. 351–369, 2024.
[62] P. Imputato, S. Avallone, M. Smith, D. Nunez, and B. Bellalta, "Beyond Wi-Fi 7: Spatial reuse through Multi-AP coordination," *Comput. Netw.*, vol. 239, pp. 1–15, Feb. 2024.
[63] M. Zulfiker Ali et al., *Multi-AP Coordinated Spatial Reuse*, Document IEEE 802.11, Piscataway, NJ, USA, Dec. 2023. [Online]. Available: https://mentor.ieee.org/802.11/documents?is_dcn=1832is_group=00bn
[64] M. K. Simon and M.-S. Alouini, *Digital Communication Over Fading Channels: A Unified Approach to Performance Analysis*. Hoboken, NJ, USA: Wiley, 2005.
[65] S. J. Johnson, *Iterative Error Correction: Turbo, Low-Density Parity-Check and Repeat-Accumulate Codes*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[66] M. Franceschini, G. Ferrari, and R. Raheli, "Does the performance of LDPC codes depend on the channel?," *IEEE Trans. Commun.*, vol. 54, no. 12, pp. 2129–2132, Dec. 2006.
[67] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.
[68] J. Žerovnik, "Heuristics for NP-hard optimization problems–simpler is better!," *Logistics Sustain. Transp.*, vol. 6, no. 1, pp. 1–10, Nov. 2015.
[69] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 6382–6393.
[70] J. Yun, Y. Goh, W. Yoo, and J.-M. Chung, "5G Multi-RAT URLLC and eMBB dynamic task offloading with MEC resource allocation using distributed deep reinforcement learning," *IEEE Internet Things J.*, vol. 9, no. 20, pp. 20733–20749, Oct. 2022.
[71] A. Gao, Q. Wang, W. Liang, and Z. Ding, "Game combined multi-agent reinforcement learning approach for UAV assisted offloading," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 12888–12901, Dec. 2021.
[72] Z. Cheng, M. Min, M. Liwang, L. Huang, and Z. Gao, "Multiagent DDPG-Based joint task partitioning and power control in fog computing networks," *IEEE IoT J.*, vol. 9, no. 1, pp. 104–116, 2022.
[73] S. Merlin, *TGax Simul. Scenarios*. Accessed: Jan., 3, 2024.Online. Available: https://mentor.ieee.org

Jaewook Jung received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Republic of Korea, in 2019. He is currently working toward the combined M.S. and Ph.D. degrees in electrical and electronic engineering with Yonsei University, where he is a Researcher of the Communications and Networking Laboratory (CNL). His research interests include MPTCP, Wi-Fi, AI, security, and 5G/6G mobile systems.

Gangwoo Lee received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Republic of Korea, in 2023, and the M.S. degree in electrical and electronic engineering with Yonsei University in 2025, where he was a Researcher of the CNL. He is currently working with Korea Telecom (KT). His current research interests include 5G/6G mobile systems, handover, security, and deep learning.

Jong-Moon Chung (Fellow, IEEE) received the B.S. and M.S. degrees in electronic engineering from Yonsei University, Seoul, Republic of Korea, and the Ph.D. degree in electrical engineering from the Pennsylvania State University, Pennsylvania, PA, USA. Since 2005, he has been a Professor with the School of Electrical and Electronic Engineering and Director of CNL with Yonsei University, where he is also the Associate Dean of the College of Engineering and Professor of the Department of Emergency Medicine in the College of Medicine with Yonsei University Seoul, South Korea. He is an Eta Kappa Nu (HKN) member. He is also a member of the National Academy of Engineering of Korea (NAEK) and served as Secretary General of the NAEK Electrical & Electronic Engineering Division from 2023 to 2024. From 1997 to 1999, he was an Assistant Professor and instructor with the Pennsylvania State University in the Department of Electrical Engineering. From 2000 to 2005, he was with the Oklahoma State University (OSU) as a tenured Associate Professor in the School of Electrical and Computer Engineering.