

## 무선 통신 환경에서의 강화 학습 기반 빔포밍 최적화 기술 동향

김수빈, 박수현\*  
숙명여자대학교

rlatnqls0822@sookmyung.ac.kr, \*soohyun.park@sookmyung.ac.kr

## Recent Advances in Reinforcement Learning-Based Beamforming Optimization for Wireless Communication

Subin Kim, Soohyun Park\*  
Sookmyung Women's University

## 요 약

본 논문은 5G 및 차세대 6G 무선통신 환경에서의 빔포밍 최적화 문제를 강화학습 (Reinforcement Learning, RL) 기반으로 접근한 주요 연구 동향을 분석한다. Massive MIMO, mmWave, IRS (지능반사표면) 등 다양한 무선 채널 구조에서 심층 강화학습 (Deep RL, DRL) 알고리즘이 적용되어 에너지 효율 (Energy Efficiency), 스펙트럼 효율 (Spectral Efficiency), 보안성(Secrecy Rate) 등을 극대화하는 방법이 제시되어 왔다. 본 연구는 이러한 연구들의 공통점 및 한계점을 분석하고, 양자강화학습(Quantum Reinforcement Learning, QRL) 기반 접근의 필요성을 논의한다.

## I. 서 론

다중 입출력 (Multiple-Input Multiple-Output, MIMO) 기술은 송수신 단에 다수의 안테나를 배치하여 전송 용량 및 신호 신뢰도를 향상시키는 핵심 기술이다. 특히 Massive MIMO 는 수십~수백 개의 송신 안테나를 통해 빔포밍 (Beamforming) 기법을 수행함으로써, 특정 사용자에게 에너지를 집중시켜 간섭을 최소화하고 높은 전력 효율을 달성한다.

기존의 빔포밍 설계는 Zero-Forcing (ZF), Minimum Mean Square Error (MMSE) 등 폐형식 해 (closed-form solution)에 의존하였으나, 현실적인 채널 비선형성, 사용자 수의 증가, IRS-NOMA 등의 복잡한 구조에서는 계산 복잡도가 급증해 적용에 한계가 존재한다. 반면, RL 기반의 빔포밍 최적화는 환경과 상호작용하며 최적 정책을 학습함으로써, Channel State Information (CSI)가 불완전하거나 고차원인 상황에서도 준최적 해를 효율적으로 탐색할 수 있다는 장점을 가진다.

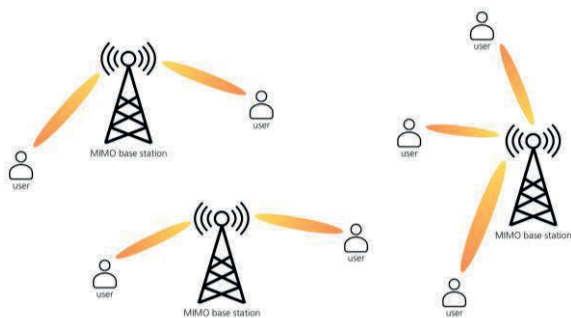


그림 1. MIMO 환경에서의 빔포밍

빔포밍 문제에서 코드북 (CodeBook) 기반의 빔 벡터 선택은 이산 공간에서 이루어진다. 예를 들어, Discrete Fourier Transform (DFT) 기반의 코드북에서 가능한 빔의 조합 수는 안테나 수와 동일하게 구성된다. 이 때, 전력 레벨, 사용자 매칭 등이 추가되면 액션 공간의 차원 폭발 (curse of dimensionality) 문제가 발생한다. 또한, 실제 무선 하드웨어의 특성상 이러한 변수들은 연속적인 값으로 조정되기 어렵고, 제한된 분해능 내에서만 제어가 가능하다. 예를 들어, 위상 시프터 (Phase Shifter)는 일반적으로 B-bit 양자화된 이산 위상만 구현할 수 있다. 이러한 제약으로 인해 빔포밍 설계의 탐색 공간은 이산적으로 구성되어야 한다. 따라서 RL 설계시 액션을 이산적으로 정의하고, 탐색 효율을 높이기 위한 DRL 및 계층형 구조가 필요하다.

최근 전력 소비와 탄소 배출 간의 상관관계를 다룬 연구에 따르면, 전력 소비량의 증가는 탄소배출량 증가와 유의미한 양의 상관관계를 보인다고 보고되고 있다 [1],[2]. 무선 통신 시스템은 기지국의 전력 소모가 전체 네트워크 에너지의 대부분을 차지하므로, 빔포밍 효율의 향상은 곧 통신 인프라의 전력 소모 절감으로 직결된다. 따라서 빔포밍 최적화는 불필요한 전력 낭비를 줄이고, 통신 시스템의 에너지 효율을 높여 결과적으로 이산화탄소 배출량 저감에 기여할 수 있다.

## II. 본론

최근 강화학습의 행동 공간을 이산적으로 설정하고, 보상함수를 에너지 효율로 정의한 다양한 대규모 무선 통신 환경에서의 빔포밍 최적화 연구가 활발히 이루어지고 있다. 먼저, 다중 기지국 (Multi-Input Single Output, MISO) 환경에서 저해상도 위상 시프터 기반의 아날로그 빔포밍을 DRL 로 최적화한 연구가 제안되었다 [3]. 이 연구는 지역 채널 정보 및 인접 셀의 간섭량

등을 바탕으로, 위상 벡터와 전력 레벨 등을 고려하였다. DRL 은 각 셀의 국소 학습(local learning)을 통해 중앙 집중형 CSI 교환 없이도 높은 에너지 효율을 달성했다. 특히 전통적 Fractional Programming (FP) · Weighted Minimum Mean Square Error (WMMSE) 기반 방법 대비 15~20%의 에너지 효율 향상을 보였다. 한편, Massive MIMO-NOMA (Non-Orthogonal Multiple Access) 다운링크 환경에서는 DFT 코드북을 기반으로 한 DQN (Deep-Q-Network) 기반 빔 선택 및 사용자 그룹핑 기법이 제안되었다 [4]. 이 연구는 에너지 효율 외에도 합 전송률도 보상함수에 포함해 공동 최적화를 진행하였다. 10dB 의 전송 Signal-to-Noise Ratio (SNR)에서 기존 K-means 기반 NOMA 대비 42%의 에너지 효율 향상을 달성하였다. IRS-assisted Massive MIMO 시스템에서 수행된 DQN 기반 하이브리드 빔포밍은 BS-IRS-사용자 간의 복합 채널 (cascade channel)을 이용해 빔-사용자 쌍을 선택하고, IRS 반사 위상 행렬을 조정하였다 [5]. 이 DRL 기반 설계는 기존 Greedy 대비 스펙트럼 효율 7.52%, 에너지 효율 16.47%가 개선되었다. 또한, mmWave 네트워크에서는 Advantage Actor-Critic (A2C) 구조를 적용하여 Grid-of-Beams (GoB) 기반 빔 선택 및 전력 제어를 통합적으로 수행한 연구가 있다 [6]. 각 사용자의 Signal-to-Interference-plus-Noise Ratio (SINR)을 고려하여 빔 방향 서브셋 및 전력 수준을 동적으로 조정하였으며, 커버리지 제약으로 포함시켜 점진적인 에너지 효율 향상을 유도했다. 단순 DQN 대비 안정적 수렴을 보였으며, 전송 전력 최적화 포함시 에너지 효율이 약 두 배 향상되었다. 마지막으로, 6G mmWave 채널에서 DRL 기반 빔 트레이닝 연구도 제안되었다 [7]. 이 연구는 에너지 효율 최적화형 (DRL-EE)과 스펙트럼 효율 최적화형 (DRL-SE)의 상태와 보상함수가 다른 2 가지 모델을 하였다. 환경 변화를 학습하며 빔 트레이닝 횟수와 활성 RF 체인 수를 동적으로 조절해 차단(blockage) 상황에서도 장기 정책을 학습해 스펙트럼 효율 손실을 최소화하고, 전력-성능 균형형 보상 모델을 통해 유연한 트레이드오프 조정이 가능함을 보였다.

이러한 연구들을 통해 강화학습이 빔포밍 최적화에서 높은 잠재력을 지니고 있음이 입증되었으나, 여전히 몇 가지 공통적 한계가 존재한다. 첫째, Massive MIMO 및 IRS 환경에서의 상태, 행동 공간의 폭발적 증가로 학습 수렴이 느리고 계산 복잡도가 매우 높다. 둘째, 채널 불확실성 및 노이즈 환경에서 DRL 이 불안정하게 수렴하거나 지역 최적(Local Optimum)에 머물 수 있다. 따라서 향후 연구에서는 이러한 한계를 극복하기 위한 새로운 형태의 학습 프레임워크가 요구된다.

### III. 결론

본 논문에서는 5G 및 차세대 6G 무선통신 환경에서 RL 기반 빔포밍 최적화의 연구 동향과 한계점을 종합적으로 분석하였다. 기존 연구들은 DRL 을 이용해 하이브리드 빔포밍, IRS 반사 제어, 전력 조절, 사용자 그룹핑을 효과적으로 최적화하였으나, 상태, 행동 공간의 고차원성 등의 문제로 인해 한계가 존재했다. 이러한 한계를 보완하기 위해 QRL 기반 연구가 새로운 대안으로 주목된다. QRL 은 양자 중첩과 병렬 탐색 특성을 활용하여 기존 DRL 보다 더 빠른 정책 탐색과 높은 전력 효율을 기대할 수 있다. 실제로 QRL 을 적용한 5G HWN 연구에서는 피코 엣지(pico edge)

사용자에 대한 보안성과 에너지 효율을 동시에 최대화하도록 설계되었다 [8]. 에이전트는 양자 상태 벡터를 기반으로 행동을 선택하며, 기존 DRL 대비 수렴 속도 30% 단축 및 전력 소비 15% 감소를 달성하였다. 향후 연구에서는 QRL 기반 빔포밍 최적화를 통해 기존 DRL 의 구조적 한계를 보완하고, 대규모 MIMO 및 IRS 환경에서도 효율적으로 적용 가능한 학습 프레임워크를 구축할 필요가 있다. 이를 통해 보다 높은 에너지 효율을 확보하고, 통신 인프라의 전력 소모를 줄여 탄소 배출량 감소에 기여할 것으로 기대된다.

### ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터사업의 연구결과로 수행되었음(IITP- 2024-RS- 2024-00436887).

### 참 고 문 헌

- [1] 진보영, 김근우, 박중구, “한국의 경제성장, 전력소비량, 이산화탄소배출량 및 환경규제 간 인과관계 분석,” 에너지공학, vol. 29, no. 1, pp.1- 12, March 2020.
- [2] A. A. Agaslim, “Empirical findings on the relationship of energy consumption, gross domestic product per capita and carbon dioxide(co2) emissions,” International Journal of Energy Economics and Policy, vol. 14, no. 4, pp. 684- 690, May 2024.
- [3] H. Zhou, X. Wang, M. Umehira, and Y. Ji, “A deep reinforcement learning based analog beamforming approach in downlink miso systems,” in 2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring). IEEE, June 2022, pp. 1- 6.
- [4] I. Ahmed, M. K. Shahid, and T. Faisal, “Deep reinforcement learning based beam selection for hybrid beamforming and user grouping in massive mimo-noma system,” IEEE Access, vol. 10, pp. 89 519- 89 533, 2022.
- [5] I. Ahmed, K. Shahid, and H. Khammari, “Drl based beam selection and hybrid beamforming for intelligent reflective surface assisted massive mimo system,” in 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring). IEEE, August 2023, pp. 1- 6.
- [6] Y. Dantas, P. E. Iturria-Rivera, H. Zhou, M. Bavand, M. Elsayed, R. Gaigalas, and M. Erol-Kantarci, “Beam selection for energy-efficient mmwave network using advantage actor critic learning,” in IEEE International Conference on Communications (ICC). IEEE, May, June 2023, pp. 5285- 5290.
- [7] Narengerile, J. Thompson, P. Patras, and T. Ratnarajah, “Deep reinforcement learning-based beam training with energy and spectral efficiency maximisation for millimetre-wave channels,” EURASIP Journal on Wireless Communications and Networking, vol. 2022, no. 1, p. 110, November 2022.
- [8] H. Sharma, G. Sharma, N. Kumar et al., “Secrecy maximization for pico edge users in 5g backhaul hwns: A quantum rl approach,” in 2022 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS). IEEE, December 2022, pp. 207- 212.