Received: 14 June 2022 | Revised: 8 September 2022 | Accepted: 15 September 2022

IET Communications

DOI: 10.1049/cmu2.12501

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Deep learning integrated reinforcement learning for adaptive beamforming in B5G networks

Geoffrey Eappen[1] | John Cosmas[1] | Shankar T[2] | Rajesh A[3] | Rajagopal Nilavalan[1] | Joji Thomas[4]

[1]Electronic and Computer Engineering, College of Engineering, Design and Physical Sciences, Brunel University, London, UK

[2]SENSE, Vellore Institute of Technology, Vellore, Tamil Nadu, India

[3]School of Electrical and Electronics Engineering, SASTRA University, Thanjavur, Tamil Nadu, India

[4]Department of Mechanical Engineering, Saveetha Engineering College, Tamil Nadu, India

Correspondence
John Cosmas, Electronic and Computer Engineering, College of Engineering, Design and Physical Sciences, Brunel University, London, UK.
Email: john.cosmas@brunel.ac.uk

Abstract

In this paper, a deep learning integrated reinforcement learning (DLIRL) algorithm is proposed for comprehending intelligent beamsteering in Beyond Fifth Generation (B5G) networks. The smart base station in B5G networks aims to steer the beam towards appropriate user equipment based on the acquaintance of isotropic transmissions. The foremost methodology is to optimize beam direction through reinforcement learning that delivers significant improvement in signal to noise ratio (SNR). This includes alternate path finding during path obstruction and steering the beam appropriately between the smart base station and user equipment. The DLIRL is realized through supervised learning with deep neural networks and deep Q-learning schemes. The proposed algorithm comprises of an online learning phase for training the weights and a working phase for carrying out the prediction. Results confirm that the performance of the B5G system is improved considerably as compared to its counterparts with a spectral efficiency of 11 bps/Hz at SNR = 10 dB for a bit error rate performance of 10^{-5}. As compared to reinforced learning and deep neural network with a deviation of $\pm 3^{o}$ and $\pm 5^{\circ}$, respectively, the DLIRL beamforming displays a deviation of $\pm 2^{o}$. Moreover, the DLIRL can track the user equipment and steer the beam in its direction with an accuracy of 92%.

1 | INTRODUCTION

The increased usage of wireless communication-based applications and the huge demand for data rates have resulted in the development of B5G networks. These future networks are required to meet the demands of high data rates and enhanced user experience. In this paper, Artificial Intelligence (AI) is incorporated with beamforming and millimetre Wave (mmWave) enabling intelligent beamsteering based on Channel State Information (CSI), thus enabling a high data rate and better user experience. The novel deep learning integrated reinforcement learning (DLIRL) algorithm is proposed for the beamforming solution to overcome the problems associated with mmWave like blockage impacting the coverage, reliability of highly mobile links, latency overheads associated with high-speed mobile devices in dense mmWave scenarios that require frequent hand-offs [1, 2]. From the research gap, it can be visualized that most of the works for beamforming for B5G networks were carried out employing the conventional schemes, but the employed schemes are not effective enough to steer the beam based on the user locations [12]. Also, the works in [20, 21] for beam direction estimation based on vehicle motion were carried out on a single domain considering only the vertical direction. But implementing the B5G networks for the real time scenario, it is important to consider beam direction from 3D point of view [11].

To have an enhanced beamforming for the real time scenario based B5G networks, we are proposing in this paper an AI-based novel DLIRL efficient beamforming scheme. The proposed scheme is trained via pilot symbols received at the coordinated integrated access and backhaul (IAB) nodes with negligible training overheads. These symbols represent the footprints of pilot information communication with the nearby scenarios. Utilizing this information for training the DLIRL to produce the efficient beamforming is the idea behind this work. In this implemented coordinated beamforming system, the UE performs a transmission of an upstream pilot arrangement, which is cooperatively acknowledged by the organizing IAB nodes. The acknowledged information carries valuable statistics regarding the nearby scenarios because of their communication to the nearby scenarios. Through deep learning, the prototype studies enhances the beamforming vector at IABs by means of the statistics so as to improve the signal to interference noise ratio at user equipment.

The DLIRL constructed beamforming acquires the use of principal patterns to envisage the ideal beamforming vectors in IABs. Here, the beamforming vector through DNN exploitation of isotropic communication from UE devices is arranged in a lattice over the exposed region. The estimated impulse response at each of the IAB receiver functions acts as input data to a DNN model and acquires state space for reinforcement learning (RL). The RL steers accurate and adaptive beams through its experienced learning. This projected method affords an inclusive solution for high mobile B5G scenarios with improved network connectivity, reduced latency, and minimized training overheads.

The long short-term memory (LSTM) based hotspot prediction in small cell 5G networks has been discussed in [25]. Here, adaptive beamforming is carried out to adjust the beam towards the small cell for traffic aggregation between the small cell and macro-cell. The requirement to adjust narrow beams in mmWave communication by leveraging machine learning and using concurrent and multiarmed bandit techniques to establish robust links has been explored in [26]. The requirement of deep learning techniques for 5G systems at the physical layer, network layer and applications layer to process information and automate decisions has been detailed in [27]. The requirement of deep learning in channel estimation for cooperative beamforming in massive multiple input multiple output (MIMO) 5G systems has been discussed in [28].

An adaptive hybrid beamforming for 5G MIMO mmWave networks has been detailed in [29]. Here, to reduce the signalling overhead and blocking probability, analogue-digital beamforming approach is used to generate beams based on the traffic demands. The exploration of narrow beamwidths and adaptive steering of signals to reduce interreference and energy consumption has been discussed in [30]. The importance of beamforming for enhanced signal quality, improved network capacity, frequency reuse, mitigation of multipath effects, estimation of angle of arrival and tracking of mobile devices have been explored [31]. The requirement of machine learning algorithms to track communication scenarios and handle big data in 5G MIMO systems has been discussed in [32]. Here, machine learning based classification models have been developed for beam selection.

The deep learning framework to allocate resources for TV multimedia service in 5G scenario has been detailed in [33]. Here, an LSTM-based deep learning model has been developed to model the traffic pattern for resource allocation under Quality of Service (QoS) requirements. A learning-based adjustable beam number training (LABNT) algorithm has been developed in [34] for optimal beam direction and reduced training overhead. The trade-off between beam alignment accuracy and spectral efficiency in beamforming training for non-line-of-sight mmWave systems has been demonstrated. The concept of deep transfer learning to explore the beamforming vector in massive MIMO systems has been explored in [35]. Here, the trade-off between the number of training data and uncertainty of real-time channel has been discussed. Further, the requirement of an effective deep learning model to train 5G system with less overhead and latency has been highlighted.

2 | RESEARCH GAP

From the literature, it can be inferred that effective schemes have been developed for beamforming [1–3]. The proposed algorithm in [1] is an initial work on adaptive beamforming. In the proposed scheme, several sensors are considered to obtain correlation. However, it is not practical to mount too many sensors on the user equipment. In [2], user location-based transmit and receive beamforming is proposed considering only the Line of Sight (LoS) condition, which is not always feasible for a real-time operation in real environments where the LoS path may be blocked due to human activity. Moreover, the proposed work is considered only for the known user locations. In addition, ref. [4] proposed robust beamforming with an assumption that Angle of Arrival (AoA) and CSI is known. Nevertheless, considering known AoA and CSI is complicated and often unrealizable [3].

In [5], hybrid beamforming with gradient iterative algorithm is proposed considering instantaneous CSI. However, the proposed gradient algorithm is prone to get stuck at local optima solutions instead of global optima [10]. Beamforming considering unknown user location and instantaneous CSI is proposed in [6]. The major challenge in [6] is the huge training overhead associated with large array beamforming vectors. To tackle the problem of training overheads, authors in [7] proposed deep learning-based beamforming. Deep learning-based predictive beamforming is also proposed in [8] considering location awareness. Employing recurrent Long Short-Term Memory (LSTM) has inefficient exploration hampering its prediction abilities [9]. The work in [10] proposed a promising RL with Q learning for a joint optimized beamforming scheme. However, the work considered known optimized beamform vectors and user locations, which are unattainable in a real scenario.

The deep learning-based beamforming scheme is an effective way of achieving efficient beamforming [11]. The DL scheme is also a fast way of beamforming for the mmWave channels [12]. However, the existing DL network requires a huge amount of training data to achieve a good performance via beamforming. Moreover, this DL network is based on mere learning and thus it is not easy to comprehend the output [13]. Therefore, in this work, we propose a DL plus RL network for beamforming capable of getting trained with fewer data. The proposed scheme is not merely based on the training of the DL network but also on the experience provided by the RL network in improving the beamforming performance. Recently, RL for multimedia data segregation has been investigated for healthcare sector using fuzzy algorithms to improve the quality of service in fog computing [22]. Alternatively, multimedia data segregation using k-fold random forest has been developed for reducing the latency from Internet of Things (IoT) devices in healthcare environments [23]. Further, Ant Colony Optimization (ACO) algorithm has been utilized to offload data in resource constrained IoT scenarios [24].

2.1 | Motivation

Achieving high capacities is currently being considered with multiuser MIMO and MASSIVE multi-user MIMO schemes. However, these schemes will require a very high number of antennas at the base station to meet capacity requirements. This paper focuses on beam steering toward users, which can enable the use of communication resources effectively while keeping the number of antennas at the base stations at manageable levels. In this paper, investigations to the question whether the efficient beamforming and beamsteering can be performed using DNN and RL is answered. We proved with the simulations that the performance of the DNN for beam forming and beamsteering can be improved with the expert learning of the RL. With the proposed DLIRL it is possible to attain better pointing of beams with the user movement and enhanced beamsteering with spectral efficiency as compared to [11, 12]. Also, the main motivation behind this paper is that to the best of our knowledge till now no work has been done on utilizing machine learning for beamsteering. Therefore, there is the need to incorporate an efficient machine learning scheme for the beamsteering that is capable of learning based on the environments.

Hence the prime contribution of this work involves:

1. A novel DLIRL-based beamsteering scheme is implemented, which is capable to steer the beamform with the user movement.
2. The proposed scheme combines the performance of the DL and RL. The DL is used for preparing the optimized beamform codebook and RL is used for selecting the best beam out of the optimized beamform codebook based on the user movements.
3. The training of the DLIRL is carried out based on the channel information without requiring the user's location. Thus, reducing the training overheads.
4. A novel way of combining the DL with the RL and employing it for beamsteering applications.
5. The proposed scheme is trained in the offline mode for a particular environment simulated using the MATLAB site viewer. The trained model is then employed successfully for the beamsteering. Furthermore, the proposed scheme is sufficiently flexible for getting trained to any provided environment.

The rest of this paper is arranged as follows. In Section 3, the system modelling and the problem formulation are elaborated. We implemented the end-to-end DLIRL based beamforming in Section 4, comprising DNN architecture, its training and working period combined with the IRL. The performance analysis of the DLIRL is investigated in Section 5 with respect to spectral efficiency (SE), bit error rate (BER), and beamform.

Notations: The bold letters represent vectors and matrices for the lower and the upper case respectively. The conjugate transpose is denoted as $()^H$, $\|.\|_F^2$ indicates Frobenius norm square, $\mathbb{C}$is the indication for the complex number and N_c indicates complex normal distribution.

3 | SYSTEM MODELLING AND PROBLEM FORMULATION

3.1 | System modelling

The proposed work is designed for mmWave-based wireless communication networks. The considered system model comprises of I number of IAB nodes termed as IABs serving a UE fitted with the single isotropic antenna as shown in Figure 1. The IAB node is connected to the B5G core via an IAB donor. The B5G core carries the brain for the intelligent and adaptive beamforming in the form of a DLIRL network. Each IABs are equipped with N^T antennas communicating information symbol $s^K \in \mathbb{C}$ for the K^{th} subcarrier here $K = 1\ldots.,\ k$. Each IAB has baseband precoder vector for each K^{th} subcarrier $\mathbf{f}_{\mathbf{BB}}^{K} \in \mathbb{C}^{I\times 1}$, and for RF precoder $\mathbf{f}_{\mathbf{RF}} \in \mathbb{C}^{N^T\times 1}$. The RF precoder for the l^{th} IAB at m^{th} antenna element can be modelled as a phase shifter network which is mathematically represented as $[\mathbf{f}_{\mathbf{RF}}^{l,m}] = 1/\sqrt{N^T}\ e^{j\phi^{l,m}}$, here $\phi^{l,m}$is quantized angle [15]. The downlink transmission for the transmitted data symbol can be represented as $\mathbf{y} = \mathbf{f}_{\mathbf{RF}}^{l}\ \mathbf{f}_{\mathbf{BB}}^{K} s^K$. Here $\mathrm{E}[s^K(s^K)^H] = P^K/k$, P^K is the power associated with the K^{th} subcarrier and k is the total number of subcarriers. The constraint in the total transmission power of IAB should be $\|\mathbf{F}_{\mathbf{RF}}\mathbf{f}_{\mathbf{BB}}^{K}\|_F^2 = 1, \quad K = 1, 2, \ldots, k$.

Here, $\mathbf{F}_{\mathbf{RF}} = \mathrm{blkdiag}(\mathbf{f}_{RF}^{1} \ldots\ldots, \mathbf{f}_{RF}^{I}) \in \mathbb{C}^{IN^T\times I}$. The channel vector between the l^{th} IAB and the UE for the K^{th} subcarrier is denoted as $\mathbf{h}_{l,K} \in \mathbb{C}^{N^T\times 1}$. Then the received data stream $\boldsymbol{x}_{l,K}$ by the UE from the l^{th} IAB during the downlink at K^{th} subcarrier can be written as in Equation (1) [14]:

$$\boldsymbol{x}_{l,K} = \sqrt{P_{avg}}\mathbf{h}_{l,K}^{H}\mathbf{f}_{\mathbf{RF}}^{l,K} f_{BB}^{l,K}\, s^K + \boldsymbol{n}_{K,l}, \qquad (1)$$

where P_{avg} is the average received power at the UE. The $\boldsymbol{n}_{K,l} \sim N_c(0, \sigma^2)$ is the K^{th} subcarrier noise at the UE. The symbol notation involved in the system modelling is represented in the Nomenclature list.

3.2 | Channel modelling

The channel between IABs and the UE is considered as a wideband mmWave channel with $c = 1,\ldots C$ clusters contributing to one ray of time delay as τ_c, AoA as ϕ_c and θ_c for azimuth and elevation angles, respectively. The channel path loss between UE

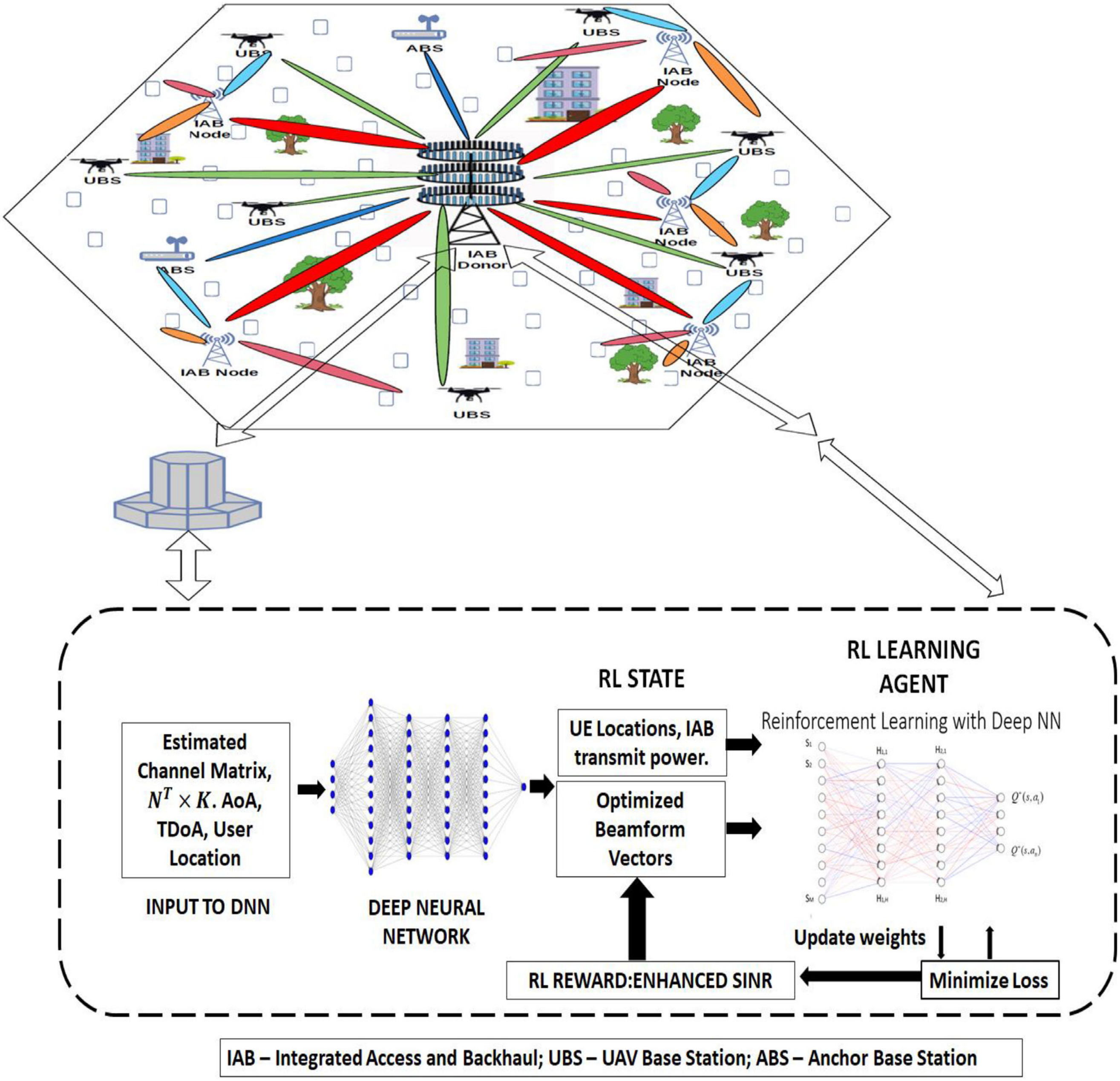

FIGURE 1 B5G system model

and the l^{th} IAB over K^{th} is represented as $L_{l,K}$. The antenna array response vector between l^{th} IAB and the UE for the K^{th} subcarrier is represented as $a_{l,K}(\theta_c, \phi_c)$. Mathematically the channel can be modelled for the K^{th} subcarrier as shown in Equation (2) [14]:

$$h_{l,K} = \sqrt{\frac{N^T}{L_{l,K}}} \sum_{c=1}^{C} \beta_c a_{l,K}(\theta_c, \phi_c), \qquad (2)$$

Here β_c is the complex gain associated with the resolvable path c. The considered channel model is assumed to remain constant over coherence time T_c as it is a block fading channel.

3.3 | Problem formulation

The challenges associated with 5G standards are to reach the goals of higher data rate, lower latency, better coverage, and high mobility. To achieve this, it is important to have the most flexible and controlled beamforming scheme. The existing beamforming techniques [1–6] tried to achieve flexibility and control via dedicated transmit/receive for each element. Considering massive MIMO-based wireless communication systems, building this type of architecture is highly difficult due to extensive cost, power, and space-based limitations [12], thus, hindering the design budget. With this motivation, this work is dedicated to utilising an effective AI training scheme based on DL and RL combinedly termed DLIRL for the beamforming strategy.

To quantify the goals achieved by the proposed scheme, the performance is evaluated by considering the beamforming system that can maximize effective spectral efficiency for mmWave-based wireless communication applications at UE. For that purpose, the achievable rate at UE for the considered hybrid beamformer, system, and channel model are evaluated as per Equation (3):

$$S_e = \frac{1}{k}\sum_{K=1}^{k} \log_2\left(1 + SNR\left|\sum_{l=1}^{I} \mathbf{h}_{l,K}^{H} \mathbf{f}_{\mathbf{RF}}^{l} f_{\mathrm{BB}}^{l,K}\right|^2\right), \tag{3}$$

Here, I is the total IABs considered. SNR represents the signal to noise ratio at UE denoted as $(Pavg/k\sigma^2)$. The objective of this work is to create an effective beamforming design and channel training so as to maximize the achievable rate at UE. The final optimization problem can be deduced as:

$$\begin{aligned} S_e\left(\mathrm{f}_{\mathbf{RF}}^{l}, f_{BB}^{l,K}\right) &= \arg\max \frac{1}{k}\sum_{K=1}^{k} \\ &\times \log_2\left(1 + SNR\left|\sum_{l=1}^{I} \mathrm{h}_{l,K}^{H} \mathrm{f}_{\mathbf{RF}}^{l} f_{BB}^{l,K}\right|^2\right), \\ &\text{s.t. } \left\|\mathrm{f}_{\mathbf{RF}}^{l} f_{BB}^{l,K}\right\|^2 = 1, \quad \forall l, \end{aligned} \tag{4}$$

4 | PROPOSED SCHEME: DLIRL-BASED BEAMFORMING

The proposed scheme comprises deep learning integrated with reinforcement learning for the beamforming design. The employed techniques make use of the learning strategy by performing mapping between beamforming weights and the environmental setup, channel estimates, AoA, and Time Difference of Arrival (TDoA). The proposed DLIRL employs a pilot signal transmitted from UE to the IAB's to learn about the channel condition and predict the optimal beamform weights. The pilot signals received at IABs are the result of the interaction of the signal with the environment during its propagation. These reflected and diffracted waves are jointly received at different IABs and carries the thumbprint of the environmental factors and channel conditions. The Environmental Thumbprint (ET) carried by these pilot signals is employed for training. The DLIRL has two periods: working and training. The DLIRL initiates with the training period. During this, DLIRL receives pilot sequences transmitted from the predefined UEs positions. The UE transmission is omnidirectional and it carries ET. The DLIRL then maps the received sequence with the training process and learns it.

The working period marks the prediction scheme of the DLIRL and performs prediction of the optimal beamform without any need for additional training. Multiple advantages achieved via the proposed scheme are that it does not require additional resources for learning during the working period.

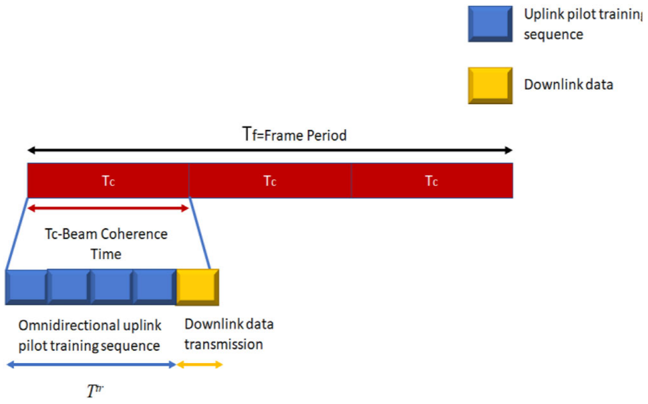
Uplink pilot training sequence
Downlink data
Tf=Frame Period
Tc
Tc
Tc
Tc-Beam Coherence Time
Omnidirectional uplink pilot training sequence
Downlink data transmission
T^{tr}

FIGURE 2 Beam coherence time interval during training process

Moreover, the proposed scheme employs an omnidirectional UE pilot sequence for the training period. Therefore, it has minimal overheads during the training period. Also, the proposed scheme can be trained for any environment including LoS and Non-Line of Sight (NLoS).

4.1 | Training period

In this period DNN observes the environment and trains the deep neural network. We consider a beam coherence time T_c, a concept introduced in [15] for mmWave-based wireless communication systems also shown in Figure 2. It can be defined as the period over which the beams are unchanging. Considering T^{tr} as the channel training period of the first T^{tr} time instants of the T_c, then Equation (4) can be re-written as in Equation (5).

$$\begin{aligned} S_e\left(\mathbf{f}_{\mathbf{RF}}^{l}, f_{\mathrm{BB}}^{l,K}\right) = \left(1 - \frac{T^{tr}}{T_c}\right) \arg\max \frac{1}{k}\sum_{K=1}^{k} \log_2 \\ \times \left(1 + SNR\left|\sum_{l=1}^{I} \mathbf{h}_{l,K}^{H} \mathbf{f}_{\mathbf{RF}}^{l} f_{\mathrm{BB}}^{l,K}\right|^2\right), \end{aligned} \tag{5}$$

Figure 1 depicts the training period for deep neural network-based learning. In each T_c UE transmits pilot matrix $\mathbf{S}_p^K \in \mathbb{C}^{N^T \times I}$ repeatedly, here $K = 1,2,3,\ldots k$. The received pilot sequence at l^{th} IAB is as in Equation (6):

$$\boldsymbol{x}_{l,K}^{p} = \mathbf{h}_{l,K}\mathbf{S}_p^K + \boldsymbol{n}_{K,l}, \tag{6}$$

The combined beamforming strategy starts with feeding the received pilot signals from all IABs to the Fusion Centre (FC) comprising DLIRL. The FC first selects the RF beamform vectors for the IABs downlink as:

$$\mathrm{f}_{\mathbf{RF}}^{l} = \underset{\mathrm{f}_{\mathbf{RF}}^{l} \in \kappa_{\mathbf{RF}}, \forall l}{\arg\max} \sum_{K=1}^{k} \log_2\left(1 + SNR\left|\sum_{l=1}^{I} \mathrm{h}_{l,K}^{H}\right|^2\right), \tag{7}$$

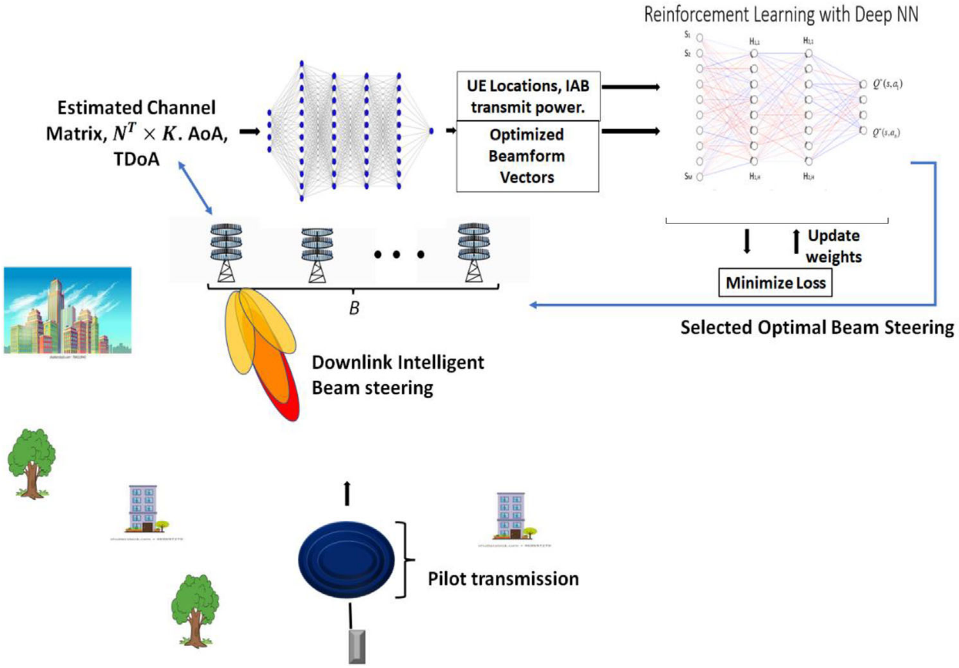
Reinforcement Learning with Deep NN
S1
S2
SM
H1,1
H2,1
H1,H
H2,H
Q*(s,a1)
Q*(s,an)
Update weights
Minimize Loss
Selected Optimal Beam Steering
UE Locations, IAB transmit power.
Optimized Beamform Vectors
Estimated Channel Matrix, $N^T \times K$. AoA, TDoA
B
Downlink Intelligent Beam steering
Pilot transmission

FIGURE 3 The DLIRL beamforming system model

Here, $\boldsymbol{\varkappa}_{\mathbf{RF}}$ is the RF beamform codebook. The FC then applies baseband precoder calculated as [14]:

$$\mathbf{f}^{*}_{BB} = \frac{\left(\boldsymbol{\hbar}^{H}_{K}\mathbf{F}_{\mathbf{RF}}\right)^{H}}{\left\|\boldsymbol{\hbar}^{H}_{K}\mathbf{F}_{\mathbf{RF}}\right\|}, \quad \forall K \tag{8}$$

The DNN is fed with the pilot symbols $\boldsymbol{x}^{p}_{l,K}$ as the input for the training model. The spectral efficiency for every RF and baseband beamform vector acts as the output shown in Equation (5). The DNN is trained during this period and learns the correlation between received pilot symbols at each IABs and the ET. Post-training the DNN moves to the working model where it performs the prediction. The detailed DNN architecture is explained in Section 4.2.

4.2 | DNN architecture

We propose a multilayer DNN structure for training the proposed model and getting adapted to the scenario. The inputs to the model are the channel impulse response (CIR), AoA, and TDoA. The CIR, AoA, and TDoA are obtained via received pilot symbols at I IABs. The DNN structure comprises four hidden layers. The first layer has 12 nodes and the remaining 3 layers have 10 nodes each. The DNN is trained end to end as a supervised learning class [16]. The DNN input is normalized based on the training dataset. The DNN architecture employed Rectifier Linear Unit (ReLU) as the activation function [17, 18].

4.3 | Working period

Based on the estimated channel, the RF and the Baseband precoding beamform codebook $\boldsymbol{\varkappa}_{\mathbf{RF}}, \boldsymbol{\varkappa}_{\mathbf{BB}}$ is respectively formed using Equations (7) and (8). The objective of the DNN is to maximize the $S_e(\mathbf{f}^{l}_{\mathbf{RF}}, \mathbf{f}^{l}_{\mathrm{BB}})$ for each IABs. The regression-based learning model is adapted such that for each l^{th} IAB, the error difference between the DNN's predicted output $S_e^{pred,p,l}$ and the desired $S_e^{des,p,l}$ output is minimum. Here, $p = 1,2,\ldots\mathrm{T}_{\mathrm{rain}}$, the $\mathrm{T}_{\mathrm{rain}}$ is the number of RF and the baseband beamforming vectors. Mathematically DNN is trained to minimize the error function $e(w)$ for the set of different weight values of the DNN. The $e(w)$ for the l^{th} IAB can be written as in Equation (9):

$$e_l(\theta) = \sum_{p=1}^{\mathrm{T}_{\mathrm{rain}}} M_{se}\left(S_e^{des,p,l}, S_e^{pred,p,l}\right), \tag{9}$$

Here $M_{se}(S_e^{des,p,l}, S_e^{pred,p,l})$ is the mean square error (MSE) between the predicted output $S_e^{pred,p,l}$ and the desired $S_e^{des,p,l}$ output. The optimal RF and the baseband precoding vectors get updated in the beamform codebook $\boldsymbol{\varkappa}_{\mathbf{RF}}, \boldsymbol{\varkappa}_{\mathbf{BB}}$ respectively. From this stage onwards, the work of the Integrated RL (IRL) begins. The IRL employs the deep Q network for fine-tuning the beamform. The IRL goes through the optimized RF and baseband beamform codebook to further fine-tune the beamform vectors and steer the beam more precisely to the user location. The deep RL network employed in this work is shown in Figure 3.

The objective of the IRL network is to maximize the received UE's SNR for the given beamform vectors. The received power at UE for the time instant t from the l^{th} IAB can be denoted as:

$$Pow^{l}_{\text{UE}}(t) = Pow^{l}_{\text{IAB}}(t)\left|\mathbf{h}^{H}_{l,K}(t)\mathbf{f}^{l}_{\mathbf{RF}}(t)\mathbf{f}^{l}_{\mathbf{BB}}(t)\right|^{2}, \tag{10}$$

Here $Pow^{l}_{\text{IAB}}(t)$ is the transmitted power associated with the l^{th} IAB at time instant t. Now the received UE's SNR for the signal from the l^{th} IAB can be written as:

$$SNR^{l}(t) = \frac{Pow^{l}_{\text{IAB}}(t)\left|\mathbf{h}^{H}_{l,K}(t)\mathbf{f}^{l}_{\mathbf{RF}}(t)\mathbf{f}^{l}_{\mathbf{BB}}(t)\right|^{2}}{\sigma^{2}_{n}}, \tag{11}$$

The maximization problem for the IRL is as shown in Equation (12) such that the transmission power of each IAB should be within the total permissible power of each IAB i.e. $Total_{power}$.

$$\begin{aligned}
&\underset{\substack{Pow^{l}_{\text{IAB}}(t)\\ \mathbf{f}^{l}_{\mathbf{RF}}(t)\mathbf{f}^{l}_{\mathbf{BB}}(t)}}{\text{maximize}} \sum_{l=1,2,3..N} SNR^{l}(t),\\
&\text{s.t. } Pow^{l}_{\text{IAB}}(t) \in Total_{power},\\
&\mathbf{f}^{l}_{\mathbf{RF}}(t)\mathbf{f}^{l}_{\mathbf{BB}}(t) \in \kappa_{\mathbf{RF}}, \kappa_{\mathbf{BB}},
\end{aligned} \tag{12}$$

In this paper, an effort is made to fine-tune beam steering via DRL fed with the DNN optimized beamform vectors, UE location, and IAB transmit power to jointly control them to maximize Equation (12). The IRL is an efficient deep learning Q network. The state-space τ associated with the IRL are $state^{l}_{1}$ indicating the transmission power of the l^{th} IAB, $state^{l}_{2}$, which is the beamform vectors associated with the l^{th} IAB. The $state^{l}_{3}$ indicates the deployed UE locations employed in the training. Action space A involves regulation in the transmission power and beamform vectors from the codebook $\kappa_{\mathbf{RF}}, \kappa_{\mathbf{BB}}$ of the respective IABs.

4.4 | Deep IRL model

The deep IRL involves training of the Q Neural Network (NN) as shown in Figure 3. For the policy po the value of the state s and action a is given as $Q_{po}(s, a)$. For converging to the optimal state-action value $Q^{*}_{po}(s, a)$ we employed the NN architecture. The NN is defined by its weight values as $\boldsymbol{we}_{t}$ for each time step t. The vec($\boldsymbol{we}_{t}$) is represented as $\mathbf{we}_{t}$. The state action value in terms of NN weights can be represented as $Q_{po}(s, a : \mathbf{we}_{t})$ as shown in Algorithm 1.

The NN architecture of the deep IRL has the activator in the form of sigmoid function as $y \rightarrow \frac{1}{1+\exp^{-y}}$ [10]. Here the objective function for the NN of the IRL is to minimize the mean square error (MSE) represented as:

$$\min_{\mathbf{we}_{t}} Err(\mathbf{we}_{t}) = \epsilon_{s,a}[(o_{t} - Q_{po}(s, a : \mathbf{we}_{t}))^{2}], \tag{13}$$

ALGORITHM 1 Deep learning integrated reinforcement learning

```
Parameter Initialization: T, j, s, a
Input: UEs' SNR
Output: Beamform steering weights
For ($j = 1; j < T; j$++)
Current $s$ observation
Choose exploitation (exi) or exploration (exo) based on $s$
if exo
Select an a randomly from the set of $a_s$
else
Obtain $a = argmaxQ_{po}(s, a : \mathbf{we}_t)$
end
Compute Reward
Obtain $SINR$
if obtained $SINR$ < threshold
Abort
end
Next $s$ is observed
Estimate
$o_t = \epsilon[reward_{s,s',a} + discount * \max_{a'} Q_{po}(s', a' : \mathbf{we}_{t-1})|s_t, a_t]$
Perform DL training
Estimate and update weights of the DLIRL
Estimate the Mean Square Error:
$\min_{\mathbf{we}_t} Err(\mathbf{we}_t) = \epsilon_{s,a}[(o_t - Q_{po}(s, a : \mathbf{we}_t))^2]$
Calculate SINR
Estimate reward based on the SINR
end
```

Here,

$$o_{t} = \epsilon[reward_{s,s',a} + discount * \max_{a'} Q_{po}(s', a' : \mathbf{we}_{t-1})|s_{t}, a_{t}], \tag{14}$$

The $reward_{s,s',a}$ is the reward for the agent post taking the action a and moving from state s at time t to state s'. The a' represents the next action to be taken. The weights of the NN are updated based on the gradient descent as:

$$\mathbf{we}_{t+1} = \mathbf{we}_{t} - step\nabla \text{Err}(\mathbf{we}_{t}), \tag{15}$$

Here, $step$ indicates the step size has a value between 0 and 1. In the proposed IRL, the value $Q^{*}_{po}(s, a)$ is estimated based on the approximation $Q_{po}(s, a : \mathbf{we}_{t})$ that minimizes the MSE $Err(\mathbf{we}_{t})$. The proposed scheme is implemented corresponding to the downlink scenario. Here, IABs are separated at an estimated distance, and user equipment is positioned at a specific geographical location within the coverage of IABs. Also, these UEs move at a particular velocity. In this work, it is fixed at 20 mph. The reward function for the IRL is estimated based on

TABLE 1 Simulation parameters

S. no.	Parameter	Specification
1	IABs/IAB count (N)	4
2	IAB antenna array	Uniform planar array
3	IAB Antenna Specification	32 × 8
4	User equipment (UE) setup	Deployed in a rectangular grid of dimension 40 m × 60 m, resolution 0.1 m
5	DNN activation unit	ReLU (Rectified Linear Unit)
6	DNN dropout rate	0.5%
7	DNN batch size	100
8	Python Libraries	Keras with Tensorflow backend
9	System bandwidth	0.5 GHz
10	OFDM subcarriers	1024
1112	Sampling FactorMultipaths	17

the performance of an action on meeting the threshold $SINR$. The maximum reward at a unit time step is assigned to the agent having the best performance.

5 | RESULTS AND DISCUSSION

The beamforming is implemented for the downlink scenario for the data transmission from IAB to UE. The simulation environment is first configured and then the proposed DLIRL beamforming is executed. The simulation parameters employed in this work are as shown in Table 1. The data set of the channel model and the channel parameters for the simulation is generated via MATAB 2021 site viewer-based ray tracing. The Shooting and Bouncing Ray (SBR) based ray-tracing model is employed for the LoS and NLoS communication. The orthogonal frequency division multiplexing (OFDM) system is employed for symbol transmission. The considered OFDM size is 1024.

The DNN architecture has a total of 6 interconnected layers including 4 hidden layers and 1 input and an output layer. The DNN has a total $l{\times}k$ number of inputs and T_{rain} number of outputs. The considered data has a set size of two hundred thousand samples and a batch size of two hundred. To have a comparative analysis of the proposed algorithm with the existing conventional beamforming techniques, we have used SE and the BER as the metrics. Figures 4 and 5 show the SE for different SNR values received at UE. The simulations were carried out for 30 runs comprising 1000 iterations each. The depicted graph values are averaged values obtained in the simulation environment. For the simulation environment, the IABs are installed on the buildings played in the x–y plane of the 3D environment. The IAB's antenna is facing the street on the y-z plane. The antenna transmit power is considered at 30 dBm. The UEs are mobile and are installed with a single antenna. For each beam coherence time, the UE locations are updated in the x–y plane.

TABLE 2 SE performance evaluation

Beamforming Technique	SE (bps/Hz) at SNR = 10 dB
DLIRL beamforming	11
Analog beamforming	5
MSE digital beamforming	8
Kalman-hybrid precoding	5.7
Minimum mean squared error (MMSE) hybrid beamforming	5.1
Zero forcing hybrid beamforming	4.2

During the training period, the UE uplink transmit power is set at 30 dB m.

From Figure 4, it is visualized that the DLIRL beamforming has achieved better spectral efficiency as compared to the existing conventional beamformers in [19]. As seen from the curves, the spectral efficiency with analogue beamforming is found to be around 2 bps/Hz, and close convergence is observed between ZF hybrid precoding, MMSE hybrid precoding, and Kalman hybrid precoding techniques. However, the MSE-based fully digital precoding displays improved spectral efficiency as compared to the above-mentioned precoding techniques. For an SNR of 5 dB, the DLIRL based beamforming technique displays an improvement of 77.5%, 60%, 50%, 50%, and 33.3% as compared to the analogue beamforming, ZF hybrid precoding, MMSE hybrid precoding, Kalman hybrid precoding, and MSE-based fully digital precoding techniques, respectively. The spectral efficiency is achieved for the multipath scenario considering both LoS and NLoS, total multipath considered for evaluation of Figure 4 is 10 and total IAB antenna elements are 256.

Moreover, as the antenna size increases the performance of DLIRL gets better as compared to the DNN and RL beamformer. Figure 5 shows the comparison between DNN, RL, and DLIRL-based beamformers for different transmitting antenna elements. For instance, for the number of IAB antenna elements equal to 10^4, the increase in spectral efficiency employing DLIRL-based beamforming is found to be 53.33% and 51.66% more efficient as compared to DNN- and RL-based beamforming techniques, respectively. The effect of BER for IAB with 4 transmit antenna elements has been displayed in Figure 6. Here, the performance of the system for different MIMO schemes is compared with DLIRL-based beamforming scheme. For a BER of 10^{-4}, the proposed DLIRL-based beamforming techniques require an Eb/N$_0$ of 7 dB. Alternatively, the system without diversity, Alamouti, and OSTBC schemes require an Eb/N$_0$ of 10 dB, 13.3 dB, and beyond 20 dB, respectively.

The quantitative analysis of the proposed scheme concerning SE and BER is presented in Tables 2 and 3 respectively. There is a drastic improvement in the SE and BER using DLIRL beamforming studied at SNR = 10 dB shown in both the tables.

To answer the question of whether the proposed scheme can learn the beamforming, we have simulated Figures 7 and 8. The simulation graphs in Figures 7 and 8 are also the report of the

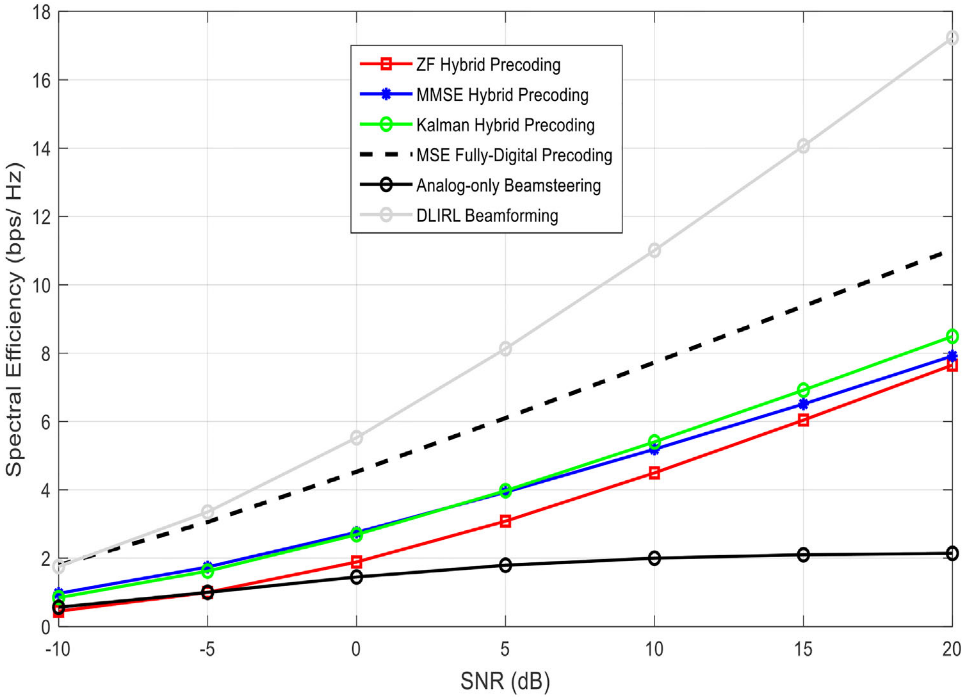

FIGURE 4 Comparative analysis of SE with reference to IAB SNR

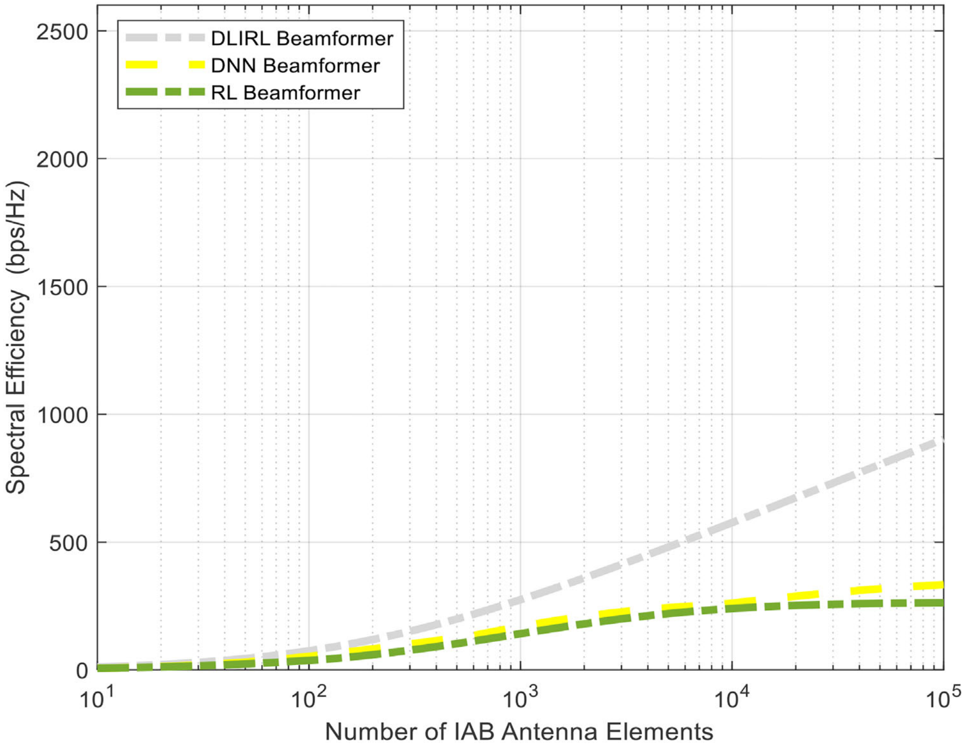

FIGURE 5 Comparative analysis of SE with reference to increase in number of antennas

average value for the 30 runs with 1000 iterations each. The proposed scheme can project the beam towards the UE positioned at 100 degrees in the northwest direction of the antenna placement. From Figures 7 and 8, it is estimated that the DLIRL based beamformer is better than its counterparts DNN and RL in steering the beam towards the UE placed at 101.5° normal to the antenna placement of IAB. The proposed DLIRL beamforming has Angle of Departure (AoD) towards UE location with a deviation of ±2°, whereas RL has a deviation of ±3° and DNN's deviation is ±5°.

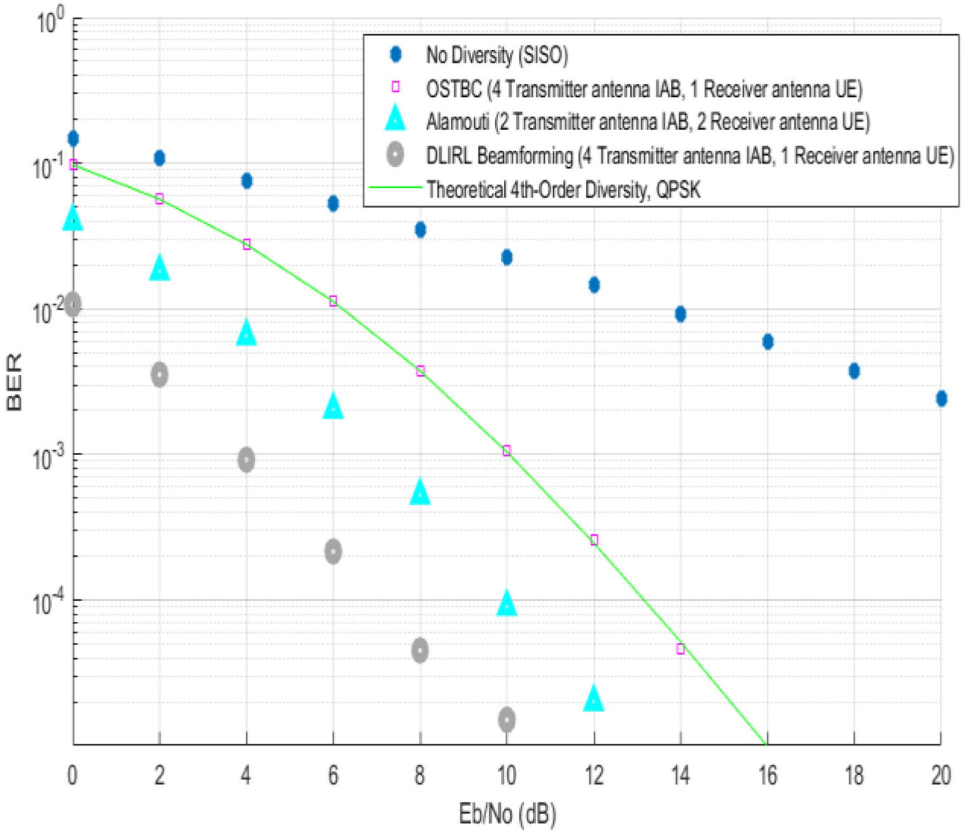

FIGURE 6 Comparative analysis using BER

TABLE 3 BER performance evaluation

Diversity scheme	BER at SNR = 10 dB
No diversity, Single input single output (SISO)	$10^{-1.8}$
Orthogonal space time block coding (OSTBC), 1 × 4 MIMO transmit diversity	10^{-3}
Alamouti, 2 × 2 MIMO diversity	10^{-4}
DLIRL beamforming, 1 × 4 Transmit diversity	10^{-5}

TABLE 4 Performance evaluation of the training algorithm

Epoch	Iteration	Validation accuracy (RL)	Validation accuracy (DNN)	Validation accuracy (DLIRL)
1	1	28.12%	28.12%	46.88%
2	10	34.38%	28.12%	59.38%
3	20	40.62%	34.38%	65.62%
4	30	46.88%	37.50%	75.00%
5	40	46.88%	37.50%	78.12%
7	50	46.88%	40.62%	71.88%
8	60	59.38%	43.75%	81.25%
9	70	59.38%	43.75%	84.38%
10	80	56.25%	43.75%	96.88%
12	90	56.25%	43.75%	96.88%
13	100	56.25%	46.88%	84.38%
14	110	56.25%	46.88%	96.88%
15	120	56.25%	46.88%	96.88%
17	130	56.25%	50.00%	96.88%
18	140	59.38%	53.12%	93.75%
19	150	59.38%	50.00%	84.38%
20	160	59.38%	53.12%	100.00%

The DLIRL is capable of performing efficient beamforming due to the effective training. It is vital to have comparative analysis of the DLIRL with existing DNN and RL algorithm in terms of training validation accuracy, training loss, number of iterations and epochs. Figure 9 sheds light on the validation accuracy of the proposed (DLIRL) and existing (DNN and RL) training algorithms. For the training we employed 20 epochs, 160 iterations, and 100 runs. Each run comprised 20 epochs and each epoch have 8 iterations. From the validation accuracy as shown in Table 4 and Figure 9 it can be inferred that the proposed DLIRL due to its optimized amalgamation of DNN and RL has better training accuracy as compared to the DNN and RL. These training accuracy results are clearly reflected in the beamforming effectiveness as shown in Figures 4–8.

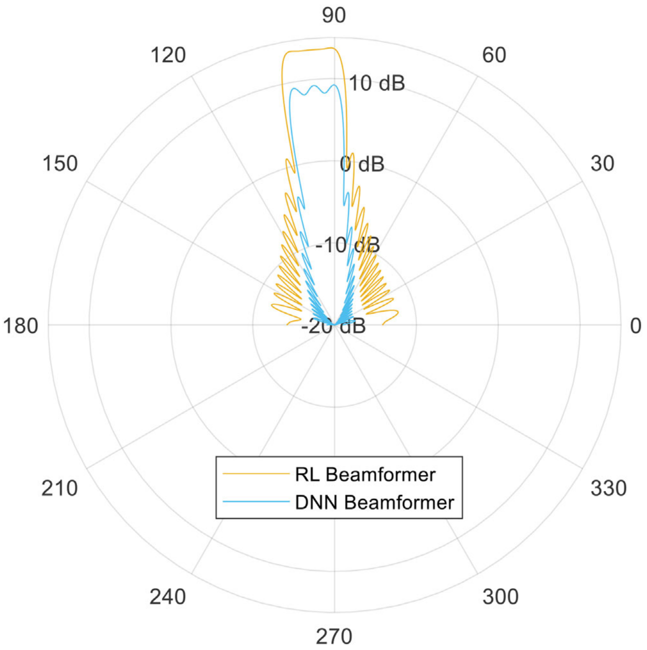

FIGURE 7 Beamform towards UE using DNN and RL separately

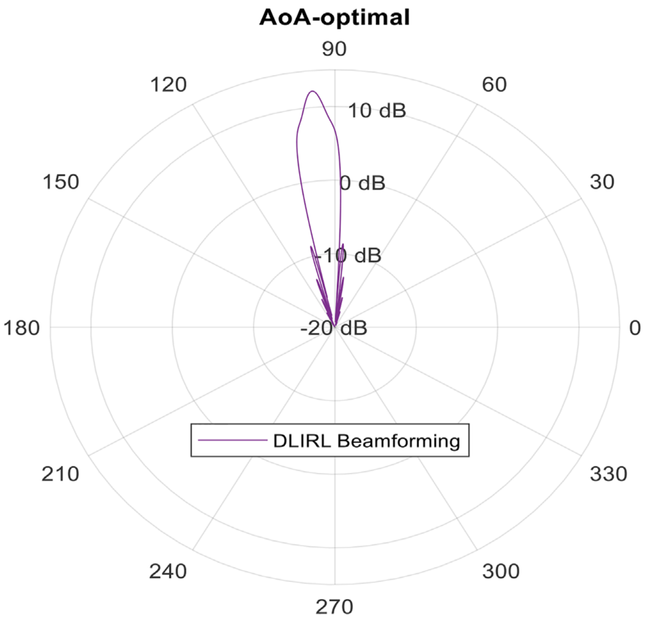

FIGURE 8 Beamform towards UE using DLIRL

The proposed DLIRL is effective in getting trained in less number of samples as shown in Figure 9. The SE of DLIRL is comparable to DNN and RL for very few samples (100) as visualized in Figure 11. Above the 100 training samples the performance of the DLIRL is better than its counterparts. Another important discussion associated with this paper is that the proposed beamforming scheme is adaptive beamforming based on the UE's movements. To test the efficacy of the DLIRL beamforming, simulation environment considered MATLAB 2021 employing a site viewer. For raytracing, Hongkong Open Street

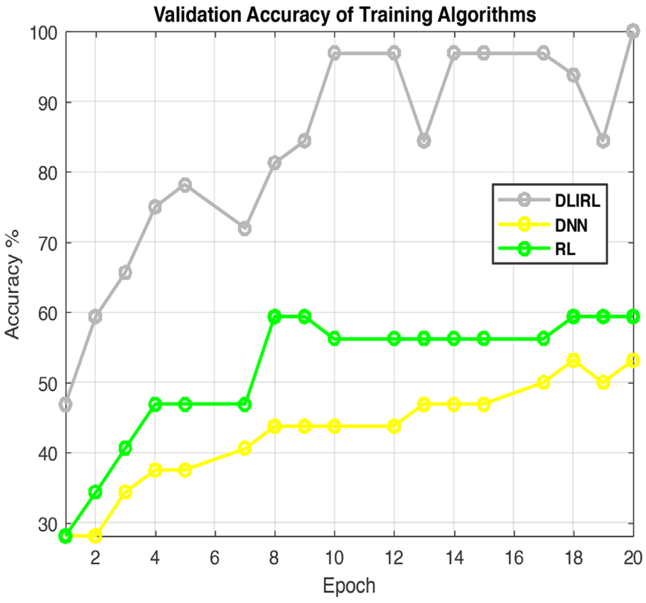

FIGURE 9 Training validation accuracy of DLIRL, DNN and RL

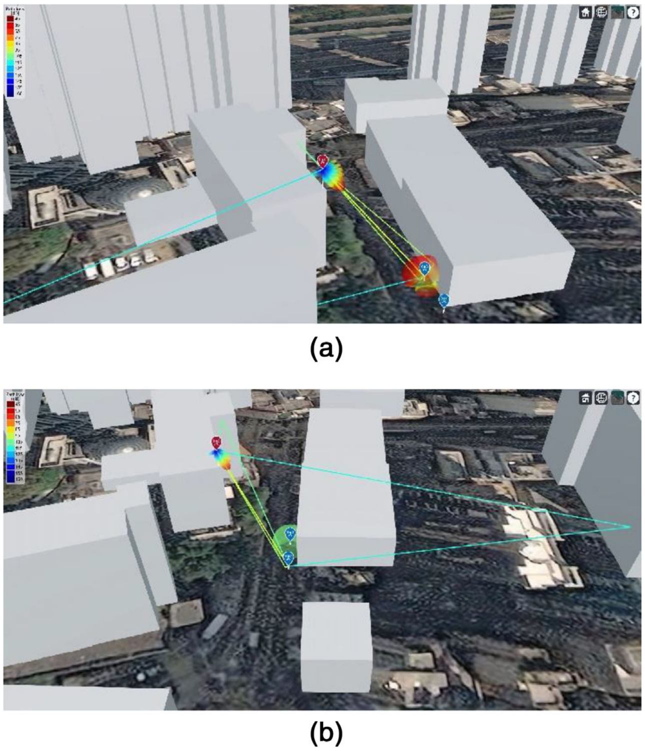

FIGURE 10 (a) Demonstration of DLIRL based adaptive beamforming using MATLAB raytracing and siteviewer. (b) DLIRL based beamsteering with user movement

Map (OSM) with the 3D building environment is employed as shown in Figure 10a, b.

The latitude and longitude associated with an IAB are 22.287495, 114.140706. The initial UE location is 22.287323, 114.140859 latitudes, and longitude respectively. The UE is moving at a constant speed of 28 kilometres/hour (kph) and

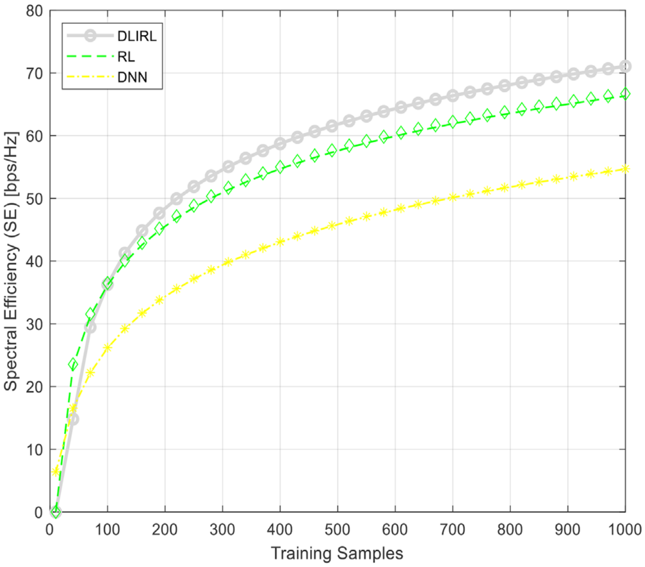

FIGURE 11 SE versus training samples for DLIRL, RL and DNN

its position changes 22.2874, 114.140859 latitude, and longitude respectively as seen in Figure 9. From the figure, the proposed method tracks the user equipment and steers the beam in its direction with an accuracy of 92%. Accuracy is measured based on the performance of DNN and RL in minimizing the error deviation achieved at the UE while moving to a new position from the initial position.

6 | CONCLUSION

A novel integration of deep neural network and reinforced learning (DLIRL) based scheme for intelligent beamforming in massive MIMO wireless communication has been investigated in this paper. The DLIRL algorithm learns from its surrounding environment and trains the network as per the requirement of B5G system. Such intelligent focusing of beams enhances the spectral efficiency between downlink integrated access and backhaul and user equipment data transmission and reception. As the DLIRL exhibits a narrow deviation as compared to reinforced learning and deep neural network-based beamforming techniques, the DLIRL may track the user equipment at very high accuracy in B5G networks. Hence, the spectral efficiency with DLIRL-based beamforming is found to be 53.33% and 51.66% more as compared to deep neural network and reinforced learning-based beamforming techniques, respectively.

NOMENCLATURE

System model symbol notations

Notations	Description
I	Number of IAB nodes
N^T	Number transmit antenna elements at an IAB
s^K	Information symbol for K^{th} subcarrier
$\mathbf{f}_{\mathbf{BB}}^K$	Baseband precoder matrix for each K^{th} subcarrier
$\mathbf{f}_{\mathbf{RF}}$	RF precoder vector
P_{avg}	Average received power
$\boldsymbol{x}_{l,K}$	The received data stream by the UE from the l^{th} IAB during the downlink at K^{th}
$\mathbf{h}_{l,K}$	Channel vector between the l^{th} IAB and the UE for the K^{th} subcarrier

AUTHOR CONTRIBUTIONS

G.E.: Conceptualization; Data curation; Formal analysis; Methodology; Software; Writing – original draft. J.C.: Supervision; Validation; Writing – review & editing. J.T.: Formal analysis; Resources

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

No data.

ORCID

Geoffrey Eappen https://orcid.org/0000-0002-4065-3626
Rajagopal Nilavalan https://orcid.org/0000-0001-8168-2039

REFERENCES

1. Cox, H., Zeskind, R., Owen, M.: Robust adaptive beamforming. IEEE Trans. Acoust. Speech Signal Process. 35(10), 1365–1376 (1987)
2. Kela, P., Costa, M., Turkka, J., Koivisto, M., Werner, J., Hakkarainen, A., Valkama, M., Jantti, R., Leppanen, K.: Location based beamforming in 5G ultra-dense networks. In: 2016 *IEEE 84th Vehicular Technology Conference (VTC-Fall)*. Montréal, Canada, pp. 1–7 (2016)
3. Rihan, M., Soliman, T.A., Xu, C., Huang, L., Dessouky, M.I.: Taxonomy and performance evaluation of hybrid beamforming for 5G and beyond systems. IEEE Access 8, 74605–74626 (2020)
4. Lin, Z., Lin, M., Wang, J.B., Huang, Y., Zhu, W.P.: Robust secure beamforming for 5G cellular networks coexisting with satellite networks. IEEE J. Sel. Areas Commun. 36(4), 932–945 (2018)
5. Jang, J., Chung, M., Hwang, S.C., Lim, Y.G., Yoon, H.J., Oh, T., Min, B.W., Lee, Y., Kim, K.S., Chae, C.B., Kim, D.K.: Smart small cell with hybrid beamforming for 5G: Theoretical feasibility and prototype results. IEEE Wireless Commun. 23(6), 124–131 (2016)
6. Zhou, B., Liu, A., Lau, V.: Successive localization and beamforming in 5G mmWave MIMO communication systems. IEEE Trans. Signal Process. 67(6), 1620–1635 (2019)
7. Zhu, X., Qi, F., Feng, Y.: Deep-learning-based multiple beamforming for 5 g uav iot networks. IEEE Network 34(5), 32–38 (2020)
8. Liu, C., Yuan, W., Wei, Z., Liu, X., Ng, D.W.K.: Location-aware predictive beamforming for UAV communications: A deep learning approach. IEEE Wireless Commun. Letters 10(3), 668–672 (2020)
9. Eappen, G., Shankar, T., Nilavalan, R.: Advanced squirrel algorithm-trained neural network for efficient spectrum sensing in cognitive radio-based air traffic control application. IET Commun. 15(10), 1326–1351 (2021)
10. Mismar, F.B., Evans, B.L., Alkhateeb, A.: Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination. IEEE Trans. Commun. 68(3), 1581–1592 (2019)
11. Alkhateeb, A., Alex, S., Varkey, P., Li, Y., Qu, Q., Tujkovic, D.: Deep learning coordinated beamforming for highly-mobile millimeter wave systems. IEEE Access 6, 37328–37348 (2018)
12. Huang, H., Peng, Y., Yang, J., Xia, W., Gui, G.: Fast beamforming design via deep learning. IEEE Trans. Veh. Technol. 69(1), 1065–1069 (2019)

13. Paszke, A., Chaurasia, A., Kim, S., Culurciello, E.: Enet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint *arXiv:1606.02147* (2016)
14. El Ayach, O., Rajagopal, S., Abu-Surra, S., Pi, Z., Heath, R.W.: Spatially sparse precoding in millimeter wave MIMO systems. IEEE Trans. Wireless Commun. 13(3), 1499–1513 (2014)
15. Va, V., Choi, J., Heath, R.W.: The impact of beamwidth on temporal channel variation in vehicular channels and its implications. IEEE Trans. Veh. Technol. 66(6), 5014–5029 (2016)
16. Li, X., Alkhateeb, A.: Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems. In: *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA, pp. 800–805 (2019)
17. Eappen, G., Shankar, T., Nilavalan, R.: Advanced squirrel algorithm-trained neural network for efficient spectrum sensing in cognitive radio-based air traffic control application. IET Commun. 15(10), 1326–1351 (2021)
18. Eappen, G., Shankar, T., Nilavalan, R.: Cooperative relay spectrum sensing for cognitive radio network: Mutated MWOA-SNN approach. Appl. Soft Comput. 114, 108072 (2022)
19. Vizziello, A., Savazzi, P., Chowdhury, K.R.: A Kalman based hybrid precoding for multi-user millimeter wave MIMO systems. IEEE Access 6, 55712–55722 (2018)
20. Va, V., Vikalo, H., Heath, R.W.: Beam tracking for mobile millimeter wave communication systems. In: *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Washington, DC, pp. 743–747 (2016)
21. Shaham, S., Ding, M., Kokshoorn, M., Lin, Z., Dang, S., Abbas, R.: Fast channel estimation and beam tracking for millimeter wave vehicular communications. IEEE Access 7, 141104–141118 (2019)
22. Kishor, A., Chakraborty, C., Jeberson, W.: Reinforcement learning for medical information processing over heterogeneous networks. Multimedia Tools and Applications 80(16), 23983–24004 (2021)
23. Kishor, A., Chakraborty, C., Jeberson, W.: A novel fog computing approach for minimization of latency in healthcare using machine learning. International Journal of Interactive Multimedia and Artificial Intelligence 6, 7–17 (2021)
24. Kishor, A., Chakarbarty, C.: Task offloading in fog computing for using smart ant colony optimization. Wireless Personal Communications 1–22 (2021)
25. Liu, Y., Wang, X., Boudreau, G., Sediq, A.B., Abou-zeid, H.: Deep learning based hotspot prediction and beam management for adaptive virtual small cell in 5G networks. IEEE Trans. Emerging Top. Comput. Intell. 4(1), 83–94 (2020)
26. ElHalawany, B.M., Hashima, S., Hatano, K., Wu, K., Mohamed, E.M.: Leveraging machine learning for millimeter wave beamforming in beyond 5G networks. IEEE Syst. J. 16(2), 1739–1750 (2022)
27. Santos, G.L., Endo, P.T., Sadok, D., Kelner, J.: When 5G meets deep learning: A systematic review. Algorithms 13(9), 208 (2020)
28. Ebrahiem, K.M., Soliman, H.Y., Abuelenin, S.M., El-Badawy, H.M.: A deep learning approach for channel estimation in 5G wireless communications. In: 2021 38th National Radio Science Conference (NRSC). Mansoura, Egypt, pp. 117–125 (2021)
29. Lavdas, S., Gkonis, P.K., Zinonos, Z., Trakadas, P., Sarakis, L.: An adaptive hybrid beamforming approach for 5G-MIMO mmWave wireless cellular networks. IEEE Access 9, 127767–127778 (2021)
30. Mohamed, K.S., Alias, M.Y., Roslee, M., Raji, Y.M.: Towards green communication in 5G systems: Survey on beamforming concept. IET Commun. 15(1), 142–154 (2021)
31. Zhou, Y., Chen, J., Zhang, M., Li, D., Gao, Y.: Applications of machine learning for 5G advanced wireless systems. In: 2021 International Wireless Communications and Mobile Computing (IWCMC). pp. 1700–1704. IEEE, Piscataway, NJ (2021)
32. Silva, D.H., Ribeiro, D.A., Ramírez, M.A., Rosa, R.L., Chaudhary, S., Rodríguez, D.Z.: Selection of beamforming in 5G MIMO scenarios using machine learning approach. In: 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). Thailand, pp. 1–4 (2022)
33. Yu, P., Zhou, F., Zhang, X., Qiu, X., Kadoch, M., Cheriet, M.: Deep learning-based resource allocation for 5G broadband TV service. IEEE Trans. Broadcast. 66(4), 800–813 (2022)
34. Shen, L.H., Chang, T.W., Feng, K.T., Huang, P.T.: Design and implementation for deep learning based adjustable beamforming training for millimeter wave communication systems. IEEE Trans. Veh. Technol. 70(3), 2413–2427 (2021)
35. Yang, H., Jee, J., Kwon, G., Park, H.: Deep transfer learning-based adaptive beamforming for realistic communication channels. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC). Jeju Island, Korea, pp. 1373–1376 (2020)

How to cite this article: Eappen, G., Cosmas, J., T, S., A, R., Nilavalan, R., Thomas, J.: Deep learning integrated reinforcement learning for adaptive beamforming in B5G networks. IET Commun. 16, 2454–2466 (2022). https://doi.org/10.1049/cmu2.12501