AI Based Beam Management for 5G (mmWave) at Wireless Edge: Opportunities and Challenges

Chitwan Arora
Hughes Systique Corporation
Gurgaon, India
e-mail: chitwan.arora@hsc.com

***Abstract*—Fast and efficient beam management mechanism is the key enabler in 5G (millimeter wave) to achieve low latency and high data rate requirements. Recent advances in Artificial Intelligence (AI) have shown that Machine Learning (ML) and Deep Learning (DL) based techniques can play a significant role in efficient beam management. These techniques can continuously learn and adapt themselves based on the highly varying traffic and channel conditions. For effective operation, it is essential that the ML and DL based beam management algorithm should be deployed at the place in network where all the relevant input parameters needed for beam management are available continuously, as well as the output of the beam management can be applied instantly. In this paper, advantages along with challenges of deploying ML and DL based beam management techniques at the wireless edge of 5G networks are explored.**

Keywords-mmWave; beam management; artifical intelligence; wireless edge.

I. INTRODUCTION

The millimeter wave (mmWave) frequencies offer the availability of huge bandwidths to provide unprecedented data rates to meet the demand for Fifth Generation (5G) applications. However, mmWave links are highly susceptible to rapid channel variations and suffer from severe free space pathloss and atmospheric absorption. To address these challenges, base stations and mobile terminals use highly directional antennas to achieve enough link budget in wide area networks. Directional links, however, require fine alignment of the transmitter and receiver beams, achieved through a set of operations known as **beam management**. They are fundamental to the performance of a variety of control tasks including (i) **Initial Access (IA)** for idle users, which allows a mobile User Equipment (UE) to establish a physical link connection with a gNB (5G base station), and (ii) **Beam tracking**, for connected users, which enables beam adaptation schemes, or handover, path selection and radio link failure recovery procedures [1][2]. Figure 1 captures the details of the beam management procedure for 5G Stand Alone (SA) scheme. In existing Long-Term Evolution (LTE) systems (using spectrum in 3-5 GHz), these control procedures are performed using omnidirectional signals, and beamforming or other directional transmissions can only be performed after a physical link is established, for data plane transmissions. On the other hand, in the mmWave bands, it is essential to exploit the antenna gains even during initial access and, in general, for control operations. Hence, there is a need for precise alignment of the transmitter and the receiver beams.

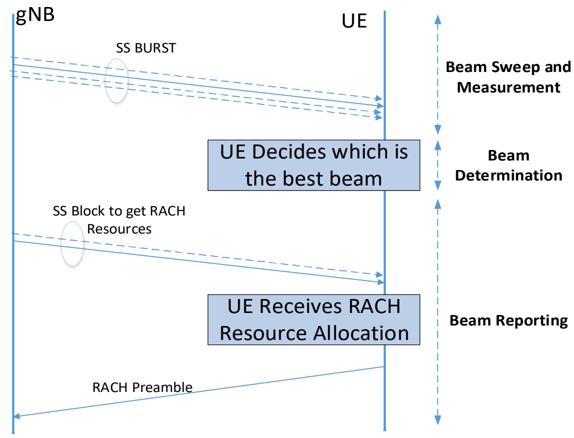


Figure 1. 5G Stand Alone beam management procedure

The initial access in 5G millimeter wave is a time-consuming search to determine suitable directions of transmission and reception. In the cell discovery phase, one approach is sequential beam sweeping by the base station that requires a brute force search through many beam-pair combinations between UE and gNB to find the optimum beam-pair i.e., the one with the highest Reference Received Signal Power (RSRP) level, as shown in Figure 2. The sequential search may result in a large access delay and low initial access efficiency. It also consumes a fair amount of energy in the receiver, which makes it unsuitable for energy constrained receivers, such as Internet of Things (IoT) endpoints.

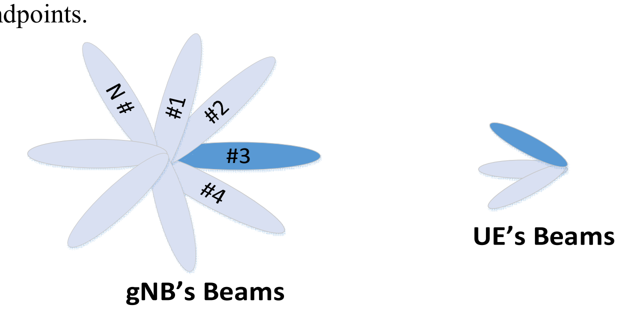


Figure 2. Sequential Beam Sweeping

 ISBN: 978-1-61208-727-6

In existing LTE systems, DL channel quality is estimated from an omnidirectional signal called the Cell Reference Signal (CRS) [3] for beam alignment and selection in connected state. CRS is regularly monitored by each UE in connected state to create a wideband channel estimate that can be used both for demodulating downlink transmissions and for estimating the channel quality [4]. In 5G mmWave networks, in addition to the rapid variations of the channel, CRS-based estimation is challenging due to the directional nature of the communication, thus requiring the network and the UE to constantly monitor the direction of transmission of each potential link. Tracking changing directions can decrease the rate at which the network can adapt and can be a major obstacle in providing robust and ubiquitous service in the face of variable link quality. In addition, UE and gNB may only be able to listen to one direction at a time, thus making it hard to receive the control signaling necessary to switch paths.

From the above description, it is apparent that 5G networks should support a mechanism by which the users and the infrastructure can quickly determine the best directions to establish the mmWave links. These are particularly important issues in 5G networks and motivate the need to extend current LTE control procedures with innovative mmWave-aware beam management algorithms and methods.

In this paper, we explore various traditional as well as upcoming ML and DL based techniques for minimizing the latency and the overhead of the initial communication process. It has been observed that online DL based techniques give better performance than offline DL based techniques. Online DL techniques efficiently adapt themselves to support high mobility in mmWave systems. Deployment strategies for the training of these deep learning algorithms are explored in this paper and we propose that the wireless edge is the appropriate place for the deployment of these DL based algorithm for beam management.

The remainder of this paper is organized as follows. Section II discusses the literature survey of traditional (non-ML/DL) beam management techniques, as well as ML/DL based mean management techniques. Section III discusses in detail different ML/DL based beam management techniques. Section IV discusses the deployment strategy of the deep learning-based beam forming algorithm and Section V presents the conclusions.

II. LITERATURE SURVEY

In this section, work related to traditional (Non-ML/DL) and ML/DL based beam management is presented.

Traditional (Non-ML/DL) based beam management: Several approaches for directional based schemes have been proposed in the literature to enable efficient control procedures for both the idle and the connected mobile terminals. Most literature on Initial Access and tracking refers to challenges that have been analyzed in the past at lower frequencies in ad hoc wireless network scenarios or, more recently, referred to the 60 GHz IEEE 802.11ad WLAN and WPAN scenarios (e.g., [5]-[7]). However, most of the proposed solutions are unsuitable for next-generation cellular network requirements and present many limitations (e.g., they are appropriate for short range, static and indoor scenarios, which do not match well the requirements of 5G systems). In [8][9], the authors propose an exhaustive method that performs directional communication over mmWave frequencies by periodically transmitting synchronization signals to scan the angular space. The result of this approach is that the growth of the number of antenna elements at either the transmitter or the receiver provides a large performance gain compared to the case of an omnidirectional antenna. However, this solution leads to a long duration of the Initial Access with respect to LTE, and poorly reactive tracking.

Similarly, in [10], measurement reporting design options are compared, considering different scanning and signaling procedures, to evaluate access delay and system overhead. The channel structure and multiple access issues are also considered. The analysis demonstrates significant benefits of low-resolution fully digital architectures in comparison to single stream analog beamforming. More sophisticated discovery techniques are reported in [11][12] to alleviate the exhaustive search delay through the implementation of a multi-phase hierarchical procedure based on the access signals being initially sent in few directions over wide beams, which are iteratively refined until the communication is sufficiently directional. In [13], a low-complexity beam selection method by low-cost analog beamforming is derived by exploiting a certain sparsity of mmWave channels. It is shown that beam selection can be carried out without explicit channel estimation, using the notion of compressive sensing. The issue of designing efficient beam management solutions for mmWave networks is addressed in [14], where the author designs a mobility-aware user association strategy to overcome the limitations of the conventional power-based association schemes in a mobile 5G scenario.

Other relevant papers on this topic include [15], in which the authors propose smart beam tracking strategies for fast mmWave link establishment. The algorithm proposed in [16] takes into account the spatial distribution of nodes to allocate the beam width of each antenna pattern in an adaptive fashion and satisfy the required link budget criterion. Since the proposed algorithm minimizes the collisions, it also minimizes the average time required to transmit a data packet from the source to the destination through a specific direction. In 5G scenarios, papers [8][9][11] give some insights on trade-offs among different beamforming architectures in terms of user communication quality. Articles [17][18] evaluate the mmWave cellular network performance while accounting for the beam training, association overhead and beamforming architecture. The results show that, although employing wide beams, initial beam training with full pilot reuse is nearly as good as perfect beam alignment.

ML/DL based beam management: The recent progress in Machine learning and Deep Learning has raised interest in applying these techniques to communication system related problem [19] – [25]. On the same line of thought as traditional beam management approaches, data-driven Deep Learning-based approaches have been used for efficient beam management. The key idea is that ML/DL is used to make recommendations of promising beam pairs based on the various system parameters as well as past beam measurements.

Papers [26] - [28] propose beam alignment techniques using Machine Learning. Position-aided beam prediction was proposed in [26][27]. Decision tree learning was used in [26], and a learning to rank method was used in [27]. The work in [26] - [28] shows that machine learning is valuable for mmWave beam prediction. A more exhaustive survey is provided in the next section.

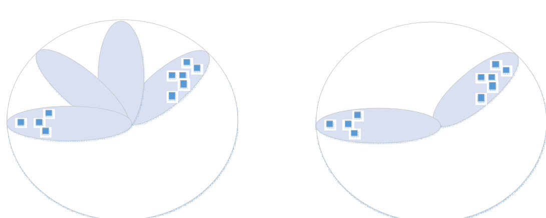
III. Insight of ML/DL Based Beam Management Techniques

This section captures the detailed analysis of challenges related to Beam sweeping, Beam alignment and Beam selection using ML/DL based techniques.

A. *Beam Sweeping*

There are various papers which focus on predicting the proposed **Beam sweeping** pattern based on the dynamic distribution of user traffic. In [29], a form of Recurrent Neural Networks (RNNs) called a Gated Recurrent Unit (GRU) has been proposed. In this paper, the spatial distribution of users is inferred from data in Call Detail Records (CDRs) of the cellular network. Results show that the user's spatial distribution and their approximate location (direction) can be accurately predicted based on CDRs data using Gated Recurrent Unit (GRU), which is then used to calculate the sweeping pattern in the angular domain during cell search. In [30] beam sweeping pattern based on GRU is compared with random starting point sweeping to measure the synchronization delay distribution. Results shows that this deep learning beam sweeping pattern prediction enables the UE to initially assess the gNB in approximately 0.41 of a complete scanning cycle with probability 0.9 in a sparsely distributed UE scenario.

Figure 3 shows that, in the sparsely distributed UE scenario, DL based techniques can help to reduce the number of beams to be traversed during beam sweeping. As a result, it will reduce the sweeping time drastically.



(a) (b)

Figure 3. Beam Sweeping in Sparsely distributed UE Scenario

B. *Beam Alignment and Selection*

Position-Aided: Position information may be leveraged for fast beam alignment in mmWave systems. Inverse fingerprinting is one approach to exploit position information [31], which works in Non-Line-of-Sight (NLOS) channels. There are multiple research papers [32]-[34] which focus on using machine learning to propose beam pairs based on the location of the UE position relative to the gNB and past beam measurements. The UE location and past beam measurements can be input into a learning algorithm that learns to rank promising beam directions. By prioritizing beam training in top-ranked directions, the training overhead can be reduced. Figure 4 shows the steps of beam management based on Position Information.

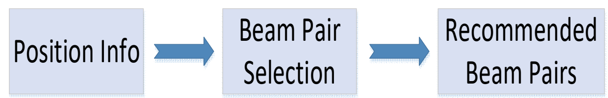


Figure 4. Beam Management based on Position Information

Paper [34] proposes UE positions-based beam alignment in the context of vehicular communication. The authors state that this inverse fingerprinting method is efficient. However, these approaches have some limitations. First, the approach is offline, which means its use is delayed until the database is collected. Second, also due to being offline, its performance depends entirely on the accuracy of the collected database, which may become stale over time. To overcome these shortcoming, online approaches have been proposed. In the online approaches, it has been proposed to keep collecting new observations during operation, making it possible to improve the database.

Situational Awareness: Machine learning tools combined with awareness of the proximity situation have been proposed in [35] to learn the beam information (power, optimal beam index, etc.) from past observations. In this paper, situational awareness that is specific to the vehicular setting including the locations of the receiver and the surrounding vehicles has been considered. The result shows that situational awareness along with machine learning can largely improve the prediction accuracy and the model can achieve throughput with little performance loss and almost zero overhead.

Coordinated Beamforming: A coordinated beamforming solution using deep learning was proposed in [36]. In this paper, the received training signals via omni reception at a set of coordinating Base Stations (BSs) are used as the input to a deep learning model that predicts the beamforming vectors at those BSs to serve a single user. These coordinated beamforming deep learning techniques are based on supervised learning techniques, which assume an offline learning setting and require a separate training data collection phase. However, there are papers which focus on online learning algorithms using the Multi Armed

 ISBN: 978-1-61208-727-6

Bandit (MAB) framework, which is a special class of Reinforcement Learning (RL).

IV. DEPLOYMENT STRATEGY AT WIRELESS EDGE

From the above studies we can see that ML/DL leverages a large amount of data samples (e.g., radio signals) to acquire accurate knowledge of the RF environment to have optimum beam management. However, the majority of the works presented above focus on centralized ML/DL (as shown in Figure 5), whose goal is to improve the communication performance assuming a well-trained ML model as well as full access to a global dataset. It also assumes massive amounts of storage and computing power are available.

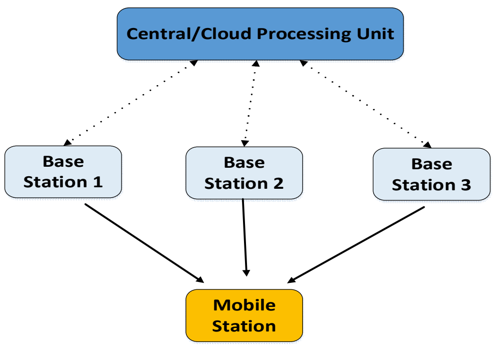


Figure 5. Centralized Deployment of ML/DL Algorithms

However, these approaches have overlooked the additional latency induced by the prior training process and the posterior inference latency. Along with that, for highly varying channel conditions, we need to regularly provide the updated input information to the ML/DL based model.

In this paper, we propose a deployment of ML/DL based algorithm for optimal beam management as a distributed solution, leveraging the Mobile Edge architecture. As we shall show, there will be numerous advantages if we deploy the ML/DL model in a more distributed way (i.e. at Wireless Edge) instead of centralized ML/DL (i.e. at the cloud), as captured in Figure 6.

In this deployment, we have assumed that the Wireless Edge will be present near to gNB. As a result, Wireless Edge will have immediate access to all the relevant data i.e. RF related data, Channel specific data, Cell specific data and User specific data. This will help to use the online learning model which will continuously train itself based on the latest UE and channel information received.

gNBs interact with each other and can have access to relevant information from the neighboring gNBs. These inputs will boost the performance of situational based and coordinated DL/ML model deployed at the wireless edge, as these models can make decisions based on the overall environmental conditions i.e., interference as well as other neighboring gNB parameters. The wireless edge can interact with central/cloud processing unit for exchanging the common information to all the gNBs.

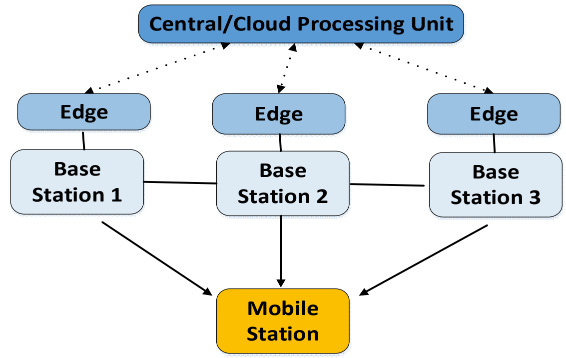


Figure 6. Deployment of ML/DL Algorithms at Wireless Edge

Based on the above description, some of the key advantages of the deployment of a DL\ML based algorithm at wireless edge are as follows:

(i) Performing inference at wireless edge reduces latency and cost of sending data to the cloud for prediction.

(ii) Rather than sending all data to the cloud for performing ML inference, inference is run directly at the wireless Edge device, and data is sent to the cloud only when additional processing is required.

(iii) Every wireless edge entity will have access to a fraction of the data and training and inference are carried out collectively. Moreover, edge devices communicate and exchange their locally trained models, instead of exchanging their private data.

(iv) Since inference results will be available with very low latency, better beam management performance will be achieved in highly mobile and dynamically changing environment conditions.

(v) Since data is present locally at the edge and not going to the cloud, it will enhance the overall reliability as well as privacy.

(vi) Higher inference accuracy can be achieved by training with a wealth of user-generated data e.g., location history, network operational status, etc.

However, there are certain challenges in deploying the ML/DL based algorithms at wireless edge, as follows:

(i) There is a lack of authentic set of data from real communication systems or prototype platforms in actual physical environments. So far, simulations results [32][33][36] prove that the recently proposed DL-based communication algorithms demonstrate a competitive performance. However, due to the lack of standardized data, benchmarking the performance is a real challenge.

ISBN: 978-1-61208-727-6

(ii) In the wireless edge-based ML/DL deployment, training data might be distributed at different wireless edge nodes and a given wireless edge node might have access to a fraction of the training data. Hence, in wireless edge based deployment, each edge device first trains the local model using its own data samples, and then exchanges the trained local model parameters among other wireless edges. Also, it is difficult to characterize the convergence behavior as well as model performance (i.e., whether the trained model is overfitted or underfitted) due to the distributed nature of the data. As a result, the complexity of networks and training phases will be increased in edge-based ML/DL deployment.

CONCLUSION

From the analysis mentioned above, we can say that emerging DL/ML based techniques can be used for efficient beam management in 5G mmWave. These AI based algorithms deployed at wireless edge can help in providing high performing networks and services that can handle data in a much more secure and faster way for 5G.

REFERENCES

[1] M. Giordani and M. Zorzi, “Improved user tracking in 5G millimeter wave mobile networks via refinement operations,” in 16th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), pp. 1-8, June 2017.

[2] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, “Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks,” IEEE Journal on Selected Areas in Communications, vol. 35, no. 9, pp. 2069–2084, September 2017.

[3] S. Schwarz, C. Mehlfuhrer, and M. Rupp, “Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE,” in 6th conference on Wireless advanced (WiAD). IEEE, pp. 1–6, 2010.

[4] M. Giordani, M. Mezzavilla, A. Dhananjay, S. Rangan, and M. Zorzi,“Channel dynamics and SNR tracking in millimeter wave cellular systems,” in 22th European Wireless Conference. VDE, pp. 1-8, 2016.

[5] T. Nitsche et al., “IEEE 802.11ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi [Invited Paper],” IEEE Communications Magazine, vol. 52, no. 12, pp. 132–141, December 2014.

[6] J. Wang, “Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems,” IEEE Journal on Selected Areas inCommunications, vol. 27, no. 8, pp. 1390–1399, October 2009.

[7] R. Santosa, B.-S. Lee, C. K. Yeo, and T. M. Lim, “Distributed Neighbor Discovery in Ad Hoc Networks Using Directional Antennas,” in The Sixth IEEE International Conference on Computer and Information Technology, September 2006, pp. 97–97.

[8] C. Jeong, J. Park, and H. Yu, “Random access in millimeter-wave beamforming cellular networks: issues and approaches,” IEEE Communications Magazine, vol. 53, no. 1, pp. 180–185, January 2015.

[9] C. N. Barati et al., “Directional cell discovery in millimeter wave cellular networks,” IEEE Transactions on Wireless Communications, vol. 14, no. 12, pp. 6664–6678, December 2015.

[10] C. N. Barati et al., “Directional initial access for millimeter wave cellular systems,” in 49th Asilomar Conference on Signals, Systems and Computers. IEEE, 2015, pp. 307–311.

[11] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, “Initial beamforming for mmWave communications,” in 48th Asilomar Conference on Signals, Systems and Computers, 2014, pp. 1926–1930.

[12] L. Wei, Q. Li, and G. Wu, “Exhaustive, Iterative and Hybrid Initial Access Techniques in mmWave Communications,” in 2017 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2017.

[13] J. Choi, “Beam selection in mm-Wave multiuser MIMO systems using compressive sensing,” IEEE Transactions on Communications, vol. 63, no. 8, pp. 2936–2947, August 2015.

[14] A. S. Cacciapuoti, “Mobility-Aware User Association for 5G mmWave Networks,” IEEE Access, vol. 5, pp. 21 497–21 507, 2017.

[15] J. Palacios, D. De Donno, and J. Widmer, “Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility,” in IEEE Conference on Computer Communications (INFOCOM). IEEE, 2017.

[16] K. Chandra, R. V. Prasad, I. G. Niemegeers, and A. R. Biswas, “Adaptive beamwidth selection for contention based access periods in millimeter wave WLANs,” in IEEE 11th Consumer Communications and Networking Conference (CCNC). IEEE, 2014, pp. 458–464.

[17] A. Alkhateeb, Y. H. Nam, M. S. Rahman, J. Zhang, and R. W. Heath, “Initial Beam Association in Millimeter Wave Cellular Systems:Analysis and Design Insights,” IEEE Transactions on Wireless Communications, vol. 16, no. 5, pp. 2807–2821, May 2017.

[18] Y. Li, J. Luo, M. Castaneda, R. Stirling-Gallacher, W. Xu, and G. Caire,“On the Beamformed Broadcast Signaling for Millimeter Wave Cell Discovery: Performance Analysis and Design Insight,” arXiv preprint arXiv:1709.08483, 2017.

[19] S. D¨orner, S. Cammerer, J. Hoydis, and S. ten Brink. “Deep learning-based communication over the air,” [Online]. Available: https://arxiv.org/abs/1707.03384, 2017.

[20] U. Challita, L. Dong, and W. Saad, “ Proactive resource management in LTE-U systems: A deep learning perspective,” https://arxiv.org/abs/1702.07031, 2017.

[21] R. C. Daniels, C. M. Caramanis, and R. W. Heath, “Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering,” IEEE Trans. Veh. Technol., vol. 59, no. 1, pp. 114–126, January 2010.

[22] S. K. Pulliyakode and S. Kalyani, “Reinforcement learning techniques for outer loop link adaptation in 4G/5G systems,” https://arxiv.org/abs/1708.00994, 2017.

[23] A. Fehske, J. Gaeddert, and J. H. Reed, “A new approach to signal classification using spectral correlation and neural networks,” in Proc. IEEE Int. Symp. New Frontiers in Dynamic Spectrum Access Networks (DYSPAN), 2005, pp. 144–150.

[24] E. E. Azzouz and A. K. Nandi, “Modulation recognition using artificial neural networks,” in Proc. Automatic Modulation Recognition of Communication Signals, 1996, pp. 132–176.

[25] M. Ibukahla, J. Sombria, F. Castanie, and N. J. Bershad, “Neural networks for modeling nonlinear memoryless communication channels,” IEEE Trans. Commun., vol. 45, no. 7, pp. 768–771, July 1997.

[26] Y. Wang, M. Narasimha, and R. W. Heath Jr, “MmWave beam prediction with situational awareness: A machine learning approach,” https://arxiv.org/abs/1805.08912, June 2018, in Proc. of IEEE SPAWC.

[27] V. Va, T. Shimizu, G. Bansal, and R. W. Heath Jr., “Position-aided millimeter wave V2I beam alignment: A learning-to-

rank approach," in Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.,October 2017, pp. 1–5.

[28] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," IEEE Access, vol. 6, pp. 37 328–37 348, June 2018.

[29] A. Mazin, M. Elkourdi, and R. D. Gitlin, "Accelerating beam sweeping in mmWave standalone 5G new radios using recurrent neural networks," in 2018 IEEE Vehicular Technology Conference (VTC), 2018. [Online]. Available: https://arxiv.org/abs/1809.01096

[30] A. Mazin, M. Elkourdi, and R. D. Gitlin, "Comparative Performance Analysis of Beam Sweeping Using a Deep Neural Net and Random Starting Point in mmWave 5G New Radio," in 9th IEEE Annual Ubiquitous Computing, Electronic and Mobile communication conference (UEMCON), 2018.

[31] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath Jr., "Inverse multipath fingerprinting for millimeter wave V2I beam alignment," IEEE Trans. Veh. Technol., vol. 67, no. 5, pp. 4042–4058, May 2018.

[32] A Klautau, P. Batista, N. Gonzalez-Prelcic, Y. Wang, and R. W. Heath Jr "5G MIMO Data for Machine Learning: Application to Beam-Selection using Deep Learning" Proc of the information theory and application workshop, February 2018.

[33] V.Shimizu, G. Bansal, and R. W. Heath, "Position-aided millimeter wave V2I beam alignment: A learning-to-rank approach", IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017.

[34] V. Shimizu, G. Bansal and R. W. Heath, "Online Learning for Position-Aided Millimeter Wave Beam Training," in IEEE Access, vol. 7, pp. 30507-30526, 2019.

[35] Y. Wang, M. Narasimha and R. W. Heath, "MmWave Beam Prediction with Situational Awareness: A Machine Learning Approach," 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, 2018, pp. 1-5.

[36] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu and D. Tujkovic, "Deep Learning Coordinated Beamforming for Highly-Mobile Millimeter Wave Systems," in *IEEE Access*, vol. 6, pp. 37328-37348, 2018.