Received 10 November 2023; accepted 27 November 2023. Date of publication 30 November 2023; date of current version 18 December 2023.
Digital Object Identifier 10.1109/OJCOMS.2023.3337850

AI Model Selection and Monitoring for Beam Management in 5G-Advanced

CHEN SUN [1] (Senior Member, IEEE), LE ZHAO [2], TAO CUI [1], HAOJIN LI [1], YINGSHUANG BAI [1], SONGTAO WU [1], AND QIANG TONG [1]

[1]Wireless Network Research Department (WNRD), Sony R&D Center, Beijing 100027, China
[2]School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

CORRESPONDING AUTHOR: C. SUN (e-mail: chen.sun@sony.com)

ABSTRACT This paper investigates the application of artificial intelligence (AI) to wireless technology, specifically in the context of beam management (BM) in the advanced 5th-generation (5G) communication system. Our focus lies in aligning our study with the ongoing discussions within the Third Generation Partnership Project (3GPP) as of December 2022. Instead of evaluating the performance of specific AI models, we take user equipment (UE) receiver (Rx) beam prediction as an illustrative example of AI-based BM. We explore various aspects of AI model management, including model selection, monitoring, and activation/deactivation operations, from a 3GPP perspective. For model selection, we propose deploying distinct AI models for different propagation environments, categorized based on base station (BS) transmitter (Tx) beam measurement results. Reference Signal Received Power (RSRP) serves as a pivotal key performance index (KPI) for model performance monitoring. Our simulation results indicate that, instead of training one all-encompassing AI model with numerous layers for universal application, transitioning between domain-specific AI models with fewer layers yields superior performance. Model activation/deactivation procedures determine whether AI-based BM or traditional BM should be employed in a given scenario. We also introduce the use of AI for predicting the performance of both AI-based BM and traditional BM. By comparing the performance of these strategies, we can ascertain whether link performance degradation results from AI output errors or UE movement into challenging propagation environments. This approach enables the effective management of model switching between AI-based BM and traditional BM. The simulation shows that we can reduce the number of unnecessary switches by 10%.

INDEX TERMS Beam management (BM), machine learning, long short-term memory (LSTM), model management, 3GPP standards.

I. INTRODUCTION

THE INTEGRATION of artificial intelligence (AI) technologies with wireless communication has become a prominent research area in recent years. This field explores two primary approaches to the development of AI-native or AI-intrinsic 6th-generation (6G) wireless networks, as discussed in [1]. The first approach, known as "wireless for AI", aims to leverage existing wireless network features or enhance current functionalities to support broad and efficient applications of AI technologies. Examples include applications in autonomous driving, as highlighted by the authors in [2]. The second approach is "AI for wireless communication", involving the application of AI algorithms to address various challenges within wireless communication systems. These challenges span from early applications like cognitive radio to more recent advancements in areas such as beamforming, interference estimation, channel estimation, modulation at the physical layer, routing at the network layer, and resource allocation for mobile operators, as discussed by in [3].

"Wireless for AI" involves enhancing wireless systems to enable the widespread deployment of AI functionalities across mobile devices, at the network edge, and in the cloud. In 2019, the Service and System Aspects (SA) technical specification group (TSG) of the Third Generation Partnership Project (3GPP) initiated an investigation into the impact of AI and machine learning (ML) model transfer (AMMT) within wireless systems [4]. This investigation led to the definition of network functions that support user selection in federated learning [5], model splitting [6], model distribution [7], and more, catering to a variety of AI model applications.

Conversely, "AI for wireless" pertains to the utilization of AI technologies to improve wireless network functions. Multiple studies have explored the application of AI technologies across various layers of communication protocols, as evidenced by [8]. As millimeter-wave (mmWave) technology plays a pivotal role in the 5th-generation (5G) system, effective beam management becomes crucial. Several studies have delved into the utilization of AI technology for beam management, as documented in [9], [10], [11], [12]. These works primarily concentrate on the deployment of AI models in various scenarios and assess their performance. In [13], reinforcement learning is explored for serving multiple beams to different vehicles, with a particular focus on mitigating blocking effects by applying reinforcement learning to joint beam allocation and relay selection [14]. Additionally, in [15], a Long Short-Term Memory (LSTM) model is employed to track the angle of arrival (AoA) based on channel information, facilitating beam management while the user equipment (UE) is in motion along a straight trajectory. Reference [16] adopts a convolutional neural network (CNN) to extract angle information from in-phase-quadrature (I/Q) signal samples. These studies harness AI technology to unveil insights into wireless channel characteristics. Furthermore, to enhance accuracy, it's possible to incorporate data beyond the radio link, as demonstrated in [17], where user location information is integrated into beam prediction. By applying deep learning models based on user movement patterns, these methods can achieve prediction accuracy exceeding 90%.

Nevertheless, a limited number of investigations have undertaken an examination of model life cycle management, a fundamental aspect in the deployment of artificial intelligence (AI) for network management. This is of paramount significance due to the inherent variability of wireless communication channels. Models designed for specific scenarios may prove inadequate in ensuring optimal performance when confronted with fluctuations in channel conditions. Consequently, there arises a necessity to delve into the systematic monitoring of AI model performance, with a focus on determining the appropriate circumstances and selection criteria for model activation.

In this paper, we present an overview of the ongoing discussions related to the standardization of AI-based Beam Management (BM) schemes [18]. Given the early stage of standardization activities and the divergence of opinions among various companies regarding the three key stages of AI deployment, we offer general perspectives on essential topics, including use cases, deployment strategies, AI model inputs and outputs, signaling, model training, performance evaluation, and model management. Following this, we delve into the time domain prediction of UE's beam prediction as an illustrative example of an AI-based BM scheme, outlining the data collection process and model training. For model management, we propose a radio propagation environment-driven model selection approach, where the environment is identified based on the measurement of BS beams. Additionally, we observe that the performance degradation in an AI-based BM scheme may result from AI prediction errors or challenging propagation environments. To prevent unnecessary deactivation of the AI model and fallback to the traditional BM scheme, we suggest utilizing AI to predict the performance of both the traditional and AI-based BM schemes. By comparing the predicted performance of these two schemes, we can make informed decisions regarding the activation and deactivation of AI-based BM.

II. PRELIMINARIES

In this section, we provide a description of the traditional Beam Management (BM), followed by an overview of the AI-based BM. We specifically delve into various aspects of AI-based BM that are currently being discussed in 3GPP. The subsequent section will focus on studying UE beam prediction from these aspects, presenting our proposal on AI model selection, and outlining AI model performance monitoring.

A. TRADITIONAL BEAM MANAGEMENT SCHEME

In 2015, the 3rd Generation Partnership Project (3GPP) initiated the development of the 5G standard [19]. A significant feature of the 5G system is its emphasis on BM, enabling the BS and UE to direct energy towards specific directions using directional beams [9], [20]. This capability holds particular significance for high-frequency communication with high pathloss. Therefore, the management of beam directions at both the BS and UE sides has become a central focus within the 5G system.

Aligning the beam directions at both transmitter (Tx) and receiver (Rx) in wireless environment for a better link quality has been topic for a long time [21], [22]. The 3GPP has developed a standard that define reference signals and measurement procedures to realize this the alignment purpose [23]. This is realized by the following procedures. Before accessing a network, a UE needs to acquire a synchronization signal and obtain system information such as the cell ID. In a Long-Term Evolution (LTE) system, the BS transmits a synchronization signal every 5 ms. In a New Radio (NR) system, commonly referred to as the 5G system, synchronization signals are transmitted together with system information and control information as a synchronization signal block (SSB). Multiple SSBs constructs a SSB burst. A core feature of NR systems is beam-space operation, in which signals are transmitted in different beam directions. Equipped with multibeam antennas, the BS can send an SSB in each beam direction consecutively over a period

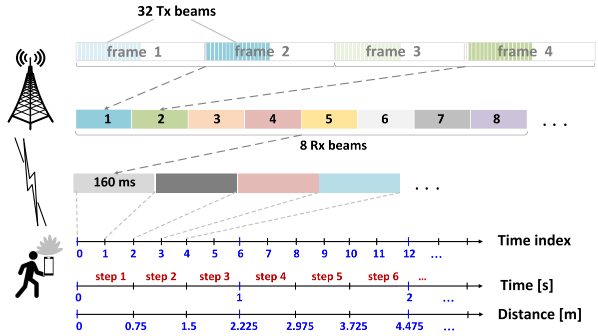

FIGURE 1. **Mapping of SSB periods, beam sweeping and UE movement distance.**

of an SSB burst. The number of SSBs in each SSB burst defines the number of beams that a BS and deploy for measurement. The maximum number of SSBs in each burst depends on the subcarrier spacing and frequency band, which determines how many Orthogonal frequency-division multiplexing (OFDM) symbols can be packed into each half of a 10 ms time frame [19], [24]. For a system that operates in frequency band below 3GHz band, each burst has 4 SSBs. For 3GHz to 6GHz band, a maximum of 8 SSBs can be supported. For frequency band above 6GHz a maximum of 64 SSBs can be supported. The period of this SSB burst transmission can be 5 ms, 10 ms, or up to 160 ms. A UE can assume a default periodicity of 20 ms during the cell search process to reduce its power consumption relative to that in an LTE system. This UE with a multibeam antenna can measure the SSB burst in each direction by repeating measurement over a few consecutive periods of SSB bursts. In this way, the best pair of BS and UE beams can be determined.

An example is shown in Fig. 1, where the BS employs 32 beams, and the UE has 8 beams. We set the SSB burst period to 20 ms. The carrier frequency is set to 28 GHz, with a 120 kHz subcarrier spacing. In each subframe of a 5 ms period, 16 SSBs are transmitted, corresponding to 16 beam directions at the BS. In total, 8 periods are measured, corresponding to 8 UE beam directions. Based on the reference signal received power (RSRP) measurements of the SSBs, the best BS beam direction and the corresponding UE beam direction are determined. In total, 160 ms is consumed for one BM procedure. In general, a pedestrian takes 3 steps in 1 s, covering 0.75 m in one step. It takes almost half a step duration of signaling measurement to complete one BM procedure. The delay and complexity will become a severe issue for applications with higher mobility. To avoid the long duration of the measurement complexity with this traditional BM scheme. The 3GPP has recently started the development of AI-based BM as one of the features of Release 18 standard [18].

B. AI-BASED BEAM MANAGEMENT SCHEME

Various approaches of realizing AI-based BM have been studied in academia, however, the lack of standardization

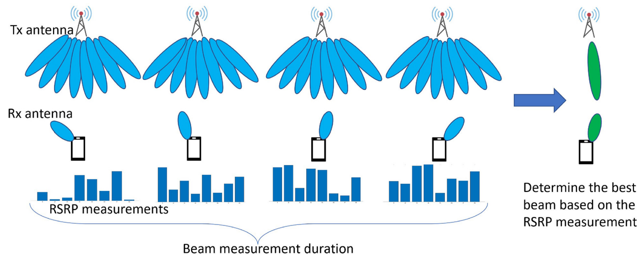

FIGURE 2. **Traditional BM scheme.**

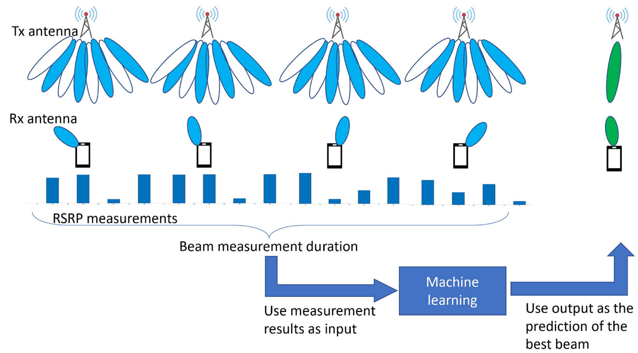

FIGURE 3. **Spatial domain AI-based BM.**

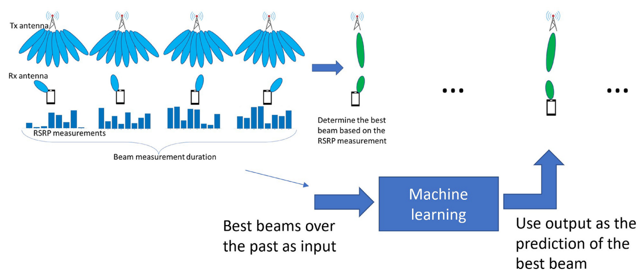

FIGURE 4. **Time domain AI-based BM.**

is one of the open issues for the mass applications of AI technologies for BM. Fortunately, the 3GPP has started working of the development of standards for AI-based BM at the end of 2020 from the following aspects.

Use Cases: Compared with the traditional BM scheme in Fig. 2. The TSG currently agrees on two major use cases as shown in Fig. 3 and Fig. 4, respectively. (1) Spatial-domain prediction case: to avoid the need to measure a large number of Tx beams, an AI model can predict one or a few of the best beams based on measurements of a subset of beams pointing in certain directions instead of measurements of all beams. (2) Time-domain prediction case: an AI model predicts one or a few of the best beams to be used in the next time instance based on a sequence of beams selected from past time instances. In addition to these two use cases, other minor use cases have also been proposed by various companies. For example, one may wish to predict the best beam based on a UE's location [25]. Discussions on this use case are ongoing, as the location information of a UE is considered private. Some companies have also proposed the prediction of mmWave beams based on the channel information in the sub-6 GHz band [26].

Deployment: An AI model can be deployed on the BS side, on the UE side, or both. Training and inference can be performed at the same location or at different locations. For example, training may be performed on the BS side, while inference is executed on the UE side. However, this requires the trained model to be transferred from the BS side to the UE side.

Input and Output: The output of an AI model can directly give the best beam to be used, in a form such as the beam index. This reduces the complexity of BM. Another approach is to provide a set of candidate beams that have a high possibility of being the best beam. In this case, an ordinary BM procedure is still needed, but a lower complexity can be achieved, as only a subset of beams need to be measured [27]. This approach provides better performance than the approach of directly using a single beam predicted by the AI model because the AI model output might be wrong. Such an incorrect result will result in the beam pointing in an unwanted direction, leading to degradation in one or more indicators of wireless link performance, for example, the RSRP.

Signaling between the BS and UE: The application of AI technologies requires a large volume of data for training, suitable arrangement of the data as the model input for inference and timely measurement for performance monitoring. Depending on whether the AI model is deployed on the BS side, on the UE side, or both, the current signaling conditions between the BS and UE must be re-examined to see if the link quality is sufficient to support the application of AI technologies. Taking the downlink direction as an example, in the traditional BM procedure, a UE will determine the four best Tx beams based on measurements and report the results to the BS. If AI is deployed at the BS to predict these four best beams, the UE might need to send measurement results corresponding to different downlink beams to the BS. This measurement report might require more data than reporting four Tx beams with their RSRPs and more frequent reporting than the traditional BM procedure. Since different AI models with different inputs and outputs will be deployed by different companies in their products, defining these signaling requirements is critically important to the interoperability of different products.

Model Training: In practical applications, an AI model is trained on data collected for a given scenario, and the performance of the AI model might change if the environment changes. For example, consider a model trained on data collected at night. When this model is applied during the day, the output will exhibit errors in beam direction prediction since the UE behavior changes and the propagation environment might also change due to reflections from cars on the street.

Model Evaluation: An important aspect of AI technology is AI model performance evaluation. The first issue of concern is the performance evaluation of an AI model. The TSG has started to define some key performance indexes (KPIs) for characterizing the inference accuracy of a model, such as the beam ID/RSRP prediction accuracy, the link-quality-related RSRP, throughput, and the signal-to-interference-plus-noise ratio (SINR). The RSRP prediction accuracy is evaluated as the RMSE between the RSRP achieved using the predicted beam and that of the RSRP achieved with the traditional BM approach. The beam ID prediction accuracy is the percentage of correctly predicted IDs among the total number of predictions.

Model Management: The objective of model management is to select AI models appropriate for a given application scenario, monitor the model performance, and activate/deactivate a model or trigger fallback from AI-based BM to traditional BM if necessary. In 2022, the TSG also started considering the overhead of signaling and measurement reports associated with such model management as well as sources of complexity such as computation and memory costs for model monitoring. For example, to check the validity of the input to an AI model, extra processing is needed to calculate the distribution of the input data. Note that although specific AI/ML algorithms and models may be studied for evaluation purposes, such AI/ML algorithms and models are specific to given implementations and are not expected to be specified in standards.

III. UE BEAM TIME DOMAIN PREDICTION

In this section, UE Rx beam time-domain prediction is used as an example to study the issues related to model management. We propose a model switching method based on the propagation environment. Using the beam accuracy and RSRP as KPIs, we examine the performance of model switching. Furthermore, we use AI to predict the performance of both AI-based BM and traditional BM for model monitoring and discuss the fallback mechanism.

A. SYSTEM MODEL

In this paper, we simulate a city scenario in which a user holds a cell phone (typically called a UE in the 3GPP context), walks along a direction within a certain range, and establishes a communication link with a BS. As shown in Fig. 5. The user starts from a location below the BS and initially moves along the road toward north and then makes a turn toward west. There are scatters placed along the road. During the movement process, the movement direction of the user might also change to the left and right, but the user will continue walking forward along the road. The elevation angle of the user's mobile phone can also change. The trace of the movement consists of four segments. The channel between the user and the BS is a line-of-sight (LoS) channel at the beginning, during the 250 m walk represented by segments 1 and 2, and it is a NLOS channel in the latter part of the trajectory, during the 200 m walk represented by segments 3 and 4.

We generate the trajectory of the UE movements following a Markov process. The change in the movement direction

FIGURE 5. Simulation scenario.

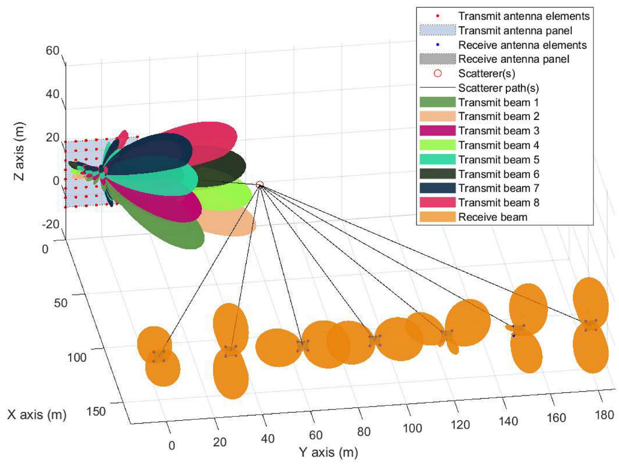

FIGURE 6. Simulated beams of BS and the beams of UE when UE is moving along the Y axis with random rotation.

TABLE 1. Simulation parameters.

Parameter	Value
Carrier frequency	28 GHz
Antenna at BS	Uniform rectangular array (URA)
Antenna at UE	Uniform rectangular array (URA)
Speed of UE	8 km/h
Number of Tx beams	32
Number of Rx beams	8
Transmitter array size	8 × 8
Receiver array size	2 × 2
Sample rate	30720000
SSB periodicity	160 ms
SNR	30 dB
Number of scatters	4

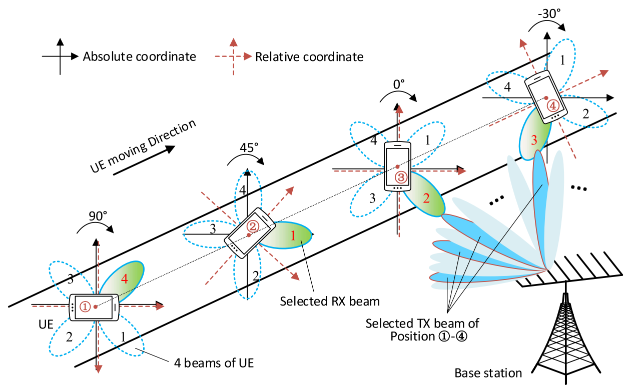

FIGURE 7. Illustration of beam direction and beam ID assignment.

of the UE is selected from a uniform distribution spanning from –60 to 60 degrees. Note that the UE's movement is always bounded inside the road. The elevation angle of the UE is set to 0 degrees for 50% of the time, emulating that most of the time, the user is walking with the mobile phone upright. For the remaining 50% of the time, the elevation angle change of the UE at the next step is uniformly selected from [–10, –20, –30, 10, 20, 30] degrees.

In this paper, we develop a location-based channel model and SSB burst transmission. The parameter settings are shown in Table 1. The antenna of the BS is a 5-meter-high uniform rectangular array (URA) with 32 beams facing north, with an azimuthal angle range of [–60 60] degrees and an elevation angle range of [-90 0] degrees. The coordinates of the BS are [0, 0, 5] meters. The UE is equipped with a URA with 8 beams covering [–180 180] degrees in the azimuthal plane and [0 90] degrees in elevation. The height of the UE is 1.8 meters. The simulated beams are shown in Fig. 6.

Referring to the mapping between the duration of the BM procedure and the human movement distance is shown in Fig. 1. Each time index represents a BM result. The 3D position of the UE antenna changes over each step of the user, which spans two consecutive time indexes. This mapping is general in that we can modify the amount of change in location for different time index to simulate different scenarios of different mobility.

For the model input and output, we consider both beam direction and beam ID values in the AI-based Beam Management (BM) scheme. As illustrated in Fig. 7, the beam direction represents the absolute angular direction in the coordinate system. In contrast, beam ID values are assigned relative to the UE. Using the AI model with beam direction requires the UE to calculate the absolute directions of each beam to generate input for the AI model. Additionally, given the predicted direction of the best beam, the UE needs to identify the corresponding beam for that direction. If the AI model with beam ID is employed, the UE's orientation and beam ID assignment must align with the UE used to train this model. In comparison to the beam direction-based model, the beam ID-based approach eliminates the need for beam direction calculation, as long as the beam ID assignment aligns with the UE used for training this model.

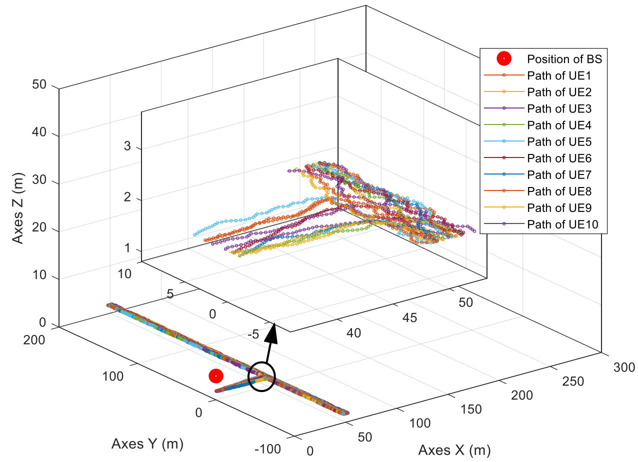

FIGURE 8. Trajectories of the UE in the simulation for training data collection.

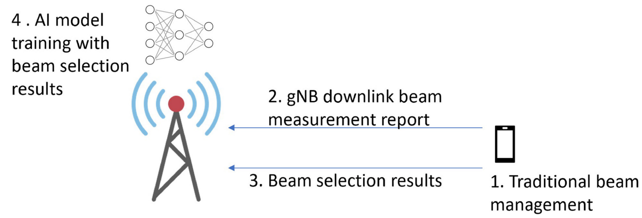

FIGURE 9. Process of model training for UE beam prediction.

B. TRAINING DATA COLLECTION

In the data construction process, we set up a simplified simulation environment. The UE starts moving directly below the BS and follows the trajectory depicted in Fig. 5. The actual movement of the UE is shown in Fig. 8. The UE's movement is modeled as a Markov process, where at each step, the direction of movement is randomly chosen. The UE's position in the next time slot is updated after moving a fixed step size in the randomly chosen direction. Based on the UE position we run the SSB (Synchronization Signal Block) beam sweeping simulation to obtain the RSRP values for different Tx/Rx beam pairs. Additionally, we determine the best beam IDs and beam directions for the UE at each position, constructing a dataset that reflects the changing beam sweeping behavior as the UE moves.

Along the trajectories of the UE, we collected beam selection results from the traditional BM scheme. Data from nine trajectories were utilized for training, while data from one trajectory was reserved for testing. We use 90% of the data for training and 10% of the data for testing. We train the LSTM model with Adam algorithm. During training, a random dropout of 25% is applied—a regularization method where input and recurrent connections to LSTM units are probabilistically excluded from activation and weight updates. This technique helps reduce overfitting and improve overall performance. Before the training process, the data normalization is performed as follows. Assuming the optimal beams from traditional BM, described by either the beam direction or beam ID, are sampled as the training data, x_t, $t = 1, 2, \ldots, T$, where t is the time index as shown in Fig. 1 and T is the total number of samples. The normalized data is [28]

$$x'_t = \frac{x_t - \bar{x_t}}{\sigma}, \tag{1}$$

where $\bar{x_t}$ and σ are the mean and standard deviation of x_t. We use x'_t as the input of the LSTM model in the training process.

C. MODEL TRAINING

In Fig. 9 we show the example of implementation of the training process. The UE performs the traditional BM scheme and reports the measurement of the BS downlink beams. Based on the reports the BS can determine the model to be trained for this UE. The UE reports the results of the selection of the receiving beam. Using the sequence of the beam selection results the BS can train the AI model to predict the UE's beam in time domain. After training, this model can be distributed to the UE the downlink as payload. Note that this training can also be done at the UE side if the computation power is sufficient.

Recurrent neural networks (RNN) are adept at processing sequential data, and LSTM is a commonly employed variant within this family. LSTM networks consist of multiple LSTM units, each featuring an input gate, a forget gate, and an output gate, which regulate the flow of information. These gates dynamically adjust the propagation and forgetting of information through learning, enabling the model to better capture long-term dependencies within sequential data. The input data comprises the results of beam sweeping, encompassing various features related to the sweeping process. These features pass through the input layer and undergo processing by the hidden layers of the LSTM network. The LSTM units are sequentially connected within these hidden layers. Throughout the training process, the model continually adjusts the weights and biases using the backward propagation algorithm, aiming to minimize the disparity between predicted values and ground truth. This iterative approach allows the model to learn patterns and regularities within the input data, enabling it to make predictions on unseen data.

This study employs an LSTM-based recurrent neural network with hidden layers to process and predict beam sweeping result data. The model effectively captures long-term dependencies within sequential data and undergoes training using the backward propagation algorithm to minimize prediction errors. The training process utilizes the Adam optimization over 500 epochs, with a training gradient threshold set at 1.5 and an initial learning rate of 0.1. Additionally, we apply a learning rate dropping factor of 0.25 with a dropping period set to 125.

D. MODEL SELECTION

In 3GPP discussion, various models have been studied including LSTM, transformer, convolutional neural network

(CNN), etc. Storage and computational complexity of these AI models are reported. The former decides the memory needed and the latter decides the inference time. For BS deployment, an LSTM model boasting a modest 80K parameters in memory and 36K floating-point operations per second (FLOPs) achieves a 70% accuracy in top beam pair prediction [29]. On the other hand, a BS-side LSTM, with a more substantial 2.755M parameters and 7.25M FLOPs computational complexity, attains an impressive 90% accuracy in beam pair prediction [30]. The work in [31] reports a UE-side LSTM model with three layers and a hidden/cell size of 128, comprising 340k parameters, yielding a 77% accuracy for the top-1 Rx beam prediction. Furthermore, a 570k encoder-only transformer model deployed at the UE achieves an 80% top-1 RX beam accuracy. Given storage constraints of a UE, it's feasible to deploy a range of AI models at the UE, allowing for the selection of models with varying parameter configurations tailored to specific environmental conditions. These diverse models can be trained at the BS and distributed to UEs before activation. The standardization body is currently designing signaling for model selection/activation and exploring the distribution of models from the BS to the UE.

Based on the conclusion given in [32], it is anticipated that deploying a single AI model for various propagation environments would yield inferior performance compared to deploying distinct models for each unique propagation environment. This study proposes AI model selection based on the measurement report of the downlink BS Tx beam. This choice is motivated by the observation that BS beam direction changes less frequently, serving as an indicator of UE environmental changes. Conversely, the UE's beam direction changes more frequently and is more closely associated with the UE's rotation.

The 5G system is designed to transmit SSBs using directional beams, facilitating the establishment of Tx and Rx beam pairs during initial access. As the UE moves, the established directional beams require periodic reevaluation. The BS has the capability to configure radio resources for beam measurement, with UEs responsible for reporting measurement results. These regular measurements assist in assessing the spatial uniqueness of a UE. For example, the sequence of the top four downlink Tx beam directions signifies the primary propagation routes, as illustrated in Fig. 1. By leveraging routine measurements, the BS can ascertain the spatial distinctiveness of a UE and select an appropriate model as shown in Fig. 10.

E. PERFORMANCE MONITORING AND MODEL ACTIVATION/DEACTIVATION

AI-driven beam prediction is prone to inaccuracies, influenced by factors such as dynamic channel conditions and interference. Industry studies have demonstrated that LSTM models of varying sizes can achieve top-1 beam prediction accuracy ranging from 70% to 90% [29], [30], [31], [33]. Despite these inherent imperfections, AI-based

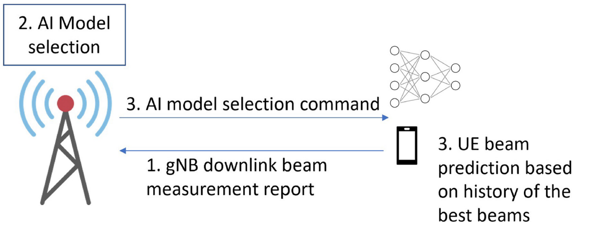

FIGURE 10. Process of model selection for UE beam prediction.

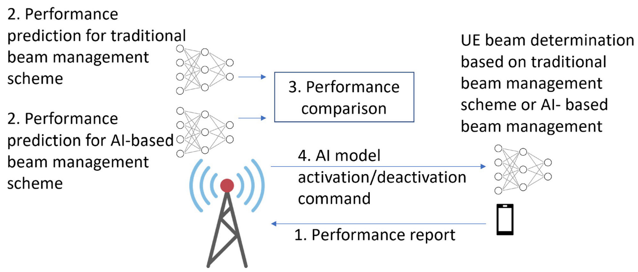

FIGURE 11. Process of model activation/deactivation.

beam management effectively streamlines system operations, optimizing resource allocation and consequently enhancing system capacity and data transmission. When evaluating beam management schemes, it is essential to consider not only prediction accuracy but also the RSRP to assess link quality.

We propose the deployment of two additional AI models to predict the RSRP performance for both the traditional BM scheme and the AI-based beam measurement scheme. The activation/deactivation of the AI-based BM scheme is determined through a performance comparison with the traditional BM scheme, as illustrated in Fig. 10. The UE reports performance metrics, such as RSRP, to the BS. It's important to note that even for the AI-based beam measurement, regular RSRP measurements are still necessary to predict the best beams, as depicted in the figure. If the predicted beam is selected, the UE can also report the associated RSRP. At the BS, the two AI models for performance evaluation predict the performance of both schemes and determine the optimal selection.

While envisioning that errors of AI-based BM can result in a drop in RSRP, we also recognize that the degradation in RSRP performance might be attributed to the UE moving into a challenging radio environment. Utilizing activation/deactivation based on the comparison of the two schemes helps prevent situations where the AI performance is underestimated in harsh environments, as switching back to the traditional scheme may not significantly improve performance. If we anticipate a decline in the performance of the AI-based beam management scheme, we assess the predicted performance of the traditional beam management scheme. The system reverts only if the predicted RSRP for the traditional scheme does not indicate performance degradation. This approach significantly minimizes unnecessary fallback instances in comparison to the current approach.

Algorithm 1: AI Based BM Switching Framework Efficiency

input : Predicted RSRP of AI BM and traditional BM
output: The percentage of invalid fallbacks decreased

Define the *3dB drop number indicator*, *drop location indicator vector* and the *number of times to avoid invalid fallbacks*;
Initialize the indicators as **0**;
for *all* **time_step do**
  Extract the RSRP value of predicted RSRP of AI BM in time_step and time_step -1;
  if *the predicted RSRP of AI BM drops by 3dB* **then**
    3dB drop number indicator $+1$;
    value of the *drop location indicator vector* in time_step $= 1$;
  Extract the RSRP value of predicted RSRP of traditional BM in time_step and time_step -1;
  if *predicted RSRP of traditional BM drops by 3dB* **then**
    3dB drop number indicator $+1$;
    value of the *drop location indicator vector* in time_step $= 1$;
  Save the *3dB drop number indicator* and *drop location indicator vector*;
for *all* **time_step do**
  if drop location indicator vector *of both AI/traditional BM predicted RSRP* $= 1$ **then**
    The RSRP drops occur at this time step, and the gain of the switch to the traditional BM is limited.;
    The *number of times to avoid invalid fallbacks* $+1$;
Obtain the *number of times to avoid invalid fallbacks*;
Calculate the percentage with *number of times to avoid invalid fallbacks* and the *3dB drop number indicator* of predicted AI BM RSRP.

This mechanism is illustrated through the following pseudo-code of Algorithm 1.

IV. PERFORMANCE

In the simulation, we focus on the downlink direction. Initially, we implement the traditional BM procedure as the UE moves along the route depicted in Fig. 5. Following the traditional BM scheme, the optimal BS Tx beam and UE Rx beam are selected based on RSRP measurements of SSBs obtained through the beam sweeping process for various UE positions. The sequence of optimal Tx and Rx beam IDs as functions of the time index is illustrated in Fig. 12. Notably, the optimal BS Tx beam IDs exhibit minimal changes, while the optimal Rx beam on the UE side undergoes substantial variations. This observation suggests that the optimal Tx beam strongly depends on the UE's location, whereas the Rx beam on the UE side is predominantly influenced by UE rotation and nearby scatters.

Utilizing the outcomes illustrated in Fig. 12, we can delineate distinct zones along the route characterized by varying values and trends in Tx beams. To address the uniqueness of the propagation environment for AI-based BM, we advocate deploying different AI models for each zone, departing from the approach of utilizing a single AI model for all scenarios. Model selection can be based on the UE's report of the best four Tx beams or the UE's position. For instance, we classify the environment into four zones corresponding to route segments 1 to 4, as depicted in Fig. 5.

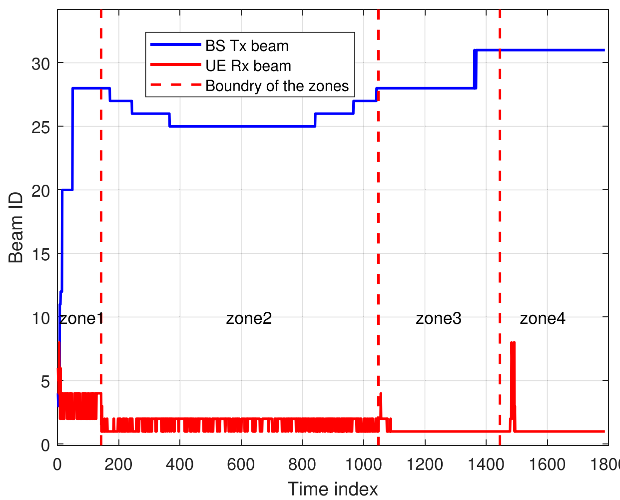

FIGURE 12. Optimal beam directions of the BS Tx antenna and the UE Rx antenna as functions of the BM time index.

Rather than focusing on the accuracy of this model selection, we emphasize the performance of this zone-specific AI-based BM, comparing it with employing only one AI model for all zones in the latter part of the simulation. Ultimately, we aim to investigate whether we can predict the performance for both traditional and AI-based BM schemes, enabling the activation/deactivation of the AI model.

We explore the implementation of AI for time-domain UE beam prediction. Employing an LSTM model for each zone,

TABLE 2. Performance AI-based BM with beam ID and beam direction.

Training data	Evaluation	zone 1	zone 2	zone 3	zone 4	all zones
Beam ID	*Accuracy of Beam ID*	47.83%	80.04%	98.35%	94.66%	83.84%
	RMSE of RSRP	0.9567	0.1132	0.1723	0.1902	0.3016
Beam Direction	*Accuracy of Beam ID by predicted Beam Direction*	50%	80.49%	86.12%	86.05%	80.65%
	RMSE of RSRP	0.8898	0.1290	0.2255	0.2220	0.2293

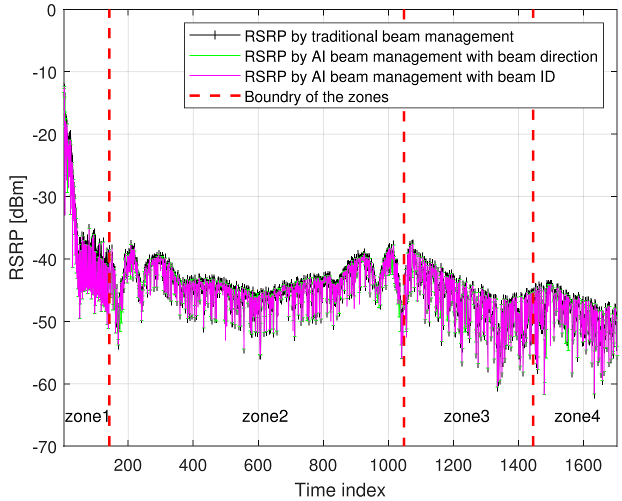

FIGURE 13. RSRPs achieved with the traditional BM scheme and the RSRPs of the AI-based BM scheme.

the architecture comprises one input layer, one output layer, and six hidden layers. To predict the optimal beam for the next time index, we utilize four consecutive optimal beams from the past 4 time indexes. Leveraging this time-domain prediction, a 20% reduction in complexity is achievable by utilizing the predicted beam instead of repeating the beam measurement procedure if the prediction proves accurate. The duration of measurements in the traditional BM procedure can also be conserved, contributing to enhanced data transmission efficiency and ultimately improving system throughput.

We compare the performance of this output from the AI model with the performance of using the optimal beam achieved through traditional BM scheme at this time index. When using the beam ID as input and output of the AI model, we compare the beam ID predicted by the AI model and the beam ID obtained through the traditional BM scheme. Whereas, when using the beam direction as the input and output of the AI model, we select the beam matching the direction predicted by the AI model and compare it with that obtained through the traditional scheme. Following the 3GPP discussion, we select two criteria to measure the performance. The link quality of each time instance of both schemes are compared the RSRPs obtained by traditional BM scheme and simulated by employing the beam predicted by the AI, respectively. In Fig. 13 we plot the RSRP achieved by the traditional scheme. The RSRPs achieved by AI with different input and output achieves very close value to those of the traditional scheme. For clarity of presentation, we plot the RSRPs of the AI-based BM scheme only on those time index where the RSRP values are different from those by the traditional BM scheme. Applying the AI-based BM scheme we can achieve similar performance with that of the traditional scheme, nevertheless there are situations where the RSRP of the predicted beam is lower than the RSRP achieved by traditional scheme. This is due to the beam prediction error. Following the 3GPP discussion, this accuracy of the prediction can be calculated as the number of correct estimations over the total number of estimations.

Table 2 lists the beam accuracy and the RMSE of the RSRP for traditional BM and AI-based BM. The results show that the errors in the LOS environment (zones 1 and 2) are larger error than those in the NLOS environment. This is because in the LOS environment, the beam selection changes more dramatically than it does in the NLOS environment (zones 3 and 4), as shown in Fig. 12. Furthermore, we can see that the beam-direction-based approach and the beam-ID-based approach yield very similar performance in terms of prediction accuracy.

Now, we compare the performance achieved using different models for different zones. To balance the training in different zones, 142 results were collected from each trajectory. Consequently, each zone has an equal training length of 1278. As shown in Table 3, we first deploy the models trained for each individual zone in other zones. We can see that the models are quite zone-specific in that each model achieves the best performance only in the zone for which it was trained. For example, when the model trained for zone 1 is used for beam ID inference in zone 1, the error is the smallest.

Subsequently, we explore the option of training a single universal model to cover all zones. Initially, the model structure mirrors that of individual AI models for each zone, featuring six layers. The comparison of beam IDs determined through traditional BM and AI-based BM is illustrated in Fig. 14. Table 3 highlights the Beam ID RMSEs of this universal model in comparison to RMSEs achieved with zone-specific AI models across different zones. It is evident that the unified model fails to outperform the zone-specific models. Moreover, an increase in the number of hidden layers to fifty does not lead to performance improvement, as depicted in Fig. 15. Notably, the performance is notably worse in zone 1 (time index from 1 to 180), attributed to the

TABLE 3. Performance comparison of different selected models in different propagation environments.

Target zone for performance evaluation	Deployed AI model	RMSE of predicted beam ID
zone 1	*model in zone 1*	**0.7772**
	model in zone 2	0.9669
	model in zone 3	1.1013
	model in zone 4	1.6376
	model in all zones	0.9528
zone 2	*model in zone 1*	1.0565
	model in zone 2	**0.8901**
	model in zone 3	1.0599
	model in zone 4	1.3936
	model in all zones	0.9107
zone 3	*model in zone 1*	0.8098
	model in zone 2	1.1292
	model in zone 3	**0.5650**
	model in zone 4	0.8959
	model in all zones	0.6094
zone 4	*model in zone 1*	1.0153
	model in zone 2	1.0649
	model in zone 3	0.6866
	model in zone 4	**0.5958**
	model in all zones	0.6890

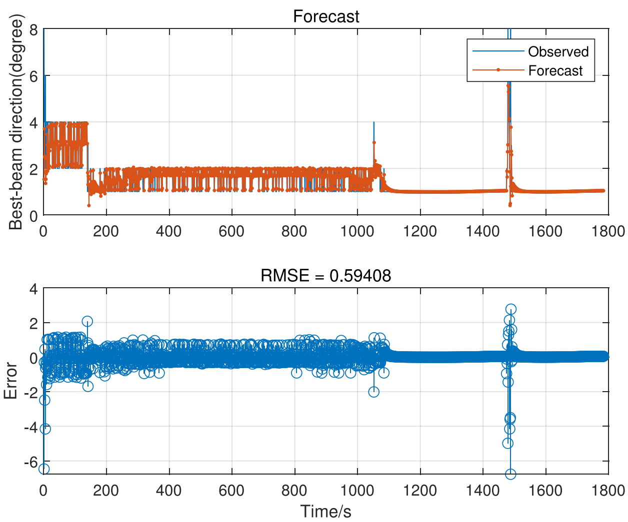

FIGURE 14. Performance achieved with AI-based BM over all zones using one LSTM model with six hidden layers.

increased likelihood of overfitting problems with a model featuring more layers [34].

In recent discussions, the TSG has delved into the issue of model performance monitoring. Monitoring the performance of an AI-based BM scheme facilitates the decision-making process to “fall back” to the traditional BM scheme when the UE encounters poor performance. Analyzing the performance depicted in Fig. 13, simulations reveal that inadequate performance may arise from either a challenging radio propagation environment where both traditional and

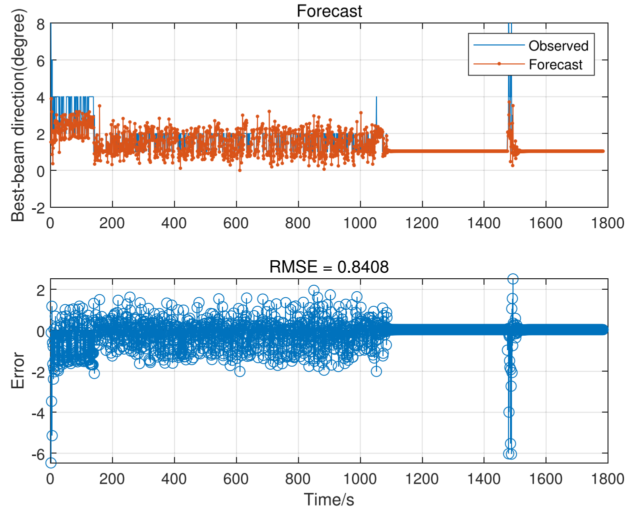

FIGURE 15. Performance achieved with AI-based BM over all zones using one LSTM model with fifty hidden layers.

TABLE 4. Accuracy of the predicted RSRPs from using LSTM with varying numbers of hidden layers.

Number of hidden layers	Evaluation criteria	AI BM based
6 layers	*RMSE*	0.6256
	Drop accuracy	49.18%
15 layers	*RMSE*	0.6355
	Drop accuracy	44.37%
50 layers	*RMSE*	0.5983
	Drop accuracy	**59.18%**
100 layers	*RMSE*	0.5868
	Drop accuracy	59.04%

AI-based BM schemes yield poor link quality in terms of RSRP. Additionally, situations arise where the RSRP achieved by the predicted beam is significantly inferior to that achieved by the traditional BM scheme, attributable to the prediction error of the AI-based BM scheme.

Motivated by this consideration, we propose using two AI models to predict the performance of both the AI-based BM scheme and the traditional BM scheme, respectively, as shown in the second step of Fig. 11. We deploy two one-layer LSTM models and use 1788 RSRP results collected along a UE trajectory as training data. In 3GPP discussion various companies have utilized a 3dB drop in RSRP to evaluate beam performance. In our simulations, we incorporate this 3dB RSRP drop as the criterion for fallback. As shown in Table 4, the LSTM with fifty hidden layers has the best performance in terms of the RMSE of the predicted RSRPs the accuracy of the predicted drops. Thus, we use the LSTM with fifty hidden layers in the model activation/deactivation mechanism to determine the fallback event. We can see that the predicted time instances where the AI-based BM performs significantly worse than the traditional BM in Fig. 16 matches the performance of both schemes in Fig. 13. This means that we can predict the

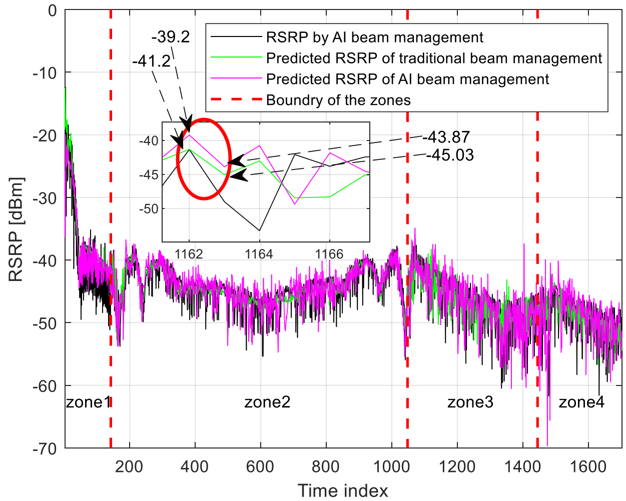

FIGURE 16. **Predicted RSRPs by traditional BM and AI-based BM compared with the RSRP by the AI-based BM.**

performance of both traditional and AI-based BM schemes quite accurately. By comparing the two predicted values, we can trigger fallback in advance only if the AI-based BM scheme has performance much worse than that of the traditional scheme, i.e., the prediction error of the AI-based BM. This allows us to avoid unnecessary fallback in harsh radio environments if we only monitor the performance of AI-based BM only whereas traditional BM cannot achieve better performance. For example, as illustrated in Fig. 16, during time indexes 1162 to 1166, the RSRP by AI BM experiences a drop, leading the existing scheme to switch over to traditional BM. In contrast, with the proposed mechanism, when predicting the upcoming drop of AI-based BM, we observe a corresponding drop in the predicted RSRP of the traditional scheme. Consequently, the system refrains from switching from AI-based BM to traditional BM. Along the entire trajectory's time indexes, we counted 294 fallbacks for the existing approach. With the proposed scheme, we experienced 264 fallbacks, avoiding 10% unnecessary system function switches. It's important to note that a system function switch will also entail changes in reference signal resource assignments, thereby increasing system complexity. Additionally, it should be noted that the intention of this paper is not to develop an advanced AI model for performance prediction but rather to propose a mechanism. It is believed that with more accurate predictions of RSRP, more precise switches can be realized.

V. CONCLUSION

This paper explores AI-based BM. We began by providing an overview of the 3GPP discussion on AI for BM. Subsequently, using UE beam prediction as an exemplar of AI-based BM, we introduced an environmental-driven model selection for beam prediction. Environmental changes are identified through BS Tx beam measurements. Additionally, we proposed performance monitoring using AI for both traditional and AI-based BM schemes.

Through simulations, we conducted a comparative analysis of beam-ID-based training and beam-angle-based training. Both methods demonstrated satisfactory performance in terms of prediction accuracy. We further illustrated that environments can be categorized based on propagation properties, utilizing indicators such as reported best BS Tx beams. Additionally, our findings indicate that switching among AI models in accordance with the environment outperforms deploying a single universal AI model for all scenarios. Finally, we demonstrated that utilizing AI to predict the performance of both traditional BM and AI-based BM schemes enables the identification of potential performance degradation, distinguishing between AI model prediction errors and harsh propagation environments. This proactive approach allows for informed fallback decisions.

The 3GPP discussion of AI-based BM is ongoing, and we have studied several aspects raised in this discussion to date, such as the inputs and outputs of AI models, AI model switching, and the model management. We hope that this study can shed light on the future academic research and the standardization process of AI for wireless communication.

In our exploration of the ongoing 3GPP discussion on AI-based BM, we have delved into various aspects discussed thus far, including the intricacies of AI model inputs and outputs, the dynamics of AI model switching, and the nuances of model management. By addressing these key facets, our study aims to contribute valuable insights to both the academic research landscape and the evolving standardization process of AI for wireless communication.

REFERENCES

[1] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019, doi: 10.1109/MCOM.2019.1900271.

[2] M. Chen, H. V. Poor, W. Saad, and S. Cui, "Wireless communications for collaborative federated learning," *IEEE Commun. Mag.*, vol. 58, no. 12, pp. 48–54, Dec. 2020, doi: 10.1109/MCOM.001.2000397.

[3] C.-X. Wang, M. D. Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 16–23, Feb. 2020, doi: 10.1109/MWC.001.1900292.

[4] "Technical specification group services and system aspects; study on traffic characteristics and performance requirements for AI/ML model transfer in 5GS (Release 18)," 3GPP, Sophia Antipolis, France, Rep. TR 22.874 V1.0.0, 2021.

[5] C. Sun, S. Wu, and T. Cui, "User selection for federated learning in a wireless environment: A process to minimize the negative effect of training data correlation and improve performance," *IEEE Veh. Technol. Mag.*, vol. 17, no. 3, pp. 26–33, Sep. 2022, doi: 10.1109/MVT.2022.3153274.

[6] S. Wang, X. Zhang, H. Uchiyama, and H. Matsuda, "HiveMind: Towards cellular native machine learning model splitting," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 2, pp. 626–640, Feb. 2022, doi: 10.1109/JSAC.2021.3118403.

[7] B. Taylor, V. S. Marco, W. Wolff, Y. Elkhatib, and Z. Wang, "Adaptive deep learning model selection on embedded systems," in *Proc. 19th ACM SIGPLAN/SIGBED Int. Conf. Lang. Compil. Tools Embed. Syst.*, 2018, pp. 31–43, doi: 10.1145/3299710.3211336.

[8] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018, doi: 10.1109/COMST.2018.2846401.

[9] Y.-N. R. Li, B. Gao, X. Zhang, and K. Huang, "Beam management in millimeter-wave communications for 5G and beyond," *IEEE Access*, vol. 8, pp. 13282–13293, 2020, doi: 10.1109/ACCESS.2019.2963514.

[10] Q. Ren, Y. Zhu, Z. Xu, and Z. Pan, "Semi-conductor pillars in 5G era and beyond," in *Proc. IEEE Asian Solid-State Circuits Conf. (A-SSCC)*, 2022, pp. 8–11, doi: 10.1109/A-SSCC56115.2022.9980537.

[11] K. Ma, Z. Wang, W. Tian, S. Chen, and L. Hanzo, "Deep learning for mmWave beam-management: State-of-the-art, opportunities and challenges," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 108–114, Aug. 2023, doi: 10.1109/MWC.018.2100713.

[12] M. Qurratulain Khan, A. Gaber, P. Schulz, and G. Fettweis, "Machine learning for millimeter wave and terahertz beam management: A survey and open challenges," *IEEE Access*, vol. 11, pp. 11880–11902, 2023, doi: 10.1109/ACCESS.2023.3242582.

[13] N. Van Huynh, D. N. Nguyen, D. T. Hoang, and E. Dutkiewicz, "Optimal beam association for high mobility mmWave vehicular networks: Lightweight parallel reinforcement learning approach," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5948–5961, Sep. 2021, doi: 10.1109/TCOMM.2021.3088305.

[14] Y. Ju et al., "Deep reinforcement learning based joint beam allocation and relay selection in mmWave vehicular networks," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 1997–2012, Apr. 2023, doi: 10.1109/TCOMM.2023.3240754.

[15] D. Burghal, N. A. Abbasi, and A. F. Molisch, "A machine learning solution for beam tracking in mmWave systems," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 173–177, doi: 10.1109/IEEECONF44664.2019.9048770.

[16] M. Polese, F. Restuccia, and T. Melodia, "DeepBeam: Deep waveform learning for coordination-free beam management in mmWave networks," in *Proc. 22nd Int. Symp. Theory, Algorithm. Found., Protocol Design Mobile Netw. Mobile Comput.*, 2021, pp. 61–70, doi: 10.1145/3466772.3467035.

[17] W. Na, B. Bae, S. Cho, and N. Kim, "Deep-learning based adaptive beam management technique for mobile high-speed 5G mmWave networks," in *Proc. IEEE 9th Int. Conf. Consum. Electron. (ICCE)*, 2019, pp. 149–151, doi: 10.1109/ICCE-Berlin47944.2019.8966183.

[18] "Technical specification group radio access networks; study on artificial intelligence (AI)/machine learning (ML) for NR air interface (Release 18)," 3GPP, Sophia Antipolis, France, Rep. TR 38.843 V0.0.0, 2022.

[19] W. Chen, P. G. Gaal, J. Montojo, and H. Zisimopoulos, *Fundamentals of 5G Communications: Connectivity for Enhanced Mobile Broadband and Beyond*. New York, NY, USA: McGraw-Hill, 2021.

[20] C. Sun, J. Cheng, and T. Ohira, Eds., *Handbook on Advancements in Smart Antenna Technologies for Wireless Networks*. Pennsylvania, PA, USA: IGI Global, 2009, doi: 10.4018/978-1-59904-988-5.

[21] C. Sun, M. Taromaru, and T. Ohira, "Employing coordinated transmit and receive beamforming in clustering double-directional radio channel," *EURASIP J. Wireless Commun. Netw.*, 2007, Oct. 2007, Art. no. 057175, doi: 10.1155/2007/57175.

[22] J. Wang et al., "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009, doi: 10.1109/JSAC.2009.091009.

[23] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 173–196, 1st Quart., 2019, doi: 10.1109/COMST.2018.2869411.

[24] A. Chakrapani, "On the design details of SS/PBCH, signal generation and PRACH in 5G-NR," *IEEE Access*, vol. 8, pp. 136617–136637, 2020, doi: 10.1109/ACCESS.2020.3010500.

[25] P. Xue, Y. Huang, D. Zhu, Y. Zhao, and C. Sun, "Reducing the system overhead of millimeter-wave beamforming with neural networks for 5G and beyond," *IEEE Access*, vol. 9, pp. 165956–165965, 2021, doi: 10.1109/ACCESS.2021.3135903.

[26] K. Ma, D. He, H. Sun, and Z. Wang, "Deep learning assisted mmWave beam prediction with prior low-frequency information," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2021, pp. 1–6, doi: 10.1109/ICC42927.2021.9500788.

[27] K. Ma, H. Zou, C. Sun, and Z. Wang, "Deep learning assisted adaptive mmWave beam tracking: A sum-probability oriented methodology," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, 2022, pp. 573–578, doi: 10.1109/GLOBECOM48099.2022.10001216.

[28] J. Grus, *Data Science From Scratch*. Sebastopo, CA, USA: O'Reilly Media, 2019.

[29] "Discussion on evaluation on AI/ML for beam management," NTT Docomo, Inc., Chiyoda City, Japan, Rep. 3GPP-R1-2209898, Oct. 2022.

[30] "Evaluation on AI beam management," ZTE Corp., Shenzhen, China, Rep. 3GPP-R1-2306797, Aug. 2023.

[31] "Evaluations on AI/ML for beam management," Qualcomm Semiconduct. Co., San Diego, CA, USA, Rep. 3GPP-R1-2307918, Aug. 2023.

[32] D. Wolpert and W. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997, doi: 10.1109/4235.585893.

[33] "Evaluation for AI/ML beam management," Intel Semiconduct. Corp., Santa Clara, CA, USA, Rep. 3GPP-R1-2306856, Aug. 2023.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

CHEN SUN (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Nanyang Technological University, Singapore, in 2005. From August 2004 to May 2008, he was a Researcher with ATR Wave Engineering Laboratories, Japan, working on adaptive beamforming and direction-finding algorithms of parasitic array antennas as well as theoretical analysis of cooperative wireless networks. In June 2008, he joined the National Institute of Information and Communications Technology, Japan, as an Expert Researcher working on distributed sensing and dynamic spectrum access in TV white space. Since then, he has been contributing to IEEE 1900.6 standard, IEEE 802.11af standard, and Wi-Fi Alliance specifications for Wi-Fi networks in TV white space. In 2012, he joined Sony China as a Research Manager working on IEEE 802.19, ETSI, and 3GPP standards development. He served as the 802.19 standards and rapporteur of ETSI standards. He is currently the Deputy Head of the Beijing Lab, Sony R&D Center. He is the author of *Handbook on Advancements in Smart Antenna Technologies for Wireless Networks*, 40 international journals, and more than 110 conference papers. He is also the first inventor of 80 granted patents in U.S., EU, Japan, and China, including nine patents that are worth more than four million dollars. His research interests include AI for beamforming, wireless federated learning, blockchain-based spectrum sharing, wireless sensing, and V2X. He received the IEEE Standards Association IEEE 1900.6 Working Group Chair Award for leadership in 2011, the IEEE 802.11af Outstanding Contributions Award in 2014, and the IEEE 802.19.1 Outstanding Contributions Award in 2018. He served as the Technical Editor for the IEEE1900.6 standards and a Contributor for IEEE 802.11af standards.

LE ZHAO received the B.S. degree in electronic information engineering and the M.S. degree in communication engineering from the Beijing Institute of Technology, Beijing, China, in 2020 and 2023, respectively, where he is currently pursuing the Ph.D. degree in communication engineering with the School of Information and Electronics. His research interests include integrated sensing and communication, beam management, reinforcement learning/deep learning for communication, machine learning, and communication networking.

TAO CUI received the B.Sc. degree in electronics and information engineering from Tianjin University, Tianjin, China, in 2012, and the M.S. degree in communication engineering from the University of Alberta, Edmonton, Canada, in 2014. He joined Datang Mobile as a Software Engineer on the PHY layer in 2015, and he was responsible for 5G prototype development. In September 2016, he joined the Beijing Lab, Sony R&D Center, as a frontend 3GPP delegate working on the standardization of Rel-15 LTE-LAA and Rel-16 NR unlicensed spectrum operation. He has currently filed more than 40 patents (two of them have been granted) and over 20 Tdocs contributed to the 3GPP RAN1 working group. He is currently leading the Sony China C-V2X Project from standardization to the interoperability test. His current research interests include vehicular ad hoc networks, C-V2X-based message layer design, and intrinsic AI for wireless networks.

HAOJIN LI received the M.S. degree in information and communication engineering from the University of China Academy of Telecommunication Technology, Beijing, China, in 2020. He joined Sony China Research Lab as a 5G Researcher in 2020, and he was responsible for 5G system-level simulator development from then on. In June 2022, he was promoted as a Deputy Principal Research and Development Engineer. His research interests include nonterrestrial networks as well as integrated communication and sensing.

YINGSHUANG BAI received the M.S. degree in information and communication engineering from Harbin Engineering University, Harbin, China, in 2020. She joined Xiaomi in April 2020 as a Standard Researcher, where her contributions primarily revolved around the advancement of New Radio link simulation techniques. In August 2022, she joined Sony China Research Lab, where her research focuses on the domain of AI for wireless, with a specific emphasis on innovations within the physical layer.

SONGTAO WU received the B.Sc. degree in electronic information engineering from Lanzhou University, Lanzhou, China, in 2009, the M.S. degree in circuits and systems from Peking University, Beijing, China, in 2012, and the Ph.D. degree from the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, in 2017. He is currently a Researcher with the AI Research Group, Beijing Lab, Sony R&D Center. His research interests include machine/deep learning, model compression, and federated learning in wireless environments.

QIANG TONG received the B.Sc. degree in automation engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2006, and the M.S. degree and the Ph.D. degree in information sciences from Tohoku University, Sendai, Japan, in 2015 and 2018, respectively. He is currently a Researcher with the AI Research Group, Beijing Lab, Sony R&D Center, China. His research interests include computer vision, machine/deep learning, IoT sensing, edge-AI, and federated learning in wireless environments.