Online Beam Learning with Interference Nulling for Millimeter Wave MIMO Systems

Yu Zhang, Tawfik Osman, and Ahmed Alkhateeb

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. However, most of the existing codebooks focus only on improving the gain of their target user, without taking interference into account. This can incur critical performance degradation in dense networks. In this paper, we propose a sample-efficient online reinforcement learning based beam pattern design algorithm that learns how to shape the beam pattern to null the interfering directions. The proposed approach does not require any explicit channel knowledge or any coordination with the interferers. Simulation results show that the developed solution is capable of learning well-shaped beam patterns that significantly suppress the interference while sacrificing tolerable beamforming/combing gain from the desired user. Furthermore, a hardware proof-of-concept prototype based on mmWave phased arrays is built and used to implement and evaluate the developed online beam learning solutions in realistic scenarios. The learned beam patterns, measured in an anechoic chamber, show the performance gains of the developed framework and highlight a promising machine learning based beam/codebook optimization direction for mmWave and terahertz systems.

I. Introduction

Millimeter wave (mmWave) and terahertz (THz) communication systems need to employ large antenna arrays to combat the severe path-loss and achieve sufficient receive signal power. Given the high cost of the mixed-signal components, these systems rely mainly on fully-analog or hybrid analog-digital architectures with much smaller number of RF chains compared to the

Yu Zhang, Tawfik Osman, and Ahmed Alkhateeb are with Arizona State University (Email: y.zhang, tmosman, alkhateeb@asu.edu). This work is supported in part by the National Science Foundation under Grant No. 1923676 and by a U. S. Army research program under contract No. W911NF21C0015. Part of this work has been accepted in the Asilomar Conference on Signals, Systems, and Computers, 2022 [1].

number of antennas [2], [3]. These architectures, however, make it hard to explicitly estimate the wireless channels, which motivated these systems to rely on pre-defined beam codebooks for both initial access and data transmission [3], [4]. 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 leads to sub-optimal performance in dense deployments or in scenarios with intended interference/jamming. This urges the research for advanced analog/hybrid beam design approaches that are *interference-aware*. Realizing that, however, is challenging because (i) these beams need to be designed online in the filed and (ii) without explicit channel knowledge which is hard to acquire for analog/hybrid architectures (especially for the interfering transmitters). With this motivation, this paper focuses on developing a beam learning framework that is able to learn interference nulling beam patterns while respecting the hardware and system operation constraints.

A. Prior work

Achieving effective interference suppression and management is a key research topic that attracted significant interest in various systems, such as network MIMO [5], [6], distributed antenna/coordinated multipoint (CoMP) systems [7]–[9], cell free massive MIMO [10], [11], and spectrum sharing and radar/communication coexistence [12]–[14]. The approaches developed in [5]–[14], however, considered conventional fully-digital MIMO transceiver architectures where the precoding and combining are implemented completely in the digital domain. With the recent evolution towards high frequency bands (e.g., mmWave and terahertz) and the utilization of even larger antenna arrays, new transceiver architectures (e.g., fully analog and hybrid analog/digital architectures) are introduced to reduce the hardware cost and power consumption. These architectures rely on implementing the beamforming in the analog domain or distribute it between the analog and digital domains. This required revisiting the precoding design problem for single-user hybrid analog/digital architectures [15]–[17], multi-user mmWave MIMO systems [18]–[20], mmWave full-duplex and relay systems [21]–[24], among others.

To address some of the limitations of the classical approaches, deep learning-based beamforming solutions have been recently proposed and shown promising gain over the conventional ones thanks to their data-driven and site-specific natures [25]–[29]. Among those, reinforcement learning-based approaches are of particular interest for their practicality and the relaxed system requirements. For instance, [28] leverages deep reinforcement learning for fast beam steering that

reduces communication overhead and achieves high energy efficiency. In [29], a multi-agent deep reinforcement learning framework is proposed for downlink sum-rate maximization problem in heterogeneous networks. In addition to the applications of deep learning, recent advancement in computing also brings surrogate model/digital twin-assisted training and optimization frameworks into attention. These novel frameworks have demonstrated promising potential in enabling a wide range of applications [30]–[34]. For instance, in [30], a digital twin-aided approach is leveraged to learn the generative process of wireless channels based on precoder matrix indicator feedback. In [34], by using the 3D model of the communication environment and ray tracing technology, the algorithm is trained in the digital replica while tested in the real world.

The existing precoding/beamforming approaches in [15]–[24] and [25]–[29], however, have certain practical limitations. First, most of the classical solutions mandate either full or partial channel state information in order to design the corresponding transmit/receive schemes. Similar requirement holds for the deep learning-based solutions where most of the algorithms require explicit channel state information (CSI) during the training stage. Acquiring this channel knowledge is challenging in realistic mmWave analog/hybrid architectures and is normally associated with large channel estimation and system coordination overhead (synchronization, information sharing among BSs or UEs, etc.). Second, the existing approaches normally lack the capability/flexibility of performing online adjustment in order to account for imperfections in the hardware fabrication/calibration, errors in estimated channels, or changes in the surrounding environment. Moreover, utilizing surrogate model/digital twin for beamforming task has received relatively limited attention. The efficient implementations of these digital replicas as well as their interactions with the design algorithms are important research problems that need further study. This motivates the development of **sample-efficient interference-aware beam learning approaches that do not require explicit channel knowledge of both the desired transmitter and the interfering transmitters, and do not any require coordination between the considered receiver and the interferers (which can not happen in practice).**

B. Contribution

Designing analog beam patterns that are aware of interference is an important problem and can be found in a variety of practical scenarios. However, the hardware constraints associated with the fully analog transceiver architectures along with the scarcity of the accurate channel knowledge make the problem quite challenging. In this paper, we propose a deep reinforcement

learning based beam pattern design framework that can efficiently adapt the beam pattern to avoid interference sources while maximizing the beamforming/combining gain of the desired user. This is done by not requiring any explicit channel knowledge of the target user or the interferers, and by only relying on the power measurements. The proposed framework also respects the key hardware constraints of the analog/hybrid transceiver architectures, such as the quantized phase shifter and constant modulus constraints, making it a hardware compatible solution. The main contributions of this paper can be summarized as follows:

- **Designing a deep reinforcement learning based beam pattern design framework** that can learn how to shape nulls towards the directions of the interference. The developed solution relies only on receive power measurements and does not require any coordination between the considered receiver and the interfering transmitters. A 3GPP standard-compatible signal power estimation method is proposed to evaluate the signal reception performance under the presence of the always-on interfering transmitters.
- **Developing a surrogate model assisted learning framework** that improves the sample efficiency. This is achieved by better leveraging the underlying signal relationship of the considered system in the surrogate model design. The improvement on the sample efficiency has the potential of reducing the beam learning overhead as well as shortening the convergence time of the proposed solution. The proposed surrogate model also provides flexibility to the practical deployment in terms of enabling data sharing and cooperation (hence better learning) in complex scenarios.
- **Extensively evaluating the proposed interference-aware beam learning solutions** using numerical simulations. This provides a comprehensive assessment of the capability of the proposed learning approaches in nulling interference without requiring any knowledge about the channel, array geometry, or user location. The results also evaluate the potential performance and practical gains of the developed surrogate-model based learning framework.
- **Demonstrating the performance using a real-world proof-of-concept prototype**. To do that, we built a hardware prototype based on mmWave phased arrays and implemented the online beam learning solutions to run in real-time. The developed solutions are tested in the field under realistic conditions and the learned beam patterns are measured in an anechoic chamber. This provides important insights about how the proposed beam learning approaches perform in real-world scenarios and with practical hardware.

The real-world prototyping results show that the developed beam learning solution, which rely only on power measurements, can effectively suppress the interference with minimal impact on the desired signal gain. The simulation results also show that proposed surrogate model assisted learning framework significantly improves the sample efficiency by integrating the underlying signal relationship into the framework design, highlighting a promising interference-avoidance path for future mmWave and THz communication systems.

The rest of this paper is organized as follows. Section II presents the system and channel models adopted in this work. Section III formulates the interference nulling beam design problem addressed in this paper. Section IV provides details of the proposed online beam learning solution for shaping interference nulling beam patterns. Section V introduces the proposed surrogate model assisted beam learning framework for improving the practicality of the proposed solution. Section VI provides the numerical results for the proposed solutions. Section VII introduces a real-world mmWave prototyping platform and shows the real measurements of the proposed online beam learning solution. Finally, Section VIII concludes the paper by discussing the main takeaways.

II. System and Channel Models

Our objective in this paper is to investigate the online design of interference-aware beam patterns, i.e., the online learning/design of beam patterns that achieve their original objectives while nulling the interference introduced by other sources in the environment. 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) in an uplink mode. Moreover, we assume that there exist K (≥ 1) non-cooperative interference transmitters[1] in the vicinity of the BS, operating at the same frequency bands and hence causing interference to the BS receiver. Such interference, for instance, could be from the sidelinks of a vehicular network communicating over the 5G NR unlicensed band at frequency range 2 [35], as illustrated in Fig. 1. Therefore, if the UE transmits a symbol $x \in \mathbb{C}$ to the BS, and the other K interference transmitters also transmit symbols $x_k \in \mathbb{C}, k = 1, \ldots, K$ at the same time and frequency slot, such that all the transmitted symbols satisfy the same average power constraint,

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

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2023.3324621

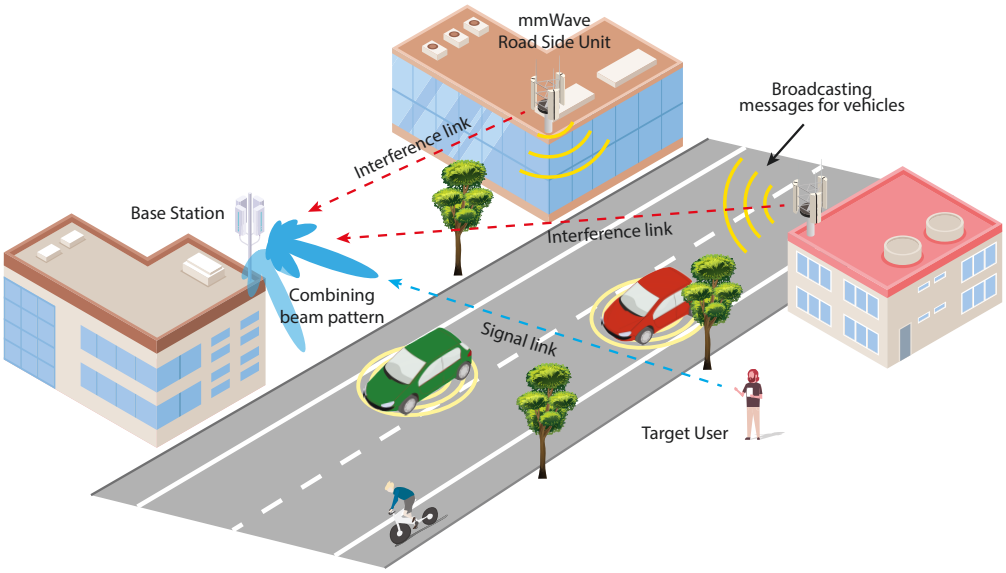


Fig. 1. The considered uplink scenario where a mmWave base station, operating in a receive mode, is communicating with its target user under the presence of non-cooperative interference transmitters. This could be the case, for instance, where the mmWave road side units of a vehicular network are broadcasting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

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

[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.

r-bit quantized phase shifters. Therefore, the combining vector at the BS can be written as

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

where each phase shift $\theta_m, \forall m = 1, \ldots, M$ 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 geometric channel model for the channel between BS and UE, as well as the interference channels between BS and any interfering transmitters. Hence, the channel between BS and its served UE takes the following form (the channel between BS and any interference transmitter takes similar form)

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

where L is the number of multi-paths. Each path ℓ has a complex gain α_ℓ, which includes the path-loss. 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 BS array response vector. 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 the analog combining/beamforming that achieves interference awareness (i.e., attempts to address the interference) without explicitly knowing any channel state information. Given the receive signal (1) at the BS, the achievable rate of its target user 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. Given the monotonicity of the logarithm function, this is equivalent to

maximize the signal-to-interference-plus-noise ratio (SINR) term in (4). Therefore, the problem of designing interference-aware beam pattern[3] can be cast as

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

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

$$\theta_m \in \mathbf{\Psi}, \ \forall m = 1, ..., M, \tag{7}$$

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 the following challenges: (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**, since it is hard to acquire CSI in practice, especially with analog/hybrid architectures, (iv) the channels of the interfering transmitters, i.e., $\mathbf{h}_k, \forall k$**, are also unknown**, since there is normally no coordination with the interfering transmitters, and (v) the possible hardware impairments are also assumed to be unknown.

Given these aforementioned difficulties, it is hard to solve (5) using the conventional optimization methods [8], [37], [38]. An important observation, however, 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 requires much less training/signaling overhead and more relaxed synchronization constraints compared to the channel estimation process. With this observation, we cast our problem as developing an online machine learning 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, $\sum_{k=1}^{K} \left|\mathbf{w}^H \mathbf{h}_k\right|^2 P_x + \sigma^2$.

In the next section, we provide a detailed description of the proposed solution for tackling the beam learning optimization problem in (5). Then, in Section V, we introduce a surrogate

[3]It is important to note that the proposed interference-aware beam learning approach can be straightforwardly extended to learning a codebook with multiple beams by, for example, using the user clustering and assignment algorithm proposed in [36].

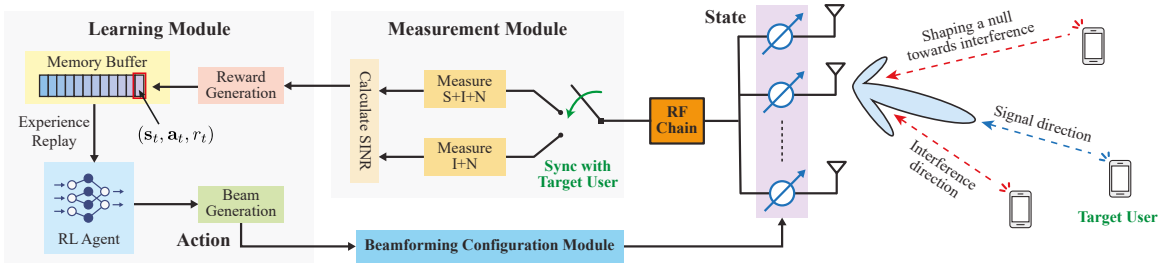


Fig. 2. An illustration of the operation flow of the proposed interference-aware beam pattern learning solution, where the signal power is estimated by configuring the desired UE to transmit the signal in an on/off fashion.

model assisted learning approach which achieves better sample efficiency as well as enhances the deployment flexibility of the proposed interference-aware beam learning framework.

IV. Online Learning of Interference Aware Beam Pattern Design

In this section, we present the proposed online reinforcement learning based interference-aware beam pattern design approach. The motivation of using reinforcement learning is twofold: First, the lack of the channel knowledge renders most of the existing beamforming design approaches, such as [15]–[24] and [25]–[29], hard to apply, as most of them assume the explicit CSI is available during training. To address this problem, by viewing the unknown channels as environment, the proposed solution learns **an implicit representation of the channel behavior** based only on power measurements. Second, the beam design problem can be regarded as a search problem over a very huge space. Hence, we consider leveraging the powerful exploration capability of reinforcement learning to efficiently navigate through this large space to find the optimal or near-optimal beam patterns. Next, we first discuss the proposed system 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 beam design optimization problem (5), which will also be used in the proposed beam learning approach. As will be discussed in Section IV-B, the proposed beam learning solution relies only on the power measurements in its operation. In particular, it needs to measure the power of the received signal from 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 to know when it is transmitting, this knowledge could be leveraged to enable the required power measurements. 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., $P_{\mathrm{I+N}} = \sum_{k=1}^{K} |\tilde{\mathbf{w}}^H \mathbf{h}_k|^2 P_x + \sigma^2$, when the target UE is not transmitting. 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., $P_{\mathrm{S+I+N}} = |\tilde{\mathbf{w}}^H \mathbf{h}|^2 P_x + \sum_{k=1}^{K} |\tilde{\mathbf{w}}^H \mathbf{h}_k|^2 P_x + \sigma^2$. The receive power of the target UE can hence be determined by subtracting the previously measured power $P_{\mathrm{I+N}}$ from the new power measurement $P_{\mathrm{S+I+N}}$, and the SINR can be approximately obtained as $(P_{\mathrm{S+I+N}} - P_{\mathrm{I+N}})/P_{\mathrm{I+N}}$. To this end, it is worth mentioning that, in practice, zero power reference signals, such as the Sounding Reference Signal (SRS) that is not scheduled for any UE in the 5G NR, could be potentially leveraged to measure the interference plus noise level [39], i.e., $P_{\mathrm{I+N}}$. In the next subsection, we present the basic idea of the proposed online reinforcement learning approach for learning interference-aware beams based solely on these power measurements.

B. Reinforcement Learning based Interference Aware Beam Pattern Design

Given the system operation in Section IV-A, we now describe our proposed reinforcement learning based solution for addressing the interference-aware beam pattern design problem (5). To do that, we first formulate our beam design problem as a reinforcement learning problem. Then, we present the proposed deep reinforcement learning architecture for solving this problem.

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 (as also illustrated in Fig. 2):

- **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, $\mathbf{s}_t = [\theta_1, \theta_2, \ldots, \theta_M]^T$. 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 $\mathbf{\Psi}$, and all the phase values in $\mathbf{\Psi}$ are within $(-\pi, \pi]$, (2) essentially defines a bijective mapping from the phase vector to the combining vector. In this paper, we will use the term "combining phase vector" to refer to this phase vector and use the term "combining vector" to refer to the actual combining vector.
- **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 $\mathbf{\Psi}$, a change of a phase represents the action that a phase shifter selects a value from $\mathbf{\Psi}$. 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.

It is worth highlighting that 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 the 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: Given the reinforcement learning formulation above for the interference-aware beam learning problem, we adopt an actor-critic based deep reinforcement learning architecture. This follows the learning framework that we proposed earlier in [36]. In summary, both the actor and critic networks are implemented using fully-connected (FC) 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.[4] 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

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

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.

Despite its full compatibility with the considered system as well as the 3GPP standards [39], the proposed interference-aware beam learning solution still has two drawbacks. First, it requires

[4]The detailed architectures and the parameters of the adopted neural networks are provided in Section VI-B.

a relatively large number of iterations to find a qualified beam pattern, especially when the number of antennas is large. As a result, this incurs a large beam learning overhead, since these iterations are done over the air. Second, as indicated by the objective function of (5), the SINR performance of a given beam is determined by two factors: (i) The desired beamforming gain and (ii) The effectiveness of suppressing the undesired interference. However, the proposed solution does not fully leverage this information as it only focuses on the overall SINR performance. It turns out that the decomposition of these two factors, as will be further discussed in the next section, makes the data sharing among the learning processes of different beams possible, which has the potential of improving the convergence behavior of the beam/codebook learning algorithm With this motivation in mind, in the next section, we introduce a modified learning framework which includes a surrogate model that assists the beam learning process. The adopted surrogate model better utilizes the underlying signal model and hence the sample efficiency (a measure for the number of real measurements) is further improved. It also provides more deployment flexibility and enables other features such as data sharing.

V. Surrogate Model Assisted Beam Learning Framework

In this section, we describe in detail the proposed surrogate model assisted interference-aware beam pattern learning framework. The motivations of introducing the surrogate model are mainly two-folds. First, it has the potential of improving the sample efficiency (i.e., reducing the number of interactions with the actual environment) of the learning process by leveraging the underlying signal models. Second, it facilitates other more complex tasks (than learning a single beamforming vector), such as data sharing (which can be very useful in learning interference-aware beam codebook) and cooperative learning[5] (among multiple BSs to avoid interfering each other). The overall objective is to have a *simulated* environment that can provide the DRL agents with *authentic* feedback as if the agents are interacting with the actual environment. Next, we first introduce the adopted surrogate model in Section V-A. Then, we provide more details about how to integrate the surrogate model with the RL beam learning algorithm in Section V-B. Finally, we discuss several practical aspects of the learning framework in Section V-C.

[5]For instance, as the system has full knowledge of its simulated environment, it can assign accurate reward to each agent. This has the potential of mitigating the non-stationary environment problem that exists in most of the multi-agent learning tasks.

A. Surrogate Model for Beam Pattern Learning

In this subsection, we introduce the proposed surrogate model that assists the learning of interference-aware beams. As can be seen from Section IV-A, in order to acquire the reward signal that is used for training the RL agent, the system needs to estimate two quantities, i.e., the signal power, $P_{\mathrm{S}} = \left|\mathbf{w}^H\mathbf{h}\right|^2 P_x$, and the interference plus noise power, $P_{\mathrm{I+N}} = \sum_{k=1}^{K} \left|\mathbf{w}^H\mathbf{h}_k\right|^2 P_x + \sigma^2$. Therefore, correspondingly, there are two major components in the considered surrogate model that provide the agent with such information, i.e., an interference predictor and a signal predictor, as will be discussed in this subsection.

1) The key idea of surrogate model: The machine learning model that virtually interacts with the agent can be considered as a **surrogate model**. This model is used to *imitate* the behavior of the actual environment, aiming to reduce the expensive (sometimes, even impossible) actual evaluations of the design. In this paper, we design the surrogate model with a particular emphasis on two important aspects:

- *Prediction accuracy:* As the name suggests, a surrogate model is essentially a prediction model which imitates (or predicts) the behavior (or response) of an unknown environment to a certain input action. Hence, being accurate is undoubtedly the most important property of the considered surrogate model.
- *Data requirement:* Another important property of a surrogate model, in the considered interference-aware beam learning task, is data requirement. This refers to the amount of data that is required by the surrogate model for the training purposes, in order to reach a certain prediction accuracy constraint. Generally speaking, a surrogate model is more valuable if it requires less data to achieve a satisfactory performance.

With these important criterions in mind, we next describe the adopted surrogate model. As mentioned before, the considered surrogate model consists of two major components, i.e., an interference prediction model and a signal prediction model. Formally, the interference prediction model predicts the interference plus noise power based on the configuration of the receive combining vector, which can be expressed as

$$\widehat{P}_{\mathrm{I+N}} = f_{\mathrm{in}}(\mathbf{w}; \boldsymbol{\Theta}_{\mathrm{in}}), \tag{9}$$

where $\mathbf{w} \in \mathbb{C}^{M\times 1}$ is the input of the model, representing the designed receive combining vector, and the output is the predicted interference plus noise power, i.e., $\widehat{P}_{\mathrm{I+N}} \in \mathbb{R}$. The model is

parameterized by $\mathbf{\Theta}_{\rm in}$. Similarly, the signal prediction model predicts the signal power of a given receive combining vector, which can be written as

$$\widehat{P}_{\rm S} = f_{\rm s}(\mathbf{w}; \mathbf{\Theta}_{\rm s}), \tag{10}$$

where $\widehat{P}_{\rm S} \in \mathbb{R}$ is the predicted signal power value and $\mathbf{\Theta}_{\rm s}$ denotes the model parameters. It is worth mentioning that the architecture of $f_{\rm in}$ and $f_{\rm s}$ is not unique and is a design choice. Next, we present two candidates that could be used in the considered beam learning task.

2) Surrogate model architecture: As mentioned before, the choice of $f_{\rm in}$ and $f_{\rm s}$ is not unique. In this paper, we study two specific designs: (i) A model-based prediction architecture, and (ii) a fully-connected neural network based prediction architecture.

Model-based architecture: The model-based architecture, as its name suggests, is inspired by the underlying *signal model*. For instance, as can be seen from the expression of the interference plus noise power, i.e., $P_{\rm I+N} = \sum_{k=1}^{K} \left|\mathbf{w}^H \mathbf{h}_k\right|^2 P_x + \sigma^2$, it takes a quadratic form of the receive combining vector $\mathbf{w}$. To see this, by defining $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]$, $P_{\rm I+N}$ can be expressed as

$$P_{\rm I+N} = \left\|\mathbf{H}^H \mathbf{w}\right\|_2^2 P_x + \sigma^2, \tag{11}$$

$$= \mathbf{w}^H \left(P_x \mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I}\right) \mathbf{w}, \tag{12}$$

$$= \mathbf{w}^H \mathbf{A} \mathbf{w}, \tag{13}$$

where $\mathbf{A} = P_x \mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I}$. The signal power can be expressed in the similar form, i.e., $P_{\rm S} = \mathbf{w}^H P_x \mathbf{h}\mathbf{h}^H \mathbf{w}$. Therefore, the interference prediction network is essentially leveraged to learn the relationship (13). Inspired by this, we design the interference prediction network with a focus on imitating the "behavior" of $\mathbf{A}$. Specifically, the interference prediction network is chosen to take the following form

$$f_{\rm in}(\mathbf{w}) = \mathbf{w}^H \mathbf{Q}_{\rm in} \mathbf{Q}_{\rm in}^H \mathbf{w}, \tag{14}$$

where $\mathbf{Q}_{\rm in} \in \mathbb{C}^{M \times r_{\rm in}}$ with $r_{\rm in}$ being a hyperparameter. Therefore, the parameter of the interference prediction network is essentially $\mathbf{Q}_{\rm in}$, i.e., $\mathbf{\Theta}_{\rm in} = \{\mathbf{Q}_{\rm in}\}$. The signal prediction network takes the similar form, i.e., $f_{\rm s}(\mathbf{w}) = \mathbf{w}^H \mathbf{Q}_{\rm s} \mathbf{Q}_{\rm s}^H \mathbf{w}$, where $\mathbf{Q}_{\rm s} \in \mathbb{C}^{M \times r_{\rm s}}$ with $r_{\rm s}$ being a hyperparameter as well, which makes $\mathbf{\Theta}_{\rm s} = \{\mathbf{Q}_{\rm s}\}$.

Fully-connected neural network based architecture: Despite being lightweight and a better fit to the signal model, the model-based architecture, fundamentally, suffers from any mismatch between the assumed signal model and the actual signal relationship. For instance, there are

normally unknown non-linearities in the practical hardware that undermine the validity of the assumed relationship between the receive combining vector and the interference plus noise power (similarly for the signal power). As a result, the signal model cannot always be met and the model-based architecture will show up certain level of residual prediction errors that are very hard to be eliminated given the less powerful expressive capability of its architecture. Motivated by this, we also investigate a more general architecture, which is built upon fully-connected neural network, given its powerful universal approximation capability [40]. Specifically, both $f_{\rm in}$ and $f_{\rm s}$ are modeled with feed-forward fully-connected neural networks. The detailed network parameters will be provided in Section VI-B.

3) Training dataset and loss function: We denote the dataset used for training the interference prediction network as

$$\mathcal{D}_{\rm in} = \left\{ \left(\mathbf{w}^{(n)}, P_{\rm I+N}^{(n)} \right)_{n=1}^{N_{\rm in}} \right\}, \tag{15}$$

where each data sample is comprised of a combining vector and its corresponding interference plus noise power value obtained from the actual environment, i.e., from the real measurement. $N_{\rm in}$ is the total number of data samples in the dataset, i.e., $|\mathcal{D}_{\rm in}| = N_{\rm in}$. And the dataset used for training the signal prediction network can be similarly denoted as

$$\mathcal{D}_{\rm s} = \left\{ \left(\mathbf{w}^{(n)}, P_{\rm S}^{(n)} \right)_{n=1}^{N_{\rm s}} \right\}, \tag{16}$$

with $N_{\rm s}$ being its size. We will discuss how to efficiently collect these datasets in Section V-C.

Since the target of these two networks is to predict the power values, we pose the learning problem as a regression problem conducted in a supervised fashion. Furthermore, we employ mean squared error (MSE) as the training loss function. Using the interference prediction network as an example, for the n-th data sample in $\mathcal{D}_{\rm in}$, the loss function is defined as

$$\mathcal{L}\left(P_{\rm I+N}^{(n)}, \widehat{P}_{\rm I+N}^{(n)} \right) = \left| P_{\rm I+N}^{(n)} - \widehat{P}_{\rm I+N}^{(n)} \right|^2, \tag{17}$$

$$= \left| P_{\rm I+N}^{(n)} - f_{\rm in}(\mathbf{w}^{(n)}; \mathbf{\Theta}_{\rm in}) \right|^2. \tag{18}$$

The loss function used for the signal prediction network is identical.

B. Surrogate Model Assisted Learning

In this subsection, we discuss how to integrate the surrogate model with the proposed RL based beam learning framework. Since the surrogate model is essentially used to provide the RL agent

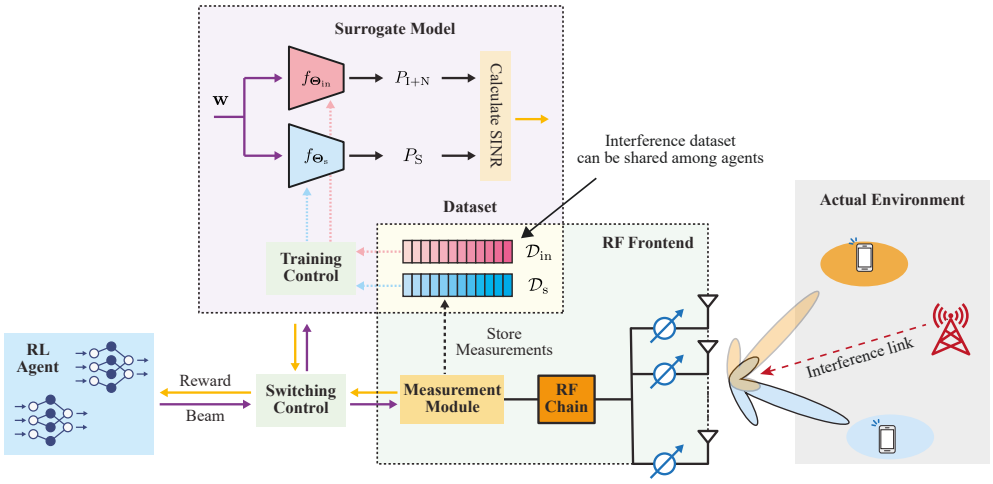


Fig. 3. An illustration of the proposed surrogate model assisted interference-aware beam pattern learning framework.

with a simulated environment to interact with, it plays the same role as the actual environment. However, in order to provide high quality *synthetic* feedback, it requires training process that relies on the *authentic* data collected from the actual environment. Based on the trained surrogate model, the system can virtually evaluate its designed beams without measuring the physical signals. Moreover, the system might require constantly switching between the surrogate model and the actual environment, triggered by the demand for the authentic data. Next, we summarize the key components of the proposed surrogate model assisted beam learning.

1) Initial interaction and data acquisition: The system starts with the normal interaction between the RL agent and the actual environment. To be more specific, upon forming a new beam $\tilde{\mathbf{w}}$, the BS follows the procedures presented in Section IV-A to estimate the interference plus noise power $P_{\text{I+N}}$ and the signal power P_{S}. The reward signal used for RL agent learning will then be generated. Moreover, these authentic power measurements together with the beam will be stored in the two datasets, i.e., $\mathcal{D}_{\text{in}}$ and $\mathcal{D}_{\text{s}}$, respectively. During this interaction process, two initial datasets are established.

2) Surrogate model training: Based on the collected initial datasets $\mathcal{D}_{\text{in}}$ and $\mathcal{D}_{\text{s}}$, the two sub-networks of the surrogate model, i.e., the interference prediction network f_{in} and the signal prediction network f_{s}, are trained in a supervised manner. After the training process saturates, the surrogate model is ready to interact with the RL agent.

3) Environment switching and virtual interaction: The switching from the actual environment to the surrogate model is triggered based on multiple factors, which will be discussed in detail in Section V-C. As a result, after the switching is finished, the reward signal required by the RL agent will be provided by the trained surrogate model instead of the actual environment. The agent keeps interacting with the surrogate model until it does not improve, which marks the saturation of the agent learning and the end of the virtual interaction process.

4) Demand based switching and active data acquisition: The system might require executing the above steps multiple times, based on the achieved performance. The motivation of such repetition can be summarized as follows. From the model training perspective, the quality of the collected datasets, i.e., $\mathcal{D}_{\text{in}}$ and $\mathcal{D}_{\text{s}}$, has significant influence on the prediction accuracy of the trained surrogate model. To be more specific, during the initial interaction process, most of the beams tried out by the agent are relatively random and hence have relatively poor quality in terms of SINR performance. This means that the datasets are, intuitively speaking, biased towards the "poor-quality" beams. As a result, the trained surrogate model will have relatively inaccurate predictions on the beams that actually have better performance. The incurred residual prediction error will in turn influence the learning of the agent, leading to unsatisfactory performance.

However, as the policy of the RL agent gets improved over time, the actions performed by the agent, i.e., the beams, are more likely to be in the beam space where the achieved SINR is high. Therefore, it is advisable to switching back to the actual environment to re-collect data (through agent-environment interaction). Such active data acquisition can enhance the training datasets with "high-quality" beams. Using those better data samples to refine the parameters of the surrogate model can help achieve higher prediction accuracy in the interested beam space, which further helps the learning of the agent. By alternatingly performing these steps, the system has higher chance to collect data samples that are more useful for the agent learning, which has the potential of further enhancing both sample efficiency and learning convergence. We show such interplay between the RL agent, actual environment and the surrogate model in Fig. 3.

C. Practical Considerations

In this subsection, we discuss some practical considerations of the proposed surrogate model assisted interference-aware beam pattern learning solution.

1) Dataset collection: An important observation is that for any given beam, its achieved signal power remains the same regardless of the presence of the interference signals. This implies that

the interference power dataset $\mathcal{D}_{\mathrm{in}}$ and the signal power dataset $\mathcal{D}_{\mathrm{s}}$, i.e., (15) and (16), can be collected in a "non-synchronized" way, which facilitates the data collection in few cases. For instance, on one hand, if the interference transmitters are not transmitting signals, the system can only collect data samples and store them into $\mathcal{D}_{\mathrm{s}}$. In this case, the interference plus noise power $P_{\mathrm{I+N}}$ becomes the noise power P_{N}, and the signal power can be measured in a similar fashion. On the other hand, if the interference happens aperiodically and sparse, the system can be dedicated to measure $P_{\mathrm{I+N}}$ whenever the interferers are present, i.e., without measuring the signal power during this period.

2) Data sharing: The aforementioned non-synchronized measurement strategy also implies that data sharing is possible. To be more specific, if the interference transmitters are fixed, the collected interference dataset $\mathcal{D}_{\mathrm{in}}$ can be reused for different target UEs. This has the potential of accelerating the learning process as well as reducing the memory requirement (i.e., to store the measurement dataset). It is also particularly interesting when learning a codebook.

3) Switching trigger: Another important problem is how to design the conditions that control the switching between the surrogate model and the actual environment. Although such criterion is a design choice and is normally determined by a variety of factors, it would be beneficial if the design can somehow reflect the intentions of introducing a surrogate model, which are: (i) To support the continuous learning (when the actual environment cannot provide immediate feedback), and (ii) to reduce the expensive evaluations. For instance, it is reasonable to switch to the actual environment when the training processes of both the surrogate model and the RL agent saturate. However, when the surrogate model can provide very accurate predictions, the switching should be avoided to reduce the unnecessary overhead.

4) Parallel computing: The proposed surrogate model assisted optimization framework can support parallel computing that has potential to accelerate learning as well as to improve the optimization performance. **To accelerate learning,** it is possible to perform the training of the surrogate model and the learning of the agent simultaneously. This can be done, for instance, by duplicating the surrogate model, where one of the duplicates (with its parameters being frozen) will be used to interact with the RL agent. The other is trained with the up-to-date datasets, and after the training finishes, its parameters can be used to update the one that is interacting with the RL agent. This implies that the second and third steps mentioned in the previous subsection can be executed in parallel, which can help improve the convergence of the proposed solution. **To improve the performance,** there could be multiple randomly initialized agents interacting

with the same surrogate model to perform the same task. The finally learned beam could be the best one among all the learning results. Through this way, it has higher chance to avoid very bad local minima, considering that the underlying optimization problem is non-convex. And importantly, such gain is attained without requesting any additional real measurements.

5) Quantized measurements: The proposed surrogate model also supports the case when the measurements are quantized power levels, which is used more commonly in practice [41]. In such case, the output layer of both interference prediction network and signal prediction network can be changed to a classification layer, and the training loss function, i.e., (17), can also be modified accordingly.

VI. Simulation Results

In this section, we numerically evaluate the performance of the proposed reinforcement learning based interference-aware beam pattern design approach. We will first describe the adopted simulation setup in Section VI-A. Then, in Section VI-B, we provide more details about the adopted architectures of the deep learning models as well as the training procedures. Finally, in Section VI-C, we present the numerical results of the proposed solutions.

A. Simulation Setup

In this simulation, we consider the case where a BS receiver adopts uniform linear array (ULA) with half-wavelength antenna spacing. Each antenna of the ULA 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 line-of-sight (LOS) connection with the BS, i.e., $L = 1$ in (3); (ii) We then learn a beam pattern assuming there is no interference and this learned beam is referred to as **"interference-unaware" beam**, since it solely 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 Models and Training Procedures

In this subsection, we provide more details about the adopted deep learning architectures for both the DRL agent and the surrogate model. We also provide the parameters regarding the training processes of these models.

1) DRL agent architecture: 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, 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 $2M$ 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. Moreover, it is worth mentioning that we adopt the same DRL architecture for both solutions, regardless of having surrogate model or not.

2) Surrogate model architecture: We describe the two different architectures of the surrogate model studied in this paper. Also, as the signal prediction network and the interference prediction network have identical architecture in both solutions (i.e., model-based solution and fully-connected neural network based solution), for brevity, we only use the interference prediction network as an example.

Signal model-based prediction network As mentioned before in (14), the interference prediction network is essentially devised to take a quadratic form of the combining vector determined by a positive semi-definite matrix $\mathbf{Q}_{\text{in}}\mathbf{Q}_{\text{in}}^H$, leaving the matrix $\mathbf{Q}_{\text{in}}$ to be the model parameter. Moreover, $\mathbf{Q}_{\text{in}}$ has a shape of $M \times r_{\text{in}}$ with M being the number of antennas and r_{in} being a hyper-parameter. The choice of r_{in} is empirically guided by the following rules: (i) r_{in} should not be too large as it will increase the model complexity and hence the required amount of training data; (ii) r_{in} should not be too small as it will limit the expressive capability of the model, leading to unsatisfactory prediction accuracy.

Fully-connected neural network based prediction network: We adopt the fully-connected neural network with two hidden layers to be the interference prediction network. The input layer of the network has M neurons, which is equal to the number of antennas. The output layer of

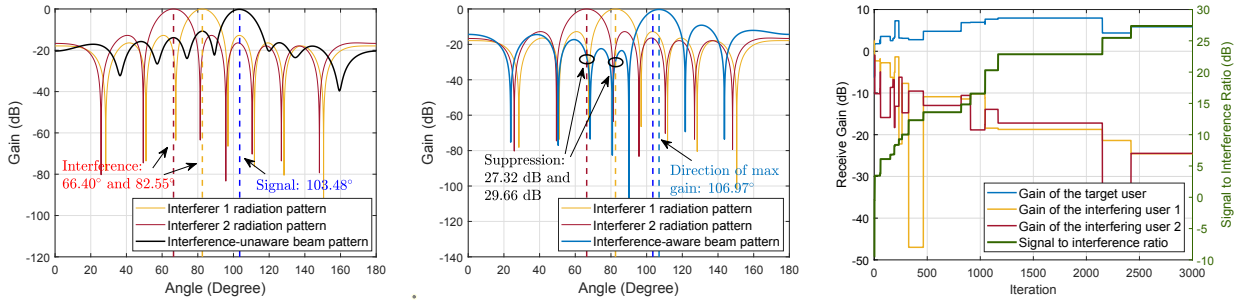
the network has only one neuron with linear activation. Both hidden layers have M' neurons. Similar to r_{in} in the model-based architecture, the selection of M' needs to strike a balance between model complexity and model expressive capability. Moreover, all the hidden layers are followed by the batch normalization layer and ReLU activation layer.

3) Training parameters: As mentioned before, the surrogate model is trained in a supervised fashion, based on the collected power datasets, i.e., $\mathcal{D}_{\text{in}}$ and $\mathcal{D}_{\text{s}}$. Moreover, the interference prediction network and the signal prediction network are independently trained, and we adopt the same training hyper-parameters for these two prediction networks. We notice that, however, when the following factors are different, it might need additional adjustments of the training schedule: (i) The model-based and FC-based architectures present noticeably different convergence behavior and hence require different training designs; (ii) When the number of antennas changes (e.g., 8 or 256 antennas); (iii) When the size of training dataset changes. The detailed hyper-parameters used for training different kinds of surrogate models with various system configurations can be found in [42].

C. Numerical Results

In this subsection, we provide the simulation results of the proposed interference-aware beam learning solutions. Moreover, to better present the results, in Section VI-C1, we first evaluate the reinforcement learning based beam design solution proposed in Section IV that keeps interacting with the actual environment. This is to demonstrate the achieved performance by the proposed beam learning algorithm without knowing the channel knowledge. Then, in Section VI-C2, we test the surrogate model assisted beam design framework proposed in Section V, with a focus on evaluating the validity and efficacy of using surrogate model to reduce the beam learning overhead, as well as comparing different surrogate model architectures.

1) Interference nulling without knowing the channels: Based on the aforementioned simulation setup and deep learning architecture, in Fig. 4, we demonstrate the learning results 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 from the BS's perspective) of the selected interfering sources. As shown in Fig. 4(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. 4(b).



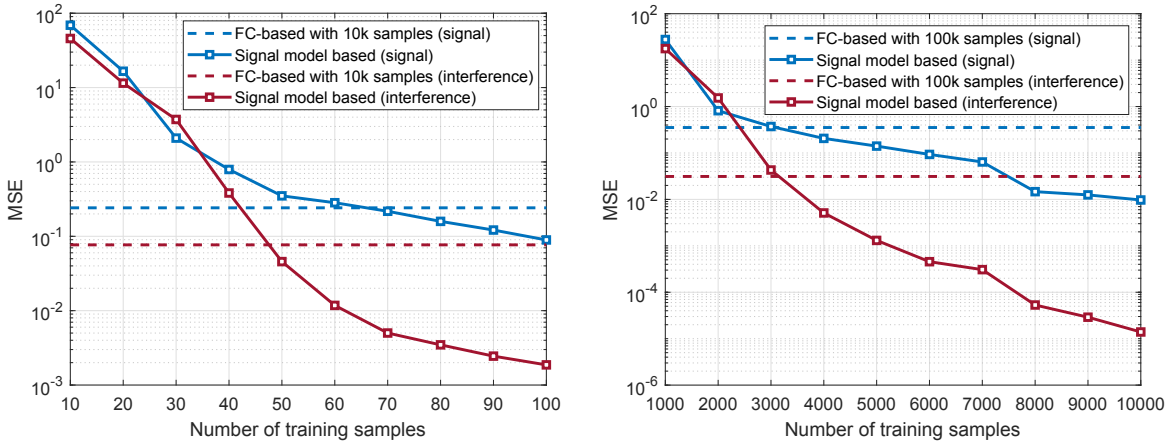
(a) Interference-unaware (b) Interference-aware (c) Learning process

Fig. 4. The beam pattern learning results in an environment with two interfering transmitters, 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.

As can be seen, **the interference-aware beam shapes nulls that have very low receive gains at the directions of the interferers, 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 transmitters. 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. 4(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. Furthermore, as can be observed from the figure, the combining gain of the target user has a high spike (achieved by a certain learned beam) at the beginning of the learning process. However, despite the good performance on the target user, that beam also incurs strong interference from other undesired transmitters, hence resulting in an unsatisfactory SIR performance. Therefore, it is finally replaced by other beams that have slightly lower combining gain to the target user but are very effective in suppressing the interference. Fig. 4(c) also shows that **with only around** 1000 **iterations, the SIR performance improved from around** -10 **dB to around** 20 **dB**, without knowing the channels (for both the target user and the interfering transmitters).

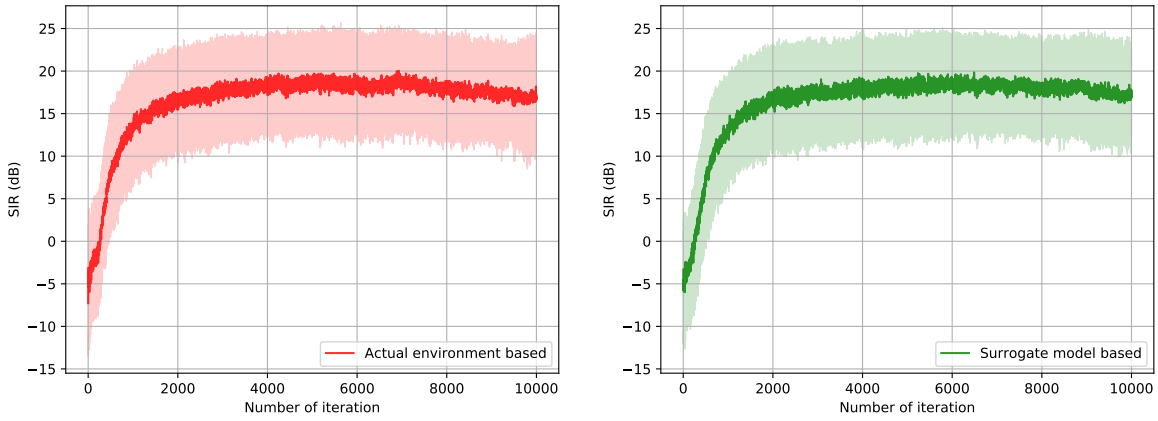
This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2023.3324621



(a) Signal and interference prediction ($M = 8$)

(b) Signal and interference prediction ($M = 256$)

Fig. 5. The prediction accuracy of different surrogate model architectures. It shows that the proposed signal model-based prediction network requires much less data samples to outperform the FC-based prediction network in both cases, i.e., the base station equipping (a) $M = 8$ antennas, and (b) $M = 256$ antennas.



(a) Interaction with the actual environment

(b) Interaction with the surrogate model

Fig. 6. The learning experience of the DRL agent when interacting with (a) the actual environment and (b) the surrogate model trained with 1000 data samples.

2) Surrogate model assisted learning: We also evaluate the performance of the surrogate model assisted learning framework, which has the potential of significantly reducing the number of interactions with the actual environment.

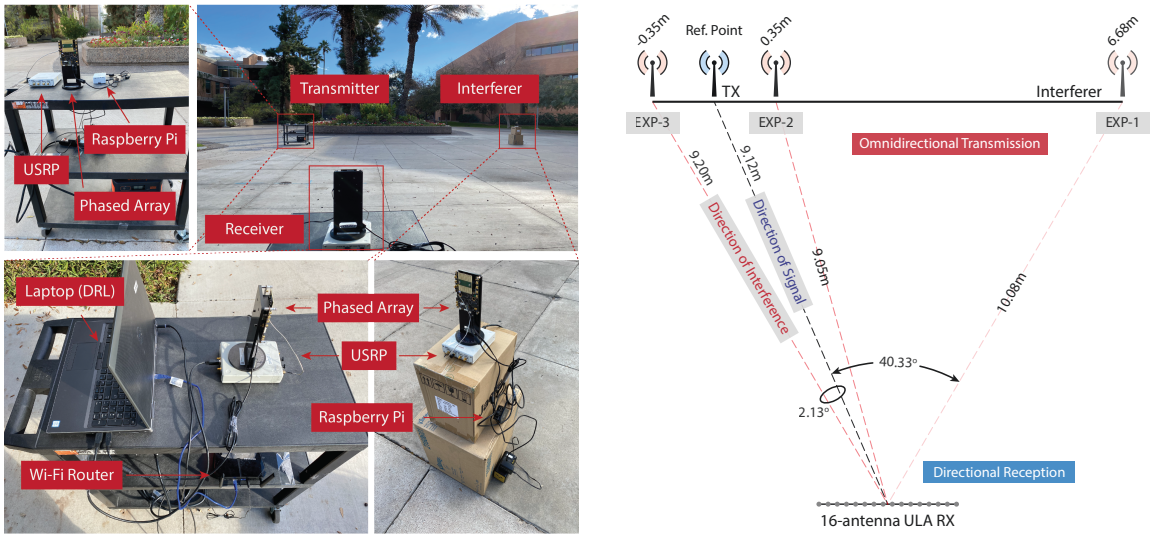
In Fig. 5, we first evaluate the prediction accuracy of the two proposed prediction network architectures, which provides insight on how much data samples are required in order to have a reasonable performance as well as the practicality of the solutions. We show the prediction

accuracy of both the signal power and the interference power. As can be seen, the signal model-based architecture requires much less data samples to achieve higher prediction accuracy than the FC-based architecture trained with much more data samples. For instance, as indicated in Fig. 5(a), with only 50 samples, the signal model-based prediction architecture can achieve even more accurate interference prediction than the FC-based architecture trained with $10,000$ samples. **This saves almost 99.5% of the measurements, yielding a more sample-efficient solution for the practical system deployment.** Moreover, as there are more data samples, the prediction accuracy of the signal model-based architecture also gets improved quite significantly. Such performance is achieved by better leveraging the underlying signal relationships and hence the model parameters are essentially searched over a much smaller space. Finally, by comparing Fig. 5(a) and Fig. 5(b), it can be observed that with more antennas, the system needs to collect more data samples to train the prediction networks in order to maintain the similar prediction accuracy when the number of antennas is small.

The trained surrogate model is utilized to interact with the DRL agent, aiming to reduce the expensive actual measurements conducted by the hardware. In Fig. 6, we show the performance of the DRL agent when interacting with the actual environment as well as the surrogate model. The training of the DRL agent is repeated for 100 times and the average performance as well as the standard deviation are reported in Fig. 6. We test the performance of a system with 8 antennas, and the surrogate model is trained using $1,000$ data samples, i.e., $|\mathcal{D}_{\text{in}}| = |\mathcal{D}_{\text{s}}| = 1000$. As can be seen, the learning experience based on the surrogate model is quite similar to that of the one based on the actual environment. This empirically shows the effectiveness of using the surrogate model in training the DRL agent. As a result, although the DRL agent requires almost a total number of $5,000$ interactions with the environment to converge, **in the surrogate model assisted learning framework, all these interactions are with the surrogate model and hence the expensive evaluations on the real hardware are avoided.**

VII. Real Measurement Results

In this section, we further evaluate the performance of the proposed interference-aware beam pattern learning algorithm in Section IV-B using a real-world mmWave prototyping platform.



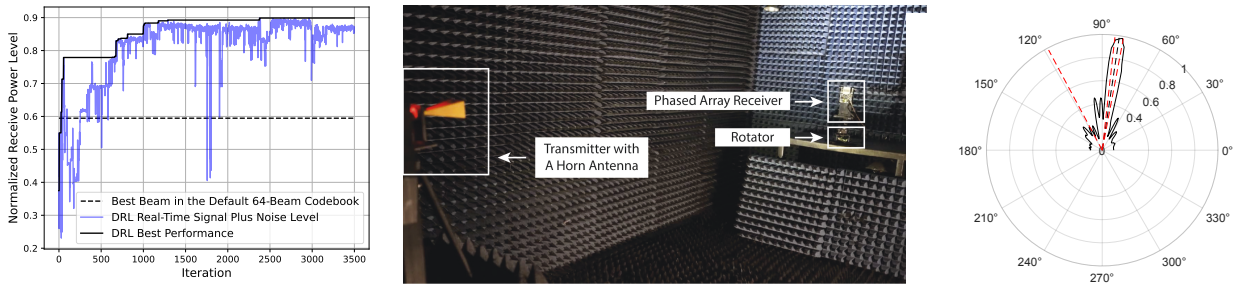
(a) Prototyping setup in an outdoor environment

(b) Measurement campaign

Fig. 7. The prototyping setup and the outdoor measurement environment for evaluating the proposed interference-aware beam pattern design algorithm. The adopted setup consists of a receiver, a desired transmitter and an interferer, as shown in (a). The upper right figure in (a) shows the EXP-1 of the conducted measurement campaign, as depicted in (b), where we provide an illustration of the relative positions of the receiver, transmitter and interferer in the outdoor measurements.

A. Hardware Platform Description

As shown in Fig. 7(a), we build a test platform comprised of a receiver, a transmitter, and an interferer. The radio frontend of all three components is the same type of mmWave phased array, which employs a 16-antenna uniform linear array (ULA) and transmits/receives signals at an operating frequency of 62.64 GHz. The control units of the transmitter and the interferer are identical, while the control unit of the receiver includes a laptop. The laptop is used for several tasks: (i) It controls the phased array at the receiver; (ii) It executes the deep reinforcement learning algorithm; (iii) It connects to a wireless router and can remotely control the transmitter and the interferer. During the measurement, it controls the on/off status of the transmitter. It is worth mentioning that although the transmitter and interferer are equipped with phased arrays, they both transmit signals in an **omnidirectional** way for an effective and fair evaluation of the proposed algorithm. For the phased array at the receiver, only 2 bits are used for the phase encoding of each phase shifter to form the **directional** beam, which means that the signal received by each antenna can only be adjusted with 4 different phase values.



(a) Power Measurement (b) Anechoic Chamber Setup (c) Beam Pattern

Fig. 8. The learning results of the interference-unaware beam pattern, where (a) shows the real-time power measurement, (b) shows the anechoic chamber setup for measuring the learning beam pattern, and (c) shows the learned beam pattern with the black dashed line representing the direction of the desired signal and the red dashed lines representing the directions of the interfering sources which will be presented later.

B. Experiment Description

In this subsection, we describe in detail how the experiments are designed to effectively evaluate the performance of the proposed algorithm. As can be seen in Fig. 7(a), we consider the scenario where both transmitter and interferer have a direct LOS connection with the receiver. To better reflect the interference suppression capability, we first turn on only the transmitter and use the algorithm proposed in [36] to learn the "interference-unaware" beam, which focuses only on maximizing the receive power from the desired transmitter. Then, we turn on the interferer and run the algorithm proposed in Section IV. This forms the "interference-aware" beam, which focuses on maximizing the SINR performance. For a better understanding of the measurement results, we also depict the relative positions of the receiver, transmitter and interferer in Fig. 7(b). It is worth pointing out that, during the measurement, we fixed the positions of the transmitter and receiver, which maintains a distance of 9.12m. We change the position of the interferer to get multiple angle differences that are of interest between the two LOS links. For example, as shown in Fig. 7(b), in the EXP-1, the distance between the receiver and the interferer is 10.08m, and the distance between the transmitter and the interferer is 6.68m, which forms a angle difference of around 40.33° from the receiver's point of view. Finally, we also visualize all the learned beams by measuring their patterns in an anechoic chamber as shown in Fig. 8(b), which provides useful information in understanding and validating the achieved performance.

TABLE I

PERFORMANCE OF THE INTERFERENCE-UNAWARE BEAM

	EXP-1	EXP-2	EXP-3
Angular Separation	40.33°	2.16°	2.13°
Signal Plus Noise Level	0.92	0.86	0.903
Interference Plus Noise Level	0.32	0.915	0.94
SIR Performance	6.96 dB	-0.33 dB	-0.22 dB
INR Performance	-0.47 dB	6.56 dB	6.29 dB

C. Measurement Results

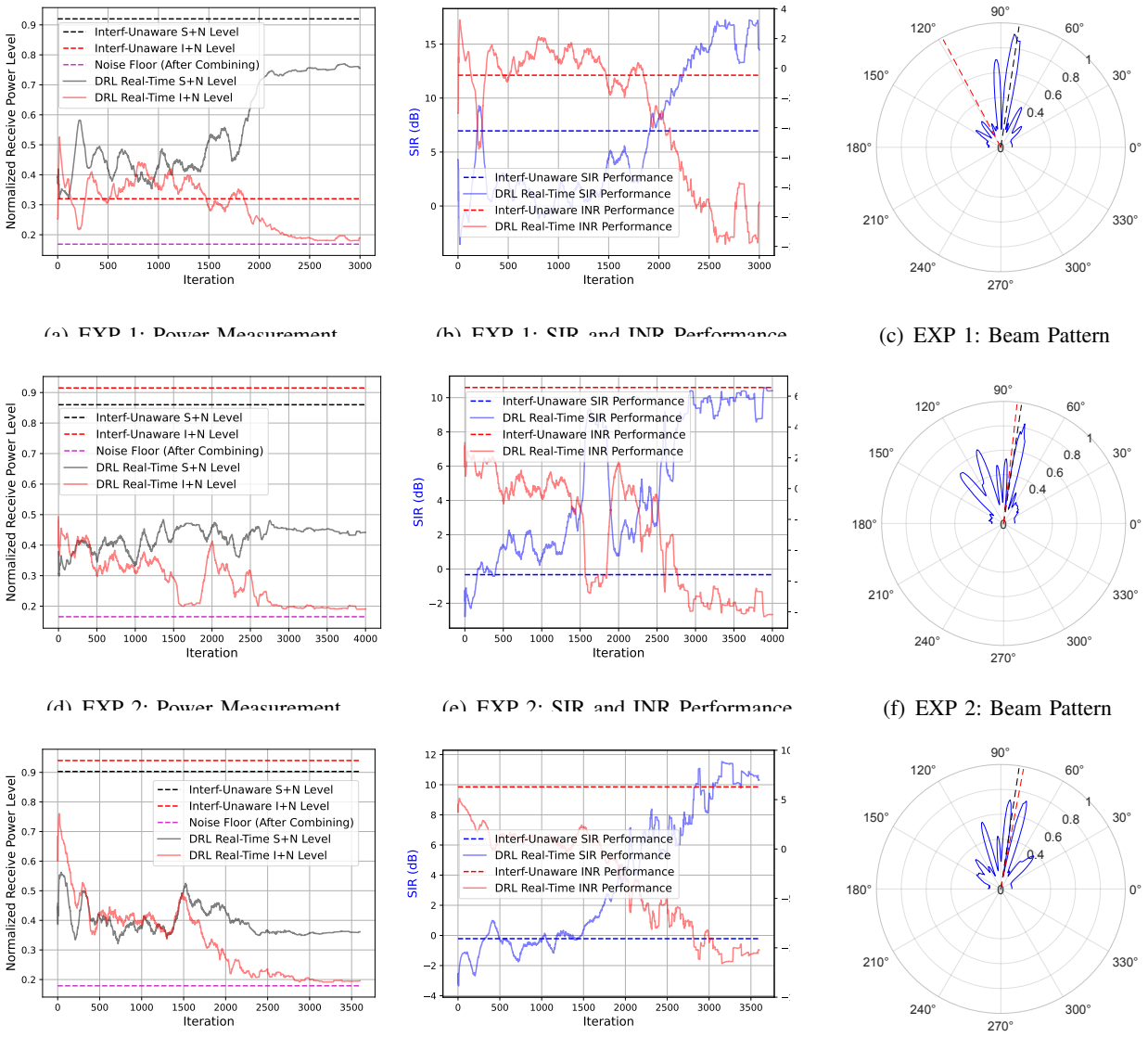
In Fig. 8(a), we plot the learning process of the interference-unaware beam pattern, where the real-time performance of the DRL-based beam pattern learning algorithm is presented. To better understand the learning process and make sure that the learning result is meaningful, we compare the performance achieved by the learned beam with a built-in beam codebook. To be more specific, the phased array adopted in the experiment includes a default codebook that has 64 beams. This codebook is essentially a beamsteering-like codebook with its beams covering -45° to $+45^\circ$ azimuth angular space. In order to find the beam with the best performance, we perform a beam sweeping and calculate the receive power after combining by each of the beams. The one that achieves the highest receive power is determined as the best beam. As can be seen in the Fig. 8(a), the learned beam finally achieves a normalized receiver power of around 0.9, significantly outperforming the best beam in the codebook. Moreover, we also measure the beam pattern of the learned interference-unaware beam (plotted in Fig. 8(c)) in an anechoic chamber as shown in Fig. 8(b). After the interference-unaware beam is learned, the beam weights are saved and the interferer is turned on. We then measure the signal and interference levels (with noise) of this learned interference-unaware beam. It is worth mentioning that the interference levels also depend on the position of the interferer. In our experiments, we select 3 different interferer positions. The measurement results of the interference-unaware beam with the different interferer placements are summarized in Table I.

The performance of the interference-aware beam pattern learning algorithm is then benchmarked with that of the interference-unaware beam. Before we delve into the detailed measurement results, it is worth pointing out that, since both the transmitter and the interferer have a direct LOS connection with the receiver, and given the channel characteristics in the mmWave

bands, it is reasonable to assume that the existing LOS link has stronger power than the other NLOS links (if any) and hence is the dominant factor in the learning process. Furthermore, due to the finite number of antennas at the receiver (which leads to the limited angular resolution), intuitively speaking, what really matters is the angle difference between the transmitter-receiver LOS link and the interferer-receiver LOS link. And in general, the "closer" the transmitter and the interferer are, the harder for the receiver to distinguish between them. Specifically, we classify whether a pair of positions (of transmitter and interferer) being well-separated or not by comparing the angle difference with respect to the receiver with the half-power beam-width (HPBW). For a ULA with half-wavelength antenna spacing, the HPBW can be approximately calculated as $\frac{1.78}{N}$ rad with N being the number of antenna elements [43]. Therefore, the HPBW of our adopted 16-antenna phased array is around $6.37°$. Next, based on the angular separation of the transmitter and the interferer of the considered scenarios (i.e., Fig. 7(b)), we divide the measurement results into two categories and discuss them respectively.

1) When the transmitter and the interferer are well-separated: We first study the case when the transmitter and the interferer are relatively well-separated, i.e., the angular separation is greater than that of the HPBW, in the experiment 1 (EXP 1 in Fig. 7(b)). In Fig. 9(a) and Fig. 9(b), we plot the learning process of the experiment 1, where the angular separation of the transmitter and the interferer is around $40.33°$. As can be seen in the figure, the performance of the interference-unaware beam is actually quite decent, yielding a SIR of 6.96 dB and a INR of -0.47 dB, thanks to the significant angular separation. However, it still introduces a certain level of interference which makes it comparable to the noise level and raises the interference plus noise level noticeably above the noise floor. By contrast, the learned interference-aware beam is able to further suppress the interference to a great extent, making the INR even below -10 dB, i.e., **achieving a nearly 10 dB gain in INR**, while only sacrificing around 10% of the desired signal power. It is also worth mentioning that such performance is achieved **with only** $3,500$ **iterations and without knowing the channels of both the desired transmitter and the interferer.** Such relaxation on the system operations (such as synchronization and channel estimation) makes the proposed solution implementation friendly in most of the practical systems.

2) When the transmitter and the interferer are extremely close: Next, we study the case when the transmitter and the interferer are extremely close, i.e., the angular separation is much smaller than the HPBW, in the experiment 2 and 3 (EXP 2 and 3 in Fig. 7(b)). The angular separations of the transmitter and the interferer in both experiments are around $2°$, which is only one third of



(a) EXP 1: Power Measurement (b) EXP 1: SIR and INR Performance (c) EXP 1: Beam Pattern

(d) EXP 2: Power Measurement (e) EXP 2: SIR and INR Performance (f) EXP 2: Beam Pattern

(g) EXP 3: Power Measurement (h) EXP 3: SIR and INR Performance (i) EXP 3: Beam Pattern

Fig. 9. The measurement results of the three experiments illustrated in Fig. 7(b), where the first column of figures shows the real-time receive power measurements and the second column of figures shows the corresponding SIR and INR performance. All these results are processed with a moving average of 100 samples to smooth out the effect of noise. Finally, the third column of figures shows the learned interference-aware beam patterns with the black dashed line representing the direction of the desired signal and the red dashed line representing the direction of the interfering source.

the HPBW of the adopted phased array receiver. It turns out that the proposed algorithm is still quite capable of suppressing the interference level. As can be seen in Fig. 9(e) and Fig. 9(h), the SIR in both experiments all finally reaches over 10 dB, and the INR level is also reduced to -8 dB and -10 dB respectively, **achieving almost the similar performance when the transmitter**

and the interferer are well-separated. However, different from the previous case, such great SIR and INR performances are traded with the significant sacrifices of the desired signal power. As indicated in Fig. 9(d) and Fig. 9(g), the signal power is only around 50% of that achieved in experiment 1, for instance. This also implies that **when the directions of the signal and interference are well aligned, the system normally needs to strike a delicate balance between SIR and SNR performances, in order to yield a meaningful SINR value.** Such observation is also empirically confirmed by the measured learned beam patterns. As can be seen from Fig. 9(f) and Fig. 9(i), the receiver intelligently shapes deep nulls towards the directions of the interference, which explains the achieved well interference suppression capability. However, as a compromise, the main-lobes of the beams are no longer pointing towards the desired transmitter, leaving only the side of the main-lobes leveraged to serve the target transmitter. This makes the receive signal power much weaker than that of the interference-unaware beam. In summary, the real-world prototype confirms the effectiveness and robustness of the proposed solution in learning interference nulling beam patterns based solely on the power measurements. It also shows the promising gains brought by the intelligent online beam learning solution in realistic scenarios when compared with the off-the-shelf beams.

VIII. Conclusion and Future Work

In this paper, we developed a sample-efficient online reinforcement learning based approach that can efficiently learn interference-aware beams. The proposed solution learns how to design beam patterns that can effectively manage interference, relying only on the 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. The results show that the proposed solution is capable of shaping nulls towards the interfering directions while maximizing the reception quality of the desired signal. When tested on a hardware proof-of-concept prototype based on real-world measurements, the proposed interference-aware beam learning framework also demonstrating efficient beam pattern optimization performance. Specifically, the developed solution was shown to improve the SNR and INR performance by at least 10 dB compared to the interference-unaware beams in all the tested scenarios. This is particularly important when the interferer is close to the transmitter. These SNR/INR gains can be translated to more than double the data rate in the considered scenarios. For the future work, it is important to extend the proposed framework

to network settings with multiple basestation and decentralized beam learning capabilities. It is also interesting to investigate the application of the developed beam learning approaches to full-duplex networks and cell-free massive MIMO architectures.

REFERENCES

[1] Y. Zhang and A. Alkhateeb, "Online Beam Learning for Interference Nulling in Hardware-Constrained mmWave MIMO Systems," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*, 2022.

[2] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[3] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid Beamforming for Massive MIMO: A Survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[4] Y. Heng, J. G. Andrews, J. Mo, V. Va, A. Ali, B. L. Ng, and J. C. Zhang, "Six Key Challenges for Beam Management in 5.5G and 6G Systems," *IEEE Communications Magazine*, vol. 59, no. 7, pp. 74–79, 2021.

[5] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming Intercell Interference in Indoor Wireless Systems," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, 2007, pp. 83–87.

[6] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On Feasibility of Interference Alignment in MIMO Interference Networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4771–4782, 2010.

[7] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 69–73, 2007.

[8] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "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.

[9] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.

[10] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-Free Massive MIMO Versus Small Cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.

[11] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, "Local Partial Zero-Forcing Precoding for Cell-Free Massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4758–4774, 2020.

[12] R. Zhang and Y.-C. Liang, "Exploiting Multi-Antennas for Opportunistic Spectrum Sharing in Cognitive Radio Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 88–102, 2008.

[13] B. Li, A. P. Petropulu, and W. Trappe, "Optimum Co-Design for Spectrum Sharing between Matrix Completion Based MIMO Radars and a MIMO Communication System," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4562–4575, 2016.

[14] J. Qian, M. Lops, L. Zheng, X. Wang, and Z. He, "Joint System Design for Coexistence of MIMO Radar and MIMO Communication," *IEEE Transactions on Signal Processing*, vol. 66, no. 13, pp. 3504–3519, 2018.

[15] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[16] 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.

[17] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, 2016.

[18] 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.

[19] F. Sohrabi and W. Yu, "Hybrid Digital and Analog Beamforming Design for Large-Scale Antenna Arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, 2016.

[20] J. Zhan and X. Dong, "Interference Cancellation Aided Hybrid Beamforming for mmWave Multi-User Massive MIMO Systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2322–2336, 2021.

[21] K. Satyanarayana, M. El-Hajjar, P.-H. Kuo, A. Mourad, and L. Hanzo, "Hybrid Beamforming Design for Full-Duplex Millimeter Wave Communication," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1394–1404, 2019.

[22] I. P. Roberts, J. G. Andrews, and S. Vishwanath, "Hybrid Beamforming for Millimeter Wave Full-Duplex Under Limited Receive Dynamic Range," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 7758–7772, 2021.

[23] L. Zhu, J. Zhang, Z. Xiao, X. Cao, X.-G. Xia, and R. Schober, "Millimeter-Wave Full-Duplex UAV Relay: Joint Positioning, Beamforming, and Power Control," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2057–2073, 2020.

[24] Y. Zhang, M. Xiao, S. Han, M. Skoglund, and W. Meng, "On Precoding and Energy Efficiency of Full-Duplex Millimeter-Wave Relays," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1943–1956, 2019.

[25] H. J. Kwon, J. H. Lee, and W. Choi, "Machine Learning-Based Beamforming in K-User MISO Interference Channels," *IEEE Access*, vol. 9, pp. 28 066–28 075, 2021.

[26] J. Shi, W. Wang, X. Yi, X. Gao, and G. Y. Li, "Deep Learning-Based Robust Precoding for Massive MIMO," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7429–7443, 2021.

[27] F. Sohrabi, K. M. Attiah, and W. Yu, "Deep Learning for Distributed Channel Feedback and Multiuser Precoding in FDD Massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4044–4057, 2021.

[28] M. Fozi, A. R. Sharafat, and M. Bennis, "Fast MIMO Beamforming via Deep Reinforcement Learning for High Mobility mmWave Connectivity," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 127–142, 2022.

[29] Z. Zhang, J. Hou, X. Chu, H. Zhou, G. Wei, and J. Zhang, "Multi-Agent Deep Reinforcement Learning Based Downlink Beamforming in Heterogeneous Networks," *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 4247–4263, 2023.

[30] Y. Li, K. Li, L. Cheng, Q. Shi, and Z.-Q. Luo, "Digital Twin-Aided Learning to Enable Robust Beamforming: Limited Feedback Meets Deep Generative Models," in *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2021, pp. 26–30.

[31] A. Masaracchia, V. Sharma, B. Canberk, O. A. Dobre, and T. Q. Duong, "Digital Twin for 6G: Taxonomy, Research Challenges, and the Road Ahead," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2137–2150, 2022.

[32] L. U. Khan, W. Saad, D. Niyato, Z. Han, and C. S. Hong, "Digital-Twin-Enabled 6G: Vision, Architectural Trends, and Future Directions," *IEEE Communications Magazine*, vol. 60, no. 1, pp. 74–80, 2022.

[33] A. Alkhateeb, S. Jiang, and G. Charan, "Real-Time Digital Twins: Vision and Research Directions for 6G and Beyond," 2023.

[34] S. Jiang and A. Alkhateeb, "Digital Twin Based Beam Prediction: Can we Train in the Digital World and Deploy in Reality?" 2023.

[35] 3GPP, "Study on architecture enhancements for 3GPP support of advanced Vehicle-to-Everything (V2X) services; Phase 2," Tech. Rep. TR 23.776, 2021, version 17.0.0.

[36] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Reinforcement Learning of Beam Codebooks in Millimeter Wave and Terahertz MIMO Systems," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 904–919, 2022.

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

[38] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1748–1759, 2010.

[39] 3GPP, "Study on new radio access technology: Physical layer aspects," Tech. Rep. 38.802, 2017, version 14.2.2.

[40] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[41] 3GPP, "Requirements for support of radio resource management," Tech. Rep. 38.133, 2020, version 16.4.0.

[42] [Online]. Available: https://github.com/YuZhang-GitHub/Interference_Nulling_Beamforming.

[43] M. A. Richards, *Fundamentals of radar signal processing*. McGraw-Hill Education, 2014.