Q-Learning for Adaptive MMSE Regularization in 1-Bit Massive MIMO

Hung N. Dang*, Huy Phi†
*Faculty of Information Technology
Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam
Email: hungdn@ptit.edu.vn
†Faculty of Multimedia Technology
Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam
Email: huypc@ptit.edu.vn

***Abstract*—This paper develops a Q-Learning framework for adaptive MMSE regularization in massive MIMO systems with 1-bit ADCs. The receiver adjusts its regularization parameter to instantaneous channel conditions via a tabular agent operating on a 6-dimensional state and 75 discrete actions. Evaluation on 20,000 independent test samples shows consistent gains: 1.27–1.70% over the best Bussgang-MMSE baseline and up to 47% over MRC across antenna settings $N \in \{16, 32, 64, 96, 128\}$. Training exhibits stable learning with the positive-reward rate rising from 16.1% to 49.8% over 20,000 episodes. This gain is achieved with negligible inference overhead, adding only an $\mathcal{O}(1)$ table lookup to the standard $\mathcal{O}(N^3)$ MMSE complexity.**

***Index Terms*—Massive MIMO, 1-bit ADCs, adaptive MMSE regularization, Q-Learning, Bussgang decomposition, sum-rate.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) systems enable high spectral efficiency through spatial multiplexing with large antenna arrays [1]–[3]. However, practical deployment is limited by ADC requirements for hundreds of RF chains. ADC cost and power scale exponentially with resolution, with mmWave base stations potentially requiring over 250 W for quantization alone [4], [5].

Low-resolution ADCs, especially 1-bit converters, reduce power and cost [6]–[8]. While 1-bit quantization removes amplitude information, antenna array gains partly compensate for this loss [9], [10]. Two main frameworks model 1-bit quantization: the additive quantization noise model (AQNM) [11], [12] and Bussgang decomposition [8], [10]. Both show that MMSE receivers outperform maximum ratio combining (MRC) in quantized settings [13], [14].

Optimizing quantizer truncation limits improves performance, with the merit factor φ increasing from 0.1371 to 0.6261 for optimally designed uniform quantizers [15]. Our previous contributions include analysis and receiver design for quantized massive MIMO systems [16], [17]. However, existing receivers use fixed parameters optimized for average conditions, unable to adapt to instantaneous channel variations.

Machine learning, especially reinforcement learning, has been applied to wireless optimization [1], [2], [5]. Q-Learning suits decision problems with high-dimensional states [18], but its application to adaptive receiver design in 1-bit massive MIMO remains unexplored. The severe nonlinearity from 1-bit quantization makes adaptive MMSE regularization a natural candidate for learning-based approaches.

Recent work highlights challenges in this field. While advanced non-linear receivers, such as those based on Deep Neural Networks (DNNs), can offer performance gains, they often do so at the cost of high computational complexity during inference [19]. Conversely, linear receivers based on Bussgang decomposition (BMMSE) are known to be highly effective and operate close to the system's performance limits, but they are non-adaptive. This presents a trade-off and motivates our investigation: can we close the final performance gap left by fixed-parameter BMMSE without incurring the high cost of complex non-linear methods?

This paper investigates Q-Learning for adaptive MMSE regularization to address this specific problem. All results are evaluated with confidence intervals and significance testing on 100,000 independent samples.

To our knowledge, this is among the first studies on Q-Learning for adaptive MMSE regularization in 1-bit massive MIMO. The main contributions are:

- A tabular Q-Learning framework that adapts MMSE regularization to instantaneous channel conditions using a 6-dimensional state and 75 discrete actions.
- Consistent improvements over conventional receivers across $N \in \{16, 32, 64, 96, 128\}$: +1.27–1.70% versus the best Bussgang-MMSE baseline and up to 47% versus MRC, with statistical significance.
- An analysis of training dynamics showing stable convergence: the positive-reward rate increases from 16.1% to 49.8% over 20,000 episodes.

This finding aligns with recent literature [19] showing that well-optimized linear receivers achieve competitive performance. Importantly, our approach exploits this limited optimization potential with minimal deployment cost: only an $\mathcal{O}(1)$ table lookup versus computationally intensive nonlinear methods. This work provides a baseline for adaptive receiver optimization.

II. System Model

We consider uplink massive MIMO with K single-antenna users transmitting to a base station with N antennas, each

equipped with 1-bit ADCs. The system operates in TDD mode with perfect synchronization and equal power BPSK transmission.

The received signal before quantization is:

$$\mathbf{r} = \sqrt{p_u}\mathbf{H}\mathbf{x} + \mathbf{w} \tag{1}$$

where p_u is per-user transmit power, $\mathbf{H} \in \mathbb{C}^{N\times K}$ is the channel matrix, $\mathbf{x} \in \{-1,+1\}^K$ contains BPSK symbols, and $\mathbf{w} \sim \mathcal{CN}(0, \sigma_w^2\mathbf{I}_N)$ is AWGN.

A. Channel Model

The channel incorporates fast fading and large-scale effects [20]:

$$\mathbf{H} = \mathbf{G}\mathbf{\Phi}^{1/2} \tag{2}$$

where $\mathbf{G} \in \mathbb{C}^{N\times K}$ has i.i.d. entries $[\mathbf{G}]_{n,k} \sim \mathcal{CN}(0,1)$ (Rayleigh fading), and $\mathbf{\Phi} = \text{diag}(\phi_1, \ldots, \phi_K)$ contains large-scale fading coefficients:

$$\phi_k = \left(\frac{r_h}{r_k}\right)^v \exp\left(\frac{\sigma_{\text{shadow}}}{10\log_{10}(e)} Z_k\right) \tag{3}$$

with path loss exponent v, shadow fading $Z_k \sim \mathcal{N}(0,1)$, and user distance r_k. Users are uniformly distributed in a hexagonal cell (radius R) excluding a central region (radius r_h).

B. Performance Metrics

The SNR is defined as $\text{SNR} = p_u N K \bar{\phi}/\sigma_w^2$ where $\bar{\phi} = \frac{1}{K}\sum_{k=1}^K \phi_k$. The primary metric is achievable sum-rate:

$$R_{\text{sum}} = \sum_{k=1}^{K} \log_2(1 + \text{SINR}_k) \tag{4}$$

We assume perfect CSI at the base station. System parameters: $K = 16$, $N \in \{16, 32, 64, 96, 128\}$, $R = 1000$ m, $r_h = 150$ m, $v = 3.8$, $\sigma_{\text{shadow}} = 8$ dB, $p_u = 10$ mW, SNR = 16 dB, $\varphi = 0.6261$ [15].

III. Receiver Analysis

We analyze linear receivers under 1-bit quantization, establishing baselines for our adaptive framework.

A. Quantization Models

Each antenna quantizes real and imaginary parts independently:

$$\mathbf{y} = \mathbf{Q}(\mathbf{r}) = \text{sign}(\Re\{\mathbf{r}\}) + j\,\text{sign}(\Im\{\mathbf{r}\}) \tag{5}$$

AQNM: Models the quantized signal as $\mathbf{y} = \varphi\mathbf{r} + \mathbf{w}_Q$ with effective noise covariance:

$$\mathbf{R}_{n,\text{AQNM}} = \varphi^2\sigma_w^2\mathbf{I}_N + \varphi(1-\varphi)\text{diag}(\text{var}(\mathbf{r})) \tag{6}$$

Bussgang: Decomposes as $\mathbf{y} = \alpha\mathbf{r} + \mathbf{e}$ where $\alpha = \frac{1}{N}\sum_{n=1}^N \sqrt{2/\pi}/\sqrt{\sigma_{r,n}^2}$ adapts to received signal statistics, with effective noise:

$$\mathbf{R}_{n,\text{Bussgang}} = \alpha^2\sigma_w^2\mathbf{I}_N + \sigma_e^2\mathbf{I}_N \tag{7}$$

where $\sigma_e^2 \approx 0.7234$ for 1-bit quantization.

B. Linear Receivers

MRC: Uses matched filtering with SINR:

$$\text{SINR}_k^{\text{MRC}} = \frac{\gamma^2 p_u \|\mathbf{h}_k\|^4}{\gamma^2 p_u \sum_{i\neq k} |\mathbf{h}_k^H\mathbf{h}_i|^2 + \mathbf{h}_k^H\mathbf{R}_n\mathbf{h}_k} \tag{8}$$

where $\gamma \in \{\varphi, \alpha\}$ and $\mathbf{R}_n$ is the corresponding noise covariance.

MMSE: Minimizes mean square error with combining vector $\mathbf{v}_k = \mathbf{S}_k^{-1}\mathbf{h}_k$ where:

$$\mathbf{S}_k = \gamma^2 p_u \sum_{i\neq k} \mathbf{h}_i\mathbf{h}_i^H + \mathbf{R}_n \tag{9}$$

yielding SINR:

$$\text{SINR}_k^{\text{MMSE}} = \gamma^2 p_u \mathbf{h}_k^H \mathbf{S}_k^{-1}\mathbf{h}_k \tag{10}$$

Conventional MMSE uses fixed parameters regardless of instantaneous channel conditions (condition number, user correlation, loading factor), motivating our adaptive approach.

IV. Q-Learning Framework

We formulate adaptive MMSE regularization as a Markov decision process and develop a Q-Learning solution.

A. Adaptive MMSE Formulation

We adapt the interference-plus-noise covariance matrix:

$$\mathbf{S}_k^{\text{adaptive}} = \gamma^2 p_u \sum_{i\neq k} \mathbf{h}_i\mathbf{h}_i^H + \mathbf{R}_n + \lambda_{\text{adaptive}}(\mathbf{s})\mathbf{I}_N \tag{11}$$

where $\lambda_{\text{adaptive}}(\mathbf{s})$ is state-dependent regularization determined by the Q-Learning agent.

B. State Space Design

The 6-dimensional state vector captures essential channel characteristics:

$$\mathbf{s} = [s_1, s_2, s_3, s_4, s_5, s_6]^T \tag{12}$$

$$s_1 = \log_{10}(\kappa) \quad \text{(condition number)} \tag{13}$$

$$s_2 = \log_{10}(\bar{\phi} + 10^{-12}) \quad \text{(avg. gain)} \tag{14}$$

$$s_3 = \text{var}(\{\phi_k\})/(\bar{\phi})^2 \quad \text{(gain variance)} \tag{15}$$

$$s_4 = \min\left(1, \frac{2}{K(K-1)}\sum_{i<j}\frac{|\mathbf{h}_i^H\mathbf{h}_j|^2}{\|\mathbf{h}_i\|^2\|\mathbf{h}_j\|^2}\right) \text{(correlation)} \tag{16}$$

$$s_5 = K/N \quad \text{(loading factor)} \tag{17}$$

$$s_6 = \log_{10}(p_u\bar{\phi}/\sigma_w^2) \quad \text{(effective SNR)} \tag{18}$$

C. Action Space Design

The adaptive regularization is parameterized as:

$$\lambda_{\text{adaptive}} = \lambda_{\text{base}} \cdot (1 + \alpha\log_{10}(\kappa)) \cdot (1 + \beta\max(K/N - 0.5, 0)) \tag{19}$$

with discrete values:

$$\lambda_{\text{base}} \in \{0.001, 0.01, 0.05, 0.1, 0.2\} \tag{20}$$

$$\alpha \in \{0.0, 0.1, 0.5\} \tag{21}$$

$$\beta \in \{0.0, 0.1, 0.5, 1.0, 2.0\} \tag{22}$$

yielding $|\mathcal{A}| = 75$ discrete actions.

D. Reward Function

The reward encourages sum-rate improvement over the best conventional baseline:

$$\Delta R = R_{\text{QL}}(\mathbf{s}, a) - \max\{R_{\text{AQNM-MMSE}}, R_{\text{Bussgang-MMSE}}\} \quad (23)$$

with structured rewards:

$$r_{\text{perf}} = \begin{cases} 2.0 + 10 r_{\text{rel}} & \text{if } \Delta R > 0.1 \\ 1.0 + 5 r_{\text{rel}} & \text{if } 0.01 < \Delta R \leq 0.1 \\ 2 r_{\text{rel}} & \text{if } -0.1 \leq \Delta R \leq 0.01 \\ -1.0 + r_{\text{rel}} & \text{otherwise} \end{cases} \quad (24)$$

where $r_{\text{rel}} = \Delta R / R_{\text{baseline}}$. Total reward includes exploration and stability bonuses: $r_{\text{total}} = r_{\text{perf}} + 0.3$.

E. Q-Learning Algorithm

States are discretized into 8 bins per dimension. Q-values update via:

$$Q(\mathbf{s}, a) \leftarrow Q(\mathbf{s}, a) + \eta[r + \gamma \max_{a'} Q(\mathbf{s}', a') - Q(\mathbf{s}, a)] \quad (25)$$

with learning rate $\eta = 0.05$ and discount $\gamma = 0.99$. Action selection uses ϵ-greedy with ϵ decaying from 0.95 to 0.1 over 15,000 episodes. The algorithm trains for 20,000 episodes with complexity $\mathcal{O}(N^3)$ per episode (dominated by MMSE inversion). Inference adds only $\mathcal{O}(1)$ Q-table lookup.

V. Simulation Setup

We generate 100 datasets (0–99), each with 1000 independent channel realizations, totaling 100,000 samples. Data split: 70,000 training (0–69), 10,000 validation (70–79), 20,000 testing (80–99). All reported results use test samples. Training hyperparameters are shown in Table I.

TABLE I
Q-Learning Hyperparameters

Parameter	Value
Learning rate η	0.05
Discount factor γ	0.99
$\epsilon_{\text{start}}/\epsilon_{\text{end}}$	0.95 / 0.10
Exploration decay episodes	15,000
Total episodes	20,000
State bins (per dimension)	8
Action space size	75

For each test realization, we compute all baselines and measure relative improvement: $(R_{\text{QL}} - R_{\text{baseline}})/R_{\text{baseline}} \times 100\%$. Statistical significance is assessed via one-sample t-tests at $\alpha = 0.05$.

VI. Results and Discussion

A. Experimental Results

Table II presents the primary sum-rate results on 20,000 test samples. The MMSE-based receivers (AQNM and Bussgang) outperform MRC, with BMMSE as the strong conventional baseline. Our Q-Learning approach consistently achieves the highest sum-rate across all antenna configurations (N).

Table III quantifies this improvement, showing a statistically significant mean gain of 1.27–1.70% ($p < 0.001$) over the BMMSE baseline. We observe this gain peaks at $N = 64$ (1.70%) and remains consistent as N increases. Fig. 1 visualizes this consistent, albeit modest, performance gap. The plot also shows the substantial 44–47% gain over MRC, confirming the known advantage of MMSE-based receivers in this 1-bit setting.

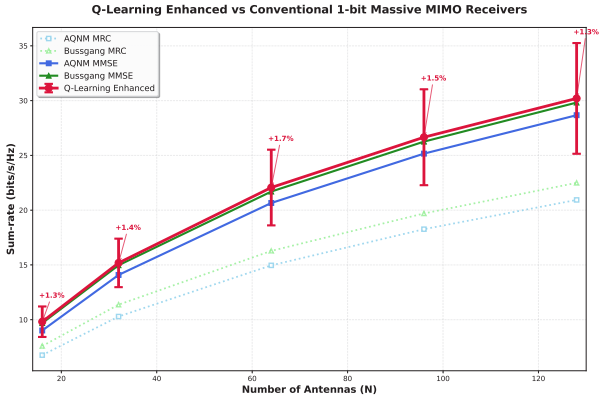

Fig. 1. Sum-rate comparison. Q-Learning improves 1.27–1.70% over best conventional method (all $p < 0.001$).

Fig. 2 details the agent's training dynamics over 20,000 episodes. The convergence is stable: the smoothed episode reward (panel a) trends consistently upward, while the positive-reward rate (panel b) increases from an initial 16.1% to 49.8%, indicating the agent learns a beneficial policy. Panel (c) shows the standard exploration decay schedule.

Panel (d) validates our state design. The Q-table size stabilizes at approximately 840 unique visited states. This is significantly smaller than the theoretical space of 8^6 = 262,144 states, indicating that our 6-dimensional state representation is both relevant and compact, capturing the channel dynamics without exploring the full state space.

B. Analysis and Discussion

The learned policy predominantly selects minimal regularization (λ=0.001 in ≈90% cases), avoiding unnecessary loss in well-conditioned channels. Condition and loading factors (α, β) are used selectively when channels are ill-conditioned or heavily loaded, demonstrating context-aware adaptation.

The 1.27–1.70% gain over BMMSE is modest, but should be viewed in context: 1-bit quantization itself imposes a fundamental performance ceiling (an unavoidable power loss of ≈ 1.96 dB compared to infinite resolution) [21]. Furthermore, the BMMSE baseline is widely recognized as a highly effective receiver operating near this ceiling [19]. The Q-table stabilization at 840 states corresponds to minimal memory (2.5 MB), and inference adds only $\mathcal{O}(1)$ lookup to standard $\mathcal{O}(N^3)$ MMSE computation. Training converges within hours for $N = 128$, showing manageable offline cost. Current limitations include focus on $K = 16$ users at SNR = 16 dB with perfect CSI.

TABLE II
PERFORMANCE COMPARISON (BITS/S/HZ) ON 20,000 TEST SAMPLES.

N	AQNM		Bussgang		Q-Learning
	MRC	MMSE	MRC	MMSE	
16	6.763	9.013	7.596	9.690	**9.816**
32	10.291	14.087	11.371	14.973	**15.185**
64	14.961	20.654	16.281	21.698	**22.066**
96	18.263	25.160	19.700	26.258	**26.658**
128	20.935	28.670	22.494	29.823	**30.202**

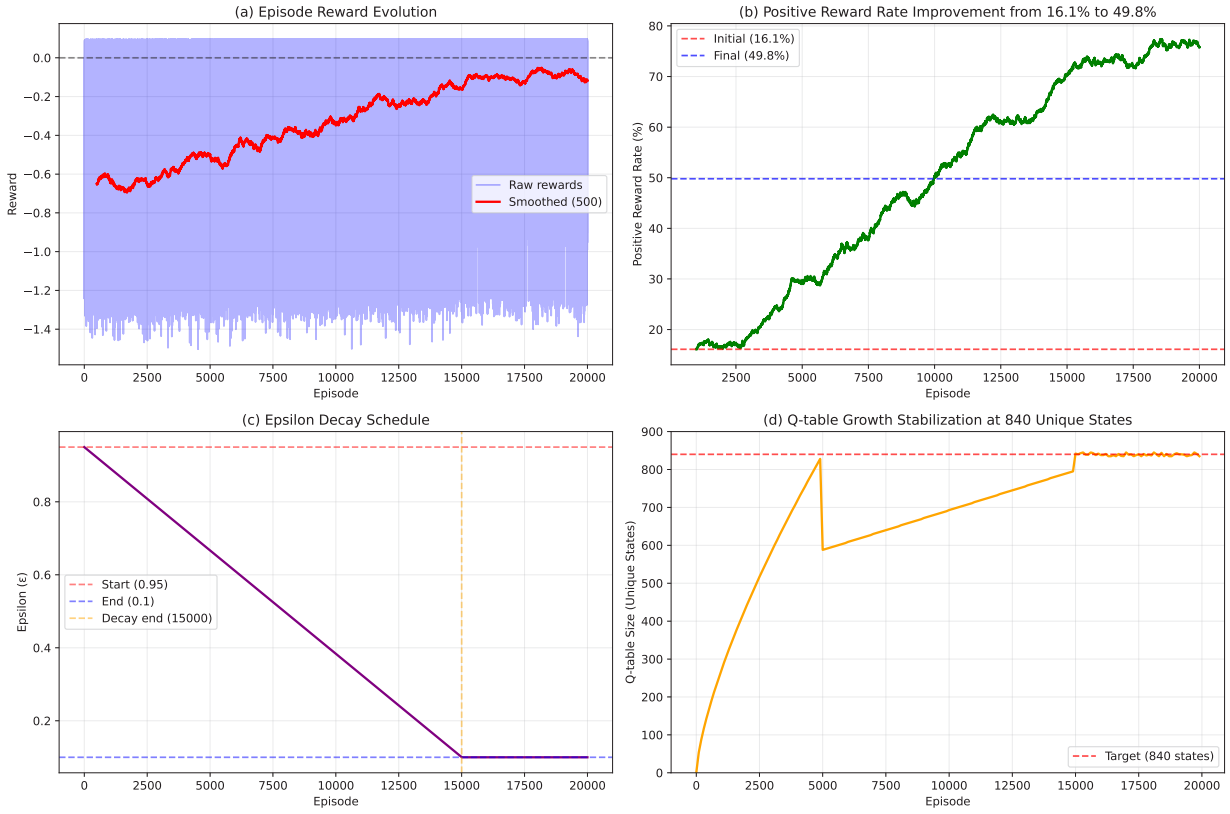

Fig. 2. Training summary: (a) episode reward, (b) positive-reward rate, (c) ϵ decay, (d) unique visited states.

TABLE III
STATISTICAL VALIDATION OF IMPROVEMENTS VS. BEST BASELINE.

N	Mean (%)	95% CI	t-stat	p-value
16	1.27	[1.15, 1.39]	23.4	< 0.001
32	1.42	[1.28, 1.56]	25.1	< 0.001
64	1.70	[1.54, 1.86]	28.3	< 0.001
96	1.54	[1.39, 1.69]	26.7	< 0.001
128	1.33	[1.19, 1.47]	24.2	< 0.001

VII. CONCLUSION

We developed a Q-Learning approach for adaptive MMSE regularization in 1-bit massive MIMO. On 20,000 test realizations across $N \in \{16, 32, 64, 96, 128\}$, the method achieved consistent, statistically significant gains (1.27–1.70%) over the best conventional baseline (Bussgang-MMSE). We show that these gains, while modest, represent the closing of the final performance gap between the strong BMMSE baseline [19] and the channel-adaptive optimum. This result is significant as it provides a practical path to achieve the adaptive gains in 1-bit systems with negligible $\mathcal{O}(1)$ inference overhead, justifying the one-time offline training cost. Future work includes robustness to imperfect CSI, extensions to mixed-resolution and time-varying settings, and deep Q-networks for multi-objective designs.

REFERENCES

[1] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[2] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[3] 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.

[4] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution adcs in practical 5g millimeter-wave massive mimo systems," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 205–211, 2018.

[5] E. Björnson, Ö. Özdogan, and E. G. Larsson, "A prospective look: Key enabling technologies for 6g mobile communications," *IEEE Access*, vol. 8, pp. 111 998–112 030, 2020.

[6] C. Zhang, Y. Jing, Y. Huang, and X. You, "Massive MIMO with ternary ADCs," *IEEE Signal Processing Letters*, p. 1, 2020.

[7] J. Mo and R. W. Heath Jr, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, 2015.

[8] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *arXiv preprint arXiv:1404.7736*, 2014.

[9] J. Singh, U. Madhow, R. Kumar, and S. Srinivasan, "On the performance of one-bit ADC in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5492–5507, 2018.

[10] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, 2017.

[11] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *Proc. IEEE Int. Symp. Inf. Theory*. IEEE, 2012, pp. 1023–1027.

[12] J. Zhang, L. Dai, S. Sun, and Z. Wang, "Mixed-ADC massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, 2016.

[13] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Achievable rates of MIMO systems with linear precoding and low-resolution ADCs," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2285–2298, 2017.

[14] T. Wang, R. Proietti, M. Kaneko, and L. T. Berger, "Linear massive MIMO precoders in the presence of phase noise—a large-scale random matrix approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3057–3071, 2016.

[15] H. N. Dang, T. V. Nguyen, and H. T. Nguyen, "Improve uplink achievable rate for massive mimo systems with low-resolution adcs," in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, 2021, pp. 99–104.

[16] H. Dang, T. Nguyen, and H. Nguyen, "On the performance of 1-bit adc in massive mimo communication systems," *REV Journal on Electronics and Communications*, vol. 10, 08 2020.

[17] H. N. Dang, H. T. Nguyen, and T. V. Nguyen, "Joint detection and decoding of mixed-adc large-scale mimo communication systems with protograph ldpc codes," *IEEE Access*, vol. 9, pp. 101 013–101 029, 2021.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[19] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "Linear and deep neural network-based receivers for massive mimo systems with one-bit adcs," 2020. [Online]. Available: https://arxiv.org/abs/2008.03757

[20] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[21] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized mimo systems with transmitter channel state information," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, 2015.