Caltech
Student-Faculty Programs

Session L Abstracts

Sub-grid black holes: Learning the multi-scale dynamics of accretion feedback
Yash Pincha
Mentors: Anima Anandkumar, Chuwei Wang, Elias R. Most, and Haiyang Wang

Supermassive black holes co-evolve with the galaxies that host them through a two-way feeding-and-feedback loop, but the physics involved spans nine orders of spatiotemporal magnitude. Near the event horizon, turbulent magnetized gas accretes and launches relativistic jets and winds; far out, at galactic and intergalactic distances, those same outflows deposit energy that regulates how gas cools and stars form. Simulating the whole system at once with a single high-fidelity magnetohydrodynamic model is computationally infeasible, since resolving the black hole forces an extremely small timestep everywhere. Existing workarounds instead inject a fixed, hand-tuned feedback recipe near the black hole, but such static recipes are not dynamically consistent with the surrounding flow and cannot reproduce the flickering, time-variable behavior of real feedback. We develop a machine-learning alternative: two neural operators that decouple time evolution from spatial structure. One operator evolves the flow state on a boundary sphere forward in time, trained on short simulated trajectories yet able to run stably for tens of thousands of steps beyond that training window. A second operator, the focus of this work, reconstructs the full three-dimensional interior flow from a single boundary snapshot, conditioned on global accretion diagnostics. Alternating between the two forms a "cyclic zoom" that couples small and large scales without the mismatch of fixed recipes.

Parameter-conditioned neural operators for size-generalizable dynamics in driven Hubbard–Holstein systems
Miles M. Waugh
Mentors: Anima Anandkumar and Chuwei Wang

Simulating the dynamics of strongly correlated quantum materials is computationally challenging due to the exponential growth of the Hilbert space with system size, confining exact simulations to lattices far smaller than those studied experimentally. Fourier neural operators (FNOs), originally developed for solving partial differential equations, naturally accommodate inputs of varying system sizes while capturing nonlocal correlations. Here, we develop a parameter-conditioned Fourier neural operator, trained only on small lattices generated by a hybrid solver that treats electrons exactly and phonons variationally, to predict the time evolution of physical observables in driven Hubbard–Holstein systems at sizes beyond the reach of conventional numerical methods. This approach aims to provide a framework for finite-size extrapolation in strongly correlated quantum materials that generalizes across Hamiltonian and pulse parameters.

Elucidating the design space of consistency flow maps for time-dependent PDEs
Lufang A. Chiang
Mentors: Anima Anandkumar and Jiachen Yao

Learning the evolution operator of time-dependent partial differential equations (PDEs) offers a direct route to fast and accurate surrogates for scientific simulation. Instead of composing many small solver or rollout steps, one can learn a transition kernel that advances the solution over an arbitrary time span in a single step. Recent methods realize this kernel as a consistency flow map along physical time, and train it by enforcing trajectory consistency equations. However, it remains unclear how these learning equations affect the model accuracy and long-term stability, and fair cross-method comparison is still limited. In this paper, we elucidate the design space of consistency flow maps, spanned by the form of the consistency equation and the source of the supervising velocity, and run a systematic study across PDE families. From this space we propose IconFlow, an integral consistency flow map trained with a decoupled, derivative-free multi-point objective. On representative PDE benchmarks, IconFlow leads the Pareto front in few-step prediction and stays robust under temporal extrapolation, providing a more efficient and reliable surrogate for time-dependent PDEs.

PosteriorBench: From point estimates to posterior matching in evaluating generative inverse solvers
Zi-Siang Hsu
Mentors: Anima Anandkumar and Jiachen Yao

Generative models are increasingly used to solve scientific inverse problems, but existing evaluations still focus primarily on whether a method can produce a single plausible reconstruction. This is insufficient for ill-posed problems, where multiple solutions may be consistent with the same sparse or noisy observations. In these settings, a method can achieve strong pointwise accuracy while still failing to capture the true posterior through mode collapse, overconfident uncertainty, or averaging incompatible solutions.

We introduce PosteriorBench, a benchmark for evaluating the distributional accuracy of generative inverse solvers. PosteriorBench evaluates four physics-based inverse problems: Darcy flow inversion, Poisson source recovery, carbon capture and storage, and light transport material inference. For each task, we construct high-fidelity reference posteriors using computationally heavy but established procedures such as rejection sampling and Markov chain Monte Carlo, enabling direct assessment of whether solvers recover the full set of solutions rather than the single best sample.

The benchmark spans sparse sensing, low-resolution observations, nonlinear forward models, varying noise levels, and multimodal priors, with a unified pipeline for distribution matching, uncertainty quantification, and hyperparameter calibration. Our experiments reveal substantial distribution-matching gaps across current solvers, while showing that neural operators improve resolution robustness and that guidance weights and generation noise are key to posterior-variance calibration.

Neural operators and stochastic methods for PDE-based navigation
Michael H. Chen
Mentors: Anima Anandkumar, Chuwei Wang, Hong Chul Nam, and Xi Deng

PDE-based navigation is a technique that produces global fields which represent environment geometry. The gradients of these global fields define local motion towards a target across complex domains, such as robot control with many degrees of freedom. However, due to the curse of dimensionality, conventional numerical methods turn prohibitively expensive to solve higher-dimensional configuration spaces. Accordingly, neural operators provide an elegant, scalable way to learn mappings from boundary conditions and domain geometry to these fields. In particular, we combine stochastic representations of elliptic PDE solutions with neural operators to learn continuous navigation fields. Deterministic numerical solver fields are used as references, whilst GPU-accelerated stochastic computation generates training data from maps diverse in geometry and dimension. Experiments show that local structure, boundary treatment, learned correction, and local control improve field quality and navigation success across varied environments. By amortizing stochastic PDE estimation across a family of environments, learned operators produce navigation fields substantially faster than conventional full-domain numerical methods.

Simulation-trained neural reconstruction of ultrasound data for detecting brain micrometastases
Tanish Rao
Mentors: Anima Anandkumar and Ryan Han

Metastatic tumors in the brain are usually detected only after they have grown well beyond the size at which treatment is most effective. Ultrasound imaging with gas vesicles, gas-filled protein nanostructures that scatter sound and can be expressed by tumor cells, offers a route to earlier detection, but skull-induced aberration degrades conventional beamforming of small, sparse targets. We are developing a learned reconstruction that maps raw radiofrequency data from cross-amplitude modulation acquisitions directly to gas vesicle concentration and location, bypassing beamforming entirely. Because training data of this kind cannot be acquired experimentally at scale, we generate it by full-wave acoustic simulation: coronal mouse skull slices from micro-computed tomography are fitted with a polymethylpentene acoustic window matched to the experimental setup, converted to speed-of-sound maps with a delineated brain cavity, and seeded with gas vesicle inclusions from which the returning radiofrequency signal is computed. A Fourier neural operator trained on these simulations recovers the position and concentration of inclusions across the full skull, including cluster configurations not represented in training. Ongoing work extends the simulation to heterogeneous cluster morphologies and validates the model against experimentally acquired radiofrequency data.

SU(2)-equivariant neural networks for magnetic molecules
Erh-Wei Sheng
Mentors: Anima Anandkumar, Beom Seok Kang, Danish Khan, and Yuchao Lin

Equivariant graph neural networks for electronic structure are built from integer-rank tensors, the representations of SO(3). In magnetic molecules, spin-orbit coupling (SOC) ties integer orbital angular momentum to spin 1/2, so the relevant states carry half-integer angular momentum and transform in representations of SU(2), the double cover of SO(3) –– spinors, paired by time reversal into Kramers doublets.

Existing methods regress the Hamiltonian in a fixed orbital basis, append spin as an auxiliary scalar, or impose time reversal on otherwise spinless features. All contract the half-integer degrees of freedom to integer rank, so none propagates a spinor through the network. The obstruction is gauge freedom: a spinor is fixed only up to a phase, and within a Kramers doublet only up to a U(2) rotation, with no smooth global choice.

Rather than fix a gauge, we classify the operations compatible with both SU(2) covariance and Kramers gauge freedom, and obtain message passing in which spinors are transported between sites along a connection built from local eigenbases, with every readout gauge invariant by construction. On a NequIP backbone we compare full spinor propagation (FullSU2) against an ablation that recontracts to invariants at each layer (LocalSU2). For mononuclear Co(II) zero-field-splitting tensors and Dy(III) crystal-field parameters, FullSU2 and LocalSU2 both outperform the SO(3) baseline. Extension to polynuclear systems is limited by data, not formalism.

A survey on relational database course pedagogy
Robert R. Walker
Mentor: Ethan Ordentlich

Unlike some Computer Science courses that may come in one or two forms, relational database courses have many freedoms that every institute can take in their own version. Caltech's version of such a course is CS 121. Since we are only one of the many schools that teach relational databases, what is the "right" way to implement a relational database course? To do this, we inspected industry leaders in Computer Science such as CMU, Berkeley, UChicago, etc. With a deeper look into these universities and why their courses are designed the way they are, we hope to better understand the pedagogically "correct" way to implement a relational database class and, if necessary, apply these changes to our course to improve our student's academic and professional careers.

Capturing rare events: A theory-motivated rethinking for attention kernels in transformers
Kailen A. Hargenrader
Mentors: Andrew M. Stuart and Bohan Chen

The transformer architecture underlies frontier models (ChatGPT, Gemini, Claude, etc.). These models have achieved incredible feats in practical tasks such as language modeling. Our work studies a task where transformers can be less successful. The defining feature of the transformer architecture is the attention map. We study the attention map in the distribution mapping setting, where samples from an input distribution are transported through the attention map to become samples from an output distribution. Previous work has shown that transformers are universal operator approximations. Given an operator and an error \$\epsilon\$, there exists a fixed size transformer that matches the operator to \$\epsilon\$ precision. This approximation considers the attention map as a mapping between distributions, and is agnostic to samples. When considering the attention map as a mapping between samples, the universal operator approximation collapses to a comparable guarantee only for distributions that decay faster than a gaussian. Our work focuses on distributions with heavy tails, where this guarantee does not apply. Heavy-tailed distributions frequently appear in applications such as economics, warranting the study of attention maps between distributions with heavy tails. We devise simple operators on heavy-tailed distributions for which the traditional softmax attention map struggles. Additionally, alternative attention kernels have been proposed in recent literature primarily with the goal of improving computational efficiency. In particular, linear attention variants factor the attention kernel to reduce $O(n^2)$ to $O(n\log n)$ or $O(n)$ runtime. We investigate whether these kernels can outperform softmax in modeling heavy tails.

Trimodal contrastive learning for joint conditional probability estimation in fluid dynamics
Huy N. Nguyen
Mentors: Andrew M. Stuart, Claire Valva, and Hojjat Kaveh

Standard multimodal contrastive learning frameworks, such as Contrastive Language–Image Pre-training (CLIP), rely on aligning pairwise embeddings across modalities. Applications include matching text-image pairs and Lagrangian data assimilation. When extended to three or more modalities, these methods involve optimizing a summation of pairwise alignment losses. However, pairwise objectives only model one-to-one conditional distributions p(u | v) but fail to capture joint conditional distributions p(u | v, z). This gap is especially significant in physical systems where individual modalities can carry marginal predictive power but provide strong information jointly.

To resolve this, we present a trimodal contrastive framework based on a novel trimodal similarity metric that directly quantifies the joint alignment among three latent representations. We then formulate a corresponding trimodal contrastive loss function and validate it in Gaussian test settings, where learned representations can be benchmarked directly against exact analytical conditional distributions. Finally, we apply our framework to fluid dynamics by recovering Eulerian velocity fields given Lagrangian flow trajectories and passive tracer concentrations data. Our final goal is to accurately reconstruct the original streamfunction even with sparse or noisy data, with potential applications in data assimilation for fluid dynamics, extending to active tracers like temperature, salinity and other measurements.