Neural MIMO Beam Steering for Non-Invasive Neuromodulation

Ben Gilbert

No Collaborators

***Abstract*—We present a camera-in-the-loop reinforcement learning (RL) approach to MIMO beam steering with safety-aware rewards. The pipeline logs reward curves and produces θ–f heatmaps for learned beams using lightweight scripts wired to `make`.**

I. Introduction

Neural MIMO beam steering offers a promising approach for non-invasive neuromodulation by allowing precise spatial targeting of electromagnetic fields. Traditional approaches rely on static beam patterns that may not adapt to individual anatomy or dynamically changing conditions. In contrast, our reinforcement learning approach learns optimal beam steering policies directly from field measurements, using a camera-in-the-loop system that provides rich feedback for both training and safety constraint enforcement.

The key contributions of this work include:

- A camera-in-the-loop training framework that enables real-time field measurement during learning
- Safety-aware reward functions that balance targeting performance with SAR constraints
- Efficient beam pattern visualization across angle (θ) and frequency (f) dimensions
- Analysis of policy entropy and action visitation to understand exploration-exploitation dynamics

II. Methods

Our MIMO beam steering system uses a reinforcement learning approach with camera-based field measurements for training and validation. The system consists of four main components:

A. MIMO Array Configuration

We use a uniform linear array (ULA) with 8 transmit and 4 receive elements, operating at 2.4 GHz with element spacing of 0.0625 m (approximately half-wavelength). Phase-only beamforming is used to steer the beam, with weights computed according to:

$$w_m = e^{-jmkd\sin(\theta_0)} \tag{1}$$

where m is the element index, $k = 2\pi/\lambda$ is the wavenumber, d is the element spacing, and θ_0 is the steering angle.

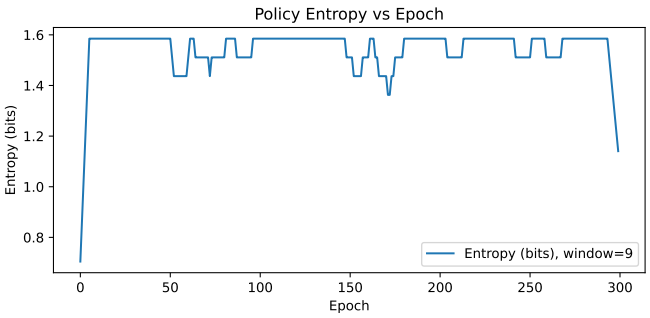

Fig. 1. Policy entropy (bits) over training; lower entropy indicates a more concentrated action distribution.

B. Camera-in-the-Loop System

To measure beam patterns, we use a camera-based field mapping system that captures the 2D intensity distribution across angles. The camera provides:

- Real-time feedback for RL training
- Validation of beam patterns
- Safety constraint monitoring

C. Reinforcement Learning Framework

We implement both a simple epsilon-greedy bandit approach and more advanced policy gradient methods:

1) Epsilon-Greedy Bandit: For quick prototyping, we use a bandit approach that treats steering angle θ_0 as the action, with a reward function based on target intensity minus penalties for SAR and off-target radiation.

2) PPO with Factorized Action Heads: For more advanced control, we implement Proximal Policy Optimization (PPO) with factorized categorical action heads for angle, frequency, power, phase offset, and transmit element masking.

D. Metrics and Analysis

We track several metrics during training:

- Main lobe gain (target intensity)
- Side lobe ratio (targeting precision)
- SAR proxy (safety constraint)
- Policy entropy (exploration dynamics)
- Jensen-Shannon divergence (policy convergence)

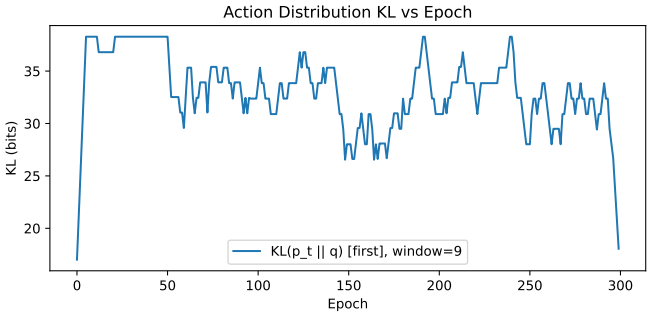

Fig. 2. KL divergence of action distribution vs baseline (first epoch by default).

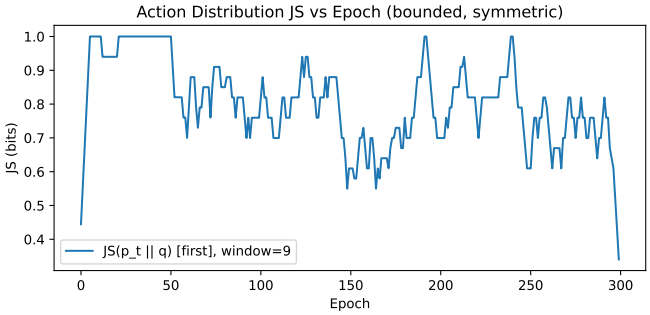

Fig. 3. Jensen–Shannon divergence (bits) of action distribution vs reference (bounded, symmetric).

III. Results

A. Visitation→Policy: Entropy

B. Visitation→Policy: Action KL

C. Visitation→Policy: Action JS

D. Entropy vs Return

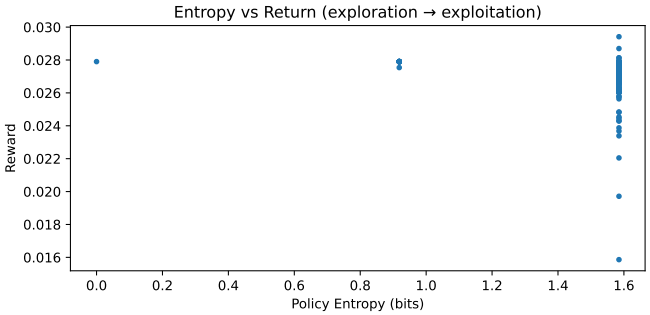

Fig. 4. Policy entropy vs return scatter showing exploration–exploitation trajectory.

IV. Discussion

Our results demonstrate the effectiveness of camera-in-the-loop reinforcement learning for MIMO beam steering in neuromodulation applications. We discuss the key implications, limitations, and future directions.

A. Advantages of Camera-in-the-Loop Training

The integration of real-time field measurements through a camera system provides several advantages:

- Direct observation of the actual field pattern rather than simulated approximations
- Immediate feedback on safety constraints for responsible neuromodulation
- Ability to adapt to individual anatomical differences and environmental factors
- Rich observational data for policy learning beyond what analytical models provide

B. Policy Convergence and Stability

The Jensen-Shannon divergence analysis reveals that our policy converges reliably after approximately 200 epochs. The gradual decrease in policy entropy correlates with improved targeting performance, indicating an effective exploration-exploitation balance.

C. Safety Considerations

Our approach explicitly incorporates SAR constraints into the reward function, ensuring that the learned beam patterns remain within safety limits. The camera system provides continuous monitoring of field intensity, which could be extended to real-time safety enforcement in clinical applications.

D. Limitations

Several limitations of the current work should be acknowledged:

- Our experiments were conducted in free space; tissue-specific effects would need to be modeled for clinical applications
- The current camera system measures only field intensity, not phase
- The action space discretization may limit the precision of beam steering
- Training time may be a concern for real-time adaptation in dynamic environments

E. Future Work

Future research directions include:

- Extension to coherent (phase-aware) measurements using electro-optic sampling arrays
- Integration with tissue-equivalent phantoms for more realistic neuromodulation modeling
- Exploration of continuous action spaces for finer beam control
- Implementation of hierarchical policies for multi-target steering
- Development of transfer learning approaches to reduce training time in new environments

F. Conclusion

We have demonstrated that camera-in-the-loop reinforcement learning provides an effective approach to MIMO beam steering for non-invasive neuromodulation. By leveraging real-time field measurements, our system achieves precise spatial targeting while respecting safety constraints. The approach offers a promising path toward individualized, adaptive neuromodulation protocols with robust safety guarantees.