A Reinforcement Learning Approach for Interference Management in Heterogeneous Wireless Networks

https://doi.org/10.3991/ijim.v15i12.20751

Akindele Segun Afolabi (✉)
University of Ilorin, Ilorin, Nigeria
afolabisegun@unilorin.edu.ng

Shehu Ahmed
The Nigerian Television Authority, Ilorin, Nigeria

Olubunmi Adewale Akinola
Federal University of Agriculture, Abeokuta, Nigeria

Abstract—Due to the increased demand for scarce wireless bandwidth, it has become insufficient to serve the network user equipment using macrocell base stations only. Network densification through the addition of low power nodes (picocell) to conventional high-power nodes addresses the bandwidth dearth issue, but unfortunately introduces unwanted interference into the network which causes a reduction in throughput. The purpose of this paper is to develop a model for controlling the interference between picocell and macrocell users of a cellular network so as to increase the overall network throughput. In order to achieve this, a reinforcement learning model was developed which was used in coordinating interference in a heterogeneous network comprising macrocell and picocell base stations. The learning mechanism was derived based on Q-learning, which consisted of agent, state, action, and reward. The base station was modeled as the agent, while the state represented the condition of the user equipment in terms of Signal to Interference Plus Noise Ratio. The action was represented by the transmission power level and the reward was given in terms of throughput. Simulation results showed that the trend of values of the learning rate (e.g., high to low, low to high, etc.) plays a major role in throughput performance. It was particularly shown that a multi-agent system with a normal learning rate could increase the throughput of associated user equipment by a whopping 212.5% compared to a macrocell-only scheme.

Keywords—Heterogeneous Network, Q-Learning, Macrocell, Picocell, Interference

1 Introduction

Mobile broadband usage has increased dramatically in the last couple of years due to new types of terminals such as smart phones and tablet computers [1, 2]. The traditional homogeneous networks [3, 4],comprising of only macrocell base stations (BSs), have become insufficient to meet the high traffic demands and stringent quality of service (QoS) requirements of mobile broadband communications [5]. A key method to fulfill the traffic demands is by network densification which involves adding smaller low power nodes, such as picocells, to traditional high power macro nodes. This results in what is termed "Heterogeneous Networks", or simply HetNets [3]-[7]. HetNets are expected to boost capacity and coverage beyond macrocells. They have been regarded as a promising paradigm to provide mobile users with high quality experience [2, 5]. However, network densification through the addition of picocells introduces harmful interference into the network [2, 4, 8]. Therefore, the influence of picocell densification on the network performance is obviously of large interest and the use of sophisticated inter-cell interference management techniques is very crucial. This paper aims at developing a learning model for coordinating inter-cell interference existing between a picocell and macrocell base stations for the purpose of improving network throughput.

The ability of learning new behaviours and adapt to the temporal dynamics of the system is associated with reinforcement learning (RL). Q-learning (QL) is a basic example of RL, which is proposed in this paper. The scenario of Q-learning is related to Markov decision method, where the learning agents interact with their environment to achieve the desired goals (rewards). Q-learning models have a set of states S, actions A, and rewards R. The learning cycle is a state-action-reward process. On learning, an agent takes an action $a \in A$ that interacts with the environment. The agent goes into a state $s(t) \in S$and receives a reward $r(s(t)) \in R$. The objective is to select actions at each state s, based on maximized reward r [9]-[12]. The agent should observe the state or environment and take actions that affect that state. Moreover, a goal must be introduced relating to the state of the environment. Learning can be performed using a centralised (single agent) [9, 10] or a distributed approach (multi-agent) [10, 11, 12]. A decentralized approach of learning is effective for solving complex problems. In this case, each agent in a multi-agent system is specialized at solving a particular problem. A multi-agent system is therefore useful, if a model can be developed for the agents' behaviour in terms of desires and goals. The performances of single and multi-agent systems are compared in this paper.

2 Related Works

The problem interference poses to heterogeneous networks has recently dominated discussions in the research community [13]-[26]. In heterogeneous cellular networks, a user equipment (UE) at the cell border always experiences high interference from neighbouring transmitters as their distance dependent attenuation is a critical issue [27]. Spectrum splitting is a mechanism used by the operator in a multi-tier network to split the available sub-bands among the cells in a cellular network to mitigate interference. Spectrum splitting can be carried out using a centralized approach. Splitting spectrum, in a centralized fashion, assigns sub bands to the macrocell base station (MBS) and small cells by means of a controller, which achieves efficient resource utilization at the expense of complexity and signalling overhead. The authors in [28] proposed an interference coordination method in which resource partitioning operation is done centrally, such that, sub-bands are given to the base stations in the network based on a weighted vertex colouring operation executed by a central controller.

The authors in [29]-[31] used beam-forming technique to mitigate interference. Specifically, [29] proposed a method called "dynamic interference steering", in which, interference is steered in an optimum direction where its impact on an interference victim is minimized. Ref. [30] applied a cross-layer approach where interference coordination is applied both at the Physical and MAC (Media Access Control) layers. Beam-forming is used to suppress interference at the physical layer while an optimisation problem is solved at the MAC layer to determine the set of users that will be accommodated by a resource block such that interference is reduced. In Ref [31], a beam selection scheme known as "beam skipping" is used to optimise a performance utility in a way that reduces inter-beam interference.

Self-organization and self-configuration are useful features that are usually exploited during interference coordination. Self-Organizing Networks (SONs) [32] attempt to minimize human intervention, where they use measurements from the network to minimize the cost of installation, configuration, and maintenance of a network [33]. Base stations can be made to self-organise by learning from their environment. Several research works on Q-Learning based interference coordination exist in the literature, in which, agents (usually base stations) self-organise based on network measurements. A self-organised method was proposed for mitigating interference in [9]. It considers a vehicular network, in which, a base station agent selects optimal resource control policy during each action policy interval of the learning and running phases. During the learning phase, the agent is trained to maximize expected future reward by updating Q elements till the attainment of convergence. The running phase involves the agent choosing the action that yields the highest expected reward from updated Q.

Ref. [34] presents a multi-agent deep reinforcement learning system where femtocell and macrocell base stations act as agents whose goal is to maximize network capacity. Neural network used in the system enhances its ability to process a large amount of state information. The parallel operation of multiple agents ensures that the overall network interference is reduced in order to achieve an enhancement in capacity. In [35] a downlink reinforcement learning-based interference control algorithm is presented. The algorithm employs convolutional neural network to estimate Q values which as a result reduces the size of the state space. After a sufficient number of power control iterations, the network throughput is significantly increased. The authors in [36] developed a reinforcement learning algorithm for optimal configuration of interference coordination parameters, which have stochastic characteristics, such as, location of users, traffic demands, and strength of received signals.

3 System Model and Formulations

In this model, learning based strategy for interference coordination in an environment, where macrocell and picocell co-exist is considered. Our focus is on the analysis of a network deployment with a picocell underlaying a macrocell network. The total bandwidth (BW) of the network is divided into sub-carriers, with each sub-carrier having a bandwidth of Δf (15 kHz). Resource blocks are grouped using orthogonal frequency division multiplexing (OFDM) symbols as shown in Figure 1. Both macrocell and picocell operate in the same frequency band, and they have access to the same set of resource blocks.

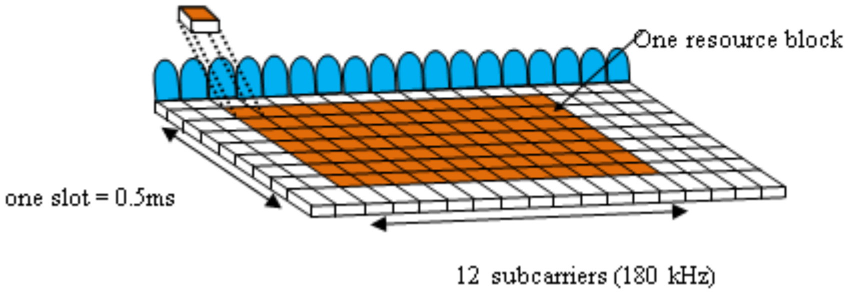


Fig. 1. LTE downlink physical resource based on OFDM

When picocell and macrocell utilise the same spectrum, inter-cell interference problem emerges. A typical collocation scenario of picocell and macro-cell is shown in Figure 2. In this scenario, the downlink transmissions from the MBS or picocell base station (PBS) will create a strong interference at a nearby macrocell user equipment (MUE) or picocell user equipment (PUE) and may cause the received macrocell or picocell signal at the MUE or PUE to be degraded. Hence, inter-cell interference hampers a successful macrocell and picocell co-existence.

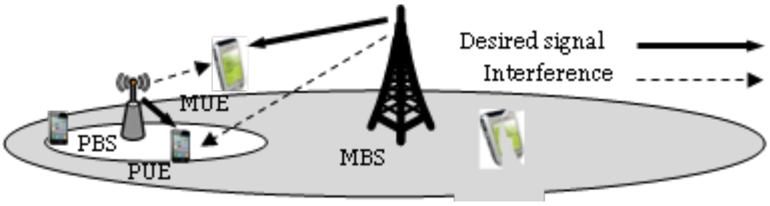


Fig. 2. A heterogeneous scenario

3.1 Computation of signal to interference plus noise ratio

We consider that each MBS has a set of MUEs associated to it and the MUEs periodically report the quality of each resource block (RB) in terms of signal to interference plus noise ratio (SINR) to their serving MBS in order to facilitate a channel-aware scheduling. PBSs in the vicinity of an MUE constitute interference sources to the downlink signal of the MUE, and in a similar way, MBS interferes with the downlink signal of PUE. Without the loss of generality, only downlink transmission is considered in this work, since interference in the downlink is usually more severe than the uplink, especially when the interfering bases station is in close proximity to the victim UE. It can be observed in Figure 2 that the interference between macrocell and picocell is mutual; which means that, the transmissions of MBS interfere with PUE's received signal and also, the transmissions of the PBSs interfere with MUE's received signal. The SINR of MUE *i* is computed as:

$$\gamma_{\mathrm{MUE}i} = \frac{P_{\mathrm{MBS}d} h_{\mathrm{MBS}d,\mathrm{MUE}i}}{\sum_{f=1, f \neq d}^{|M|} P_{\mathrm{MBS}f} h_{\mathrm{MBS}f,\mathrm{MUE}i} + \sum_{g=1}^{|\psi|} P_{\mathrm{PBS}g} h_{\mathrm{PBS}g,\mathrm{MUE}i} + \delta^2} \tag{1}$$

where $P_{\mathrm{MBS}d}$ denotes the transmit power from serving MBS *d* to the *i*-th MUE,
$P_{\mathrm{MBS}f}$ denotes the transmit power from interfering MBS *f* to the *i*-th MUE,
$\mathrm{P}_{\mathrm{PBS}g}$ denotes the transmit power from interfering PBS *g* the *i*-th MUE,
$h_{\mathrm{MBS}d,\mathrm{MUE}i}$ denotes the link gain between serving MBS *d* and the *i*-th MUE,
$h_{\mathrm{MBS}f,\ \mathrm{MUE}i}$ denotes the link gain between interfering MBS *f* and the *i*-th MUE,
$h_{\mathrm{PBS}g,\ \mathrm{MUE}i}$ denotes the link gain between interfering PBS *g* and the *i*-th MUE,
δ^2 denotes the noise power. Similarly, the SINR of PUE *j* is computed as:

$$\gamma_{\mathrm{PUE}j} = \frac{P_{\mathrm{PBS}k} h_{\mathrm{PBS}k,\mathrm{PUE}j}}{\sum_{f=1}^{|M|} P_{\mathrm{MBS}f} h_{\mathrm{MBS}f,\mathrm{PUE}j} + \sum_{g=1, g \neq k}^{|\psi|} P_{\mathrm{PBS}g} h_{\mathrm{PBS}g,\mathrm{PUE}j} + \delta^2} \tag{2}$$

where$P_{\mathrm{PBS}k}$ denotes the transmit power from serving PBS *k* to the *j*-th PUE,
$h_{\mathrm{PBS}k,\ \mathrm{PUE}j}$ denotes the link gain between serving PBS *k* and the *j*-th PUE,
$h_{\mathrm{MBS}f,\ \mathrm{PUE}j}$ denotes the link gain between interfering MBS *f* and the *j*-th PUE,
$h_{\mathrm{PBSg},\ \mathrm{PUE}j}$ denotes the link gain between interfering PBS *g* and the *j*-th PUE.

By applying Shannon's capacity formula, the data rate achieved on an RB with SINR γ scheduled by the base station is computed as:

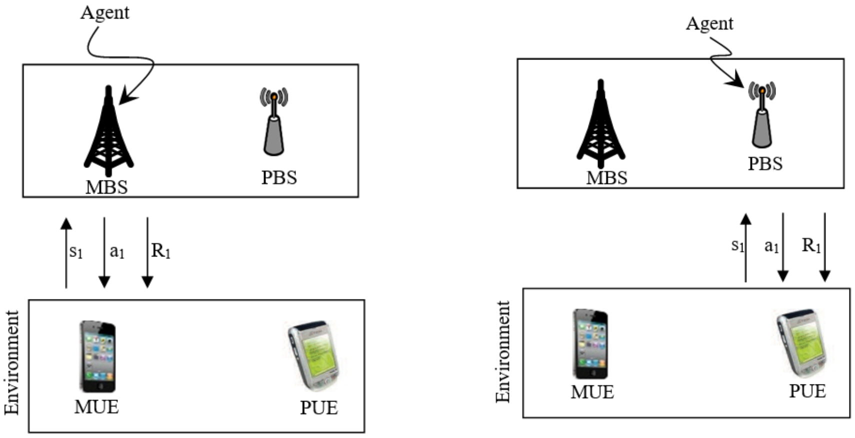
$$C = BW_{\mathrm{RB}} \log_2 (1+\gamma) \tag{3}$$

where BW_{RB} denotes resource block bandwidth (in Hertz). The throughput of a UE is a function of the SINR and is expressed as:

$$TP_{\mathrm{user}} = f(\gamma) \tag{4}$$

3.2 Model formulation

This section presents a single agent, and also, multi-agent learning approach in order to solve the inter-cell interference problem in HetNets. In this study, the base-station is modeled as the learning agent as shown in Figures 3(a) and (b). It learns the condition or state of the UE in terms of interference level before taking an action of power allocation on RBs of UEs, while ensuring that the best reward in terms of the throughput is realised. We consider both single and multi-agent Q-learning models, such that, in the former, either PBS or MBS acts as an agent (not both concurrently), while in the latter, they concurrently both act as agents. Throughout the rest of this paper, the terms, "Q-learning" and "reinforcement learning" are used interchangeably.



a) A single agent learning model where macrocell BS is the learning agent.

b) A single agent learning model where picocell BS is the learning agent.

Fig. 3. Agent learning model

In the next section, we present our proposed learning method for mitigating the aggregate interference generated between the picocell and the macrocell of a HetNet. The section also introduces the concept of learning rate.

Enhanced inter-cell interference coordination based on single agent Q-learning model: A single agent, which is either a picocell or macro-cell BS (but not both) learns the condition of the SINR of the resource blocks of its UE. Typical single agent scenarios are illustrated in Figures 3(a) and (b). Let *D* represent a set of similar base stations; in this context, we are assume that picocell BSs are similar to each other but are dissimilar to macrocell and vice versa; then, a single agent is $x \in D$ such that

$$D \cup D' = \psi \cup M \,. \tag{5}$$

The actions of the learning agent $x \in D$, the associated states, and reward functions are explained next:

- Agent: This is base station x, which is a member of the set of similar base stations D that satisfies Eqn. (5).
- State: The state represents the condition of a UE within the cell of agent $x \in D$ based on the SINR seen on RB r of its UE. The set of states of a UE for all N RBs can be represented mathematically by:

$$\boldsymbol{S}^{x} = \left\{ s_{r}^{x} \right\}_{r \in \{1,\ldots,N\}} \tag{6}$$

where,

$$s_r^x = \begin{cases} 0 \text{ if } \gamma_r^x < \gamma_\mathrm{T} & \text{bad state} \\ 1 \text{ if } \gamma_r^x \geq \gamma_\mathrm{T} & \text{good state} \end{cases} \tag{7}$$

where γ_r^x is the instantaneous value of the SINR reported by a UE on a resource block *r* served by a learning agent $x \in D$, while γ_T is the SINR threshold used in classifying an RB as either being in a good or bad state. Depending on what the state $s_r^x \in \boldsymbol{S}^x$ of an RB of its UE is, the agent takes an action. For instance, if the agent observes that an RB *r* of a UE has an SINR less than the threshold value γ_T, it takes an appropriate action which will be different from the action it would take when the SINR is above this threshold. The actions of the BS are described next.

- Action*:* The action is the power level allocation by the single agent $x \in D$ to the resource block $r \in \{1,2,\ldots,N\}$ of its served UE. The possible set of actions is represented mathematically as:

$$\boldsymbol{A}^x = \{a_r^x\}_{r\in\{1,\cdots,N\}} \tag{8}$$

where,

$$a_r^x = \begin{cases} 0 \quad \text{if } \; s_r^x = 0 & \text{zero power will be loaded on resource block } r \\ 1 \quad \text{if } \; s_r^x = 1 & \text{full power will be loaded on resource block } r \end{cases} \tag{9}$$

where a_r^x is the transmitted power level of agent $x \in D$ on resource block *r*. In this paper, we consider two possible levels of transmitted power for each resource block. They are maximum power level and zero power level. Maximum power is loaded when the reported SINR for an RB is above the threshold value γ_T, which indicates a state of 1, while zero power is loaded for a state of 0.

- Reward*:* The reward is the capacity achieved by the single-agent $x \in D$ on resource block *r* when it is transmitting at a power level to a UE associated to it. It is represented mathematically as:

$$R_r^x = C_r^x \tag{10}$$

where C_r^x of agent $x \in D$ is computed according to Eqn. (3) such that, whenever zero transmission power (bad state of resource block of UE) is loaded on a resource block *r*, the capacity will be zero for that resource block; meaning that the reward R_r^x is zero and vice versa. A reward of 0 is regarded as a penalty, and by learning an optimal policy, agent *x*, after some time, will be able to avoid actions causing zero rewards but instead, will take the ones that yield higher rewards.

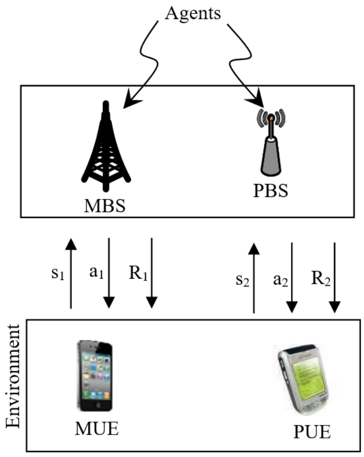


Fig. 4. A multi-agent learning model where both macro-cell BS and picocell BS are learning agents.

Enhanced inter-cell interference coordination based on multi-agent Q-learning model: In this section, the multi-agent Q-learning model in which, both picocell and macrocell serve as learning agents that learn the condition of the UE is introduced. This is called a multi-agent Q-learning approach where multiple agents (picocell and macrocell) aim at carrying out the learning process by repeatedly interacting with the environment to provide the best reward to their associated user equipment. A typical multi-agent learning scenario is illustrated in Figure 4. The set of learning agents in this case comprise all picocell and macrocell base stations represented as $\psi \cup M$. The learning agent, actions, states, and reward functions are designed and explained as follows:

- Agent*:* An agent y is a member of $\psi \cup M$
- State*:* The state represents the condition of a UE within the cell of agent $y \in (\psi \cup M)$ based on the SINR seen on RB *r* of the UE. The set of states of a UE for all N RBs can be represented mathematically by:

$$\boldsymbol{S}^{y} = \{s_r^y\}_{r \in \{1,\ldots,N\}} \tag{11}$$

where

$$s_r^y = \begin{cases} 0 \;\text{ if }\; \gamma_r^y < \gamma_\mathrm{T} & \text{bad state} \\ 1 \;\text{ if }\; \gamma_r^y \geq \gamma_\mathrm{T} & \text{good state} \end{cases} \tag{12}$$

where γ_r^y is the instantaneous value of the SINR reported by a UE on a resource block *r* served by a learning agent $y \in (\psi \cup M)$. $s_r^y \in \boldsymbol{S}^y$ represents the state of an RB *r* of the UE in the cell of agent $y \in (\psi \cup M)$, such that the state takes value of 0 if the SINR of the user falls below a certain threshold value γ_{T}, but takes a value of is 1, if otherwise. MBS and PBS, as agents, have the capability of jointly observing interference levels through periodical SINR reports received from their associated UEs. If the reported SINR falls below a threshold γ_{T}, the BSs identifies the RB as occupied (i.e., bad state) and takes a subsequent action which is explained next.

- Action*:* For a multi-agent scenario*,* the action is defined as:

$$\boldsymbol{A}^{y} = \{a_r^y\}_{r\in\{1,\ldots,N\}} \tag{13}$$

where

$$a_r^y = \begin{cases} 0 \quad \text{if} \quad s_r^y = 0 \quad \text{zero power will be loaded on resource block } r \\ 1 \quad \text{if} \quad s_r^y = 1 \quad \text{full power will be loaded on resource block } r \end{cases} \tag{14}$$

where a_r^y, just like the single agent case, is the transmitted power level of agent $y \in (\psi \cup M)$ on resource block *r.*

- Reward: In this paper, the reward is the capacity achieved by the multi-agent *y*, while transmitting to a UE in its cell. It is represented mathematically as:

$$R_r^y = C_r^y \tag{15}$$

where C_r^y is computed according to Eqn. (3). The rationale behind this reward function is that the Q-learning model aims to select optimum power level capable of improving the capacity of UEs associated to agent $y \in (\psi \cup M)$.

Algorithm of Q-learning for inter-cell interference coordination scenario: To achieve interference coordination, exploration is performed and the Q-learning equation is updated as:

$$Q^*(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha(R + \beta(\max Q(s',a') - Q(s,a))) \tag{16}$$

where $Q^*(s,a)$ is the learnt or updated *Q*-value corresponding to the Quality value of state (interference) and the action (allocated power level) that gave the best reward in terms of throughputs of the UE. $Q(s,a)$ is the previous *Q*-values corresponding to the quality value of state (interference) and the action (power level allocation) that was previously learnt by the agent (base station) that does not result into an optimum reward in terms of throughputs to the UE [34]. $Q(s',a')$ is the next optimal *Q*-value learnt by the agent (base station) after observing that $Q(s, a)$ is not optimal to give a correct update. Action is denoted as a, that is, transmitted power level of the agent (or base station). α and β are the learning rate and discount factor, respectively. Note that x, y, and r are not included in Eqn. (16) in order to reduce nomenclature complexities. The algorithm for computation of the Q-value and associated parameters is given in Algorithm 1 [34].

```
Algorithm 1
Initialisation:
    For each $s \in S$, $a \in A$ do
    |   Initialize $Q^*(s,a) \leftarrow Q(s,a)$
    End For
    Evaluate the starting state $s$
Learning:
    Loop
    |   Select the action $a$
    |   Execute $a$
    |   Receive an immediate reward $R$
    |   Observe the next state $s'$
    |   Select the action corresponding to the maximum $Q$-value in state $s'$
    |   Update the $Q$-value as follows:
    |   $Q^*(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha(R + \beta(\max Q(s',a') - Q(s,a)))$
    |   Update $s = s'$
    End loop
```

Learning rate (α) model: Learning rate implies willingness of the agent to learn from its environment. Three types of learning rate are considered in this paper. These are:

Normal learning rate:
$$\alpha = \frac{1}{\Theta} \tag{17}$$

Logarithm learning rate:
$$\alpha = \log\left(\frac{\Theta+1}{\Theta}\right) \tag{18}$$

Polynomial learning rate:
$$\alpha = \frac{\Theta}{1+\Theta^2} \tag{19}$$

where Θ is the state learning indicator and the update of α is constrained by Eqn. (20) which is computed as:

$$\alpha \leftarrow \begin{cases} 0+\varepsilon & \text{if } \alpha \le 0 \\ 1-\varepsilon & \text{if } \alpha \ge 1 \\ \alpha & 0<\alpha<1 \end{cases} \tag{20}$$

where ε is a small positive number greater than 0. Equation (20) ensures that α is always maintained in the interval (0,1) during any learning episode.

Analysis of learning rate model: Consider Eqn. (16), it is observed that 1-*α* and *α* are the weighting factors of $Q(s,a)$ and $(R+\beta(\max Q(s',a')-Q(s,a)))$, respectively. This indicates that when *α* is high, $(R+\beta(\max Q(s',a')-Q(s,a)))$ contributes significantly to the updated value of $Q^*(s,a)$. This allows the system to explore new state and action pairs that have the tendency of yielding higher rewards. On the other hand, when *α* is low, $Q(s,a)$ contributes significantly to the updated value of $Q^*(s,a)$, making the system to potentially adopt already known values (i.e., exploitation is favoured). Figure 5 graphically illustrates the relationship between learning rate and state learning indicator.

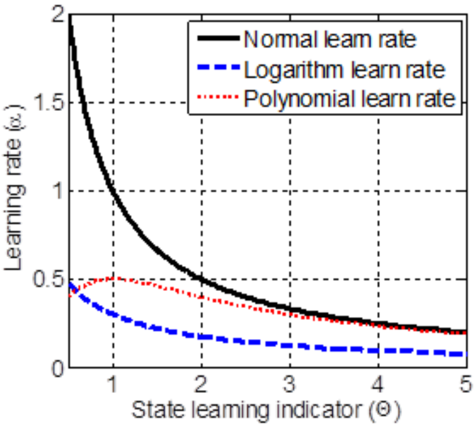


Fig. 5. Illustration of *α* as a function of *Θ*

It can be observed that for Normal learn rate and Logarithm learn rate, the learning rate is a monotonically decreasing function of the state learning indicator. In the case of the Polynomial learning rate, the learning rate initially increases and subsequently decays as the state learning indicator is increased. If we consider 2 extreme values of *Θ* (i.e., 1.0 and 5.0) the behaviour of the learning system can be summarised as shown in Table 1.

Table 1. Summary of the Behaviour of the Learning System

Model	State learning indicator (Θ)	Learning rate (α)	Implication on system	
			Exploration	***Exploitation***
Normal learning rate	1.0	1.0	Very high	No
	5.0	0.22	Low	High
Polynomial learning rate	1.0	0.5	Moderate	Moderate
	5.0	0.22	Low	High
Logarithm learning rate	1.0	0.3	Low	High
	5.0	0.09	Very low	Very high

4 Results and Discussions

4.1 Simulation setup

Three scenarios were evaluated in the simulation and all of them will be discussed shortly. The simulation parameters used are shown in Table 2.

Table 2. Simulation Parameters

Simulation Parameter	Value
Cellular layout	Hexagonal cell
Carrier frequency	2 GHz
Bandwidth	10 MHz
Number of resource blocks(RBs)	50
Number of MBS	1
Number of PBS	1
Maximum MBS transmit power	46 dBm
Maximum PBS transmit power	30 dBm
Macrocell path loss model (*d* in km)	128.1 + 37.6 $\log_{10}$ (d) dB
Picocell path loss model (*d* in km)	140.7 + 37.6 $\log_{10}$ (d) dB
Target SINR (γ_T)	18 dB
Thermal noise density	-174 dBm/Hz
Discount factor (β)	0.8
Number of MUEs	20
Number of PUEs	20
Macro-cell radius	1 km
Pico-cell radius	114 m
Scheduling type	Proportional fair

4.2 Average UE throughput

Figure 6 shows the throughput performances of under-laying a picocell on a macrocell network. As can be observed in the figure, when there is no picocell base station present, only the macrocell base station served the UEs in the network. Due to the high loading of the MBS system, UE throughput suffers. However, adding picocell resulted in an improvement in the overall average UE's throughput; for instance, from

0.32 Mbps to 0.43 Mbps. This is a 34.4% increase in throughput which indirectly translates to a 34.4% increment in data transmission speed. When the picocell is introduced to the macro-cell layout, some of the UEs that are now served by the picocell are able to get some throughput increment. However, the deployment of picocell comes at a cost, which is the introduction of unwanted interference into the network. It is therefore important to address this interference in order to ameliorate its negative impact on the overall network. This can be achieved through reinforcement learning-based inter-cell interference coordination technique that is being proposed in this paper and the simulation results of the proposed scheme will be discussed next.

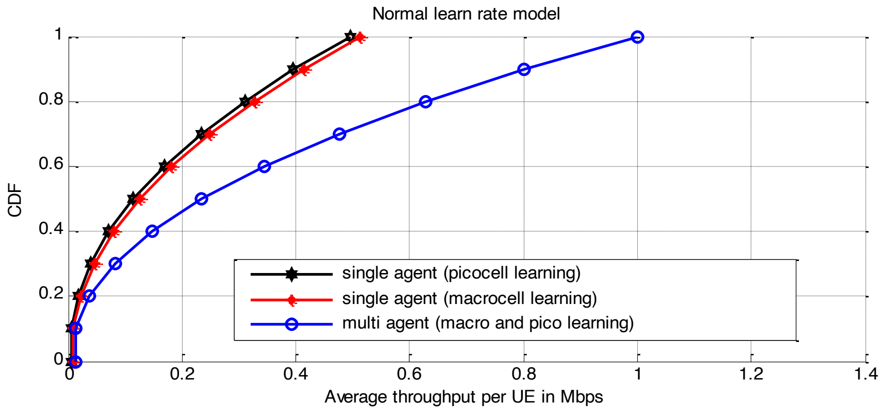


Fig. 6. Graph of Cumulative Distribution Function (CDF) against average UE throughput

4.3 Inter-cell interference coordination using the proposed Q-learning model

The reinforcement learning model discussed in Section 2.2 was applied to the network and the result in terms of cumulative distribution function (CDF) of average UE throughput is shown in Figure 7. As can be observed in the figure, multi-agent reinforcement learning model outperforms single agent reinforcement learning when normal learning rate (i.e., when Eqn. (17) is applied) is used, while the single agent macrocell learning model is slightly better than single agent picocell learning model.

The lower performance of the single agent system stems from the fact that the agents have a limited knowledge of the radio environment in terms of interference level, which makes them to take arbitrary decisions (or actions) that subsequently caused harmful interference to other network users in their vicinity, thereby reducing the overall throughput gain. The higher performance of multi-agent system shown in the figure is because the agents are allowed to learn from each other and this in turn improves cooperation among them, thereby yielding again in throughput. Figure 8 shows a comparison of the performances of multi-agent and single agent systems. As can be observed in the figure, multi-agent reinforcement learning model outperforms single agent reinforcement learning model in terms of CDF of average UE throughput when logarithm learning rate (i.e., when Eqn. (18) is applied) is used, while the single agent macrocell learning model is slightly better than single agent picocell learning model.

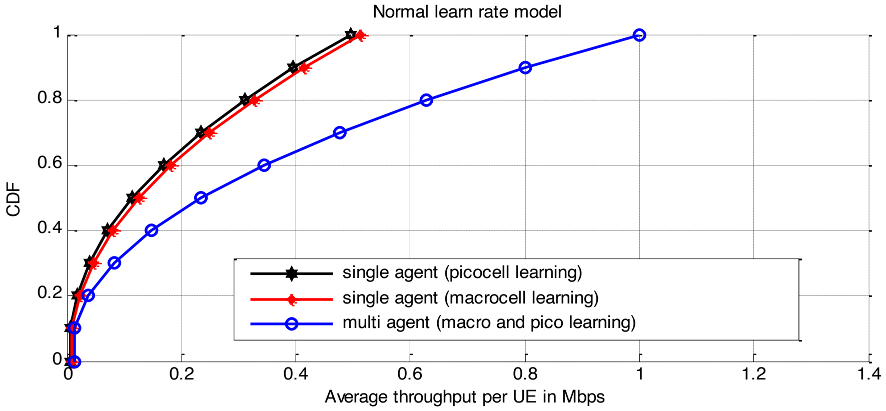


Fig. 7. Comparing effects on throughput of UEs due to single agent and multi-agent Q-learning model using a normal learn rate

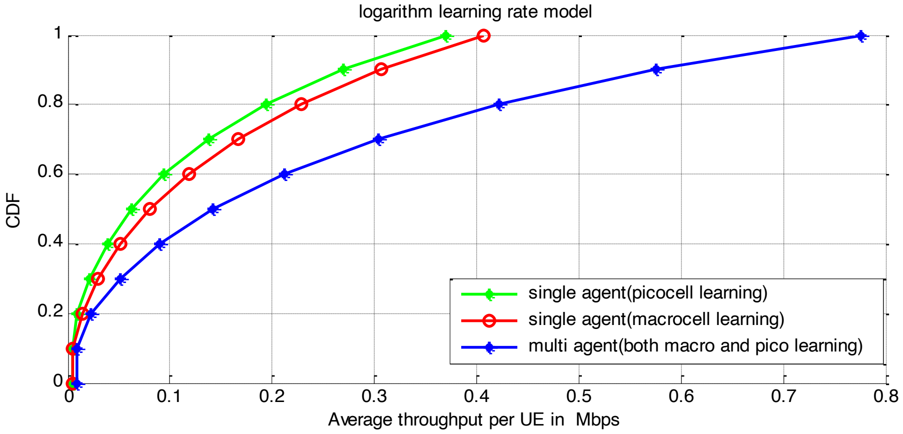


Fig. 8. Comparing effects on throughput of UEs due to single agent and multi-agent Q-learning model using a logarithm learn rate.

Figure 9 compares the performances of multi-agent and single agent systems when polynomial learning rate (i.e., when Eqn. (19) is applied) is used, and again, it can be observed that multi-agent reinforcement learning model outperforms single agent reinforcement learning model in terms of CDF of average UE throughput, while the single agent macrocell learning model is much better than single agent picocell learning model. The single agent picocell learning model showed the lowest performance. Based on the results shown in Figures 7 through 9, it can therefore be inferred that UEs benefit more from multi-agent learning schemes than from single-agent learning schemes.

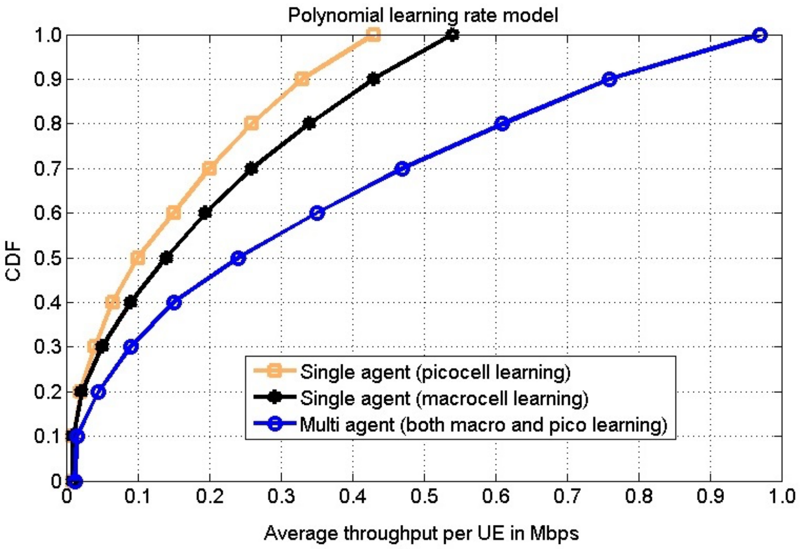


Fig. 9. Comparing effects on throughput of UEs due to single agent and multi-agent Q-learning model using a polynomial learn rate.

4.4 Comparing throughput performance of multi-agent Q-learning model in terms learning rate

The normal, logarithm, and polynomial learning rates are applied to the proposed multi-agent system and the results in terms of CDF of average UE throughput are compared as shown in Figure 10. The figure shows that the normal learning rate achieved higher average UE throughput compared to logarithm and polynomial learning rate, while polynomial learning rate has much better performances than the logarithm learning rate. In particular, normal learning rate shows a 28.2% throughput gain over normal learning rate, while polynomial learning rate has a 25.6% gain over logarithm learning rate. The higher throughput achieved by the normal learning rate is as a result of the fact that the system was able to first explore the environment due to low initial value of state learning indicator and correspondingly high value of learning rate as was illustrated in Figure 5 and Table 1. After a number of episodes, the value of the state learning indicator would have sufficiently increased to result in low value of learning rate. By this time, the system must have sufficiently explored the environment and exploitation based on learnt values can be done. The early exploration and later exploitation by the system ensures that the system is not trapped in a local optimum region. In the case of polynomial learning rate, both exploration and exploitation are performed together with the same preference at the initial stage while the system switches to more of exploitation at a later time. This results in UE throughput that is lower than that of normal learning rate but higher than logarithm learning rate. The values of learning rate for polynomial learning rate model can be observed in Figure 5. The logarithm learning rate performed worst in terms of average UE throughput since low values of learning rate are always selected which prevents the system from engaging in sufficient exploration and possibly causing the system to be trapped in a local optimum region which yields lower average UE throughput as shown in Figure.10

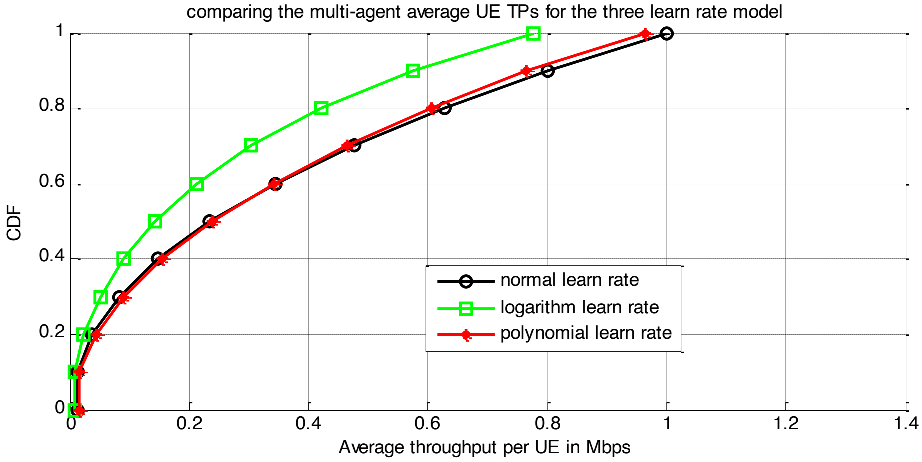


Fig. 10. Comparing effects on throughput of UEs due to multi-agent Q-learning model

4.5 Comparing the multi agent Q-learning scheme results to no learning scenario

The no learning scenario, which consisted of macrocell deployment when under-laid with picocell, is used as the reference scheme to compare with the proposed Q-learning schemes. The results are compared based on the cumulative distribution function (CDF) of average UE throughput as shown in Figures 11 through13. Figure 11 shows the comparison of the CDF of the average UE throughput between a macrocell only, macrocell under-laid with a picocell, and macro+picocell with multi agent Q-learning using logarithm learning rate. As can be observed in the figure, deploying picocell helped to increase the average UE throughput from 0.32 Mbps to 0.43 Mbps, which is a 34.4% throughput increment. The overall performances, as shown in Figure 11, is even further boosted by the introduction of multi-agent Q-learning (logarithm learn rate) scheme. It can be observed that the maximum throughput obtained with the introduction of Q-learning (logarithm learning rate) is 0.77 Mbps as against 0.32 Mbps in a macrocell-only deployment. This is a 140.6% throughput increment.

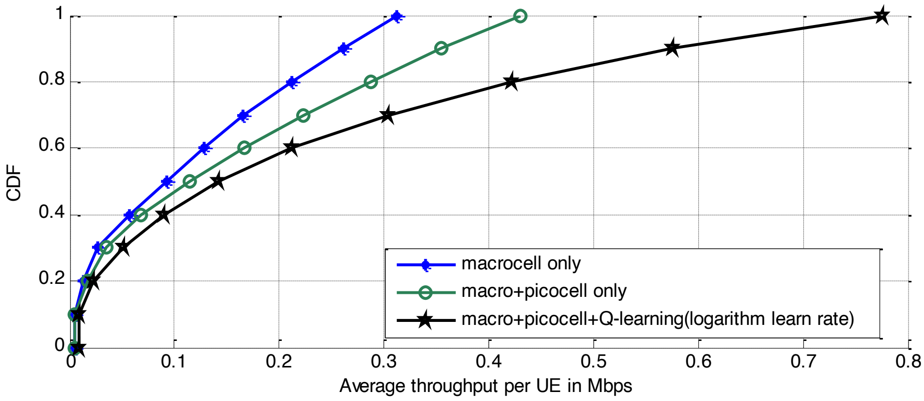


Fig. 11. Comparing between the throughput of UEs based on multi agent Q-learning scheme (logarithm learning rate) and no learning scenario.

Figure 12 illustrates the comparison of the CDF of the average UE throughput of between macrocell-only, macrocell underlaid with a picocell, and macro+picocell with multi agents Q-learning using polynomial learning rate. The overall performance, as shown in the figure, is boosted by the introduction of multi-agent Q-learning (polynomial learning rate) scheme. The scheme yielded a whooping 200% increase in the average UE throughput (i.e., 0.96 Mbps) compared to macrocell-only deployment (i.e., 0.32 Mbps).

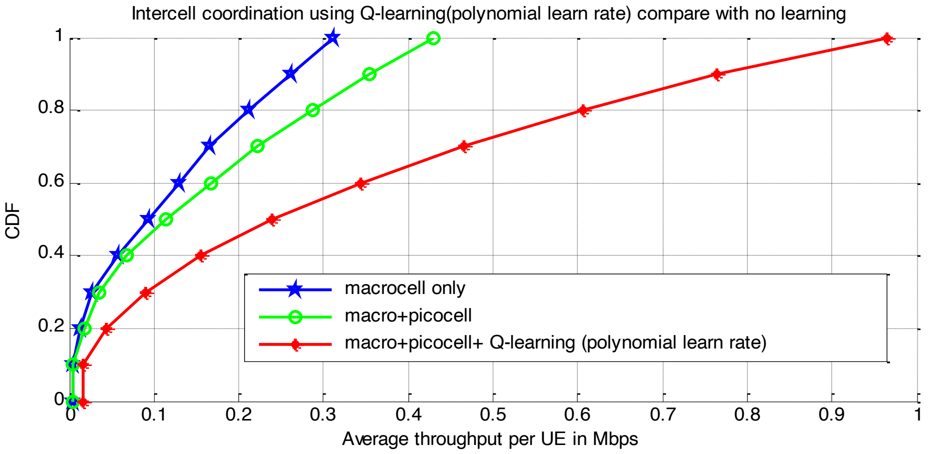


Fig. 12. Comparison between throughput of UEs based on multi agent Q-learning scheme (polynomial learning rate) and no learning scenario.

Figure 13 illustrates the comparison of the CDF of the average UE throughput between a macrocell-only, macrocell underlaid with a picocell, and macro+picocell with multi-agent Q-learning using normal learning rate. It can be observed that the multi-agent Q-learning (normal learning rate) scheme yielded a whooping 212.5% increase in the average UE throughput (i.e., 1 Mbps) compared to macrocell-only deployment (0.32 Mbps).

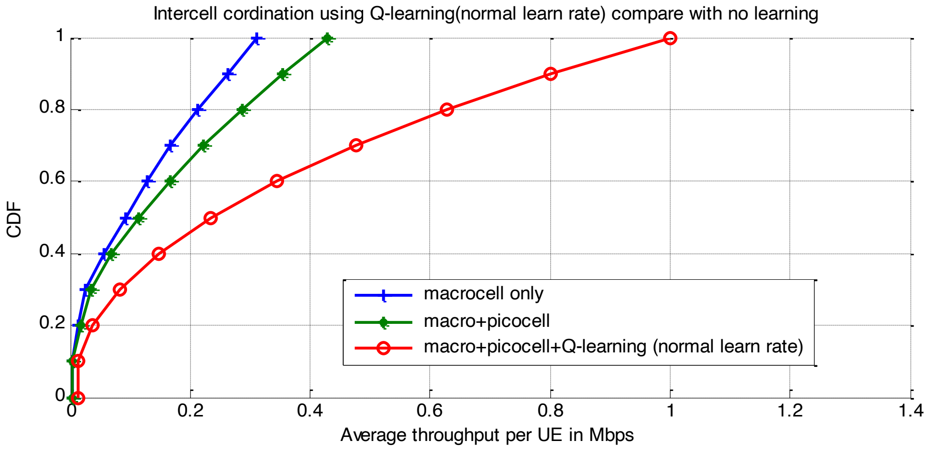


Fig. 13. Comparison between throughput of UEs based on multi agent Q-learning scheme (polynomial learning rate) and no learning scenario.

5 Conclusion

In this paper, single agent and multi-agent Q-learning models have been analyzed from the perspective of their effectiveness in interference coordination in heterogeneous wireless networks. The results show that Q-learning model, using multi-agent system, outperforms the single agent scenario since the multi-agent Q-learning scheme significantly improved spectrum utilization compared to the single agent Q-learning counterpart.

Three learning rates were proposed, namely, Normal learning rate, Polynomial learning rate, and Logarithm learning rate. For the normal learning rate, the initial values of the learning rate are very high which ensures that the system accorded more preference to environment exploration than exploitation of known state-action pairs. As time progresses, the value of the learning rate decays which causes the system to accord more preference to exploitation than exploration. These particular settings made the normal learning rate to perform much better that the other 2 learning rates proposed.

The polynomial learning rate, on the other hand, has an initial medium value of learning rate. This causes the system to initially accord the same preference to exploration and exploitation. The throughput performance of polynomial learning rate is not as good as the normal learning rate but is better than logarithm learning rate which consistently selected low values of learning rate throughout the experiment. Simulation experiment shows that a multi-agent system based on normal learning rate could achieve a throughput gain of up to 212.5% compared to a macrocell-only reference scheme.

For future work, a hybrid of two or more of the proposed learning rates could be developed and tweaked for possible performance improvement.

6 References

[1] Kumar, d., Chellappan, C., Vani, B. B. J. (2009). Dynamic Resource Management in MC-CDMA Based Cellular Wireless Networks. International Journal of Interactive Mobile Technologies, 3(2): 51-57. https://doi.org/10.3991/ijim.v3s2.890

[2] Xiao, L., Li, Z., Zhen, L., Zhang, Xu, B. Y., and Liu, J. (2019). Taichi: Signal Dodging and Interference Guiding. 2019 International Conference on Networking and Network Applications (NaNA), South Korea, pp. 39-44, https://doi.org/10.1109/nana.2019.00016.

[3] Khan, H. Z., Ali, M., Naeem, M., Rashid, I., Siddiqui, A. M., Imran, M., and Mumtaz, S. (2020). Joint admission control, cell association, power allocation and throughput maximization in decoupled 5G heterogeneous networks. Telecommunication Systems, 76: 115-128. https://doi.org/10.1007/s11235-020-00707-4

[4] Moreira, J., Silva, É. R., and Guardiero, P. R. (2020). eICIC Optimization Improvements in Downlink Resource Allocation in LTE-A HetNets. Journal of Communication and Information Systems, 35(1): 15-24. https://doi.org/10.14209/jcis.2020.2

[5] Dai, M., Su, Z.,Xu, Q., and Chen, W. (2019). A Q-Learning Based Scheme to Securely Cache Content in Edge-Enabled Heterogeneous Networks. IEEE Access, 7: 163898-163911. https://doi.org/10.1109/access.2019.2946319.

[6] Supriadi H. and Putri H. (2020). Range expansion method on heterogeneous network to increase picocell coverage. TELKOMNIKA Telecommunication, Computing, Electronics and Control, 18(5): 2341-2351, https://doi.org/10.12928/telkomnika.v18i5.14640.

[7] Ghorbani, K. and Falahati, A. (2020). Improving symbol error rate of heterogeneous network by clustering, IA, and interference detection. IET Electronics Letters, 56(8): 408-410. https://doi.org/10.1049/el.2019.3736

[8] Zou, Y. (2018). Intelligent Interference Exploitation for Heterogeneous Cellular Networks against Eavesdropping. IEEE Journal on Selected Areas in Communications, 36(7): 1453-1464. https://doi.org/10.1109/jsac.2018.2824258

[9] Zhou, Y., Tang, F., Kawamoto, Y., and Kato, N. (2020). Reinforcement Learning Based Radio Resource Control in 5G Vehicular Network. IEEE Wireless Communications Letters,9(5): 611-614. https://doi.org/10.1109/lwc.2019.2962409

[10] Wang, J., Jiang, C., Zhang, K., Hou, X., Ren, Y., and Qian, Y. (2020). Distributed Q-Learning Aided Heterogeneous Network Association for Energy-Efficient IIoT. IEEE Transactions on Industrial Informatics, 16(4): 2756-2764. https://doi.org/10.1109/tii.2019.2954334.

[11] Ding, H., Zhao, F., Tian, J., Li, D., and Zhang, H. (2020). A deep reinforcement learning for user association and power control in heterogeneous networks. Ad Hoc Networks, 102:1-9. https://doi.org/10.1016/j.adhoc.2019.102069 .

[12] Shi, D., Tian, F., and Wu, S. (2020). Energy Efficiency Optimization in Heterogeneous Networks Based on Deep Reinforcement Learning. 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Ireland, pp. 1-6, https://doi.org/10.1109/iccworkshops49005.2020.9145404.

[13] Liu, H., Lin, Z., Chen, Y., and Xin, P. (2020). Elite User Clustering-Based Indoor Heterogeneous VLC Interference Management and Sub-Channel Allocation Strategy. IEEE Access, 8: 43582-43591. https://doi.org/10.1109/access.2020.2978135.

[14] Aihara, N., Adachi, K., Takyu, O., Ohta, M., and Fujii, T. (2020). Generalized Interference Detection Scheme in Heterogeneous Low Power Wide Area Networks. IEEE Sensors Letters, 4(6): 1-4. https://doi.org/10.1109/lsens.2020.2992723.

[15] Liu, H., Pu, X., Chen, Y., Yang, J., and Chen, J. (2020). User-Centric Access Scheme Based on Interference Management for Indoor VLC-WIFI Heterogeneous Networks. IEEE Photonics Journal, 12(4): 1-12. https://doi.org/10.1109/jphot.2020.3002246.

[16] Gu, Z., Shen, T., Wang Y., and Lau, F. C. M. (2020). Efficient Rendezvous for Heterogeneous Interference in Cognitive Radio Networks. IEEE Transactions on Wireless Communications, 19(1): 91-105. https://doi.org/10.1109/twc.2019.2942296

[17] Ono, K., Akimoto, K., Kameda, S., and Suematsu, N. (2020). Dual-CTS: Novel High-Efficiency Spatial Reuse Method in Heterogeneous Wireless IoTSystems.31stIEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, United Kingdom, pp. 1-6, https://doi.org/10.1109/pimrc48278.2020.9217244.

[18] Yang, R., Zhang, Zhang, W., Deng, L., and Yang, H. (2020). Resource Allocation for Hybrid Visible Light Communications (VLC)-WiFi Networks. IEEE Access, 8: 176588-176597. https://doi.org/10.1109/access.2020.3026388.

[19] Lin, K., Li, C., Rodrigues, J. J. P. C., Pace, P., and Fortino, G. (2020). Data-Driven Joint Resource Allocation in Large-scale Heterogeneous Wireless Networks. IEEE Network, 34(3): 163-169. https://doi.org/10.1109/mnet.001.1900291.

[20] Haroon, M. S., Muhammad, F., Abbas, Z. H., Abbas, G., Ahmed, N., and Kim, S. (2020). Proactive Uplink Interference Management for Nonuniform Heterogeneous Cellular Networks. IEEE Access, 8: 55501-55512. https://doi.org/10.1109/access.2020.2981631.

[21] Ahmed, S., Benaya A. M., and Elsabrouty, M. (2020). Dynamic Quantization based IA for Homogeneous and three-tier Heterogeneous Cellular Systems. International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), Aswan, Egypt, pp. 272-277. https://doi.org/10.1109/itce48509.2020.9047751.

[22] Ding, H., Zhao, F., Tian, J., and Zhang, H. (2020). Performance Analysis of MISINR User Association in 3-D Heterogeneous Cellular Networks. IEEE Transactions on Vehicular Technology, 69(4): 4119-4129. https://doi.org/10.1109/tvt.2020.2976120.

[23] Saha, R. K. (2020). Modeling Interference to Reuse Millimeter-wave Spectrum to In-Building Small Cells Toward 6G. IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), Canada, pp. 1-7, https://doi.org/10.1109/vtc2020-fall49728.2020.9348747.

[24] Younes, B., Mohammed, F., Said, M., Bekkali, M. E. (2021). 5G uplink interference simulations, analysis and solutions: The case of pico cells dense deployment. International Journal of Electrical and Computer Engineering (IJECE), 11(3): 2245-2255, https://doi.org/10.11591/ijece.v11i3.pp2245-2255

[25] Li Z., Shin K. G., and Zhen L. (2017). When and how much to neutralize interference? IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, Atlanta, pp. 1-9, https://doi.org/10.1109/infocom.2017.8057211.

[26] Fujisawa, K., Kemmochi, F. and Otsuka, H. (2019). Personal picoCell Scheme Using Adaptive Control CRE for Multicarrier HetNets. IEEE 90th Vehicular Technology Conference (VTC2019-Fall), USA, pp. 1-5, https://doi.org/10.1109/vtcfall.2019.8891193.

[27] Al-Ani1, M., and Al-Sawalmeh, W. (2009). Simulation and Proposed Handover Alert Algorithm for Mobile Communication Networks. International Journal of Interactive Mobile Technologies, 3: 6-11. https://doi.org/10.3991/ijim.v3s2.838

[28] Liu, C.-J., Huang, P., Xiao, L., and Esfahanian, A.-H. (2020). Inter-femtocell Interference Identification and Resource Management. IEEE Transactions on Mobile Computing, 19(1): 116-129. https://doi.org/10.1109/tmc.2019.2892138

[29] Li, Z., Guo, F., Shu, C., Shin, K. G., and Liu, J. (2018). Dynamic Interference Steering in Heterogeneous Cellular Networks. IEEE Access, 6, 28552–28562. https://doi.org/10.1109/access.2018.2836221

[30] Marzi, Z. and Madhow, U. (2019). Interference Management and Capacity Analysis for mm-Wave Picocells in Urban Canyons. in IEEE Journal on Selected Areas in Communications, 37(12): 2715-2726. https://doi.org/10.1109/jsac.2019.2947819.

[31] Masson, M., Altman, Z., and Altman, E. (2020). Multi-User collaborative scheduling in 5G massive MIMO heterogeneous networks. IFIP Networking 2020 Conference, France, pp. 1-5, https://hal.inria.fr/hal-02566253.

[32] Li, G., and Su, Y. (2017). Intelligent Building Control System Based on Mobile Wireless Internet of Things. International Journal of Interactive Mobile Technologies, 5(1): 63-72. https://doi.org/10.3991/ijoe.v13i10.7746

[33] Amiri, R., Almasi, M. A., Andrews, J. G., and Mehrpouyan, H. (2019). Reinforcement Learning for Self-Organization and Power Control of Two-Tier Heterogeneous Networks. IEEE Transactions on Wireless Communications, 18(8): 3933-3947. https://doi.org/10.1109/twc.2019.2919611.

[34] Su, Q., Li, B., Wang, C., Qin, C., and Wang, W. (2020). A Power Allocation Scheme Based on Deep Reinforcement Learning in HetNets. 2020 International Conference on Computing. Networking and Communications (ICNC), USA, pp. 245-250, https://doi.org/10.1109/icnc47757.2020.9049771.

[35] Xiao, L., Zhang, H., Xiao, Y., Wan, X., Liu, S., Wang, L.-C., and Poor, H. V. (2020). Reinforcement Learning Based Downlink Interference Control for Ultra-Dense Small Cells. IEEE Transactions on Wireless Communications, 19(1): 423-434, https://doi.org/10.1109/twc.2019.2945951.

[36] Alcaraz, J. J., Ayala-Romero, J. A., Vales-Alonso, J., and Losilla-Lopez, F. (2020). Online reinforcement learning for adaptive interference coordination. Transactions on Emerging Telecommunications Technologies. https://doi.org/10.1002/ett.4087

7 Authors

Akindele Segun Afolabi is a lecturer at the Department of Electrical and Electronics Engineering, University of Ilorin, Ilorin, Nigeria.

Shehu Ahmed is an electrical engineer and works with the Nigerian Television Authority, Ilorin, Nigeria.

Olubunmi Adewale Akinola is a lecturer at the Department of Electrical and Electronic Engineering, Federal University of Agriculture, Abeokuta, Nigeria.

Article submitted 2021-01-18. Resubmitted 2021-04-08. Final acceptance 2021-04-09. Final version published as submitted by the authors.