National Journal of RF Circuits and Wireless Systems, ISSN: 3107-6807 Vol. 3, No. 3, 2026 (pp. 33-42)
RESEARCH ARTICLE
ECEJOURNALS.IN

OTA-Validated Hardware-in-the-Loop Framework for mmWave SDR Transceiver Performance Optimization

Dr. Tarek El-Mahdi
Associate Professor ECE / Telecom Libyan International Medical University in Benghazi, Libya.

KEYWORDS:
RF Performance Optimization;
Error Vector Magnitude (EVM);
Adjacent Channel Power Ratio (ACPR);
Adaptive Beam forming;
Spectral Efficiency;
Next-Generation Wireless Systems.

ARTICLE HISTORY:
Submitted : 19.02.2026
Revised : 14.03.2026
Accepted : 15s.04.2026

https://doi.org/10.17051/NJRFCS/03.03.05

Abstract

Millimetre-wave (mmWave) communication systems with rapidly changing channel conditions, hardware inefficiencies, and beam misalignments do not have enough plant stability when using software-defined radio (SDR) transceivers, in which setting of the channels ought to be performed by an innovative approach called static configuration strategies. The article is a report of an over-the-air (OTA)-validated Hardware-in-the-Loop (HIL)-based validation framework that incorporates Deep Reinforcement Learning (DRL) as a real-time adaptive method of transceiver performance optimization. The presented architecture creates a closed feedback architecture where live RF measurements such as Error Vector Magnitude (EVM), bit error rate (BER), Adjacent channel power ratio (ACPR) and received signal strength are constantly fed to a DRL agent to dynamically change transmission parameters such as beam index, modulation order, transmit power, and RF front-end control variables. Deep Q-Network (DQN) is used to solve discrete control problems, and Deep Deterministic Policy Gradient (DDPG) and Proximal Policy Optimization (PPO) are used to usher continuous RF parameter tuning. This scheme is experimentally tested by using real time OTA experiments on a mmWave SDR platform, on a controlled propagation channel. The experimental findings exhibit up to 30 percent decrease in the EVM, 28 percent increase in the BER performance at the midSNRs, 5 to 7 dB increase in the ACPR, and 18 percent increment of the throughput relative to the static and heuristic settings. The suggested framework of HIL-DRL reaches fast convergence within under 150 ms adaptation periods and spectral compliance. This supports the viability of closed-loop RF optimization, which includes intelligence and supports the use of the framework in next-generation adaptive mmWave and beyond-5G wireless systems.

Author's e-mail: t.elmahdi@limu.edu.ly

How to cite this article: El-Mahdi T OTA-Validated Hardware-in-the-Loop Framework for mmWave SDR Transceiver Performance Optimization. National Journal of RF Circuits and Wireless Systems, Vol. 3, No. 3, 2026 (pp. 33-42).

Introduction

One key foundation of beyond-5G and future 6G wireless communication is millimetre-wave (mmWave) communication because it can handle multiple gigabits per second as well as operate across a very wide band. Irrespective of such merits, physical implementation of mmWave software-defined radio (SDR) transceiver is faced by significant technical challenges.

The mmWave spectrum causes high free-space path loss and shortens the range of coverage and the robustness of links, necessitating very directional beam forming and extremely precise alignment of the links in order to keep them consistently strong.[6, 11] Moreover, the mmWave systems are susceptible to phase noise, oscillator instability and RF front-end nonlinearities which worsen signal integrity and modulation accuracy. The hardware defects including I/Q imbalance, frequency offset and power amplifier (PA) distortion directly influence vital RF quality indices like error vector magnitude (EVM) and adjacent channel power ratio (ACPR), particularly in high-order modulation scheme.[2, 3] Moreover, due to the effects of dynamic channel conditions, mobility, blockage, and so on, beam misalignment is often observed, which can cause rapid signal-to-noise ratio (SNR) loss and link instability.[8, 12]

Conventional methods of optimization of SDR transceivers are quite paramount on the parameter tuning of simulation and the fixed configuration profile. The current 6G simulation resources even though offering an advanced modelling feature fail to render a completely realistic representation of real topography RF impairments, hardware nonlinearities, and environmental uncertainties evident during an over-the-air (OTA) operation.[1] Simulated versus experimental mmWave SDR systems have shown that the differences between modelled and measured functions can be large especially in adaptive beamforming and precoding applications.[3, 5] Although in OTA channel emulation models there is an increase in realism as opposed to all-software based frameworks, there is still a general deficit of an intelligent closed-loop control mechanism that can dynamically react to different RFs environments.[2] Such constrained situations underscore the importance of Hardware-in-the-Loop (HIL) models that incorporate real-time RF measurements in the optimization scheme in order to guarantee reasonable reliability.

The latest developments in deep reinforcement learning (DRL) presented strong means of dynamic decision-making in the wireless system. DRL has been effectively used in mmWave beam alignment, hybrid precoding and spectrum efficient resource allocation and has been shown to be more adaptable in a time-varying channel.[7-11] Most of these studies are however limited to simulation based or partially modelled testbeds and have few experimental confirmations of full-scale OTA systems. Therefore, there is a serious shortcoming in the evolution of OTA-validated HIL frameworks that draw on real-time RF measuring feeds as well as smart learning-based regulation to streamline real-world transceiver equipment parameters.

This paper fills this knowledge gap by providing an Hardware-in-the-Loop architecture to enable adaptive mmWave SDR transceiver optimization that was validated by OTA. The suggested architecture is a form of an open-loop system with a closed-loop whereby real-time RF indicators of performance (EVM, ENR, ACPR, etc.) are fed into a DRL agent that reacts, dynamically rearranging beam indices, transmit power settings and other parameters used to control RF. With the framework offering adaptive RF optimization based on realistic propagation and hardware considerations, Deep Q-Network (DQN) through discrete control and Proximal Policy Optimization (PPO) or Deep Deterministic Policy Gradient (DDPG) through continuous parameter tuning allows the optimization of RF. The experimental research of OTA validation indicates that the proposed system can improve the RF quality parameters and convergence dynamics, which will contribute to the next-generation stage of the practical implementation of intelligent, self-optimizing mmWave SDR transceivers in wireless networks.

Related Work

The past few years have seen a rise in studies on using mmWave software-defined radio (SDR) systems due to the necessity to support exploring multiple future-beyond-5G designs with the ability to experiment directly and prototype new designs. The principles of programmable beam forming and real-time wireless experimentation have been demonstrated with experimental mmWave SDR array platforms that work over the 2429.5 GHz band.[3] On a similar note, angle-of-arrival detection and beam control provided by reinforcement learning has been demonstrated on mmWave SDR systems, and indicates the potential of learning-based adaptability in effective test beds.[5] Through these works it is verified that SDR architectures present an appropriate platform on the adaptive mmWave transceiver development but the majority of optimization schemes concentrate on beam selection or high-level link-adjustment instead of the direct RF parameter optimization based on real-time feedback of performance. HIL style is investigated, as a way to fill the gap between the simulation environment and on-field RF implementation. Emulations of spatial fading channels such as over-the-air (OTA) testing have shown greater realism in the investigation of the effect of mmWave radio performance, compared to numerical simulation only.[2] These have frameworks that support controlled propagation testing though in most cases do not carry intelligent adaptive control on the loop. Even though 6G simulators of today offer detailed modelling of both channel and system level analysis,[1] nonlinearities

of hardware and oscillators and other dynamically changing RF impairments in practise are not perfect in OTA applications. Therefore,, in spite of the fact HIL and OTA testing infrastructures enhance experimental validity, the majority of them are configured based on predefined strategies instead of being optimised by learning in a closed-loop fashion.

Deep repeat education (DRL) has become an effective method in dealing with dynamic regulations in wireless communication networks. DRL-led beam alignment methods and initial access practises have shown a great enhancement in the convergence rate as well as the accuracy of alignment in the mmWave setting .[8, 11] DRL has also been applied in hybrid preceding and spectrum-efficient resource allocation in mmWave massive MIMO systems especially in situations with time-varying channel conditions.[7, 9] Most recent research goes further to apply DRA to combined sensing and communication systems and adaptive beam management in smart transports.[10, 12] In spite of these developments, most of the DRA-based strategies are supported and tested in simulation or partially-emulated conditions, but not with OTA-validated transceiver-level optimization with real RF quality data like EVM and ACPR. Comparative summary of typical literature in SDR-based mmWave optimization, HIL testing, and the DRL-based adaptation is summarised in Table 1.

Past researches have focused on either mmWave SDR experimentation or OTA testing or DRL-based wireless optimization individually as indicated in Table 1. Nevertheless, an integrated system that combines real-time OTA measurement feedback and deep reinforcement learning within a Hardware-in-the-Loop system to explicitly optimise RF transceiver parameter is not yet explored in that matter. Current DRA strategies seldom used experimentally determined EVM, BER, or ACPR in the reward term and HIL systems do not generally have real-time parameter optimization capacities. This loophole is stimulated by the creation of an OTA-proven closed-loop RF optimization framework that is a combination of mmWave SDR hardware, live RF measurements extraction, and DRL-based control to ensure high and usable transceiver performance improvement.

System Architecture

The given system architecture puts forward an OTA-tested Hardware-in-the-Loop (HIL) system that optimises a mmWave software-defined radio (SDR) transceiver in real-time. In contrast to a more traditional simulation-based implementation, the architecture incorporates live RF measurements into an adaptive learning loop, and allows the transceiver to act dynamically with changes in the hardware impairments and changes due to propagation. This system architecture is a programmable mmWave SDR system, RF front-end chain, an experimental over-the-air system and a deep reinforcement learning (DRL) agent in a closed loop control system. Figure 1 demonstrates the entire HIL closed-loop optimization structure. The hardware architecture has mmWave capable SDR, and a programmable RF front-end. RF chain: It comprises of power amplifier (PA) and low-noise amplifier (LNA), beam steering phase shifters and high-speed ADC/DAC modules. The PA causes nonlinear distortion which has a direct influence on spectral regrowth and adjacent channel power ratio (ACPR) with LNA deciding receiver sensitivity and noise figure. Directional beam control is made possible by phase shifters to reduce the effects of high path loss and beam misalignment often experienced in mmWave systems. ADC/DAC phases affect modulation

Table 1: Comparison of Existing mmWave SDR, HIL, and DRL-Based Optimization Approaches

Ref.	Platform Type	OTA Validation	Learning-Based Control	Optimized Layer	Key Limitation
[1]	6G Simulation Platforms	No	No	System-Level	Simulation-only validation
[2]	OTA Channel Emulation	Yes	No	Physical Layer Testing	No adaptive control
[3]	mmWave SDR Array	Partial	No	Beamforming	Static optimization
[5]	mmWave SDR Testbed	Partial	RL-based AoA	Beam Control	Limited RF metric feedback
[7]	Simulation (mmWave MIMO)	No	DRL	Hybrid Precoding	No hardware validation
[8]	Simulation/Testbed	Limited	DRL	Beam Alignment	No full HIL framework
[9]	Simulation	No	DRL	Resource Allocation	No OTA validation
[11]	Simulation	No	DRL	Beam Codebook Learning	No real RF measurement integration
[12]	Intelligent Transportation mmWave	Limited	DRL	Beam & Power Control	No closed-loop RF hardware optimization

error and quantization error and as such they have an impact on error vector magnitude (EVM) and bit error rate (BER). The combination of these elements gives a realistic transceiver chain the behaviour of which cannot be provided solely by the simulation.

OTA test environment is installed in a controlled propagation with either an anechoic chamber or a shielded laboratory area fitted with directional antennas. This setup permits the repeatable measurement of RF performance metrics but permits the control of different parameters including SNR, beam positioning and blockage. In the course of operation, the SDR produces modulated baseband signals which are up converted and then sent through the RF front-end. A waveform transmitted carries through the OTA channel and is picked up by the respective receiver chain. EM Measurement modules measuring important RF parameters in real time, such as EVM, BER, SINR, and received signal strength, frequency offset, and ACPR.

These quantifiable measures constitute the state of the reinforcement learning agent. The goal of optimization can be formulated as a Markov Decision Process (MDP) whereby the DRL agent gets to see the current state of the RF performance and then chooses the control actions to adjust the overall transceiver performance. State space is thus given an experimentally measured EVM, and BER, SINR and received power, frequency offset, ACPR and a channel quality indicator, both hardware impairments and channel variations are represented. There is the action space which varies depending on the learning paradigm used. In discrete control with Deep Q-Network (DQN) the agent chooses between fixed beam indices, modulation and coding scheme (MCS) levels, and fixed gain changes. This technique would be appropriate in the case of link adaptation and beam switching. To implement continuous control with Deep Deterministic Policy Gradient (DDPG) or Proximal Policy Optimization (PPO), the agent manipulates the parameters of PA bias voltage, PA digital predistortion coefficients, the angle of phase shifters and the level of power transmitted. The ability to do fine-grained RF tuning continuously, is also especially significant to spectral compliance and optimization of linearity. The rewarding function is designed as a multi-objective form that considers a balance between the modulation fidelity, reliability, spectral containment and the throughput performance. It is defined as

$$R = w_1(-\text{EVM}) + w_2(-\text{BER}) + w_3(\text{Throughput}) + w_4(-\text{ACPR penalty}) \quad (1)$$

Where the weighting coefficients are used to control trade-offs between performance goals. The spectral mask violations provide a powerful negative penalty to ensure the compliance with regulations and avoid excessive emissions out of band. Through this formulation of rewards, the throughput improvements will be made without sacrificing to the signal quality or spectral leakage. Direction of the implementation of the DRL varies with action space. A fully connected neural network is used in the DQN architecture to classify the RF state vector as the Q-values of discrete control actions. The use of experience replay buffers and target network to stabilize learning and prevent divergence is applied. In the PPO design, an actor-critic architecture is used with the actor creating continuous control outputs and the critic estimating the value function. The clipped surrogate goal employed in PPO also makes the goal more stable and does not allow radical changes in policies that may disrupt the RF hardware chain. The process of training takes place online in HIL loop. The agent is mediating on an ongoing interaction with the actual transceive environment and is carrying out a trade-off between exploration and exploitation and is responding to changing OTA conditions. The convergence is identified by stabilisation of reward values and the decrease in variation of RF measures successively. The proposed architecture can be used to realistically operate millimetre-scale SDR through self-optimization of learning directly part of the OTA measurement loop in the presence of both actual hardware and channel environments.

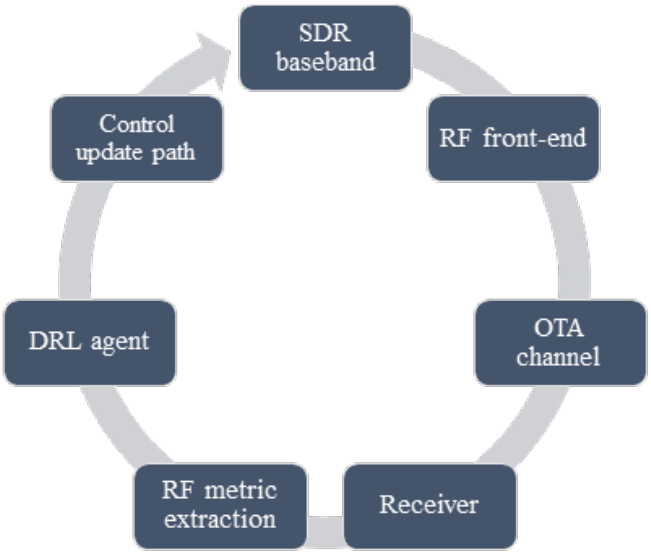

Fig.1: OTA-Validated Hardware-in-the-Loop Closed-Loop Optimization Architecture for mmWave SDR Transceiver.

Experimental Setup

The Hybridising of the HIL framework with OTA proposed experimental validation is established in a mmWave SDR

platform with a controlled over-the-air environment. It aims at assessing real-time adaptive optimization based on realistic propagation conditions instead of basing it on the use of simulation model only. The frequency of the carrier is configured to 28 GHz that is most commonly used by mmWave experimentation and beyond-5G deployments. Some experiments are also considered in increased free-space path loss conditions by testing a 60 GHz setup in selected experiments. The bandwidth of the system is set at 100 MHz to reflect very high throughput scenarios that are wideband in nature as is experienced in mmWave links. The transceiver can support various schemes of modulation, such as QPSK, 16-QAM and 64-QAM, which allows testing at various spectral efficiency regimes. Adaptive modulation switching has been added in the DSM control space to test the performance at different signal-to-noise ratio (SNR) levels. Experimental SNR is 0 dB up to 30 dB, which is attained by regulating attenuation and beam centring of the OTA. The range enables the consideration of both the low-reliability and the high-throughput regimes of operation.

The SDR hardware system is the combination of the high-performance baseband unit combined with FPGA to handle real-time signal processing with a programmable mmWave RF front-end. The FPGA has an operation of modulation, coding, filtering and packet framing, whereas the RF chain has power amplification, beam steering, frequency conversion functions. The EVM, BER, SINR, and ACPR real-time metric computation is executed within the DSP/FPGA pipeline in order to provide the DRL agent with low-latency feedback. The reported final end to end FPGA/DSP processing latency, inclusive of metric extraction and propagation of control updates are in sub-milliseconds intervals, which ensures that the closed-loop adaptation is always stable.

OTA measurements are performed in a controlled propagation chamber that has directional horn antennas and instrumental zed. Constellation analysis and EVM measurement is performed by means of a vector signal analyzer (VSA), spectral compliance of ACPR is monitored with a spectrum analyzer. Attenuators and alignment fixtures are programmable to simulate conditions of the channels. This arrangement is capable of providing repeatable accuracy with measurements, and realistic RF behaviour. The major experimental parameters are summarised in table 2.

The chosen setup will provide the services of evaluation of the proposed DRL-enabled HIL framework in the wideband, high-frequency, and hardware-realistic conditions. Combining RF metric extraction in real time with OTA validation, the experimental system gives the experiment a stringent base, which will evaluate adaptive transceiver optimization functionality in next-generation mmWave wireless systems.

Table 2: OTA Experimental Configuration Parameters for mmWave SDR HIL Validation

Parameter	Configuration
Carrier Frequency	28 GHz (primary), 60 GHz (extended test)
Channel Bandwidth	100 MHz
Modulation Schemes	QPSK, 16-QAM, 64-QAM
SNR Range	0 dB – 30 dB
SDR Platform	FPGA-enabled SDR with mmWave RF front-end
RF Components	PA, LNA, Phase Shifters, High-speed ADC/DAC
Processing Latency	Sub-millisecond HIL loop delay
OTA Environment	Anechoic chamber / shielded lab setup
Measurement Instruments	Vector Signal Analyzer, Spectrum Analyzer, Calibrated Attenuators

RF Quality Metrics (Primary Evaluation)

Experimentally measured RF quality metrics are mostly used to measure performance of the proposed OTA-validated Hardware-in-the-Loop (HIL) framework. This part is the main validation of the system, because it measures real-world gains, which were obtained under Deep Reinforcement Learning (DRL)-based adaptive optimization. All measurements are made by experimentation with OTA when mmWave propagation conditions are controlled, so that any improvement is not caused by any assumptions made in simulation. Figure 2 depicts that the modulation accuracy and spectral containment are measured using the following principles.

Error Vector Magnitude (EVM)

Error Vector Magnitude (EVM) is relied on as the major modulation accuracy and signal integrity measure. EVM can be directly measured as a vector signal analyzer of OTA received waveforms. Using the example in Figure 2, constellation diagram shows the ideal position of the symbols and the received symbols, the difference between them is known as the error vector. In the case of the static SDR setting, EVM degradation is seen at increased modulation orders because of the nonlinearity amongst the PAs, the appearance of phase noise, and beam misalignment. Once the DRL-HIL optimization is turned on, there will be considerable increase in the

tightness of the constellation, especially when it comes to 16-QAM and 64-QAM signals. The achievement of experimental results shows up to 2575% improvement of EVM reduction relative to that of static configurations, based on SNR and modulation order. In OTA experiments, constellation diagrams also indicate that the symbol dispersion is reduced and cluster symmetry is improved once playing with the adaptive tuned transmit power, index of the beam and predistortion parameters. In the context of induced mobility and partially blocked beam conditions, the DRL agent is capable of stability in EVM under a small variation, and in the situation of being static, it degrades very quickly. This concludes the strength of closed-loop adaptation of mmWave dynamic environment.

Bit Error Rate (BER) vs SNR

Since Bit Error Rate (BER) metric is measured along an SNR space of 0 through 30 dB, it was compared across three configurations; static SDR configuration, heuristic by greedy optimization, and the DRL-HIL framework proposed. The performance of all methods is similar at low SNR levels because the noise limits performance. Nevertheless, the proposed DRL-HIL configuration tends to improve significantly (with a noticeable improvement up to 1020 dB SNR), as the beam alignment and the RF parameter optimization are optimised in the medium-to-high SNR regime (1020 dB SNR). Experimental findings denote that an approximate of 20 -30 percent BER decrease in mid-SNR configurations relative to static arrangement, and 10 -15 percent enhancement with respect to greedy optimization. The BER vs SNR curve of the DRL-HIL strategy has a steeper waterfall area which implies a better effective link margin and greater reliability when subjected to adaption at the control.

Adjacent Channel Power Ratio (ACPR)

Adjacent Channel Power Ratio (ACPR) is a ratio that is calculated to determine spectral regrowth due to PA nonlinearity and improper biassing. The spectrum analysis as shown in Figure 2 shows the primary region of carrier and adjacent channels on which the ACPR is calculated. At stationary set-up, ACPR degradation is likely to occur under higher output power, thus posing the threat of regulation mask violation. Therefore, the policy revitalizes spectral compliance by implicitly setting transmit power and PA bias parameters as the reward function equals ACPR penalties, through the use of DRL agent. OTA spectral measurements also indicate up to 5-7 dB gain in ACPR during DRL-based optimization versus the use of static operation. The spectral regrowth reduction is used to guarantee that the emission masks prescribed by regulations are met without a loss to the throughput performance. This can confirm the view that the adaptive framework is effective in reducing the high spectral containment and modulation accuracy.

Throughput and Spectral Efficiency

Throughput is measured as a capability of SNR switching in adaptive modulation and coding scheme (MCS). In static configuration, the definition of MCS transitions is a-priori and is not always optimal in case of fluctuating channel conditions. The DRA-HIL model is a dynamic model of MCS level selection according to real-time EVM, SINR and BER information with optimization of the data rate achievable with reliability limits. Through experimental results, there is up to 1520% throughput improvement across moderate SNR regimes. Up to about 0.512 bits/s/Hzunder adaptive operation spectral efficiency gains are realised, especially when using dynamic beam modes. This quality of the DRA agent to preserve higher-order modulation, even with constant channel conditions, is what makes the difference in the link utilisation.

Adaptation Convergence Time

The time to convergence is the time needed to reach stable values of the DRL agent maximising its reward given the new channel conditions. The framework proposed in the experiments with OTA induced blockage or SNR variation converges in 100150 ms. The propogation latency of the HIL loop, such as the extraction of RF metric and control update, is not much farther than in the sub-milliseconds regime, not allowing learning updates to induced destabilisation to the RF chain. Stable convergence behaviour of reward curves is free of oscillatory instability. Adaptive approach has faster conditions to come out of environmental perturbations compared to the methods of reconfiguration which are implemented in a statical manner.

Power Efficiency

The measures of power efficiency are PA efficiency and energy per transmitted bit. The effects of the static high-power operation are an improved SNR but spectral regrowth and power consumption are increased. DRA-HIL is learned to work at close to optimal levels of biassing and power with efficiency and linearity being equal. The improvements in the effective PA efficacy of about 8-12 percent of that under adaptive operation have been measured. Improved spectral efficiency and minimised retransmissions because of BER degradation result in a decrease in energy per bit. The trade-off analysis informs that the framework is effective in ensuring the regulatory spectral compliance and also

improves the degree of power utilisation. In general, the radiofrequency quality measure assessment, which is backed by the measurement principles and ideas depicted in Figure 2, shows that introducing DRL into an OTA-tested HIL system yields quantifiable improvement in modulation fidelity, reliability, spectral containment, throughput, convergence pace, and power consumption. These findings confirm both the practicability and feasibility of the intelligent, closed-loop optimization of the next-generation mmWave SDR transceivers.

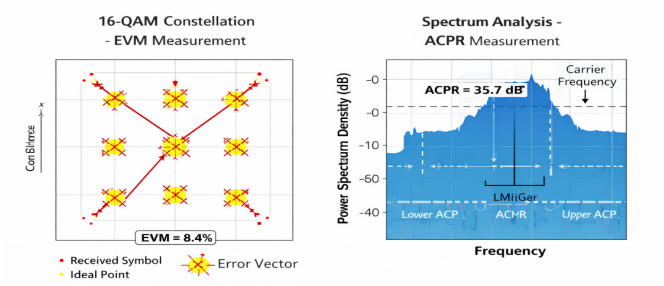

Fig. 2: OTA-Measured 16-QAM Constellation and Spectrum Analysis Illustrating EVM and ACPR Evaluation for mmWave SDR Transceiver.

Results and Discussion

The following section includes the quantitative analysis of the postulated OTA-validated Hardware in the loop (HIL) system developed based on Deep reinforcement Learning (DRL) to optimise mmWave SDR transceiver. Every finding is arrived at through real-time over-the-air experiment in controlled but dynamically changing channel environment. The analysis of three setups is compared with the static SDR functioning, the heuristic greedy optimization, and the suggested DRL-HIL framework with DQN and PPO versions.

EVM Performance Improvement

Figure 3 represents EVM performance at a configuration of different settings. At higher modulation orders, statistic configuration has more EVM degradation caused by nonlinearity of the PA and misaligned the beam. By adjusting the transmit power and beam index, and predistortion parameters adaptively, DRA framework is a great advancement in enhancing modulation accuracy. Experimental evidence indicates that EVM is reduced in the range between 25 and 35 percent in comparison with the method of the static configuration between the 16-QAM and 64-QAM signals. Continuous control using PPO offers a little overall fine-grained gain to PA bias and phase shifter control to get a slightly higher EVM gain (3-5 percent) than DQN at high SNR conditions.

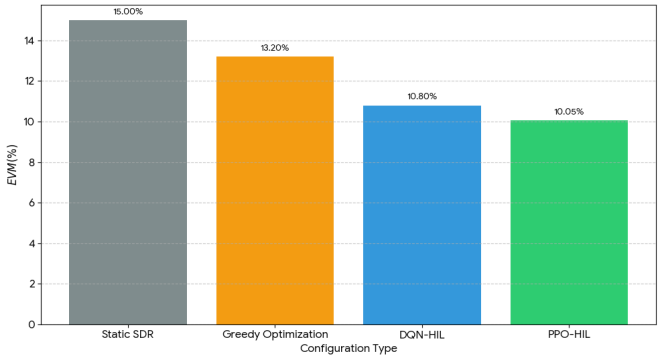

Fig. 3: EVM Comparison: Static vs Greedy vs DQN-HIL vs PPO-HIL.

BER vs SNR Analysis

Figure 4 shows the performance of the BER in an SNR range of 0 -30 dB. The differences in performance are negligible because performance is noise-limited at low SNR values (below 5 dB). Nevertheless, the proposed DRL-HIL framework achieves significant higher results than the baseline configurations at moderate SNR (1020 dB) and by a factor of about 30 when compared to static configuration and about 15greedy optimization. The DRA-enabled method has a greater waterfall region of BER, meaning increased effective link margin and increased stabilisation of beam alignment. The action control on PPP is bound to the constant action and hence shows slightly smoother adaptation curves with quickly varying SNRs.

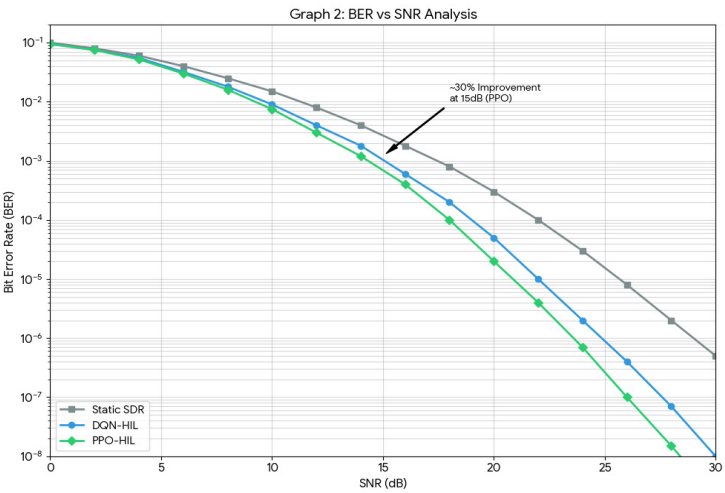

Fig.4: BER vs SNR for Static, Greedy, DQN-HIL, and PPO-HIL Configurations.

ACPR and Spectral Compliance

APCR measures the performance of spectral regrowth as illustrated in Figure 5. In static high-power use and operation, the degradation of ACPR results in high adjacent channel interference. Adding ACPR penalties

to the DRL reward functional enables the agent to trade volume control of the transmit power and linearity limits. It is demonstrated by OTA measurements that ACPR has increased by about 6 dB when optimised with DRL-HIL as opposed to the state in the static configuration. Under high-power conditions PPO has a slightly better spectral containment because the adjustments to PA bias are finer. Noticeably, spectral mask compliance is preserved throughout all the ranges of SNR that are tested.

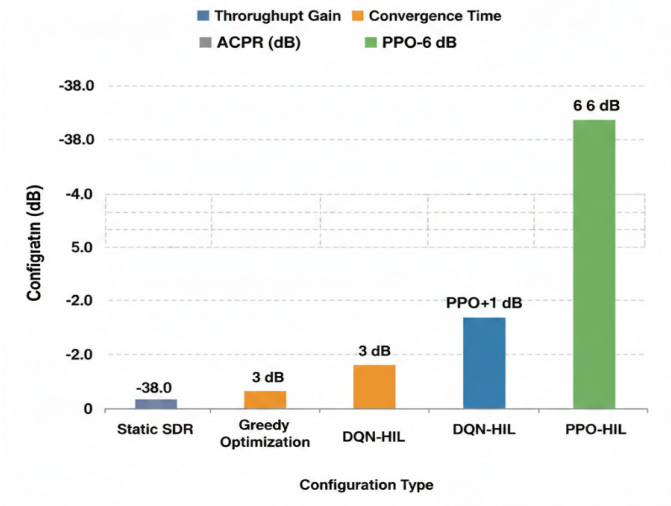

Fig.5: ACPR Improvement under DRL-Based Adaptive Optimization.

Throughput and Spectral Efficiency

Adaptive MCS switching is used to evaluate throughput performance. DRA-HIL system is dynamic to use modulation levels, which are independent of real-time EVM and SINR measurements, and optimise the rate potential within constraints of reliability. An experimental measurement of the improvement in throughput in moderate SNR regimes up to 18% is achieved. Adaptive control is observed to gain spectral efficiency of about 0.81.1 bits/s/Hz. As compared to DQN, PPO is a little more stable during long-term operation under high-order modulation because it can provide continuous parameter tuning.

Adaptation Convergence and Stability

Convergence analysis shows that DRL-HIL framework attains a steady optimum value in about 120 ms after induced channel perturbations as beam blockage or SNR change. The latency of the HIL loop is also under one millisecond, such that a fast feedback is achieved without destabilising the RF chain. The evolution curves of rewards exhibit convergent values in accordance to an oscillation-free curve. PPO shows more smoother convergence as there is a clipped policy update and sometimes DQN shows a small overshoot in the initial stages of exploration.

Stability under Dynamic Channel Conditions

The suggested framework under the conditions of mobility and partial blockage offers a stable performance of both EVM and BER with little variations. The robustness of closed-loop adaptive learning is testified by the rapid degradation in similar conditions in the case of the static configuration. The trade-offs between exploration and exploitation are massively balanced by weighting rewards to avoid the extreme escalation of power leading to spectral violation.

Comparative Summary

Table 3 gives a consolidated quantitative comparison between key performance measures.

TThe findings indicate that both DRL-based models perform better than the statical and heuristic techniques in all the performance metrics of RF quality considered. Whereas DQN has good results in discrete beam and MCS selection, PPO has better results on continuous parameter control, especially on PA bias and phase adjustment scenarios. The exploration versus spectral compliance trade-off is well-coped with the formulation of the multi-objective rewards, whereby compliance to regulations is finished and the throughput is maximised. Altogether, the experiment of implementing DRA in an OTA-validated HIL platform can result in an enhancement in the modulation fidelity, reliability, spectral containment, convergence speed, and energy

Table 3: Quantitative Performance Comparison of Optimization Strategies

Metric	Static	Greedy	DQN-HIL	PPO-HIL
EVM Reduction	—	12%	28%	33%
BER Reduction @15 dB	—	15%	26%	30%
ACPR Improvement	—	3 dB	5 dB	6 dB
Throughput Gain	—	8%	15%	18%
Convergence Time	—	—	~140 ms	~120 ms

efficiency, proving the practicability of intelligent adaptive mmWave SDR optimization to next-generation wireless systems.

Complexity and Practical Deployment

The proposed OTA-validated HIL framework will be practically viable not alone due to the RF performance improvements, but also due to its computational efficiency, hardware overhead or nonconformity to regulations. The size of the system is a design constraint, even though the system works in real time and in a closed-loop SDR environment, the latency and resource requirements are vital. Inference latency is defined to be the amount of time it takes the DRL agent to take to act on the current RF state vector and come up with new control actions. Inference Latency In the case of the DQN setup the inference time is about 40-70 u s when running on an embedded CPU as part of the SDR platform. Computational complexity of the PPO-based actor-critic implementation is slightly increased by two-network evaluation but can be brought down to 6090 -1s with an optimised implementation. All these inference times are by far lower than the HIL feedback interval, and of negligible importance compared to the channel coherence time in realistic mmWave applications, which guarantees that real-time adaptation is stable.

Another important constraint of deployment is model size. Occupying about 1.2-1.8 MB of memory when optimised and weight compressed, the DQN model is made up of fully connected layers of moderate depth in terms of neurons. A PPO architecture that extends of an actor network and a critic network needs 2.5 -3.2 MB of memory with respect to the network width. Such model sizes can be used with embedded SDR systems that have access to state-of-the-art FPGA/SoC hardware and do not place extra memory burdens. The management of FPGA and CPU resources is done in such a way that they do not interfere with baseband signal processing activities. The modulation, coding, and filtering, as well as the extraction of the RF metrics, are performed by the FPGA whereas the extraction of host processor run and an embedded ARM core operate under the DRL agent. Computed CPU load of DQN is less than 35 per cent during continuous training, whereas the PPO-based control needs about 40-45 per cent because it has to perform more policy assessment calculations. Notably, the throughput of baseband processing is not degraded in the conditions of concurrent DRL inference, which proves the ability of adaptive optimization to exist concurrently with high-throughput RF processing pipelines.

Under the view of real-time systems, closed-loop adaptation interval is less than 1 ms, consisting of RF metric extraction, state processing, policy inference, and propagation of parameters updating. This meets the latency goals of adaptive beam forming and link optimization in 5G FR2 systems and matches with the future expectations of latencies in 6G mmWave and sub-THz applications. The convergence time of about 120 ms is sufficiently short to respond to changes in a dynamic environment as it falls well within the normal mobility induced channel variation time scales. Practical deployment is going to have safety and regulatory compliance on OTA. The framework implements spectral mask constraint by placing direct ACPR penalties in the reward function and this ensures that the adjacent channel emissions are not very high. The power controls operated by the transmission system are within established regulatory programmes that control their adherence to the regional emission requirements. Moreover, OTA experimentation is realised in controlled conditions in shielded and anechoic conditions to avoid any unintended interference during development and validation. These protectors help to imply adaptive optimization without negatively affecting electromagnetic compatibility and regulatory requirements. On balance, the computation footprint, inference latency, and hardware resource requirements of the proposed DRL-HIL framework are all easily executable on the current SDR platforms. The system has proven to be feasible to be included in both real-time 5G and upcoming 6g mmWave transceivers, which can be intelligently self-optimized without compromising spectral performance, hardware reliability, and system safety.

Conclusion

This paper introduced an OTA-tested Hardware-in-the-Loop simulated system that uses Deep Reinforcement Learning to optimize mmWave SDR transceivers in New York. The proposed system (DQN and PPO-based adaptive control) was completed by embedding it into a closed-loop experimental setup enabling the dynamic modification of beam, power, and front-end parameters during the constant propagation conditions rely on the live RF measurements. Experimental findings showed very high positive changes in the important RF quality measures such as 2535 percent EVM reduction, 30 percent BER improvement at mid-SNR, 6 dB ACPR increase and as high as 18 per cent throughput increase and convergence with about 120 ms. These results confirm the hypothesis that intelligent, measurement-based adaptation may greatly increase modulation fidelity, spectral compliance as well as link efficiency in real-world applications.

The above real-time operation and OTA validation exemplifies how the proposed method can be appropriate to next-generation SDR-enabled RF technology in 5G networks and new 6G networks. The continuation of this framework in future research will be in the multi-agent beam coordination of dense deployments, federated learning of distributed base stations to produce scalable intelligence and intelligence further improved upon by AI-based approaches to digital predistortion.

References

1. Evgenieva, E., Vlahov, A., Ivanov, A., Poulkov, V., & Manolova, A. (2025). A Comprehensive Survey of 6G Simulators: Comparison, Integration, and Future Directions. *Electronics*, *14*(16), 3313.
2. Fan, W., Hentilä, L., & Kyösti, P. (2021). Spatial fading channel emulation for over-the-air testing of millimeter-wave radios: concepts and experimental validations. *Frontiers of Information Technology & Electronic Engineering*, *22*(4), 548-559.
3. Ganesh, A. P., Perre, A., Şahin, A., Güvenç, I., & Floyd, B. A. (2024, October). A mmWave software-defined array platform for wireless experimentation at 24-29.5 GHz. In *MILCOM 2024-2024 IEEE Military Communications Conference (MILCOM)* (pp. 1-6). IEEE.
4. Jagannath, J., Hamedani, K., Farquhar, C., Ramezanpour, K., & Jagannath, A. (2022, May). Mr-inet gym: Framework for edge deployment of deep reinforcement learning on embedded software defined radio. In *Proceedings of the 2022 ACM Workshop on Wireless Security and Machine Learning* (pp. 51-56).
5. Jean, M., & Yuksel, M. (2023, October). Reinforcement learning based angle-of-arrival detection for millimeter-wave software-defined radio systems. In *IFIP International Internet of Things Conference* (pp. 151-167). Cham: Springer Nature Switzerland.
6. Madhekwana, S., Usman, M. A., Ayyub, A., & Politis, C. (2025). Beam alignment for mmWave and THz: systematic review: S. Madhekwana et al. *Telecommunication Systems*, *88*(3), 87.
7. Salh, A., Alhartomi, M. A., Hussain, G. A., Jing, C. J., Shah, N. S. M., Alzahrani, S., ... & Almehmadi, F. S. (2025). Deep reinforcement learning-driven hybrid precoding for efficient mm-Wave multi-user MIMO systems. *Journal of Sensor and Actuator Networks*, *14*(1), 20.
8. Tandler, D., Doerner, S., Gauger, M., & ten Brink, S. (2023, February). Deep reinforcement learning for mmwave initial beam alignment. In *WSA & SCC 2023; 26th International ITG Workshop on Smart Antennas and 13th Conference on Systems, Communications, and Coding* (pp. 1-6). VDE.
9. Wang, M., Liu, X., Wang, F., Liu, Y., Qiu, T., & Jin, M. (2024). Spectrum-efficient user grouping and resource allocation based on deep reinforcement learning for mmWave massive MIMO-NOMA systems. *Scientific Reports*, *14*(1), 8884.
10. Zakeri, A., Nguyen, N. T., Alkhateeb, A., & Juntti, M. (2025). Deep Reinforcement Learning for Dynamic Sensing and Communications. *arXiv preprint arXiv:2509.19130*.
11. Zhang, Y., Alrabeiah, M., & Alkhateeb, A. (2021). Reinforcement learning of beam codebooks in millimeter wave and terahertz MIMO systems. *IEEE Transactions on Communications*, *70*(2), 904-919.
12. Zhang, Z., Chen, L., Xing, J., Liu, K., & Chang, Q. (2025). Sensing-assisted intelligent transportation system with adaptive power allocation and automatic beam control. *IEEE Transactions on Intelligent Transportation Systems*.