Lizarraga, Enrique M.; Maggio, Gabriel N.; Dowhuszko, Alexis A.

Deep reinforcement learning for hybrid beamforming in multi-user millimeter wave wireless systems

Published in:
Proceedings of IEEE 93rd Vehicular Technology Conference, VTC 2021

DOI:
10.1109/VTC2021-Spring51267.2021.9449053

Published: 15/06/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Deep reinforcement learning for hybrid beamforming in multi-user millimeter wave wireless systems

Enrique M. Lizarraga[1], Gabriel N. Maggio[1], Alexis A. Dowhuszko[2], *Senior Member, IEEE*
[1]Digital Communications Research Laboratory, National University of Cordoba, Argentina
[2]Department of Communications and Networking, Aalto University, 02150 Espoo, Finland
Email: emlizarraga@unc.edu.ar; gabriel.maggio@unc.edu.ar; alexis.dowhuszko@aalto.fi

***Abstract*—This paper proposes a Machine Learning (ML) algorithm for hybrid beamforming in millimeter-wave wireless systems with multiple users. The time-varying nature of the wireless channels is taken into account when training the ML agent, which identifies the most convenient hybrid beamforming matrix with the aid of an algorithm that keeps the amount of signaling information low, avoids sudden changes in the analog beamformers radiation patterns when scheduling different users (flashlight interference), and simplifies the hybrid beamformer update decisions by adjusting the phases of specific analog beamforming vectors. The proposed hybrid beamforming algorithm relies on *Deep Reinforcement Learning* (DRL), which represents a practical approach to embed the online adaptation feature of the hybrid beamforming matrix into the channel states of continuous nature in which the multiuser MIMO system can be. Achievable data rate curves are used to analyze performance results, which validate the advantages of DRL algorithms with respect to solutions relying on conventional/deterministic optimization tools.**

***Index Terms*—Machine learning; Hybrid beamforming; Millimeter Wave; Deep reinforcement learning; Multiuser MIMO.**

I. Introduction

Hybrid beamforming is an attractive technique to exploit the spatial degrees-of-freedom that a Multiple-Input Multiple Output (MIMO) wireless system offers in *millimeter wave* bands [1]. Hybrid beamforming can compensate the strong attenuation that millimeter wave signals experience [2], [3], enabling a reliable link-level connectivity while demanding a moderate hardware implementation complexity. The hybrid beamforming matrix can be divided into two parts: The digital precoder matrix and the analog beamforming vectors.

The size of the digital precoder matrix depends on the number of RF chains of the hybrid beamforming architecture, and its elements can be complex-valued provided its Frobenius norm is unitary. On the other hand, the elements of the analog beamforming vectors have less flexibility, and can adjust the antenna phases but not the amplitude weights. This way, the implementation complexity is kept low, even when the size of the transmit/receive antenna arrays grows. Unfortunately, the most convenient way to exploit the degrees-of-freedom that are enabled by the analog beamforming vectors remains an open problem, particularly in presence of a massive MIMO system with multiple users. Therefore, an iterative way to approximate the most convenient hybrid beamforming matrix is needed, where the adjustment of the analog beamformer elements can be done in a distributed way, using the actions informed by each receiver when being scheduled by the transmitter.

Machine learning (ML) algorithms have been proposed to solve the different optimization problems that emerge in wireless systems [4]. Many variants of Reinforcement Learning (RL) algorithms have been proposed, providing interesting performance results when used in wireless communication networks. The efficiency of ML improves notably when using the *deep learning* approach, where a neural network is used to identify hidden features/characteristics of the observation set, finding relationships between input and output variables to make decisions that are highly-rated under the target *figure of merit*. Deep Reinforcement Learning (DRL) algorithm *learns* from variables that are fed as input, making reliable predictions of the figure of merit that would be obtained as output.

In this paper, a DRL algorithm is proposed to estimate the most convenient phase update in the analog beamforming vectors to maximize the data rate of the scheduled user. The DRL algorithm knows the instantaneous phase status of the analog beamformer, as well as the achievable data rate of the lower-dimension MIMO channel that results when using this analog beamformer in transmission. Based on this, the DRL agent in the scheduled MIMO receiver identifies the most convenient action, and feeds this information back to the multiuser MIMO transmitter. After updating the phase in the analog beamforming vector elements marked in the action, the digital precoder (transmitter) and digital decoder (receiver) are updated to maximize the achievable data rate.

Similar research was done in [5], [6], where a RL algorithm with *tabular implementation* was considered. The main drawback of tabular RL is in its implementation complexity, which grows notably with the size of the MIMO scenario. Therefore, a DRL algorithm is considered instead, using a neural network that is easier to train (and assess) in a high-dimensional problem. Although a DRL algorithm was proposed in [7] to identify the most convenient hybrid beamforming matrix, its presented analysis was only addressed for single-user MIMO. Finally, since this paper focuses on a multiuser MIMO scenario, the DRL processing has been adapted to be performed distributively in the MIMO receivers, feeding back the actions to be performed in each transmission time interval.

The rest of this paper is organized as follows: Section II presents the system model and the details of the hybrid beamforming implementation. Section III derives the DRL algorithm to find the analog beamformers and digital precoders. Simulation setting and obtained results are discussed in Section IV. Finally, conclusions are drawn in Section V.

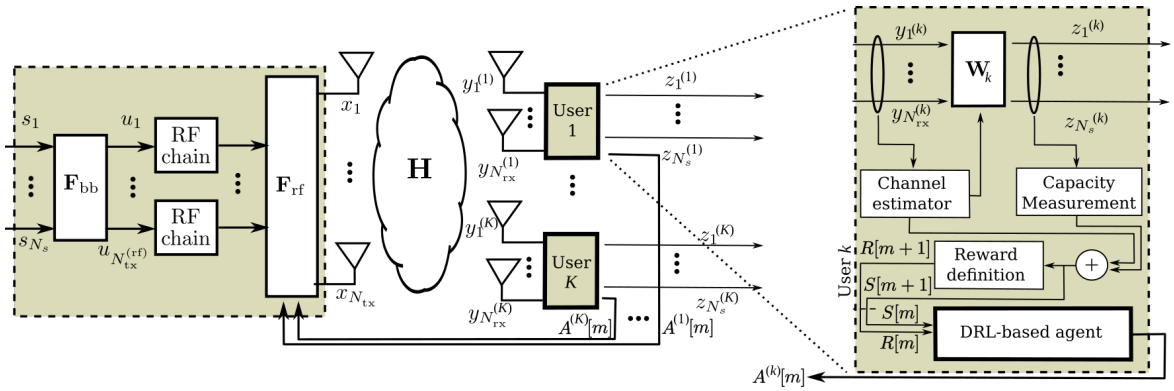

Fig. 1. Hybrid beamforming architecture for a large-scale MIMO system deploying $N_{\rm tx}$ and $N_{\rm rx}$ antennas in the transmitter and receiver, respectively. A codebook-based analog beamformer ($\mathbf{F}_{\rm rf}$) and a fully-adaptive digital precoder ($\mathbf{F}_{\rm bb}$) are used in transmission to transport $N_{\rm s}$ symbol streams. In reception, a fully-digital combiner ($\mathbf{W}_k$) is considered per user k. The DRL-based agent is implemented in the receiver, where the state information $S[m]$ and reward estimation $R[m]$ are used to estimate the most convenient action $A[m]$ on the instantaneous analog beamformer $\mathbf{F}_{\rm rf}$ while scheduling transmission to user k.

II. System Model

The multiuser massive MIMO system under analysis is presented in Fig. 1, where communication takes place from the base station (transmitter) to the mobile stations (receivers) with index k in set $\mathcal{K}$. In the transmitter, a digital precoding matrix $\mathbf{F}_{\rm bb}$ and analog beamforming matrix $\mathbf{F}_{\rm rf}$ of size $N_{\rm tx}^{\rm (rf)} \times N_{\rm s}$ and $N_{\rm tx} \times N_{\rm tx}^{\rm (rf)}$, respectively, define the hybrid beamforming matrix. In reception, a fully-digital combiner $\mathbf{W}_k$ of size $N_{\rm s} \times N_{\rm rx}$ is considered. The MIMO wireless channel for receiver k is described by a complex matrix $\mathbf{H}_k$ of size $N_{\rm rx} \times N_{\rm tx}$, whose coefficients have strong correlation and are modelled following the guidelines in [8]. Strong attenuation in millimeter wave frequency bands can be compensated in part using large antenna arrays. Simultaneously, the use of multiple antennas also calls for sum data rate optimization functions, exploiting the spatial multiplexing dimension that emerges.

In this paper, we assume a block fading channel model (*i.e.*, coefficients of $\mathbf{H}_k$ remain fixed during each transmission time interval, and vary independently between consecutive transmission time intervals). Without loss of generality, we also assume a flat frequency response in the whole millimeter wave communication bandwidth, and that the hybrid beamforming architecture is implemented only in the base station (transmitter) for the downlink direction of communication. As mobile stations (receivers) should have a moderate number of antennas ($N_{\rm rx} \ll N_{\rm tx}$), hybrid receive beamforming architectures are not considered, and received signal samples from the different antennas are entirely processed in the digital domain.

Let us first start from the classical *full-digital* beamforming problem, in which precoding matrix $\mathbf{F}$ and combining matrix $\mathbf{W}$ of size $N_{\rm tx} \times N_{\rm s}$ and $N_{\rm s} \times N_{\rm rx}$, respectively, should be determined. In this case, the coefficients of these matrices can be obtained using Singular Value Decomposition (SVD), *i.e.*,

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\rm H}, \tag{1}$$

where $\mathbf{V}$ and $\mathbf{U}^{\rm H}$ are unitary matrices whose columns contain the coefficient of the transmit precoding and receive combining vectors, respectively. Here, the elements of diagonal matrix $\mathbf{\Sigma}$ contain the singular values σ_n of the spatial streams associated with transmit precoding vector $\mathbf{v}_n$ and receive combining vector $\mathbf{u}_n$. Then, when $N_{\rm s}$ data streams are multiplexed in the spatial domain, the optimal transmit precoding matrix becomes

$$\mathbf{F}^{\rm (opt)} = \left[\mathbf{v}_1 \cdots \mathbf{v}_{\rm N_s}\right], \tag{2}$$

where $\mathbf{v}_n$ represents the full-digital transmit precoding vector of size $N_{\rm tx} \times 1$, which corresponds to the n-th (strongest) singular value σ_n of $\mathbf{H}$. Similarly, $\mathbf{W}$ is defined by $\left[\mathbf{u}_1 \cdots \mathbf{u}_{\rm N_s}\right]$, where $\mathbf{u}_n$ is a column vector obtained from unitary matrix $\mathbf{U}$.

The full-digital approach becomes impractical when $N_{\rm tx}$ grows large. In this situation, hybrid beamforming matrix

$$\mathbf{F}^{\rm (hybrid)} = \mathbf{F}_{\rm rf}\,\mathbf{F}_{\rm bb} \tag{3}$$

can be used instead, which combines the effect of the digital precoding matrix $\mathbf{F}_{\rm bb}$ and the analog beamforming matrix $\mathbf{F}_{\rm rf}$.

A. Analog beamforming matrix

The columns of the analog beamforming matrix (*i.e.*, the analog beamforming vectors) are codebook-based, such that the beamforming weights in each antenna can only take discrete values from a uniform quantization set [9]. On the other hand, the coefficients of the digital precoding matrix are not restricted to belong to a codebook and, in principle, can take any complex number with Frobenius norm $\|\mathbf{F}_{\rm bb}\|_{\rm F} = 1$.

The hybrid beamforming algorithm in this paper assumes that the phase adjustments per transmit antenna can only take two possible values: $\theta_{i,j} \in \{-\pi/2, +\pi/2\}$ for $i = 1, \ldots, N_{\rm tx}$ and $j = 1, \ldots, N_{\rm tx}^{\rm (rf)}$. This is done without loss of generality, and can be easily extended when a larger number of phase adjustments per transmit antenna is allowed [10]. Then, the goal of the proposed hybrid beamforming algorithm is to identify the most convenient digital precoder $\mathbf{F}_{\rm bb}$ and analog beamformer $\mathbf{F}_{\rm rf}$, such that the elements of $\mathbf{F}_{\rm rf}$ attain the form

$$f_{i,j} = \sqrt{1/N_{\rm tx}}\exp\left(\mathrm{j}\,\theta_{i,j}\right) \quad i = 1, \ldots, N_{\rm tx}; j = 1, \ldots, N_{\rm tx}^{\rm (rf)}. \tag{4}$$

Given the definitions of $\{f_{i,j}\}$ and $\{\theta_{i,j}\}$, it is highlighted that every candidate solution for $\mathbf{F}_{\rm rf}$ is defined on a binary

basis, which is a characteristic that is exploited in the proposed scheme. As analyzed in [5], the equivalent problem with objective function (3) has a very high dimension when dealing with antenna numerology that is relevant to millimeter wave.

B. Digital precoding matrix

As explained [5], $\mathbf{F}_{\mathrm{rf}}$ can be used to identify a lower-order wireless channel $\widetilde{\mathbf{H}}$ of size $N_{\mathrm{rx}} \times N_{\mathrm{tx}}^{(\mathrm{rf})}$, which results after combining the full-order wireless channel matrix $\mathbf{H}$ with the selected analog beamforming vector. After applying SVD,

$$\widetilde{\mathbf{H}} = \mathbf{H}\,\mathbf{F}_{\mathrm{rf}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^{\mathrm{H}} \tag{5}$$

results, where $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ are unitary matrices of reduced order. Then, the strongest eigenvalues are identified to extract column vectors of $\widetilde{\mathbf{V}}$, to construct the full-digital precoding matrix

$$\mathbf{F}_{\mathrm{bb}} = \left[\widetilde{\mathbf{v}}_1 \ldots \widetilde{\mathbf{v}}_{\mathrm{N_s}}\right] \tag{6}$$

that is used by the hybrid beamforming architecture.

Note that the lower-order equivalent channel $\widetilde{\mathbf{H}}$ defines the upper-bound performance of the system, while the estimated data rate in reception is used as figure of merit to adjust the hybrid beamformer. As expected, the selected $\mathbf{F}_{\mathrm{rf}}$ affects the achievable data rate. Then, the essence of this proposal is to define $\mathbf{F}_{\mathrm{rf}}$ with the aid of the DRL algorithm, and compute $\mathbf{F}_{\mathrm{bb}}$ (and $\mathbf{W}$) using classical (full-digital) methods.

III. Deep reinforcement learning algorithm to be implemented in the MIMO receiver

A DRL-based algorithm is implemented in each of the mobile stations (receivers) of the multi-user MIMO system, with the aim of identifying the most convenient variation of the analog beamforming vector that the base station (transmitter) should apply for data rate maximization. Like in any other RL algorithm, two main components are distinguished: The *agent* and the *environment* [11]. The agent includes inherently the learning algorithm, while the environment models the effect of the wireless channel on the problem to be solved (*i.e.*, optimizing the hybrid beamforming matrix in this multiuser massive MIMO system). The relationships between the processing parts in the whole scenario are illustrated in Fig. 1. For this RL algorithm, the environment includes all the blocks shown in Fig. 1, except from the block that wraps the agent.

The interaction between agent and environment occurs through few signals which are known as *action*, *state* and *reward*, represented by $A[m]$, $S[m]$ and $R[m]$, respectively, where m refers to the trial and error iterations that are followed by the algorithm [11]. The feedback loop between mobile stations (receivers) and base station (transmitter) is also shown explicitly in Fig. 1, as well as the approach that was followed to place the DRL agent in the receivers. Note that this is a key difference with respect to other RL proposals, such as the one presented in [5]. The new proposed architecture simplifies the channel estimation procedure, and enables the simplification of the feedback signal $A[m]$ [1]. Another difference with respect to [5] is the way in which the channel state information is observed, which is done on the reduced-size wireless channel $\widetilde{\mathbf{H}}$ and is used to generate the *state* signal $S[m]$.

[1]In case of a multiuser scenario, the k-th user sends the feedback signal $A^{(k)}[m]$ when scheduled by the base station in its transmission time interval.

A. Definition of Actions, State and Reward for the specific Deep Reinforcement Learning algorithm

Actions are elementary stimulus that the agent feeds into the environment to observe variations on its state. In our case, actions are applied to the phases $\theta_{i,j}$ defined in (4). Due to that, the actions are defined as a variable

$$A[m] \in \{0, \ldots, 2N_{\mathrm{tx}}^{(\mathrm{rf})}N_{\mathrm{tx}} - 1\}, \tag{7}$$

which indicates to set either on 0 or 1 the selected coefficients of $\mathbf{F}_{\mathrm{rf}}$. The possibility of setting the feedback indicator to 0 or 1 justifies factor 2 in (7). Given this definition, a feedback channel with $\log_2(2N_{\mathrm{tx}}^{(\mathrm{rf})}N_{\mathrm{tx}})$ bits per update signaling is needed to report $A[m]$ from the receiver to the transmitter.

The state information is obtained from two observations, as suggested in Fig. 1. On one hand, the output of the channel estimator block is utilized. On the other hand, an estimation of the achievable data rate in reception is also used. Finally, the reward is empirically defined using the following metric:

$$R[m] = \begin{cases} 1 & \text{if } (C[m] - C[m-1]) > 0 \\ -1 & \text{otherwise} \end{cases}, \tag{8}$$

where $C[m]$ represents the estimated data rate for the current MIMO system status (snapshot), while $C[m-1]$ represents the achievable data rate before $A[m]$ is sent to the transmitter. It is highlighted that this definition avoids the use of the so called *shaped* rewards, reducing the implementation complexity. In addition, this reward definition facilitates the future exploration of the attainable performance when using other RL agents proposed in the literature, which are complementary algorithmic variants with respect to the reward definition used in this paper.

The operation of the intended hybrid beamforming system is based on raising a certain number of iterations, in which the wireless channel $\mathbf{H}$ remains fixed (*i.e.*, coherence time is assumed larger than transmission time interval). Additionally, a limited number of iterations is used to enable slow variations on the transmitter and receiver beams. Based on these considerations, the tracking of the interference pattern that the target multiuser MIMO system creates on adjacent cells is simplified. Then, the DRL algorithm makes a more efficient use of the limited adjustments allowed in $A[m]$, with $m = 0, \ldots, M-1$, where M is the maximum number of iterations.

B. Deep Reinforcement Learning agent setting

The DRL agent in this paper uses a neural network with three layers, where full connection between layers is defined, activation functions are set as rectified linear unit (ReLU), and *buffer replay* is implemented following the conventional way. This processing seeks to establish general conditions to test this strategy, while more specific implementation of the DRL agent could provide improved performance results, *e.g.*, by considering *multi-agents* in RL.

The proposed DRL algorithm solves some implementation challenges that the tabular RL implementation proposed in [5]

had. For example, the channel state information is now directly obtained from the reduced-size equivalent channel $\widetilde{\mathbf{H}}$, and this information is treated in a more efficient way. Unlike in [5], where the channel state information showed its effect by means of the reward, in this algorithm the channel is tracked by means of the reward signal and the state signal. Additionally, the present algorithm uses the *deep learning* paradigm to extract characteristics of the problem that cannot be explicitly modelled in closed form. The multiple layers of the neural network are mainly responsible of this acquisition.

The refactoring of the RL engine used in this paper outperforms the previous tabular implementation in two ways:
(a) It extends the practical importance of the proposed strategy since, based on our previous paper simulations, the tabular version can only handle scenarios with $N_{\mathrm{tx}} = 9$ antennas before computational problems start to affect the convergence.
(b) It enables the approximation-based framework of DRL, extracting underlying problem characteristics and *acting* more conveniently with channel states not visited during the training. The tabular RL implementation has a finite number of rows to use when a new channel state appear. The neural network estimation, in contrast, is obtained without table search.

IV. Simulation scenario and analysis of results

The behavior of the proposed algorithm is analyzed by means of three reference scenarios, namely: low ($N_{\mathrm{tx}} = 9$), moderate ($N_{\mathrm{tx}} = 16$), and high number of transmit antennas ($N_{\mathrm{tx}} = 25$), which aims at estimating the attainable performance of the proposed DRL-based hybrid beamforming algorithm in multiuser massive MIMO scenarios. The Saleh-Valenzuela geometric channel model [8] is employed with

$$\mathbf{H}(t) = \gamma \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{j=1}^{N_{\mathrm{ray}}} \alpha_{i,j} \Lambda_{\mathrm{r}}(\phi_{i,j}^{\mathrm{r}}, \theta_{i,j}^{\mathrm{r}}) \Lambda_{\mathrm{t}}(\phi_{i,j}^{\mathrm{t}}, \theta_{i,j}^{\mathrm{t}}) \mathbf{a}_{\mathrm{r}}(\phi_{i,j}^{\mathrm{r}}, \theta_{i,j}^{\mathrm{r}}) \mathbf{a}_{\mathrm{t}}^{\mathrm{H}}(\phi_{i,j}^{\mathrm{t}}, \theta_{i,j}^{\mathrm{t}}), \quad (9)$$

being a matrix that models the channel state where $N_{\mathrm{cl}} = 8$ and $N_{\mathrm{ray}} = 10$ is set, γ is a normalization factor, and complex gain is denoted with coefficients $\alpha_{i,j} \sim \mathcal{CN}(0,1)$. For each propagation path j associated with the i-th cluster, the azimuth angles of arrival and departure are represented $\phi_{i,j}^{\mathrm{r}}$ and $\phi_{i,j}^{\mathrm{t}}$, respectively, whereas the elevation angles of arrival and departure are represented by $\theta_{i,j}^{\mathrm{r}}$ and $\theta_{i,j}^{\mathrm{t}}$. The described angles are supposed to belong to a Laplacian distribution with an angle deviation of 7.5°, centered at an uniformly distributed mean cluster angle of 0° and 90° for azimuth and elevation, respectively. Additionally, normalized planar array response and antenna element gain at the receiver (transmitter) side, is represented by $\mathbf{a}_{\mathrm{r}}(\phi_{i,j}^{\mathrm{r}}, \theta_{i,j}^{\mathrm{r}})$ ($\mathbf{a}_{\mathrm{t}}(\phi_{i,j}^{\mathrm{t}}, \theta_{i,j}^{\mathrm{t}})$) and $\Lambda_{\mathrm{r}}(\phi_{i,j}^{\mathrm{r}}, \theta_{i,j}^{\mathrm{r}})$ ($\Lambda_{\mathrm{t}}(\phi_{i,j}^{\mathrm{t}}, \theta_{i,j}^{\mathrm{t}})$), respectively, for all rays indexes j and cluster indexes i, assuming a $\lambda/2$-spacing between antenna elements. The uniform planar array model has been taken from [5].

For the receiver, a configuration with a low number of antennas is assumed (*i.e.*, $N_{\mathrm{rx}} = 4$). Following the millimeter wave channel model reported in [5], two spatial streams are multiplexed (*i.e.*, $N_{\mathrm{s}} = 2$). The maximum number of iterations is set to $M = 12$, in order to limit the number of times that the direction of the analog beamforming beam can vary when scheduling a given user. Since every user of the multiuser system should track the interference beam, this constraint simplifies the problem notably. Moreover, since a single phase bit of $\mathbf{F}_{\mathrm{rf}}$ can be modified in every iteration, mitigating the impact of flashlight interference in adjacent cells [12].

TABLE I
Agent Parameters for the DRL algorithm

First layer coefficients	200
Second layer coefficients	200
Third layer coefficients	200
ε-greedy value for Q-learning	0.3 to 0
Learn rate α in Q-learning	0.3
Neural networks adjustment step	10^{-4} and 10^{-8}

The Adam optimization algorithm has been used to adjust neural networks, reading mini-batches [13] of size equal to the number of coefficients in the first layer of the neural network. Other parameters of the experiment are summarized in Table I.

The first simulation case is proposed for $N_{\mathrm{tx}} = 9$, and its achievable data rate performance is shown in Fig. 2. Five different beamforming algorithms are analyzed in this figure, namely: baseline algorithms for upper and lower bound performance, as well as three hybrid beamforming schemes. The upper bound is obtained by means of an analog beamformer that comes from approximating the continuous-valued optimal precoder (*i.e.*, the eigenvector computed from the SVD) with a quantized vector, where all the gains are adjusted to the same value to provide unitary Frobenius norm, and the phases are the nearest element $\theta_{i,j}$ in the quantization set. The lower bound is obtained using a completely random digital precoder, where the beamforming beams are pointed into random directions, regardless the channel state information of the user. The next scheme under analysis is the one that uses hybrid beamforming with a random analog beamformer, combined with a digital precoder, which is computed from the equivalent channel model stated in (5). A tabular implementation is also included as a reference in this comparison case. Finally, the performance behavior of the proposed hybrid beamforming algorithm is also presented, which uses DRL to define/update the analog beamformer, and is accompanied by a digital precoder computed according to (5). Note that a gain on the data rate of the hybrid beamforming scheme is observed in this latter case, which becomes more notable for high Signal-to-Noise Ratio (SNR).

The achievable data rates in the second simulation scenario, which considers the same four hybrid beamforming algorithms listed above with $N_{\mathrm{tx}} = 16$, are shown in Fig. 3. Note that in this situation, the benefits of using the proposed DRL-based hybrid beamforming algorithm to identify the most convenient analog beamforming vector becomes more visible. This effect on performance gain becomes more notable when the multiuser MIMO system operates in high SNR regimes.

Finally, in Fig. 4 shows the performance of the proposed algorithm when the number of transmit antennas in the base station grows large (*i.e.*, when $N_{\mathrm{tx}} = 25$). By means of the val-

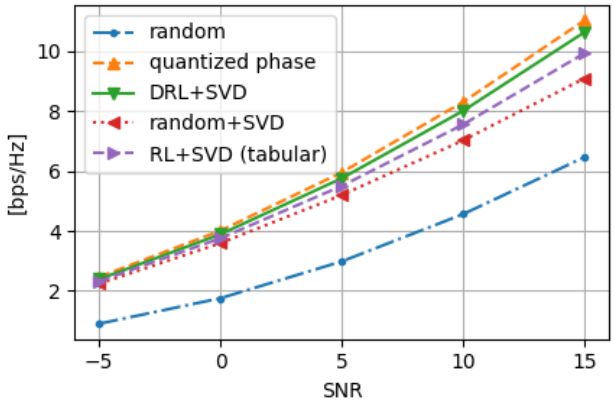

Fig. 2. Achievable data rate as function of the mean SNR for different analog beamformer selection methods with $N_{\mathrm{s}} = 2$. The millimeter wave channel gains were generated according to the extended Saleh-Valenzuela geometric channel model assuming $N_{\mathrm{tx}} = 9$ and $N_{\mathrm{rx}} = 4$.

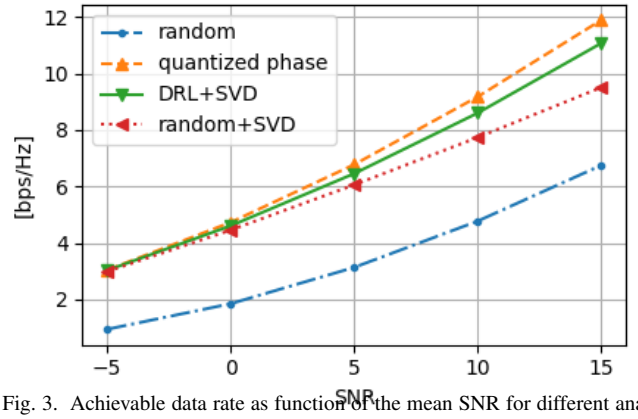

Fig. 3. Achievable data rate as function of the mean SNR for different analog beamformer selection methods with $N_{\mathrm{s}} = 2$. The millimeter wave channel gains were generated according to the extended Saleh-Valenzuela geometric channel model assuming $N_{\mathrm{tx}} = 16$ and $N_{\mathrm{rx}} = 4$.

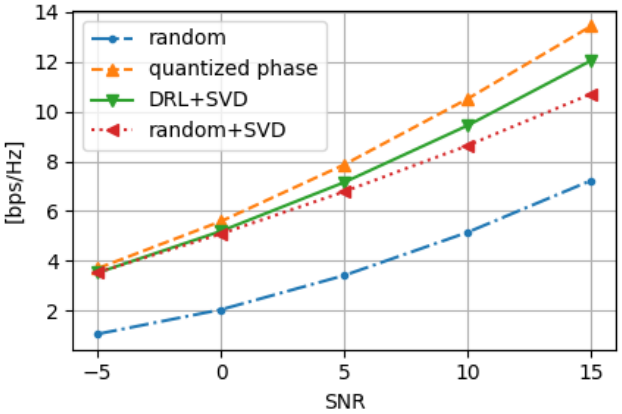

Fig. 4. Achievable data rate as function of the mean SNR for different analog beamformer selection methods with $N_{\mathrm{s}} = 2$. The millimeter wave channel gains were generated according to the extended Saleh-Valenzuela geometric channel model assuming $N_{\mathrm{tx}} = 25$ and $N_{\mathrm{rx}} = 4$.

ues that were successively given to N_{tx}, it can be noted that the achievable data rate of the multiuser MIMO system increases notably with the number of antennas that are deployed in transmission, although the implementation complexity of the analog beamformer is kept low. Meanwhile, it is also observed that the proper selection of the analog beamformer gains importance as N_{tx} grows. For example, when $N_{\mathrm{tx}} = 25$, the scheme with random analog beamformer and digital precoder computed with SVD shows a 2 [bps/Hz] deterioration, with respect to the performance achieved with the proposed DRL-based algorithm. In addition, the performance of the DRL-based hybrid beamforming is also close to the upper bound.

V. Conclusions

In this paper, a hybrid beamforming algorithm based on *Deep Reinforcement Learning* was proposed for a multiuser massive MIMO system, in order to enable the iterative update of the analog beamforming matrix when scheduling each user. This algorithm was complemented with a Singular Value Decomposition operation in a reduced-size channel matrix, which was used to compute the most convenient digital precoder matrix for the hybrid beamforming scheme. The DRL processing was concentrated in the receiver, and relied on a low-rate feedback channel to inform the most convenient update of analog beamforming matrix coefficient to serve the scheduled user. The effectiveness of the proposed DRL approach was compared to well-known lower and upper bounds when using random beamforming and brute-force search, respectively, providing a good tradeoff solution, particularly at high SNR.

References

[1] R. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[2] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] O. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[4] T. Wang, C. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Communications*, vol. 14, no. 11, pp. 92–111, Nov. 2017.

[5] E. Lizarraga, G. Maggio, and A. Dowhuszko, "Hybrid beamforming algorithm using reinforcement learning for millimeter wave wireless systems," in *Proc. Workshop Inform. Process. and Control*, Sept. 2019, pp. 253–258.

[6] T. Peken, R. Tandon, and T. Bose, "Reinforcement learning for hybrid beamforming in millimeter wave systems," in *Proc. Annual Int. Telemetering Conf.*, Jan. 2019, pp. 138–147.

[7] Q. Wang, K. Feng, X. Li, and S. Jin, "PrecoderNet: Hybrid beamforming for millimeter wave systems with deep reinforcement learning," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1677–1681, Oct. 2020.

[8] S. Buzzi, C. D'Andrea, T. Foggi, A. Ugolini, and G. Colavolpe, "Single-carrier modulation versus OFDM for millimeter-wave wireless MIMO," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1335–1348, Mar. 2018.

[9] A. Dowhuszko, G. Corral-Briones, J. Hämäläinen, and R. Wichman, "Performance of quantized random beamforming in delay-tolerant machine-type communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5664–5680, Aug. 2016.

[10] A. Dowhuszko and J. Hämäläinen, "Performance of transmit beamforming codebooks with separate amplitude and phase quantization," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 813–817, July 2015.

[11] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. The MIT Press, 2018.

[12] A. Grassi, G. Piro, G. Boggia, M. Kurras, W. Zirwas, R. SivaSiva Ganesan, K. Pedersen, and L. Thiele, "Massive MIMO interference coordination for 5G broadband access: Integration and system level study," *Computer Networks*, vol. 147, pp. 191 – 203, Dec. 2018.

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