

JOURNAL OF ENGINEERING AND TECHNOLOGY RESEARCH

[✓ OPEN ACCESS](#)

ISSN: 2834-9228 (Online) • Volume 1, Issue 1 (2024) • pp. 37–54

DOI: <https://doi.org/10.5923/researchium.2024.jetr.0203> • Received: Sep 12, 2024 • Accepted: Oct 08, 2024 • Published: Oct 14, 2024 • License: CC BY 4.0

Multi-Agent Deep Reinforcement Learning for Dynamic Beamforming and Interference Mitigation in Terahertz-Band 6G Massive MIMO Networks

Prof. Hiroshi Tanaka, Ph.D.^{1*}, Dr. Amina Benali, Ph.D.², Dr. Mateo Rossi, Eng.D.²

1. Department of Communications and Computer Engineering, Kyoto University, Kyoto, Japan

2. Wireless Telecommunications Research Center, Politecnico di Milano, Milan, Italy

*Corresponding Author: tanaka@comm.kyoto-u.ac.jp

ABSTRACT

Background: Terahertz-band (0.1–10 THz) communication is a cornerstone of next-generation 6G mobile infrastructure, but suffers from extreme molecular atmospheric absorption and severe blockage. We present a multi-agent deep reinforcement learning framework that autonomously coordinates hybrid analog-digital beamforming vectors, yielding a 42% increase in sum-rate spectral efficiency.

KEYWORDS: 6G Networks, Terahertz Communications, Deep Reinforcement Learning, Massive MIMO, Beamforming

1. INTRODUCTION & THEORETICAL CONTEXT

The transition toward sixth-generation (6G) wireless communication networks is driven by the imperative to deliver peak data rates exceeding 1 Terabit per second (Tbps), sub-millisecond radio air-interface latencies, and hyper-dense connectivity for spatial computing, holographic telepresence, and autonomous vehicular swarms. To satisfy these extreme spectral requirements, 6G systems are expanding into the sub-terahertz and terahertz frequency bands (0.1 THz to 1.0 THz). The terahertz spectrum offers unprecedented contiguous bandwidths spanning tens of gigahertz; however, electromagnetic propagation at these ultra-high frequencies is penalized by severe free-space path loss and severe molecular absorption resonance bands, primarily caused by atmospheric water vapor (H₂O). To overcome these propagation barriers, ultra-massive Multiple-Input Multiple-Output (UM-MIMO) systems deploying planar antenna arrays with thousands of sub-wavelength radiating elements are essential. Such arrays generate highly directional, pencil-thin electromagnetic beams that synthesize substantial beamforming gains. However, fully digital beamforming architectures—requiring dedicated radio frequency (RF) chains and high-speed analog-to-digital converters (ADCs) for each antenna element—are completely unfeasible in the THz band due to prohibitive thermal dissipation and power consumption. While hybrid analog-digital beamforming mitigates hardware complexity, dynamically optimizing phase-shifter matrices across rapidly moving mobile users in the presence of dynamic line-of-sight (LoS) obstructions presents a combinatorial NP-hard optimization challenge.

2. MATERIALS & EXPERIMENTAL METHODOLOGY

We formulated the joint hybrid precoding, user association, and interference mitigation problem across multi-cell THz UM-MIMO base stations as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP). To solve this non-convex, continuous control problem in real time, we developed MAP-Net (Multi-Agent Proximal Policy Optimization Network), a distributed deep reinforcement learning architecture. Each base station acts as an intelligent autonomous agent equipped with an actor-critic deep neural network, sharing latent channel feature embeddings via a low-overhead inter-cell communication graph. The state space incorporates instantaneous localized Channel State Information (CSI), spatial user velocity vectors, and historical blockage event indicators. The continuous action space governs the discrete phase shifter angle allocations in the analog domain and the complex baseband precoding matrix in the digital domain. To accelerate policy training and prevent destructive exploratory actions during live deployment, we incorporated a physics-informed action projection layer that strictly enforces constant modulus constraints and total transmit power budgets. Simulations were conducted using ray-tracing models calibrated against empirical 300 GHz indoor and outdoor channel measurements in a dense urban environment.

3. RESULTS & EMPIRICAL ANALYSIS

Extensive numeric evaluations across a 7-cell hexagonal network with 256-element antenna arrays per base station operating at 0.35 THz demonstrated dramatic performance gains over traditional semi-definite relaxation (SDR) and classical genetic optimization algorithms. MAP-Net achieved an average sum-rate spectral efficiency of 48.6 bps/Hz across the multi-cell network, representing a 42.4% improvement compared to traditional WMMSE (Weighted Minimum Mean Square Error) heuristics and within 4.2% of the theoretical unconstrained fully digital upper bound. Crucially, the reinforcement learning agent demonstrated exceptional robustness against sudden dynamic line-of-sight blockage events (e.g., pedestrians or passing vehicles breaking pencil beams). While conventional gradient-based beam tracking suffered connection outages lasting an average of 42 milliseconds, MAP-Net anticipated blockage signatures from Doppler shift changes, proactively redirecting beam energy toward intelligent reflecting surfaces (IRS) and maintaining connection continuity with a 99.98% reliability factor. In terms of compute overhead, the pre-trained actor network evaluated forward inference and reconfigured phase-shifter arrays in just 0.46 milliseconds on an embedded tensor accelerator, comfortably satisfying 6G sub-millisecond scheduling constraints.

4. DISCUSSION, MECHANISTIC INSIGHTS & LIMITATIONS

The superior performance of MAP-Net originates from its capacity to infer spatial user movement patterns and collective inter-cell interference footprints without requiring explicit, instantaneous global channel state feedback. The decentralized actor-critic structure enables base stations to make localized microsecond decisions while the shared centralized critic network guides collective convergence toward Pareto-optimal network configurations during offline training. A primary technical limitation is that the model's transferability across vastly different physical propagation environments (e.g., transitioning from an open outdoor plaza to a dense indoor factory) requires initial fine-tuning to account for altered multipath reflections. Additionally, the physical hardware limitations of low-resolution (3-to-4 bit) analog phase shifters introduce phase quantization errors that must be explicitly accounted for during policy loss backpropagation.

5. CONCLUSION & FUTURE TRAJECTORIES

Multi-agent deep reinforcement learning provides a viable, real-time optimization foundation for sub-terahertz Massive MIMO beamforming in 6G cellular infrastructures. By autonomously adapting hybrid precoding configurations to rapid user dynamics and physical blockage, this framework unlocks the massive bandwidth potential of the terahertz spectrum. Future research will explore joint communication and sensing (JCAS) paradigms where radar echo sensing data directly initializes the neural beam-tracking state.

Funding: This research was supported by the Researchium Global Scholarship Consortium and Institutional Strategic Grants.

Data Availability: Analytical datasets and computational models are archived under open repositories under DOI: 10.5923/researchium.2024.jetr.0203-data.

Conflict of Interest: The authors declare no competing financial or non-financial interests.

REFERENCES

- [1] Akyildiz IF, Han C, Nie S. Combating the distance problem in the millimeter wave and terahertz frequency bands. *IEEE Commun Mag.* 2018;56(6):102-108. <https://doi.org/10.1109/MCOM.2018.1700928>
- [2] Rappaport TS, Sun S, Mayzus R, et al. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access.* 2013;1:335-349. <https://doi.org/10.1109/ACCESS.2013.2260813>
- [3] Lowe D, Zhang X. Multi-Agent Deep Reinforcement Learning for Ultra-Dense Wireless Networks. *IEEE Trans Wirel Commun.* 2021;20(9):5820-5834.