Hey all,

Here is your disruptive ML/AI digest for the day. Here's what you folks might find interesting today.

TOP PICKS

Novel Architectures and Objectives

A Blueprint for Equilibrium-Based Differentiable Continuous-Variable Thermodynamic Computing (Lockwood et al.)

This paper proposes a genuinely novel computing paradigm: using stochastic analog superconducting circuits governed by Langevin dynamics to natively sample from energy-based models, bypassing digital simulation entirely. The key conceptual move is treating thermodynamic hardware not as an accelerator for existing algorithms but as a substrate whose physical dynamics *are* the inference procedure. The authors construct probabilistic graphical models directly in hardware-native energy potentials and analyze runtime/energy tradeoffs theoretically. Guillaume Verdon is a coauthor, known for his work bridging quantum thermodynamics and ML. This is a departure from the mainstream because it reframes sampling as a physical process rather than an algorithmic one, with potential implications for probabilistic ML at scale.

Theory and Generalization

Relevant and Irrelevant: A Renormalization Group Analysis of Transformer Attention (Haggi-Mani & Rish)

This work applies Wilsonian renormalization group theory to treat attention as a perturbation around a trained MLP fixed point, asking whether attention is a relevant, marginal, or irrelevant operator. The answer depends on the spectral structure of the data-generating process: for long-correlation Markov chains, attention is a relevant operator that drives a phase transition in representation space; for short correlations, it is irrelevant and the Transformer converges to MLP behavior. The first-layer head (L0H0) dominates the representational shift by 4x over any subsequent head, and perturbation decay reveals spectral selectivity that reverses between regimes. This is a principled physics-derived framework that makes testable predictions about *when* attention matters, rather than treating it as universally beneficial. Irina Rish is a coauthor.

Emergent Behavior and Agentic AI

Verbalizable Representations Form a Global Workspace in Language Models (Gurnee et al.)

Using a new interpretability technique called the Jacobian lens, the authors identify a "J-space" -- a small set of intermediate-layer representations that a model is poised to verbalize, exhibiting functional properties of a global workspace: they can be reported, deliberately held, used for silent reasoning steps, and passed as arguments to downstream computations. The J-space carries coherent content only in an intermediate band of layers, holds on the order of tens of concepts at a time, and is broadcast more widely than other representations. In alignment audits, it reveals strategic deliberation and evaluation awareness that never surface in outputs. The authors introduce counterfactual reflection training that improves behavior by training only on what a model would say if interrupted. This is a mechanistic account of something resembling conscious access in LLMs, grounded in global workspace theory rather than metaphor. The author list includes researchers from Anthropic.

Quantum ML and Hybrid Paradigms

Rethinking Quantum Continual Learning with Quantum Fisher Information (Hsu et al.)

This paper replaces classical Fisher information in elastic weight consolidation with quantum Fisher information (QFI) for continual learning in variational quantum circuits. The distinction is substantive: CFI measures parameter importance through measurement-dependent output statistics, while QFI quantifies intrinsic sensitivity of the parameterized quantum state manifold itself. The authors show that these impose different regularization geometries -- CFI acts selectively on measurement-sensitive directions, QFI imposes a denser state-geometric constraint. Under depolarizing noise, CFI values are suppressed by degraded measurement statistics while QFI preserves a more stable sensitivity structure. This is an information-geometric approach to a quantum ML problem that is physically motivated rather than borrowed wholesale from classical ML.

Probabilistic and Statistical Methods

Neural Non-Equilibrium Hamiltonian Monte Carlo for Corrected Boltzmann Sampling (Qian)

This paper introduces a train-then-correct learned Hamiltonian sampler where a neural network learns stochastic Hamiltonian-style paths from a base distribution toward a Boltzmann target, and then the proposals are statistically corrected using recorded non-equilibrium work. The correction is principled: the generalized work defines path-space KL divergence during training, importance weights during evaluation, and Metropolis-Hastings acceptance ratios at inference. The key insight is that the learned proposal generates complete paths with enough path-probability information for exact correction, rather than requiring the proposal to be reversible. On double-well and lattice phi^4 targets, the method gives corrected estimates when path overlap is sufficient, and honestly reports failure modes (weight degeneracy, low acceptance) when overlap is poor. This bridges learned samplers with rigorous statistical mechanics in a way that does not sacrifice correctness guarantees.

BY CATEGORY

Theory and Generalization

- A zero-one law for one-shot system identification
- Retraining Seeks Stable Signals
- A Statistical Formulation Gap for Nonlinear Multiscale Physics-Informed Learning
- Understanding Reasoning from Pretraining to Post-Training

Probabilistic and Statistical Methods

- Diffusion models recover accurate mixture weights despite score function insensitivity
- Cluster-Aware Matching via Laplacian Optimal Transport
- Prediction-Only Distillation in Linear and Logistic Regression
- Aggregation of Statistical Evidence under Exchangeability

Novel Architectures and Objectives

- From Diffusion to Reaction-Diffusion: A Dynamical-Systems View of Oversmoothing in Hypergraph Neural Networks
- (MPO)^2: Multivariate Polynomial Optimization based on Matrix Product Operators
- Fast and Scalable Caputo Fractional Gradient Descent via Perturbation-Preserving Memory Compression

Emergent Behavior and Agentic AI

- Induction in Both Directions: A Mechanistic Analysis of In-Context Learning in Masked Diffusion Language Models

Transferable Methods from Adjacent Fields

- Learning Standard Model structure from LHC data with Riemannian flow matching
- Neural spectroscopy of AlphaFold2 reveals encoded protein conformational landscapes

---

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 115 papers stat.ML: 14 papers