NEURAL FIELD TRANSFORMATIONS FOR HYBRID MONTE CARLO: ARCHITECTURAL DESIGN AND SCALING

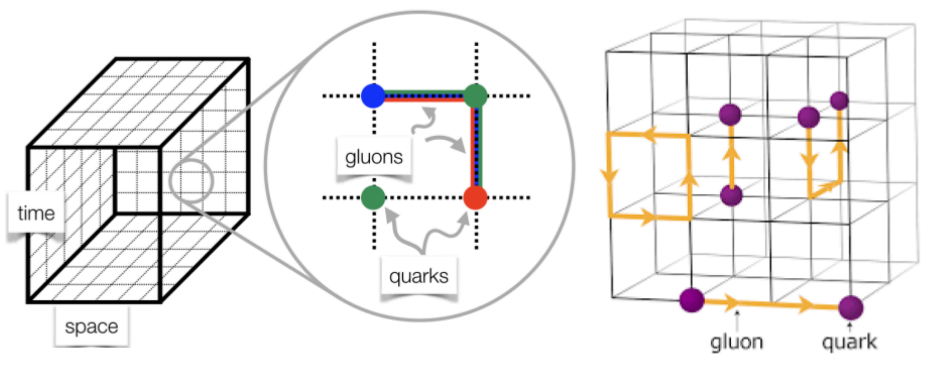
NeurIPS Machine Learning and the Physical Sciences Workshop 2025

Jinchen He[1,2], Xiao-Yong Jin[3], James C. Osborn[3] and Yong Zhao[2]

[1]Maryland Center for Fundamental Physics, University of Maryland, College Park, MD 20742, USA
[2]Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA
[3]Computational Science Division, Argonne National Laboratory, Lemont, IL 60439, USA

LATTICE GAUGE THEORY

Discretization: Lattice gauge theory (LGT) discretize the space-time into a finite lattice in Euclidean space.



Monte Carlo Sampling: Replace the path integral with Monte Carlo integration (Wick rotation to Euclidean space).

$$Z = \int \mathscr{D}A\, \mathscr{D}\psi\, \mathscr{D}\bar{\psi}\, e^{iS_{\mathrm{QCD}}[A,\psi,\bar{\psi}]} \to Z_{\mathrm{E}} = \int \mathscr{D}U e^{-S_{\mathrm{E}}^{g}[U]} \int \mathscr{D}\psi \mathscr{D}\bar{\psi} e^{-\bar{\psi}M[U]\psi} = \int \mathscr{D}U\, e^{-S_{\mathrm{E}}^{g}[U]} \det M[U]$$

Then the expectation value becomes average on configurations.

$$\langle \hat{O} \rangle = \frac{1}{Z_{\mathrm{E}}} \int \mathscr{D}U\, \mathscr{D}\psi\, \mathscr{D}\bar{\psi}\, \hat{O}\, e^{-S_{\mathrm{E}}^{\mathrm{QCD}}[A,\psi,\bar{\psi}]} = \frac{1}{N} \sum_{i=1}^{N} O[U^{(i)}]$$

Critical Slowing Down [1]: Low autocorrelation is required to ensure effectively independent samples. However, as we approach the continuum limit (larger volumes and smaller lattice spacing), autocorrelations in the Markov chain increase rapidly.

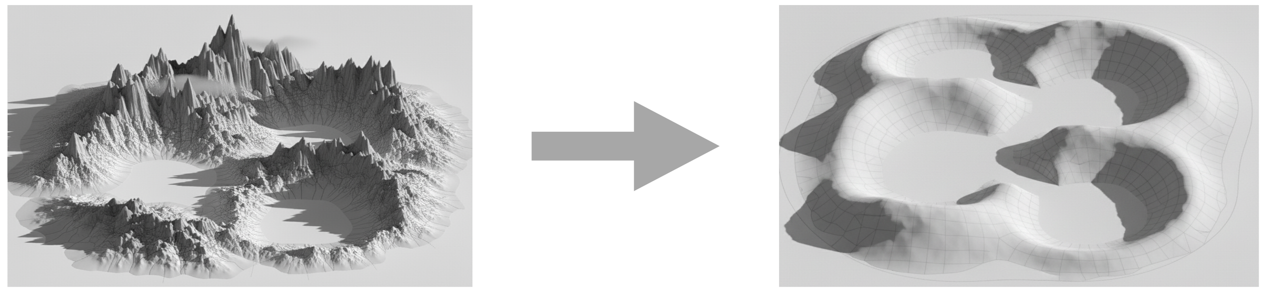
NEURAL FIELD TRANSFORMATION

Topology Freezing: One reason of critical slowing down is that Markov chain will be trapped in a topological sector when approaching the continuum limit.

Trivializing Map [2]: We can reduce autocorrelation and boost topological tunneling via field transformation.

$$S_{\mathrm{FT}}(\tilde{U}) = S\big(\mathscr{F}(\tilde{U})\big) - \log\left|\det \mathscr{F}*(\tilde{U})\right|$$

Then Monte Carlo sampling is performed in the space of auxiliary field $\tilde{U}$ with new action.



Neural-network Transformed HMC (NTHMC) [3]: Motivated by the concept of trivializing map, we can construct invertible gauge-covariant field transformations using neural network.

- ✦ Our objective function is to minimize the gradient of action;

$$\mathscr{L}(\tilde{U}) = \sum_{p\in\{2,4,6,8\}} \frac{c_p}{V^{1/p}} \left\| \frac{\partial S_{\mathrm{FT}}(\tilde{U})}{\partial \tilde{U}} \right\|_p$$

- ✦ We employ convolutional neural network (CNN) to construct an invertible, gauge-covariant and local field transformation, which can be naturally extended to larger lattices.

ARCHITECTURAL DESIGN

Hybrid Monte Carlo (HMC)[4]: Generates lattice gauge configurations by introducing conjugate momenta, evolving the fields along molecular-dynamics trajectories in an extended phase space. Inside the HMC update we employ the 2nd order minimum norm symplectic integrator [5] with 10 MD steps per trajectory, tuning the step size to achieve~80% acceptance.

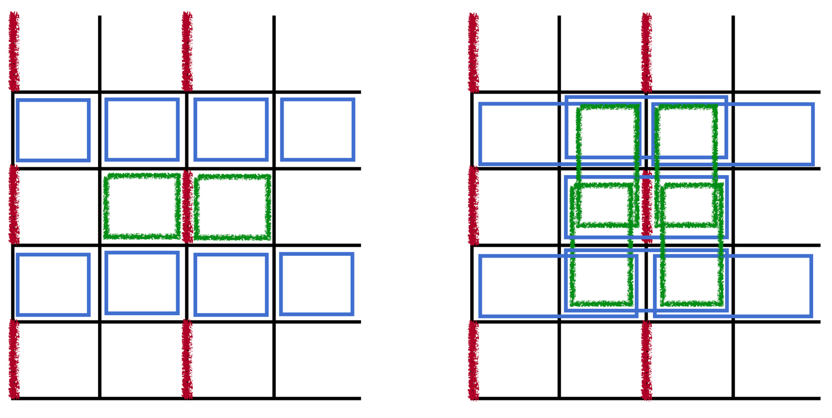
NTHMC: To maintain gauge covariance, the link variables on configurations are updated using the form

$$\mathscr{F} : \tilde{U}_{x,\mu} \mapsto U_{x,\mu} = e^{\Pi_{x,\mu}} \tilde{U}_{x,\mu}, \quad \Pi_{x,\mu} = \sum_{l} \epsilon_{x,\mu,l} W_{x,\mu,l}\ ,$$

where $W_{x,\mu,l}$ are Lie algebra projection of gauge-invariant Wilson loops and coefficients $\epsilon_{x,\mu,l} = \phi[N_l(X, Y, \ldots)]$ are parameterized using CNN. To get a tractable Jacobian $\mathscr{F}*$, the lattice links are partitioned into eight disjoint subsets and updated sequentially. In U(1) gauge theory, an invertible field transformation requires

$$\sum_{p} \left|\epsilon_{x,\mu,p}\right| + \sum_{r} \left|\epsilon_{x,\mu,r}\right| < 1\ .$$

In the figure below, red lines mark one of the eight link subsets updated sequentially, green lines indicate Wilson loops $W_{x,\mu,l}$ and blue lines show the gauge-invariant features (X,Y,...) used as CNN inputs.



Base: Two-layer CNN that processes six input channels (two plaquette, four rectangle). A first 3×3 convolution maps them to 12 hidden channels with GELU activation, followed by a second 3×3 convolution producing coefficients, which are scaled by *arctan* to $|\epsilon_{x,\mu,l}| < 1/6$;

Tanh: Replace *arctan* with *tanh*, set $|\epsilon_{x,\mu,p}| < 0.4$ and $|\epsilon_{x,\mu,r}| < 0.05$;

Resn: Add two residual blocks (two 3x3 convolutions with skip connections);

Attn: Add channel-attention module between the two convolutions in **Base**;

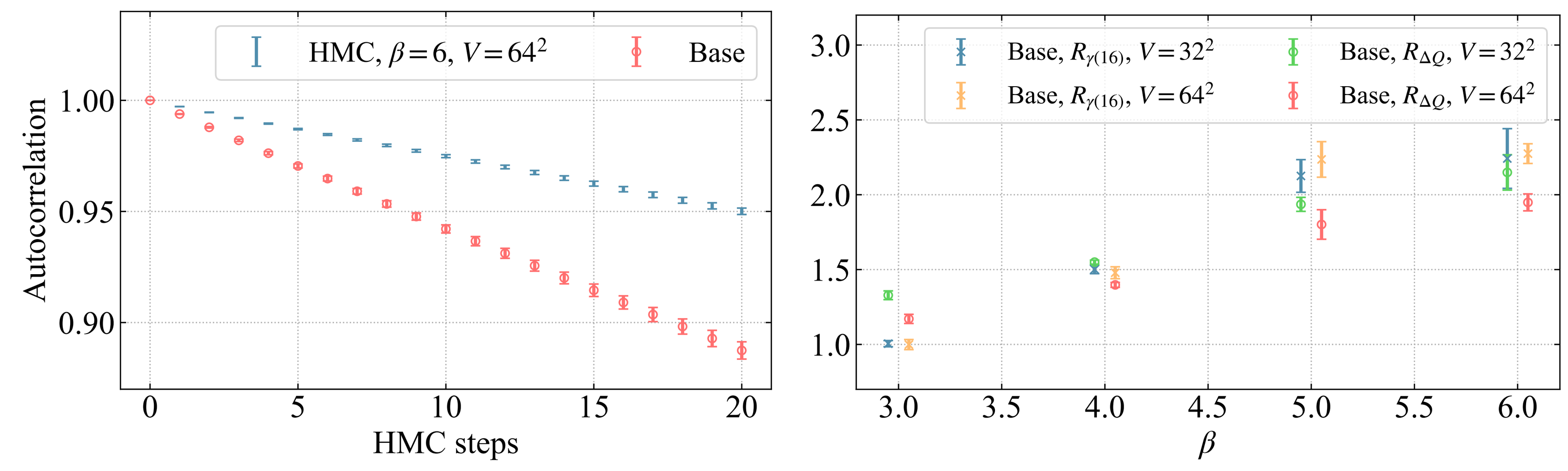
Comb: Combine architectures of **Tanh**, **Resn**, **Attn** above.

REFERENCES

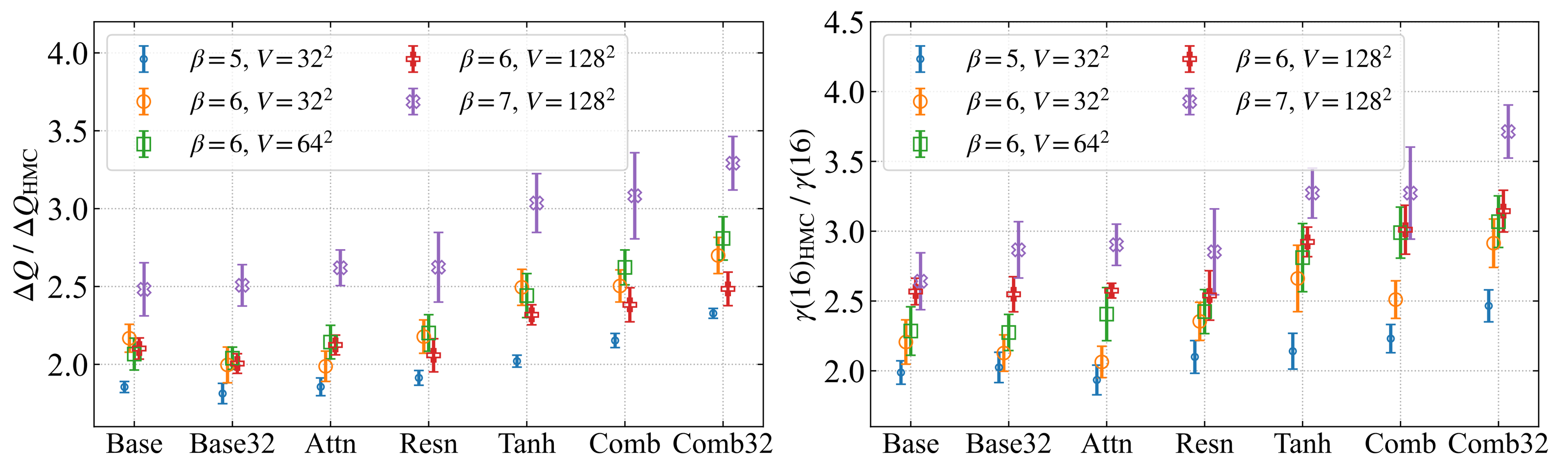
1. Ulli Wolff. Critical slowing down. Nucl. Phys. B-Proc. Sup, 17:93–102, 1989.
2. Martin Lüscher. Trivializing maps, the wilson flow and the hmc algorithm. Communications in mathematical physics, 293(3):899–919, 2010.
3. Xiao-yong Jin. Neural network field transformation and its application in hmc. arXiv preprint arXiv:2201.01862, 2022.
4. Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. Physics letters B, 195(2):216–222, 1987.
5. IP Omelyan, IM Mryglod, and R Folk. Symplectic analytically integrable decomposition algorithms: classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations. Computer Physics Communications, 151(3):272–314, 2003.

NUMERICAL RESULTS

Comparison of HMC and NTHMC with the Base model



Comparison of NTHMC with various network architectures



To assess the sampling efficiency of HMC and NTHMC, we consider two complementary metrics:

- ✦ The first is $\gamma(\delta)$, which measures the relative independence of the configuration

$$\gamma(\delta) = \frac{1}{1 - \Gamma_t(\delta)}\ , \quad \Gamma_t(\delta) \equiv \frac{\langle Q_\tau Q_{\tau+\delta} \rangle}{\langle Q^2 \rangle} = 1 - \frac{\langle (Q_\tau - Q_{\tau+\delta})^2 \rangle}{2V\chi_t^{\infty}(\beta)}$$

where $\Gamma_t(\delta)$ is the autocorrelation of the topological charge and χ_t^{∞} is the topological susceptibility;

- ✦ The second metric is the average topological change ΔQ, the mean absolute shift in topological charge per HMC step, which probes the tunneling between sectors; larger values of ΔQ indicate more frequent topological transitions and more efficient sampling.

CONCLUSIONS

- ✦ Residual blocks and channel-dependent activation (**Tanh**) lead to improvements over the **Base** model, and the combined design (Comb) yields the strongest effect, achieving nearly a 40% improvement;
- ✦ As the coupling β increases, all NTHMC models exhibit a higher sampling efficiency compared to standard HMC, and the efficiency does not show statistically significant changes with increasing lattice volume, indicating that our method can be effectively extended to larger volumes and finer lattices.
- ✦ Interestingly, wider receptive field does not hinder transferability. A possible explanation is that, as the continuum limit is approached, infrared (long distance) structures are preserved while ultraviolet (short distance) details become refined, so the transferable features are precisely the long-range correlations that benefit from a relatively wider receptive field.