Artificial Intelligence
Fall 2023
Final Project

Project title: Neural Network and Reinforcement Learning to Optimize Compressive Measurement Matrix in a Massive MIMO System

Name: Md. Saidur Rahman Pavel
&
Nazia Rahman

TUID: 915940349 & 916057947

1 Introduction

Massive multiple input multiple output (MIMO) is considered one of the most important technologies in the next generation, i.e., beyond 5G and 6G wireless technology. Equipping a large number of antenna at the base state (BS), massive MIMO system is capable of providing increased system capacity, energy efficiency, and robustness. Furthermore, in Millimeter-wave (mmWave) communication, which allows exploiting the underutilized multi-gigabit bandwidth available at the mmWave spectrum to address the ever-increasing demand for higher data rates in future wireless systems, massive MIMO system is an integral part. This is because the poor propagation characteristics of a mmWave channel can be addressed by forming highly directional beams exploiting the large antenna array from a massive MIMO system. To form the directional beams, it is very important to estimate the user signals direction of arrival (DOA).

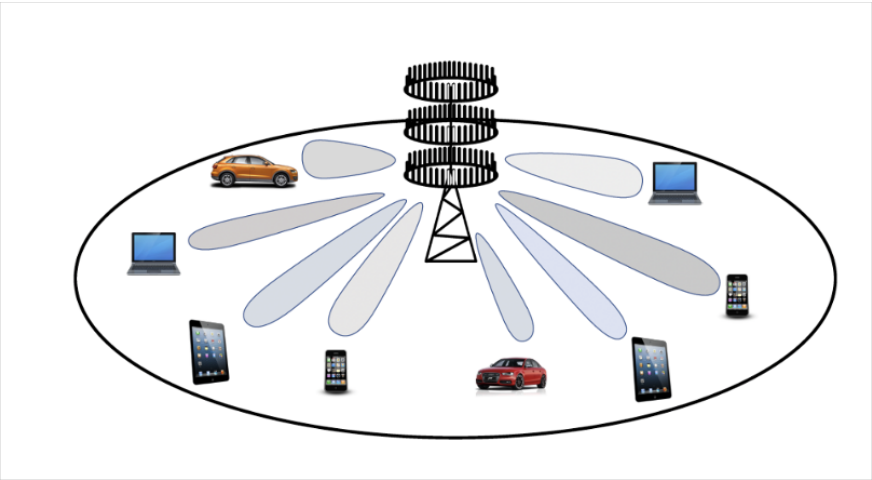


Figure 1: A massive MIMO system

Despite the advantages, the practical implementation of a massive MIMO system is challenging due to the requirement of a large number of radio frequency (RF) front-end circuits and high-resolution analog-to-digital converters (ADCs). An RF front-end chain generally consists of a band-pass filter, a low-noise amplifier, a mixer and a low-pass filter. The ADC then digitizes the analog signal to obtain the baseband digital signal to perform subsequent signal processing. These components are the limiting factor from the perspective of cost and power consumption. To reduce the number of RF chains and

ADCs, a hybrid analog-digital beamforming strategy is adopted, which will be introduced in the next section.

To implement a hybrid beamforming strategy, we consider using a neural network to optimize a compressive measurement matrix to project the high-dimensional received signal into a lower-dimensional space. The most challenging part of this process is that, the original high dimensional signal is not observable, and the optimization procedure needs to be performed by observing only the low dimensional signal. Additionally, we consider a reinforcement learning (RL) framework to enhance the estimation of the week signals.

2 Hybrid beamforming strategy

Assume D uncorrelated users impinge on a massive MIMO array equipped with N antennas from the directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_D]^{\mathrm{T}}$. The received signal in the baseband measured at the t-th sampling time , $\boldsymbol{x}(t) \in \mathbb{C}^N$ can be expressed as

$$\begin{aligned} \boldsymbol{x}(t) &= \sum_{d=1}^{D} \boldsymbol{a}(\theta_d) s_d(t) \boldsymbol{n}(t) \\ &= \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{s}(t) + \boldsymbol{n}(t), \end{aligned} \tag{1}$$

where $A(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \cdots, \boldsymbol{a}(\theta_D)] \in \mathbb{C}^(N \times D)$ denotes the array manifold matrix whose column $\boldsymbol{a}(\theta_d) \in \mathbb{C}^N$ represents the steering vector of the d-th user with DOA θ_d, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_D(t)]^{\mathrm{T}} \in \mathbb{C}^D$ represents the signal waveform vector, and $\boldsymbol{n}(t) \sim \mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I})$ represents the zero-mean additive white Gaussian noise (AWGN) vector.

As shown in the system block diagram of Fig. 2, each antenna of the receive array is equipped with a dedicated separate front-end chain. The front-end chain transforms the received RF signal to the digital baseband signal by performing low-noise amplification, down-conversion, low-pass filtering and analog-digital conversion in turn, and the result is then fed to a digital signal processor for further digital signal processing.

In massive MIMO systems, the number of users is typically much less than the number of antennas at the base station. We, therefore, consider compressive sampling of the array received signal, as depicted in the system block diagram in Fig. 3 , where the array received signals are compressed in the analog domain

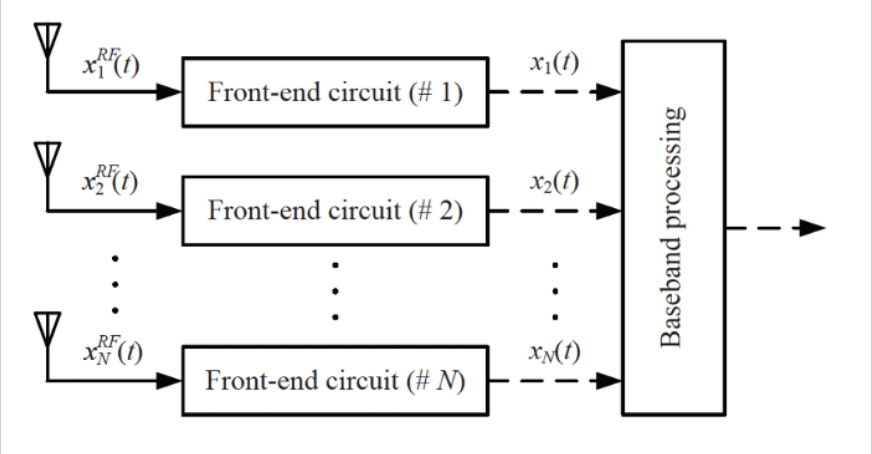


Figure 2: Full digital beamforming. Here the solid lines denote analog signals and dashed lines denote digital signals.

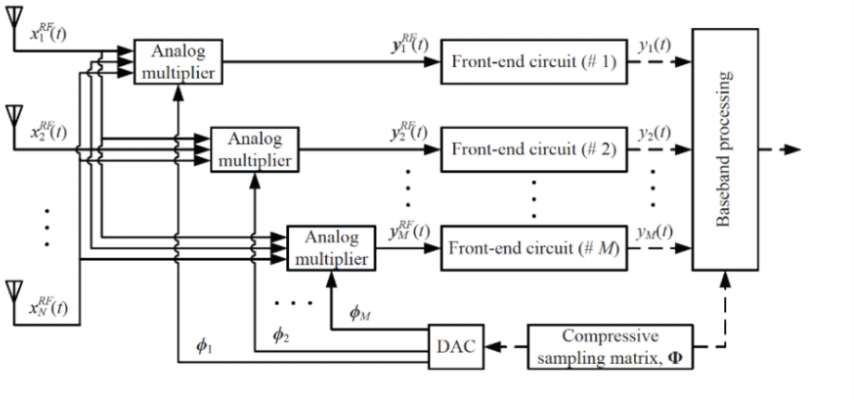


Figure 3: Hybrid analog-digital beamforming. Here the solid lines denote analog signals and dashed lines denote digital signals.

before passing through much fewer number of front-end chains. In particular, $M \ll N$ linear projections of the array received signal (in the analog domain), $\boldsymbol{x}^{RF}(t)$, onto a set of measurement kernels $\{\phi_m, m = 1, \cdots, M\}$ are computed. Stacking the M measurement kernels as a compressive sampling matrix $\boldsymbol{\Phi} = [\phi_1^{\mathrm{T}}, \phi_2^{\mathrm{T}}, \cdots \Phi_M^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{M \times N}$ yields an M-dimensional compressed measurement vector in baseband, $\boldsymbol{y}(t) = [y_1(t), y_2(t), \cdots, y_M(t)]^{\mathrm{T}} \in \mathbb{C}^M$, expressed as

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}\boldsymbol{x}(t) = \boldsymbol{\Phi}\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{\Phi}\boldsymbol{n}(t), \tag{2}$$

where $\boldsymbol{\Phi}\boldsymbol{A}(\boldsymbol{\theta}) \in \mathbb{C}^{M \times D}$ is a compressed array manifold matrix with a significantly reduced dimension because $M \ll N$.

Optimizing the matrix $\boldsymbol{\Phi}$ is particularly crucial so that the required information about user signals DOA will remain in the compressed measurement to obtain

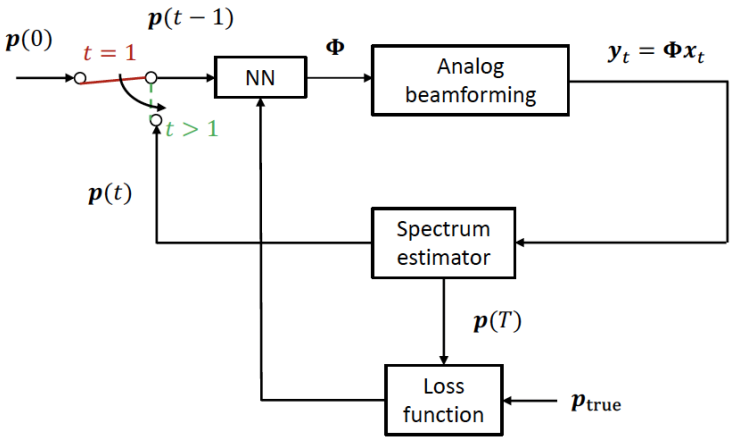


Figure 4: Neural Network-Based optimization framework

accurate beams. To optimize the matrix, we use a fully connected neural network, as described in the section.

3 Neural network approach to optimize the compressive measurement matrix

3.1 Theoritical framework

We consider a neural network framework as depicted in Fig. 4 to optimize the matrix $\boldsymbol{\Phi}$ in an iterative process. We denote $\boldsymbol{p}(t)$ as the discretized posterior distribution of the direction of arrivals (DOAs) at time sample t. The posterior of DOAs at time sample $t-1$ can be regarded as the sufficient statistic for designing the matrix $\boldsymbol{\Phi}$ at next time sample t as

$$\boldsymbol{\Phi}(t) = \hat{\mathcal{F}}(\boldsymbol{p}(t-1)), \tag{3}$$

where $\hat{\mathcal{F}}$ is a mapping function.

we consider an L-layer FC network to obtain the compressive sampling matrix $\boldsymbol{\Phi}(t)$ based on the DOA distribution $\boldsymbol{p}(t-1)$ from the previous time frame. Define $\tilde{t} = (t-1)/T$ as the normalized time index, where T is the total number of samples. We use the posterior $\boldsymbol{p}(t-1)$ and the normalized time index $\tilde{t}$ as the input to the neural network at time t, i.e., $\boldsymbol{v}(t-1) = [\boldsymbol{p}^{\mathrm{T}}(t-1)\ \ \tilde{t}]^{\mathrm{T}}$. For a total of B observations in a particular batch of the training data, the complete training dataset $\boldsymbol{G}(t-1)$ is formed by concatenating vectors $\boldsymbol{v}_b(t)$ corresponding to the

observations $b \in \{1, 2, \cdots, B\}$ as $\boldsymbol{G}(t-1) = [\boldsymbol{v}_1^{\mathrm{T}}(t-1), \boldsymbol{v}_2^{\mathrm{T}}(t-1), \cdots, \boldsymbol{v}_B^{\mathrm{T}}(t-1)]$. The neural network output provides the compressive sampling matrix for the next measurement as

$$\tilde{\boldsymbol{\Phi}}(t) = \mathcal{A}_L\left(\boldsymbol{W}_L \mathcal{A}_{L-1}(\cdots \mathcal{A}_1(\boldsymbol{W}_1 \boldsymbol{G}(t-1) + \boldsymbol{b}_1)\cdots) + \boldsymbol{b}_L\right), \tag{4}$$

where $\{\boldsymbol{W}_l, \boldsymbol{b}_l, \mathcal{A}_l\}_{l=1}^{L}$ are the weights, biases, and nonlinear activation function corresponding to the lth layer, respectively. $\tilde{\boldsymbol{\Phi}}(t)$ is an augmented matrix that denotes the real-valued representation of the complex-valued compressive sampling matrix at time t, i.e., $\tilde{\boldsymbol{\Phi}}(t) = [\mathcal{R}(\boldsymbol{\Phi}(t)) \;\; \mathcal{I}(\boldsymbol{\Phi}(t))]$. The required $\boldsymbol{\Phi}(t)$ can then be extracted from $\tilde{\boldsymbol{\Phi}}$.

Using the neural network output $\boldsymbol{\Phi}(t)$, we then form an analog beamformer to obtain compressed measurements $\boldsymbol{y}(t)$ from $\boldsymbol{x}(t)$ at time t. We use the minimum variance distortionless response (MVDR) spatial spectrum estimator based on the compressed measurement vector to find the spatial spectrum as

$$P_{\text{CS-MVDR}}^{(t)}(\theta) = \frac{1}{N} \frac{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{\Phi}^{\mathrm{H}}(t)\boldsymbol{\Phi}(t)\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{\Phi}^{\mathrm{H}}(t)\hat{\boldsymbol{R}}_{\boldsymbol{yy}}^{-1}(t)\boldsymbol{\Phi}(t)\boldsymbol{a}(\theta)}, \tag{5}$$

where $\hat{\boldsymbol{R}}_{yy}(t)$ is the sample covariance matrix of $\boldsymbol{y}$ estimated at time t.

The normalized spatial spectrum can be considered as the posterior distribution of the DOAs at time t, i.e.,

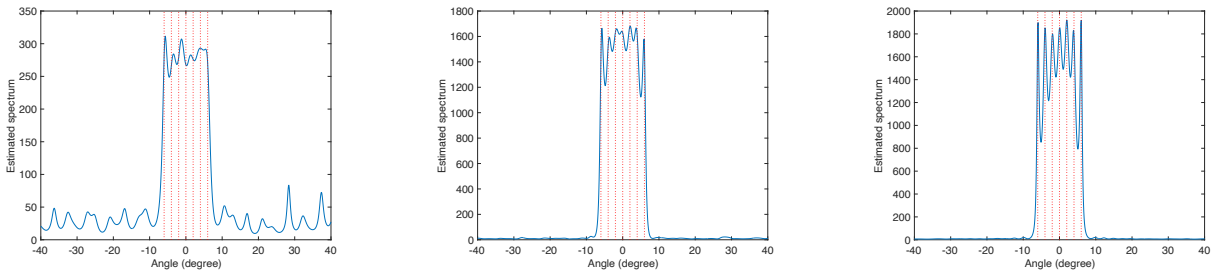
$$\boldsymbol{p}(t) = \frac{[P_{\text{CS-MVDR}}^{(t)}(\theta_1), \cdots, P_{\text{CS-MVDR}}^{(t)}(\theta_K)]}{\sum_{k=1}^{K} P_{\text{CS-MVDR}}^{(t)}(\theta_k)} \tag{6}$$

The obtained posterior distribution $\boldsymbol{p}(t)$ is then fed to the neural network again for sequential optimization of $\boldsymbol{\Phi}$. Once the iterations through all time samples are completed, we update the neural network parameters by minimizing a MSE loss function, expressed as

$$\text{Loss} = \frac{1}{BK} \sum_{i=1}^{B} \left\| \boldsymbol{p}_i(T) - \boldsymbol{p}_{i_{\text{true}}} \right\|^2. \tag{7}$$

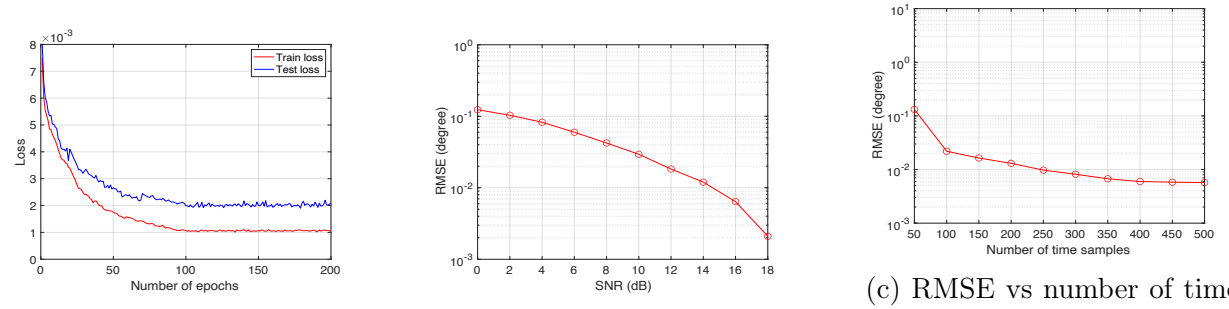
3.2 Neural Network Architecture

Our neural network is composed of 4 fully connected layers where each layer contains 500 nodes. We generate 10000 different angular distribution scenarios



(a) After 30 time sample (b) After 50 time sample (c) After 100 time sample

Figure 5: Direction of arrival (DOA) estimation performance



(a) Training and test losses (b) RMSE vs SNR (c) RMSE vs number of time sample

Figure 6: Performance curves

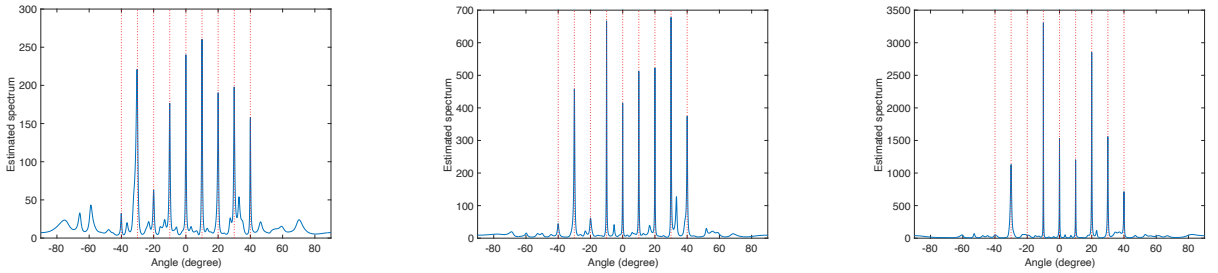
to train the network, and 1000 scenarios for testing. ADAM optimizer is used to optimize the MSE loss function with learning rate progressively decreasing from 0.1 to 0.001. $T = 100$ time samples are considered.

3.3 Simulation result

We consider $N = 50$ received antenna at the base station. We also consider a compression ratio of 5 to yield the dimension of the compressed measurement of 10.

Fig. 5 depicts the estimation performance of developed approach. For illustration, an example is considered with DOAs $-6°, -4°, -2°, 0°, 2°, 4°, 6°$. Starting from uniform distribution of DOAs, Fig. 5 shows that the posterior of DOAs converge to the actual DOAs as the time samples increase once the training is complete.

Fig. 6a demonstrates the convergence of training and test losses. We also evaluate the performance based on the root mean squared error (RMSE), defined as



(a) After 30 time sample (b) After 50 time sample (c) After 100 time sample

Figure 7: Misdetection of weak signals

$$\text{RMSE} = \sqrt{\frac{1}{QD}\sum_{q=1}^{Q}\sum_{d=1}^{D}(\hat{\theta}_{q,d} - \theta_d)^2} \tag{8}$$

where Q is the number of trials and $\hat{\theta}_{q,d}$ is estimated DOA for the dth source of the qth trial. We performed 100 trials to compute the RMSEs. From both Figs. 6b and 6c, it is evident that the RMSE decrease as the signal to noise ratio (SNR) and number of time sample increase.

4 Reinforcement learning based weak signal enhancement

The digital beamforming approach like MVDR as described in the previous section, generally favors strong signals in the presence of signals with mixed strengths. It may even progressively decrease the weak signals as the iteration progresses. As illustrated in Fig. 7, the true DOAs are $-40°, -30°, -20°, -10°, 0°, 10°, 20°, 30°, 40°$ with SNR corresponding to the $-40°$ and $-20°$ signals being 0dB, and the rest of them are 10dB, i.e., two signals are weaker. It is clear from Fig. 7 that the weak signal disappears from the spectrum as the number of iterations increases. To address this issue a reinforcement learning framework is considered to protect weak signals. Particularly, SARSA (state-action-reward-state-action), which is a on policy RL framework, is adopted here. The idea is instead of directly use the normalized power spectrum as prior for next iteration, the PMF is adjusted based on SARSA strategy. The basic steps are described in the following subsections.

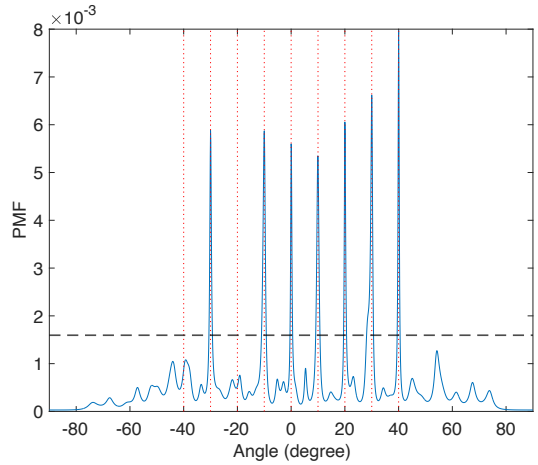


Figure 8: Selection of states

4.1 Selection of state

To obtain the current state, we empirically choose a threshold δ such that

$$\tilde{p}_t(\theta) = \begin{cases} 1, & \text{if } p_t(\theta) \geq \delta, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

$\tilde{p}^{(i)}(\theta)$ signifies candidate angular bins that likely contain signals. The number of angular bins where the spatial spectrum is above the threshold constitutes the state of the environment. As such, the state at the ith iteration denotes the total number of angular bins that likely contain signals and is given as

$$s_t = \sum_{k=1}^{K} \tilde{p}_t, \tag{10}$$

with $s_t \in \{1, 2, \cdots, K\}$. As shown in Fig. 8, 7 sources are above the threshold, so the state is 7.

4.2 Selection of Action

By considering that the other angular bins potentially contain weak signals, we enhance the power of other angular bins by a factor of 5 as

$$\hat{p}_t(\theta) = \begin{cases} p_t(\theta), & p_t(\theta) \geq \delta, \\ 5p_t(\theta), & p_t(\theta) < \delta. \end{cases} \tag{11}$$

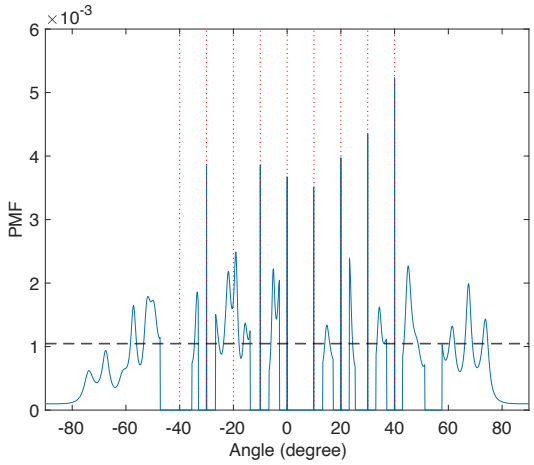


Figure 9: Selection of actions

And we denote candidate action as the number of peaks are now above the threshold. For example, in Fig. 9 after power enahncing, 24 peaks are above the threshold. So, the candidate action is $a_{\text{candidate},t} = 24$. Based on these, an action space at time t is formed, which is a set of numbers ranging from current state to the current candidate action, i.e., $\mathcal{A}_t = \{s_t, \cdots, a_{\text{candidate}}\} = \{7, \cdots, 24\}$. From these, the action is selected using the ϵ greedy strategy as

$$a_t = \begin{cases} \arg\max_{a \in \mathcal{A}} Q(s_t, a), & \text{with probability } 1 - \epsilon, \\ \text{random action}, & \text{with probability } \epsilon. \end{cases} \tag{12}$$

. The state action matrix $\boldsymbol{Q}$ is initialized randomly, and updated as

$$\begin{aligned} Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma Q(s_{t+1}, a + t + 1) \\ - Q(s_t, a_t), \end{aligned} \tag{13}$$

Once the action is chosen, the a_t angular bins having the highest power $\hat{p}(\theta)$ from Fig. 9 is utilized to generate PMF for next iteration. The overall RL based weak signal protection schemen is depicted in Fig. 10.

5 Reward function

$$r_{t+1} = \sum_{\theta_m \in \Theta_{s,t}} p_{t+1}(\theta_m) - \sum_{\theta_n \in \Theta - \Theta_{s,t}} p_{t+1}(\theta_n) - |a_t - s_t + 1|, \tag{14}$$

where $\Theta_{s,t}$ denotes the collection of angular bins based on s_t and Θ denotes the collection of all angular bins. The reward function has three parts,

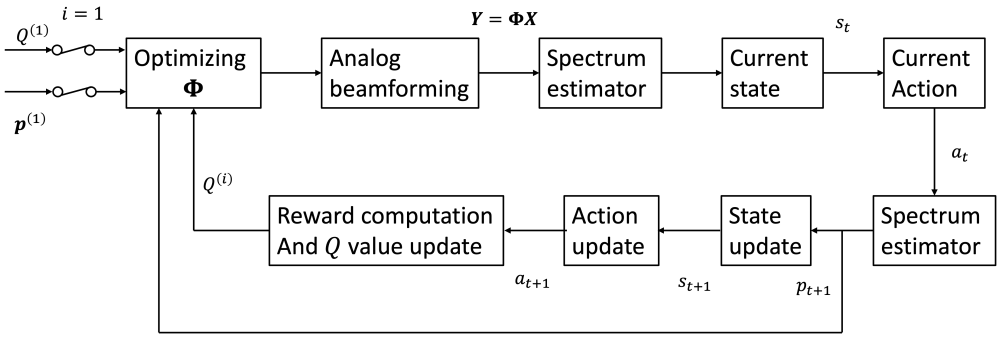


Figure 10: Weak signal protection scheme

1. A positive reward for the accurate detection of angular bins from the previous state. (Possibly strong sources)
2. Penalizes detection at other angular bins (probably false detection).
3. The third part imposes a penalty if the action from the previous step differs from the current state.
   (a) The state defines angular bins corresponding to strong sources.
   (b) The action defines angular bins corresponding to combinations of strong sources, possibly some weak sources, and possibly some unwanted sources.
   (c) This penalty imposes the removal of unwanted bins.
   (d) If a weak signal containing bin is part of the action, a high PMF assigned to this bin may cause the power spectrum at this bin to exceed the threshold, which can be detected as a state. And the first part of the reward will ensure its survival

6 Conclusion

n this project, we developed a neural network framework to optimize the compressive measurement matrix in a massive MIMO system. By exploiting this optimized measurement matrix, effective hybrid beamforming becomes possible, ultimately reducing the number of RF front-end circuits and consequently lowering the system complexity. Additionally, as the strengths of weak signas gradually decrease in an iterative process, we consider a reinforcement learning framework to enhance the detection of weak signals.

References

[1] F. Sohrabi, Z. Chen, and W. Yu, "Deep active learning approach to adaptive beamforming for mmwave initial alignment," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2347–2360, 2021.

[2] Y. Gu and Y. D. Zhang, "Compressive sampling optimization for user signal parameter estimation in massive mimo systems," *Digital Signal Processing*, vol. 94, pp. 105–113, 2019.

[3] A. M. Ahmed, A. A. Ahmad, S. Fortunati, A. Sezgin, M. S. Greco, and F. Gini, "A reinforcement learning based approach for multitarget detection in massive mimo radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 5, pp. 2622–2636, 2021.

[4] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2012.

[6] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[7] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.