Online Beam Learning for Interference Nulling in Hardware-Constrained mmWave MIMO Systems

Yu Zhang and Ahmed Alkhateeb

2022 56th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5906-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEEECONF56349.2022.10051931

***Abstract*—Employing large antenna arrays is a key characteristic of millimeter wave (mmWave) and terahertz communication systems. Due to the hardware constraints and the lack of channel knowledge, codebook based beamforming/combining is normally adopted to achieve the desired array gain. Existing codebooks, however, are typically pre-defined and focus only on improving the beamforming gain of their target user, without taking interference into account, which incurs critical performance degradation. In this paper, we propose an efficient deep reinforcement learning approach that learns how to iteratively optimize the beam pattern to null the interference. The proposed solution achieves that while not requiring any explicit channel knowledge of the desired or interfering users and without requiring any coordination with the interferers. Simulation results show that the developed solution is capable of finding a well-shaped beam pattern that significantly suppresses the interference while sacrificing negligible beamforming/combing gain, highlighting a promising solution for dense mmWave/terahertz networks.**

I. Introduction

Deploying large number of antennas is crucial in enabling millimeter wave (mmWave) and terahertz (THz) communications. By applying beamforming/combining, mmWave/THz systems are able to combat the severe path loss incurred in the high frequency bands and hence provide sufficient receive signal power. To reduce the high cost and power consumption of the mixed-circuit components, on the one hand, these systems start to seek either fully analog or hybrid architecture to achieve such potential [1]. On the other hand, the adoption of such architectures also introduces several difficulties in the following signal processing, one of which is channel estimation. As a result, pre-defined codebooks (such as beamsteering codebooks) are normally used for both initial access and data transmission. Being pre-defined, however, those beams are normally designed in a way that focuses solely on improving the beamforming/combining gain from specific directions, without taking interference into account. This raises issues in situations where there are interfering users in the surrounding, communicating at the same time-frequency slots. Those "interference-unaware" beams might incur severe interference from other users, which could possibly degrade the system performance to a great extent. Therefore, an ideal beam pattern design algorithm should be able to strike a balance between the desired user and interfering users, targeting the signal-to-interference-plus-noise ratio (SINR) as its final objective, which is the focus of this paper.

Yu Zhang and Ahmed Alkhateeb are with Arizona State University (Email: y.zhang, alkhateeb@asu.edu). This work is supported by the National Science Foundation under Grant No. 1923676.

Prior work: Designing analog beamforming or hybrid precoding for MIMO systems has been an important topic for quite some time [2]–[6]. In fully analog systems, pre-defined beamsteering codebooks are normally adopted to simplify the design process, as in [2], [3], where sharp and directional beams are used for acquiring the desired beamforming/combining gains. In [4], a neural network based analog beam codebook design approach is proposed, although the authors focus only beamforming gain and the potential interference problem is ignored. In hybrid analog/digital systems, a variety of hybrid precoding strategies are proposed to improve the multiplexing and sum-rate performance, as in [5], [6]. However, the common problems in these solutions include the need for either complete or partial channel knowledge of all the user channels, as well as the relaxation of the quantized analog phase shifter constraints during the design stage.

Contribution: In this paper, we propose a deep reinforcement learning based beam pattern design framework that can efficiently adapt the beam pattern to avoid interference from surroundings while maximizing the beamforming/combining gain of the desired user. This is done by not requiring the channel knowledge of both target user and the interferers, and by only relying on the power measurements. The proposed framework also respects the key hardware constraints such as quantized phase shifter constraint, making it a hardware compatible solution. Simulation results show that the proposed solution is capable of forming a beam pattern that can strike a balance between the beamforming/combining gain of the target user and the suppression gain of the surrounding interferers.

II. System and Channel Models

Our objective in this paper is to investigate the design of interference-aware beam patterns. To study this problem, we consider a communication system where a mmWave MIMO base station (BS), equipped with M antennas, is communicating with a single-antenna user equipment (UE). Moreover, we assume that there exists K (≥ 1) non-cooperative interference transmitters[1] in the vicinity of the BS, operating at the same frequency bands and hence causing inevitable interference to the considered communication link. More specifically, we consider the uplink transmission where the BS will receive signal from the UE, together with the interference signals transmitted from the interference transmitters. Therefore, if the UE transmits a symbol $x \in \mathbb{C}$ to the BS, and the other K

[1]For ease of exposition, each interference transmitter is also assumed to have single-antenna. This means that the interference signals are being transmitted omni-directionally.

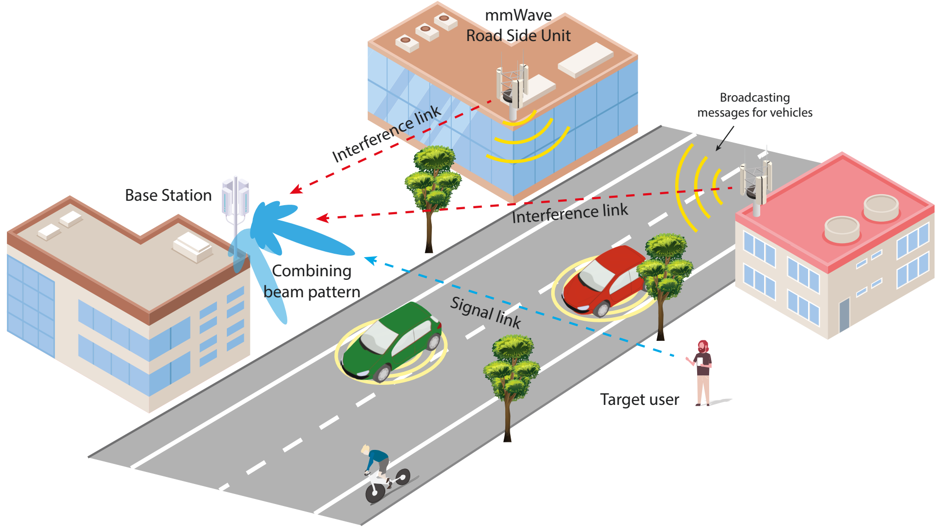


Fig. 1. The considered scenario where a mmWave base station is communicating with its target user under the presence of non-cooperative interference transmitters. This could be the case, for instance, where the mmWave road side units of a vehicular network are broadcasting traffic messages to the surrounding vehicles, which interferes the civilian data communication link, as depicted in the figure.

interference transmitters also transmit symbols $x_k \in \mathbb{C}, k = 1, \dots, K$ at the same time and frequency slot, such that all the transmitted symbols satisfy the same average power constraint, i.e., $\mathbb{E}[|x|^2] = P_x$ and $\mathbb{E}[|x_k|^2] = P_x, \forall k$, the received signal at the BS after combining can then be expressed as

$$y = \mathbf{w}^H \mathbf{h} x + \sum_{k=1}^{K} \mathbf{w}^H \mathbf{h}_k x_k + \mathbf{w}^H \mathbf{n}, \tag{1}$$

where $\mathbf{h} \in \mathbb{C}^{M \times 1}$ is the channel between the BS and the UE, $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$ is the channel between the BS and the k-th interference transmitter. It is worth pointing out here that for clarity, we subsume the factors such as path-loss and transmission power into the channels. $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ is the receive noise vector at the BS with σ^2 being the noise power and $\mathbf{w} \in \mathbb{C}^{M \times 1}$ is the combining vector used by the BS. Furthermore, given the high cost and power consumption of the mixed-signal components, we consider a practical system where the BS has only one radio frequency (RF) chain[2] and employs analog-only beamforming/combining using a network of r-bit quantized phase shifters. Therefore, the combining vector at the BS can be written as

$$\mathbf{w} = \frac{1}{\sqrt{M}} \left[e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_M} \right]^T, \tag{2}$$

where each phase shift $\theta_m, \forall m = 1, \dots, M$ is selected from a finite set $\mathbf{\Psi}$ with 2^r possible discrete values drawn uniformly from $(-\pi, \pi]$. The normalization factor $M^{-1/2}$ is to make sure the combiner has unit power, i.e., $\|\mathbf{w}\|_2^2 = 1$.

We adopt a narrowband geometric channel model for the channel between BS and UE, as well as the interference channel between BS and any interferer. Hence, the channel between BS and its served UE takes the following form[3]

$$\mathbf{h} = \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}(\phi_\ell, \vartheta_\ell), \tag{3}$$

where we assume that the signal propagation between BS and UE consists of L multi-paths. Each path ℓ has a complex gain α_ℓ, which subsumes the factors such as path-loss, transmission power, etc. The angles ϕ_ℓ and ϑ_ℓ represent the ℓ-th path's azimuth and elevation angles of arrival respectively, and $\mathbf{a}(\phi_\ell, \vartheta_\ell)$ is the array response vector of the considered BS to the signal with such arriving angles. The exact expression of $\mathbf{a}(\phi_\ell, \vartheta_\ell)$ depends on the array geometry and possible hardware impairments.

III. Problem Formulation

In this paper, we investigate the design of analog combiner that achieves interference awareness without knowing the explicit channel knowledge.[4] Given the receive signal (1) at BS, the achievable rate of its target UE can be written as

$$R = \log_2 \left(1 + \frac{|\mathbf{w}^H \mathbf{h}|^2 P_x}{\sum_{k=1}^{K} |\mathbf{w}^H \mathbf{h}_k|^2 P_x + \sigma^2} \right). \tag{4}$$

The objective is to design the combining vector $\mathbf{w}$ such that the achievable rate of the target user, i.e., (4), can be maximized. This is equivalent to maximize the SINR term in

[2]It is very important to note that the RF precoder in a system with hybrid architecture is normally constructed using pre-defined codebooks that have pre-determined beams. Therefore, the learned beams in this paper can be included in such codebooks and be used in the hybrid analog/digital architectures as well.

[3]The channel between BS and any interference transmitter takes similar form.

[4]It is very important to note that the proposed interference-aware beam pattern learning approach in this paper can be straightforwardly extended to learning a codebook with multiple beams, by using the user clustering and assignment algorithm proposed in [7].

(4). Therefore, the problem of designing interference-aware beam pattern can be cast as

$$\mathbf{w}^\star = \arg\max_{\mathbf{w}} \frac{|\mathbf{w}^H \mathbf{h}|^2 P_x}{\sum_{k=1}^{K} |\mathbf{w}^H \mathbf{h}_k|^2 P_x + \sigma^2}, \tag{5}$$

$$\text{s. t. } w_m = \frac{1}{\sqrt{M}} e^{j\theta_m},\ \forall m = 1, \dots, M, \tag{6}$$

$$\theta_m \in \boldsymbol{\Psi},\ \forall m = 1, \dots, M, \tag{7}$$

where w_m is the m-th element of the combining vector $\mathbf{w}$. The interference-aware beam pattern design problem formulated in (5) has several defining characteristics: (i) The constraint (6) requires constant-modulus on all the elements of the combining vector, which is a non-convex constraint, (ii) to respect the discrete phase shifter hardware constraint, w_m can only take finite number of values based on all the possible phase shifts given by (7), (iii) the target UE's channel $\mathbf{h}$ **is assumed to be unknown**, due to the difficulties encountered in the CSI acquisition in a practical mmWave system with fully-analog transceiver architectures, (iv) the possible hardware impairments are also assumed to be unknown, and (v) the channels of the interference transmitters, i.e., $\mathbf{h}_k, \forall k$, **are also unknown**, which is mainly because normally there is no coordination between the interference transmitters and the considered BS receiver.

Given all these aforementioned difficulties, (5) is very hard to be solved using the conventional optimization based methods [8]–[10]. However, an important observation is that for a given combining beam $\mathbf{w}$, evaluating the SINR requires only the power values (after combining) of the desired and interference signals, and does not require explicit knowledge about the channel vectors. Fortunately, it is less hard and more robust to acquire the receive power measurements for both the desired and interference signals, which require much less control signaling compared to the complex channel estimation process. With this observation, we cast our problem as developing a machine learning based approach that learns how to design an interference-aware beam pattern $\mathbf{w}$ that optimizes (5), **given only the receive power measurements** for the signal plus interference and noise, $\left|\mathbf{w}^H \mathbf{h}\right|^2 P_x + \sum_{k=1}^{K} \left|\mathbf{w}^H \mathbf{h}_k\right|^2 P_x + \sigma^2$, and the interference plus noise, $\sum_{k=1}^{K} \left|\mathbf{w}^H \mathbf{h}_k\right|^2 P_x + \sigma^2$.

IV. Reinforcement Learning of Interference Aware Beam Pattern Design

In this section, we present the proposed reinforcement learning based interference-aware beam pattern learning approach. The motivation of using reinforcement learning is mainly two-fold. First, the lack of the explicit channel knowledge makes most of the existing approaches, such as [4], [11], invalid. Second, the beam design problem is essentially a search problem over dauntingly huge space. Hence, we consider leveraging the powerful exploration capability of reinforcement learning to efficiently navigate through such very large space to find the optimal or near-optimal beam patterns. Next, we first discuss how the system is supposed to be operating in practice, and how the formulated problem is fully compatible with such operation in Section IV-A. Then, we provide the details of the proposed solution in Section IV-B.

A. Practical System Operation

In this subsection, we discuss how to acquire the power measurements that are used for evaluating the objective function of the formulated optimization problem (5). As can be seen from (5), in order to estimate the SINR performance, the system needs to know the receive power measurements of the target user as well as the interference power incurred from the other undesired transmitters. Given that the BS can coordinate with its served UE, this can be achieved if the BS knows when the UE is transmitting or not transmitting signals. To be more specific, to estimate the SINR performance of a certain beam $\tilde{\mathbf{w}}$, the BS first measures the interference plus noise level, i.e., $P_{\text{I+N}} = \sum_{k=1}^{K} |\tilde{\mathbf{w}}^H \mathbf{h}_k|^2 P_x + \sigma^2$, by "muting" the target UE. Then, when the target UE starts transmitting reference signals, the BS uses the same beam to measure the signal plus interference plus noise level, i.e., $P_{\text{S+I+N}} = |\tilde{\mathbf{w}}^H \mathbf{h}|^2 P_x + \sum_{k=1}^{K} |\tilde{\mathbf{w}}^H \mathbf{h}_k|^2 P_x + \sigma^2$. We depict such "on/off" measurement method in Fig. 2. The receive power of the target UE can hence be determined by subtracting the previously measured power $P_{\text{I+N}}$ from the new power measurement $P_{\text{S+I+N}}$, and the SINR can be approximately obtained as $(P_{\text{S+I+N}} - P_{\text{I+N}})/P_{\text{I+N}}$. To this end, it is worth mentioning that, in practice, zero power reference signals, such as Zero Power (ZP) Channel State Information Reference Signal (CSI-RS), are normally used to measure the interference plus noise level [12], i.e., $P_{\text{I+N}}$. This means that the formulated problem requires nothing more than what is already supported in the current 5G NR system to obtain all the quantities needed to perform the beam learning task. In the next subsection, we present the idea of how to leverage reinforcement learning for designing the interference-aware beam pattern based on these acquired power measurements.

B. Reinforcement Learning based Interference Aware Beam Pattern Design

In this subsection, we present our proposed reinforcement learning based solution for addressing the interference-aware beam pattern design problem (5).

Reinforcement Learning Formulation: To solve the problem with reinforcement learning, we first fit all the ingredients of problem (5) into a general reinforcement learning framework as follows:

- **State:** We define the state $\mathbf{s}_t$ as a vector that consists of the phases of all the phase shifters at the t-th iteration, that is, $\mathbf{s}_t = [\theta_1, \theta_2, \dots, \theta_M]^T$. This phase vector can be converted to the actual combining vector $\mathbf{w}$ by applying (2). Since all the phases in $\mathbf{s}_t$ are selected from $\boldsymbol{\Psi}$, and all the phase values in $\boldsymbol{\Psi}$ are within $(-\pi, \pi]$, (2) essentially defines a bijective mapping from the phase vector to the combining vector. Therefore, for simplicity, we will use the term "combining vector" to refer to both this phase vector and the actual combining vector (the conversion is given by (2)), according to the context.
- **Action:** We define the action $\mathbf{a}_t$ as the element-wise changes to all the phases in $\mathbf{s}_t$. Since the phases can only take values in $\boldsymbol{\Psi}$, a change of a phase represents the action that a phase shifter selects a value from $\boldsymbol{\Psi}$.

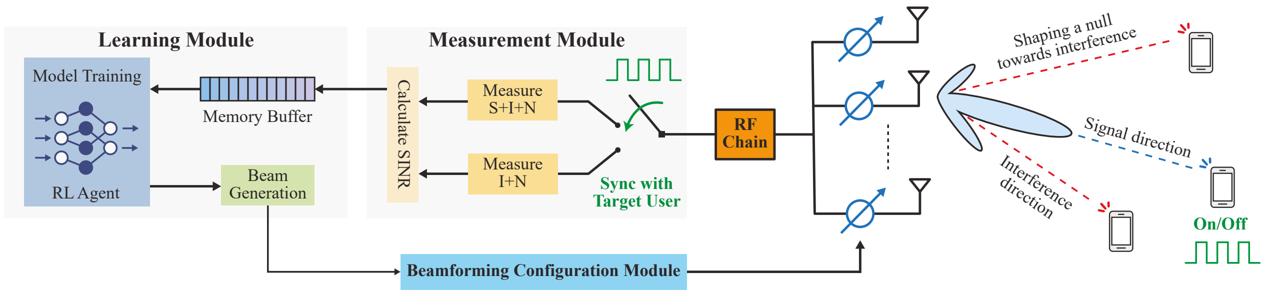


Fig. 2. An illustration of the operation flow of the proposed interference-aware beam pattern learning solution.

Therefore, the action is directly specified as the next state, i.e., $\mathbf{a}_t = \mathbf{s}_{t+1}$, which can be viewed as a deterministic transition in the Markov Decision Process (MDP).

- **Reward:** We define a binary reward mechanism, i.e., the reward r_t takes values from $\{+1, -1\}$. Since the objective of (5) is to maximize the SINR performance, we compare the SINR achieved by the current combining vector, denoted as SINR_t, with the previous one, i.e., SINR_{t-1}. The reward is determined according to the following rule: $r_t = +1$, if $\mathrm{SINR}_t > \mathrm{SINR}_{t-1}$; $r_t = -1$, otherwise.

The above reinforcement learning formulation is fully compatible with the original problem (5) in the following aspects. First, since the state and action are directly specified as phase shifts of the discrete analog phase shifters, the constraints (6) and (7) are automatically satisfied. Second, to obtain the reward, the objective function of (5), i.e., the SINR performance, needs to be evaluated, which can be done in a way that does not rely on channel state information of both the target user and the interfering transmitters, the details of which has been provided in Section IV-A.

Deep Reinforcement Learning Architecture: We adopt an actor-critic based deep reinforcement learning architecture. More details about this learning framework can be found in [7]. To put it in simple words, both actor and critic networks are implemented using elegant fully-connected feed-forward neural networks. The input of the actor network is the state and the output is the action, while the critic network takes in the state-action pair and outputs the predicted Q value. Moreover, to respect the discrete phase shifter hardware constraint (7), we perform an element-wise quantization to make the predicted action a valid one. To be more specific, assume that $\widehat{\mathbf{a}}_t$ is the predicted action from the actor network at time t. Then, the action that finally gets implemented to the system is given by

$$[\mathbf{a}_t]_m = \arg\min_{\theta \in \mathbf{\Psi}} \left| [\widehat{\mathbf{a}}_t]_m - \theta \right|, \ \forall m = 1, \dots, M. \tag{8}$$

It is worth emphasizing that such quantization operation is only activated when the system is actually implementing the predicted action by the actor network to obtain reward. It is not involved in the training process of the actor network due to its non-differentiability. The detailed architectures and the parameters of the adopted neural networks are provided in Section V-B.

V. Simulation Results

A. Simulation Setup

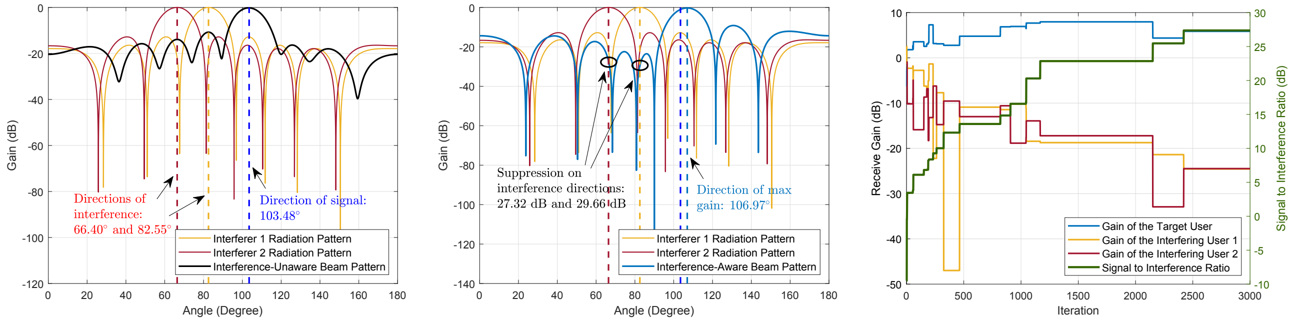
In this simulation, we consider a BS equipped with uniform linear array that has 8 antenna elements and half-wavelength antenna spacing, where each antenna is followed by a 3-bit analog phase shifter. Besides, for a better demonstration, we adopt the following simulation steps: (i) We generate the channel of the target user based on (3), where, for simplicity, we consider the case when the user only has a LOS connection with the BS, i.e., $L = 1$ in (3); (ii) We then learn a beam pattern when there is no interference and this learned beam is referred to as **"interference-unaware" beam** since it focuses on maximizing the combining gain of the desired signal; (iii) After this beam is learned, we *intentionally* position the interfering transmitters at the directions aligning with the strongest side lobes of the learned beam and also assume that they only have LOS channels with the considered BS, which causes non-negligible interference; and (iv) We finally take the interference into account and re-design an **"interference-aware" beam** that learns how to manage the interference in such a way that improves the SINR performance.

B. Deep Learning Architectures

Since the input of the actor network is the state and the output is the action, the size of both the input and output of the actor network is $M = 8$, i.e., the number of antennas. The critic network takes in the state-action pair and outputs the predicted Q value and hence it has an input size of $2M = 16$ and an output size of 1. Both the actor and critic networks have two hidden layers in our proposed architecture, with the size of the first hidden layer being 16 times of the input size and the size of the second hidden layer being 16 times of the output size in both networks. All the hidden layers are followed by the batch normalization layer for an efficient training experience and the Rectified Linear Unit (ReLU) activation layer. The output layer of the actor network is followed by a Tanh activation layer scaled by π to make sure that the predicted phases are within $(-\pi, \pi]$ interval. The output layer of the critic network is a linear layer.

C. Numerical Results

Based on the above simulation setup and deep learning architectures, in Fig. 3, we demonstrate the learning results



(a) Interference-unaware (b) Interference-aware (c) Learning process

Fig. 3. The beam pattern learning results in an environment with two interfering sources, where (a) shows the learned beam pattern when ignoring the surrounding interfering transmitters, and (b) shows the interference-aware beam pattern. (c) shows the interference-aware beam pattern learning process.

when there are two interfering transmitters. We show the beam patterns learned with and without taking the interference into account, together with the receive patterns (i.e., the distribution of receive power strength in angular domain at the BS) of the selected interfering sources. As shown in Fig. 3(a), the two interferers are present at the directions aligning with the two most strongest side-lobes of the interference-unaware beam, which incurs significant interference and causes performance degradation. The learned interference-aware beam is plotted in Fig. 3(b). Clearly, unlike the interference-unaware beam, **the interference-aware beam shapes nulls that have very low receive gains at the directions where the interferers are, which nearly eliminates the severe interference.** To be more specific, in the interference-unaware case, the signal-to-interference ratio (SIR) levels are 10.56 dB and 13.71 dB with respect to the two interfering users. By contrast, the SIR levels are improved to 28.63 dB and 26.28 dB when using the interference-aware beam, which only incurs a loss of 0.8348 dB for the combining gain of the target user.

In Fig. 3(c), we show how the combining gains of the received signals from the target user and the interfering transmitters are changing as the learning proceeds, as well as the overall SIR performance. As can be seen, the combining gain of the target user and the combining gains of the two interfering transmitters start from almost the same level, since a random beam is used as the starting point. As learning proceeds, the combining gain of the target user maintains, generally speaking, an increasing trend, while the combining gains of the two interfering transmitters are gradually decreasing. The overall SIR, however, maintains a monotonically increasing trend. Fig. 3(c) also shows that **with only around** 1000 **iterations, the SIR performance is able to be improved from around** -10 **dB to around** 20 **dB**, without knowing the channels (for both target user and the interfering transmitters).

VI. Conclusions and Discussions

In this paper, we developed a deep reinforcement learning based approach that can efficiently learn interference-aware beams. The proposed solution learns how to design the beam pattern to shape nulls towards the interfering directions relying only on the receive power measurements and without any channel knowledge. This solution also relaxes the coherence/synchronization requirements of the system and respects the key hardware constraints of practical mmWave transceiver architectures. Simulation results show the effectiveness of the proposed solution in learning beam pattern that achieves satisfying interference suppression performance.

References

[1] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, “MIMO Precoding and Combining Solutions for Millimeter-Wave Systems,” *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, 2014.

[2] Junyi Wang and Zhou Lan and Chang-woo Pyo and Baykas, T. and Chin-sean Sum and Rahman, M.A. and Jing Gao and Funada, R. and Kojima, F. and Harada, H. and Kato, S., “Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems,” *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, 2009.

[3] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks,” *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, 2013.

[4] M. Alrabeiah, Y. Zhang, and A. Alkhateeb, “Neural Networks Based Beam Codebooks: Learning mmWave Massive MIMO Beams that Adapt to Deployment and Hardware,” 2020.

[5] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, “Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[6] A. Alkhateeb and G. Leus and R. W. Heath, “Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, 2015.

[7] Zhang, Yu and Alrabeiah, Muhammad and Alkhateeb, Ahmed, “Reinforcement Learning of Beam Codebooks in Millimeter Wave and Terahertz MIMO Systems,” *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 904–919, 2022.

[8] Lorenz, R.G. and Boyd, S.P., “Robust minimum variance beamforming,” *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1684–1696, 2005.

[9] Gesbert, David and Hanly, Stephen and Huang, Howard and Shamai Shitz, Shlomo and Simeone, Osvaldo and Yu, Wei, “Multi-Cell MIMO Cooperative Networks: A New Look at Interference,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, 2010.

[10] Dahrouj, Hayssam and Yu, Wei, “Coordinated beamforming for the multicell multi-antenna wireless system,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1748–1759, 2010.

[11] A. Alkhateeb and R. W. Heath, “Frequency selective hybrid precoding for limited feedback millimeter wave systems,” *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1801–1818, 2016.

[12] 3GPP, “Study on new radio access technology: Physical layer aspects,” Tech. Rep. 38.802, 2017, version 14.2.2.