Radboud Universiteit

PhD Research Proposal

Name: Sagar Prakash Barad

Topic: Theory of Learning in Artificial and Biologically Inspired Neural Networks

Position: PhD student in Neuroscience

Institute: Donders Institute for Brain, Cognition and Behaviour, Radboud University

Prior publications

Articles

Sagar Prakash Barad, Sajag Kumar, and Subhankar Mishra. *Estimation of Electronic Band Gap Energy From Material Properties Using Machine Learning*. 2024. arXiv: `2403.05119 [cond-mat.mtrl-sci]`. URL: `https://arxiv.org/abs/2403.05119`.

Rucha Bhalchandra Joshi et al. *Graph Neural Networks at a Fraction*. 2025. arXiv: `2502.06136 [cs.LG]`. URL: `https://arxiv.org/abs/2502.06136`.

Research Project

Efficient and constrained learning in neural circuits: structure, energy, and transfer

Abstract

Biological neural circuits learn under stringent structural and energetic constraints that fundamentally shape their representational geometry and adaptability. I propose a three phase program that integrates analytic order-parameter theory with reproducible computational experiments: (A) characterize how wiring constraints drive transitions in representational geometry and network phase, (B) derive principled energy accuracy tradeoffs for synaptic plasticity and identify optimal hybrid plasticity schedules, and (C) establish architecture-aware task-similarity bounds that predict transfer and generalization. Each phase yields testable analytical predictions, open benchmarks, and public code, ensuring immediate empirical verifiability.

Motivation & Problem Statement

Random-matrix and spiked-covariance theory show that structured low-rank signals produce outlier eigenvalues that undergo a sharp detectability transition as noise increases (BBP transition), a natural mathematical foundation for "phase boundaries" in representation geometry. [3] Population-level dimensionality measures (participation ratio, effective dimensionality) are widely used to quantify neural representations in both experimental and model systems. [4] At the same time, the brain's energetic budget places a strong cost on signaling and plasticity, motivating models that explicitly trade off metabolic cost and memory/accuracy. [1] Finally, empirical task-taxonomies and robust representation-similarity measures (SVCCA / CKA) demonstrate that task relationships are measurable and exploitable for transfer learning, suggesting we can formalize transferability in architecture-dependent bounds. [6]

Research Questions & Hypotheses

1. **Connectivity constraints and representational geometry.** *Research question:* How do low-rank, sparse, or modular wiring patterns shape the geometry of learned representations, and where are the phase boundaries between low-dimensional, task-aligned, and chaotic regimes? *Hypothesis:* Such transitions arise at analytically predictable thresholds derived from spiked-random-matrix theory and small-DMFT (dynamical mean-field theory) extensions.

2. **Dynamics and stability of learning under structural constraints.** *Research question:* How do recurrent feedback and structural sparsity affect the stability and convergence of learning dynamics in biologically constrained circuits? *Hypothesis:* Constrained architectures exhibit critical points where learning dynamics shift from stable fixed-point convergence to chaotic or oscillatory regimes, identifiable via Lyapunov spectra and mean-field order parameters.

3. **Energetic limits and optimal plasticity policies.** *Research question:* Under explicit metabolic or energetic cost constraints, what are the optimal update policies for synaptic plasticity, and how do they trade off accuracy against energy expenditure? *Hypothesis:* Hybrid plasticity schedules that mix fast labile updates with slow consolidation achieve Pareto-optimal energy–accuracy tradeoffs and outperform constant-update rules in energy-limited continual-learning settings.

4. **Architecture-aware transfer and task similarity.** *Research question:* How do architectural constraints (e.g., low-rank bottlenecks or modular wiring) interact with task similarity to determine the limits of transfer and generalization? *Hypothesis:* A task-similarity kernel combined with architectural constraints yields provable upper bounds on transfer gains; these bounds can be empirically calibrated using representational-similarity metrics such as CKA or SVCCA.

Methods (Analytic + Computational Pipeline)

Project A: Phase Diagram of Representational Geometry

Analytic: Start from linear or feedforward models and low-rank-plus-noise ensembles; use BBP/spiked-matrix results to derive thresholds for outlier survival and participation-ratio scaling; extend via perturbative DMFT or order-parameter calculations to simple non-linear and recurrent dynamics [3].
Computational: Controlled simulations: (i) toy linear maps with rank/noise sweeps and participation-ratio heatmaps; (ii) low-rank RNNs trained on simple curricula to measure ID, alignment with task axes, and generalization. Reproduce analytic phase boundaries and measure deviations.

Project B: Energetic Rate–Distortion for Synaptic Learning

Analytic: Formulate learning as minimization of expected loss + $\lambda\times$ energy(cost per update). Derive optimal update scheduling as a stochastic control / rate–distortion problem and compute approximate closed-form policies under simplified dynamics (e.g., linear regression, single-layer perceptron). Ground the energy term with empirical brain signaling cost estimates [1].
Computational: Implement hybrid plasticity rules (labile fast updates + slow consolidation) in continual-learning benchmarks (small curricula such as permuted/rotated MNIST or synthetic signal tasks). Measure accuracy vs energy (update counts × per-update cost) and compare to baselines.

Project C: Task Kernel & Transfer Bounds

Analytic: Define a task kernel combining input-distribution alignment and label-geometry; derive upper bounds for transfer improvement under linear-feature-transfer models and for networks with low-rank connectors.
Computational: Empirically estimate task kernels via representational-similarity metrics (SVCCA/CKA) and validate bound tightness on curated task pairs (synthetic tasks + a small Taskonomy subset). Propose simple regularizers (e.g., architecture-level low-rank adapters) guided by the bounds and test worst-case transfer improvements [6].

Deliverables & Success Criteria

- **Analytic deliverables:** closed-form BBP-derived thresholds for simple models (Project A); near-closed-form optimal consolidation policies for simplified learning (Project B); provable transfer bound statements for constrained architectures (Project C).

- **Computational deliverables:** reproducible notebooks and scripts reproducing (i) ID vs rank/noise heatmaps, (ii) energy vs accuracy curves comparing hybrid vs baseline plasticity, (iii) calibration plots of transfer bounds vs empirical transfer. All code will be public with README and small demo data.

- **Paper targets:** one consolidated methods+experiments manuscript (A+B) and one focused submission on transfer bounds (C), or a combined conference submission depending on results.

Reproducibility & Resources

All derivations, notebooks, and plotting scripts will be published in a public repository (`radboud-application`) with small, runnable demo notebooks and instructions for scaling to HPC (Slurm + CUDA). I have experience running large-scale experiments on multi-GPU HPC and producing reproducible, well-documented code.

Fit and Preparation

My academic and research trajectory has been shaped by a consistent focus on information-theoretic approaches to learning, compression, and interpretability. Trained as a physicist, I approach learning systems as processes of information flow and constraint, where structure and efficiency emerge from the way information is represented and transmitted. My recent work, *InfoGate: Information Theoretic Gating for Continual Learning*, introduced entropy and confidence based gating mechanisms for selective memory updates in transformers. That framework provided both a theoretical grounding in information compression through the Information Bottleneck and Minimum Description Length principles and practical expertise in large-scale neural training, GPU optimization, and benchmarking. In parallel, I have engaged with rule-based and symbolic frameworks such as Classy and S-Classy, exploring how information-theoretic objectives can guide the emergence of interpretable structures. These experiences have prepared me to bridge theory and implementation, designing systems that are both mathematically principled and computationally executable.

The proposed PhD directly continues this line of inquiry. Stockholm University's Department of Computer and Systems Sciences (DSV) is a natural fit: since its interest in developing interpretable multimodal systems capable of reasoning over hetrogenrous data, the precise intellectual context in which this work belongs. The department's emphasis on reproducibility, hybrid modeling, and information-centric design aligns closely with my own perspective, that interpretability is not a post-processing step, but a natural consequence of how information is organized in learning systems. This alignment means I can contribute productively from the outset, bringing a mature understanding of information-theoretic learning, interpretable modeling, and multimodal architectures, while continuing to grow in the theoretical and human-centered aspects of the discipline. My long-term aim is to help formalize a unified information-theoretic framework for transparency in AI, a goal I see as strongly resonant with DSV's vision of explainable, trustworthy, and efficient learning systems.

References

[1] David Attwell and Simon B. Laughlin. "An energy budget for signaling in the grey matter of the brain". In: *Journal of Cerebral Blood Flow and Metabolism* 21.10 (2001). Official journal of the International Society of Cerebral Blood Flow and Metabolism, pp. 1133–1145. DOI: 10.1097/00004647-200110000-00001. URL: https://doi.org/10.1097/00004647-200110000-00001.

[2] Sagar Prakash Barad, Sajag Kumar, and Subhankar Mishra. *Estimation of Electronic Band Gap Energy From Material Properties Using Machine Learning*. 2024. arXiv: 2403.05119 [cond-mat.mtrl-sci]. URL: https://arxiv.org/abs/2403.05119.

[3] Alex Bloemendal and Bálint Virág. "Limits of spiked random matrices I". In: *Probability Theory and Related Fields* 156.3–4 (Sept. 2012), pp. 795–825. ISSN: 1432-2064. DOI: 10.1007/s00440-012-0443-2. URL: http://dx.doi.org/10.1007/s00440-012-0443-2.

[4] John P Cunningham and Byron M Yu. "Dimensionality reduction for large-scale neural recordings". In: *Nature Neuroscience* 17.11 (Nov. 2014), pp. 1500–1509. ISSN: 1546-1726. DOI: 10.1038/nn.3776. URL: https://doi.org/10.1038/nn.3776.

[5] Rucha Bhalchandra Joshi et al. *Graph Neural Networks at a Fraction*. 2025. arXiv: 2502.06136 [cs.LG]. URL: https://arxiv.org/abs/2502.06136.

[6] Amir Zamir et al. *Taskonomy: Disentangling Task Transfer Learning*. 2018. arXiv: 1804.08328 [cs.CV]. URL: https://arxiv.org/abs/1804.08328.