Learning Neural Manifold Representations for Articulated Robotics

ECE594 Research Project Proposal

Project leads: Mathilde Papillon and Francisco Acosta, Geometric Intelligence Lab @ UCSB

Motivation. Recent work in computational neuroscience (NeuroAI) has shown that recurrent neural networks (RNNs) trained on *state-estimation objectives for navigation* (i.e., RNNs trained to predict position of a 2D point in space given information on its velocity through time) learn structured internal representations. Specifically, the internal representations of the data (or *embeddings*) closely resemble biological grid cells. Grid cells are neurons in the mammalian brain with highly structured activity patterns, that perform comptuations necessary for spatial navigation. Researchers have found that these representations (embeddings) can improve downstream reinforcement learning (RL) models whose goal is to locate and remember rewards in space. In other words, when an RL model is trained with these embeddings, instead of the raw data, it performs better. These findings suggest a general principle: state-estimation objectives can induce low-dimensional, highly structured internal representations that are both interpretable and functionally useful. However, nearly all existing work focuses on point-agent spatial navigation in 2D space. Whether similar representational structure emerges for physical, articulated bodies remains an open question with wide-spanning applications to robotics and embodied agents.

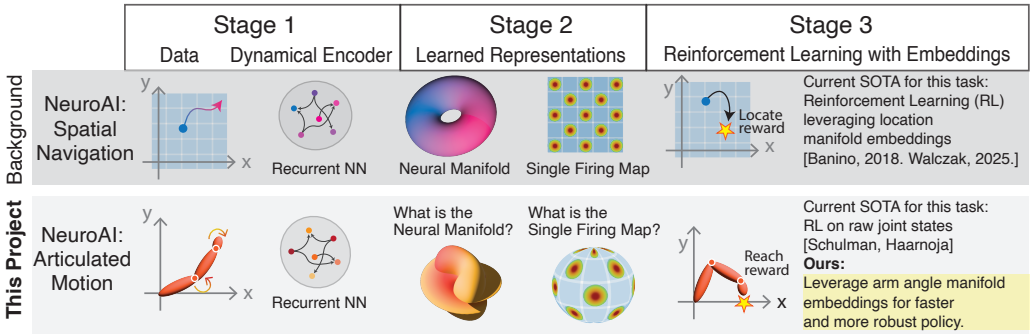

Figure 1: **Overview of our proposal.** Details explained below.

Project goal. The goal of this project is to test whether the NeuroAI paradigm of state estimation $\rightarrow$ structured internal representations $\rightarrow$ improved control (RL) extends to a minimal articulated motion system: a robotic arm. We consider a fixed-base articulated arm with two rotational joints, each parameterized by a full 3D orientation. The configuration space of this system is $\mathcal{Q} = SO(3) \times SO(3)$, making it the simplest nontrivial analogue of spatial navigation for articulated motion in 3D. In contrast to navigation in Euclidean space or planar joint systems, $\mathcal{Q}$ here is a product of non-commutative rotation manifolds, introducing fundamentally new geometric and topological structure beyond periodic position variables.

Methodology. The project consists of three stages, explained below and visually in Fig. 1

(1) Body-state estimation via Dynamical Encoder. We train a dynamical Encoder (RNN or LSTM) to perform *path integration*. The model receives joint angular velocities as input and must integrate them over time to estimate joint configuration. Rather than regressing joint states directly, the network is trained to predict a population code of *pose place cells* tiled over $\mathcal{Q}$, using a loss that respects the geometry of the configuration space. This explicitly separates state estimation from control and mirrors the NeuroAI navigation setup.

(2) Visualizations and Analysis of Learned Representations. To understand what is learned by the state-estimation network, we perform lightweight analyses of the recurrent hidden activity. This includes visualizing latent trajectories using dimensionality reduction techniques (e.g., PCA and UMAP) and examining how latent coordinates vary with underlying joint configurations (e.g., visualize tuning curves of neurons with respect to task variables). The goal is to assess whether the learned latent space exhibits structured, low-dimensional organization aligned with the underlying configuration space.

(3) Use Representations as Embeddings for Reinforcement Learning. We evaluate the functional usefulness of the learned representation by using it as the state input to a downstream reinforcement learning agent on the standard `Reacher-v5` benchmark. Performance is compared against relevant baselines using raw joint states. Metrics include sample efficiency, convergence speed, robustness to noise, and generalization across target locations.

Expected outcomes and impact. This project provides a clean, falsifiable test of whether articulated state-estimation objectives induce structured neural representations analogous to grid codes in navigation. A positive result would establish a principled bridge between NeuroAI and robotics, demonstrating how structured representations can emerge naturally and improve control without being hand-designed. The scope is well matched to a 10-week graduate research project, while producing results that are suitable for submission to NeurIPS 2026.

Project structure and student involvement. Students will work closely with the project leads, Mathilde Papillon and Francisco Acosta, who will provide an initial Python codebase via a shared GitHub repository. Students are expected to contribute Python code, follow clean coding practices, and maintain clear documentation in the collaborative repository. The project leads will also provide and maintain an Overleaf document during the paper-writing phase. Students will have the opportunity to serve as co-authors on the resulting NeurIPS submission. The project leads will be available for regular meetings ($\sim$ 1 /week) and provide ongoing communication via Slack.