AI/ML-inspired RMHMC kernels—CPN model study

Nobuyuki Matsumoto (Boston U, RBRC)

In collab w/ Ryan Abbott, Peter Boyle, Luchang Jin, Christopher Kelly, Christoph Lehner++, in prep.

Lattice 2026

(Scidac5, USQCD)

General background

- Critical slowing down has been the major hindrance in Markov chain Monte Carlo methods

  - Fourier acceleration

    Parisi 1984, Batrouni Katz Kronfeld Lepage Svetitsky Wilson 1985

  - Trivializing map

    Lüscher2009

  Efforts to replace the MCMC w/ normalizing flow **See Dan's Plenary**

The nature of a fictitious thermal process leaves a big room of ideas for improvement

- Nontrivial historical developments have led to the state-of-the-art QCD algorithm:

  - HMC **Duane Kennedy Pendleton Roweth 1987**
  - Omelyan, force-gradient **Omelyan Mryglod Folk 2003; Takaishi de Forcrand 2006**
  - Pseudofermion **Weingarten Petcher 1981**
  - Sexton-Weingarten **Sexton Weingarten 1992**
  - Hasenbusch **Hasenbusch 2001; Hasenbusch Jansen 2003**
  - RB precond, RHMC, EOFA, and even more related to the solvers **See, e.g., Peter's talk**

    DeGrand Rossi 1990, Clark Kennedy 2004, 2007, Chen Chiu 2014

  Uncertain whether an artificial intelligence can outpower (at least in short-term)
  and flip over the accumulation of investigation
  that took the forerunners of the field decades in research

- Nonetheless, AI/ML should be capable of accelerating *future evolution* of the algorithm

  Philosophy of this work:

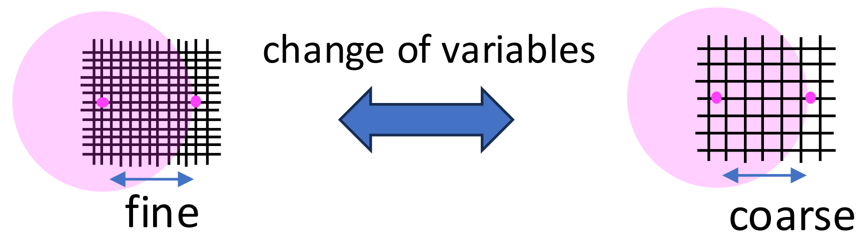
  Stick to the historical HMC algorithm, follow the good ideas advocated so far,
  and improve them with the aid of AI/ML

- Gauge symmetry restricts us to focus on Wilson loops and their associated staples
- Anticipate that AI/ML can find a better linear combination of them by thorough linear regression search

Perfect action: Holland Ipp Müller Wenger 2025

Field-transform HMC (FTHMC)

$$\int (dU)\, e^{-\int d^4x\, \mathcal{L}_{\mathrm{lat}}} = \int (dU')\, e^{-\int d^4x\, \mathcal{L}'_{\mathrm{lat}}}$$



- Successful with simple smearing kernels

→Shuhei's talk

- Need to evaluate the Jacobian *matrix* of the transformation

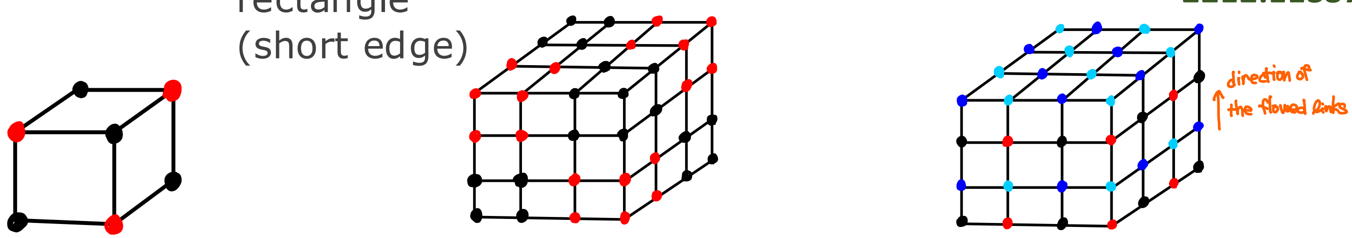
Masking schemes

plaquette

rectangle (short edge)

(long edge)

NM, LATTICE 2022 2212.11387



Full evaluation of the 2nd derivative of the Wilson loops in the adjoint space is a costly endeavor for complicated loops

Strategy

- Riemannian manifold HMC (RMHMC) **See Nguyen-Boyle-Christ--Jang-Jung LATTICE2021**

$$H_{\mathrm{HMC}} \equiv \frac{1}{2} p^T K(U) p + S(U) - \frac{1}{2} \log \det K(U)$$

- Design $K(U)$ to accelerate; a generalization of the Fourier acceleration

- Can construct the kernel by a linear transformation; good synergy with the general AI/ML developments

E.g., preconditioners:
Lehner Wittig 2023, 2023
Knüttel Lehner Wittig 2026

Reasonable application of the AI/ML framework
that can fully leverage the existing AI/ML softwares

GPT: **Lehner Bruno Richtmann Schlemmer <u>Lehner</u> Knüttel Wurm Jin Bürger Hackl Klein**

https://github.com/lehner/gpt/tree/2024-10

HP Cortex: Chris Kelly

https://github.com/giltirn/HPCortex

1. Iterative implicit solves — *persistent but improvable (e.g., Aitken acceleration; thanks to Claude)*
2. Evaluation of the (inverse of the) kernel
3. Force from the Jacobian determinant

To be here addressed

When combined with the multilevel integrator, the overhead is additive to the fermion CG

Nguyen-Boyle-Christ--Jang-Jung LATTICE2021

Algorithm 1 Multilevel implicit leapfrog

From a given (w,p), set $w_0 \equiv w$ and

$$p_0 = p - \frac{\tau}{2} F_f(w). \tag{17}$$

for $i = 0, \dots, n-1$ **do**

$$p_{i+1} = p_i - \frac{\tau}{2n}\left[F_K(w_i, p_{i+1}) + F_g(w_i) + F_d(w_i)\right], \tag{18}$$

$$w_{i+1} = w_i + \frac{\tau}{2n} K^{-1}(w_i)\, p_{i+1}. \tag{19}$$

end for

Set $w'_0 \equiv w_n (\equiv w^{1/2})$ and $p'_0 \equiv p_n (\equiv p^{1/2})$.

for $i = 0, \dots, n-1$ **do**

$$w'_{i+1} = w'_i + \frac{\tau}{2n} K^{-1}(w'_i)\, p'_{i+1}, \tag{20}$$

$$p'_{i+1} = p'_i - \frac{\tau}{2n}\left[F_K(w'_{i+1}, p'_{i+1}) + F_g(w'_{i+1}) + F_d(w'_{i+1})\right]. \tag{21}$$

end for

Set $w' \equiv w'_n$ and

$$p' = p'_n - \frac{\tau}{2} F_f(w'). \tag{22}$$

(w', p') is the updated phase space configuraion.

Coupling layers

Idea: Mitigate the kernel-related overhead (2&3) by utilizing the coupling layers

See, e.g., Albergo, Kanwar, Shanahan 2019
Albergo et al. 2021

$$\begin{cases} \pi'_r = \exp[f_r(U)] \odot \pi_r + \sum_{s \neq r} g_{rs}(U)\pi_s, \\ \pi'_s = \pi_s \quad (s \neq r). \end{cases}$$

$$A \equiv A^{(n)}_{r_1} \circ \cdots \circ A^{(n)}_{r_k} \circ \cdots A^{(1)}_{r_1} \circ \cdots \circ A^{(1)}_{r_k}.$$

$$\begin{aligned} H &\equiv \frac{1}{2}\pi_A^2 + S(U) - \log\det A(U) \\ &= \frac{1}{2}\pi^T(A^T A)\pi + S(U) - \log\det A(U). \end{aligned}$$

- No CG solves for the inverse
$$\begin{cases} \pi_r = \exp[-f_r(U)] \odot \left(\pi'_r - g_{rs}(U)\pi'_s\right), \\ \pi_s = \pi'_s \quad (s \neq r). \end{cases}$$
- Controlled/no determinant
$$\log\det A_r(U) = \mathrm{tr}\, f_r(U).$$
- Well-developed methodologies in ML, e.g., the force propagation

$$S_{CP^{N-1}} \equiv -N\beta \sum_{x,\mu} \left[z^{\dagger}_{x+\mu}\, e^{-i\phi_{x,\mu}}\, z_x + z^{\dagger}_{x}\, e^{i\phi_{x,\mu}}\, z_{x+\mu} \right] \qquad \left(\begin{array}{l} z \in \mathbb{C}^N,\ |z|^2 = 1 \\ \phi_{x,\mu} \in \mathbb{R} \end{array} \right)$$

- Asymptotic freedom
  Classically plausible degenerate vacua, but no SSB in 2D → nontrivial massive state "σ"

- To the leading order in large-N,
  the $U(1)$ Lagrange-multiplier field becomes dynamical Maxwell in the eff theory

- $z \in \mathbb{C}^N, z \in S^{2N-1}$
- $S^{2N-1} \simeq SU(N)/SU(N-1)$ $\rightarrow z = Uz_0$ $(U \in SU(N),\ z_0 \in S^{2N-1})$
  zero modes assoc. w $SU(N-1)$
  that leaves z_0 invariant

Good toy model towards QCD application

- Group variables on the *site*
- Scalar variables on the *link*
(not originating from fermions but Lagrange multipliers)

Smearing kernel

$$A_\theta^{eo}: \begin{pmatrix} p'_e \\ p'_o \end{pmatrix} = \begin{pmatrix} 1 & g_\theta(U) \\ & 1 \end{pmatrix} \begin{pmatrix} p_e \\ p_o \end{pmatrix}$$

$$p'_x \equiv p_x + \epsilon \sum_\mu \left(p_{x+\mu} + p_{x-\mu} \right) \simeq (1 + \underbrace{2\epsilon}_{\text{rescaling}}) \underbrace{e^{\epsilon \partial^2}}_{\text{smearing}} p_x$$

(Applied to both group and scalar variables)

- Generic: Trivial determinant; the most straightforward UV remover
- CPN: Gauge-invariant already thanks to $U(1)$

We don't need to *fully align* the speed of updates for all the collective modes in the system;
we only need to extract the important part/eliminate the UV noise
and use the resulting margin to effectively update the relevant modes

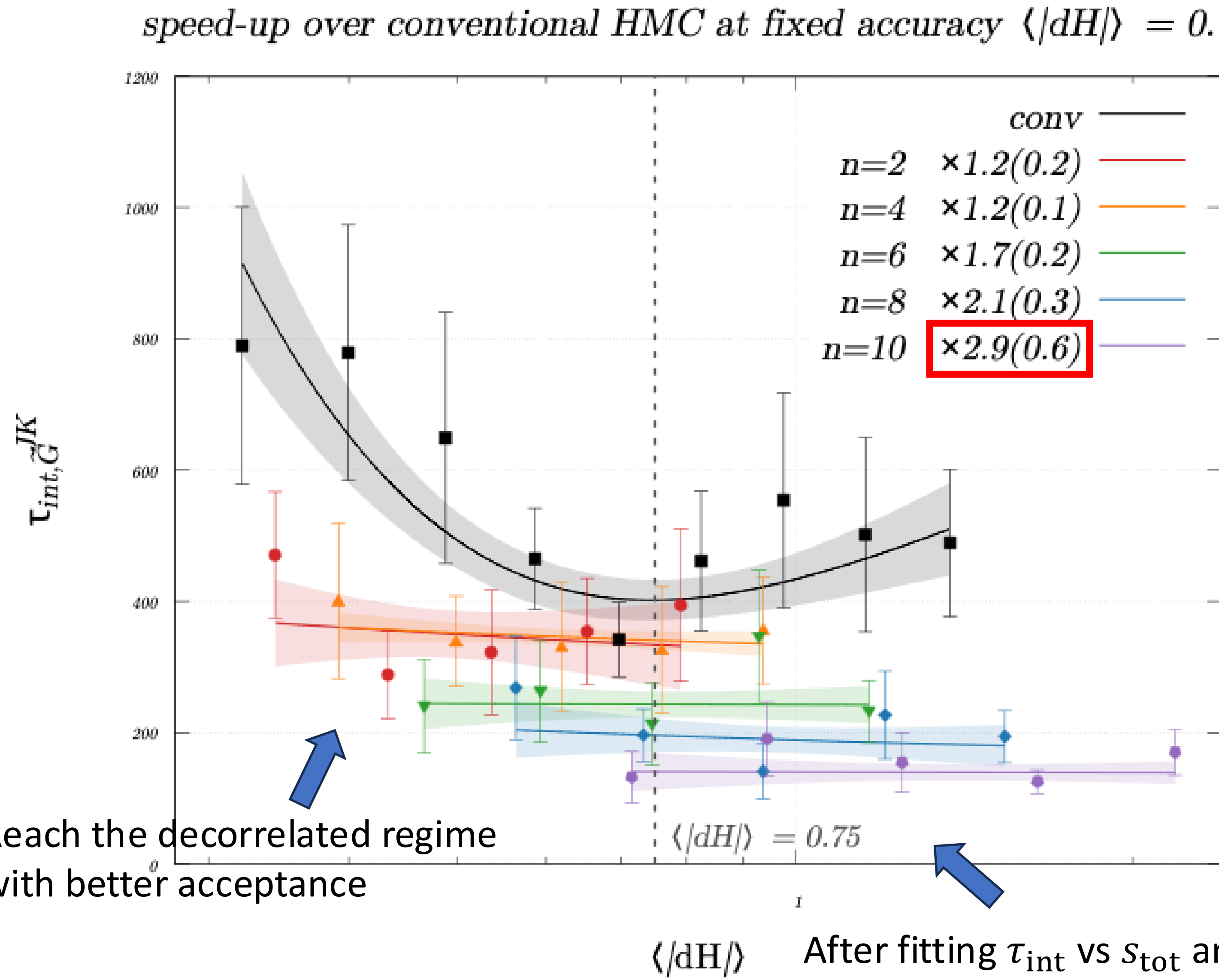
- $128^2,\ N = 3,\ \beta = 1.4$
- HMC trajectory length $s_{\mathrm{tot}} = 1.0 - 1.6$; $N_{\mathrm{step}} = 8$ (fixed); minimum norm
- Smearing kernel: increment $\eta = 0.025$, up to steps $n = 10$
- Sample size: 100k-500k
- Code written in Julia **A. Bryutkin, J. Glaubitz, Y. M. Marzouk, NM++**

  https://github.com/nobuyukimatsumoto68/2024_latticeCPN_dev/tree/main
  (work-in-progress fork of https://github.com/jglaubitz/2024_latticeCPN_dev)

Zero mode of the correlator

$$G(x-y) \equiv \left\langle \left(zz^{\dagger}\right)_x \left(zz^{\dagger}\right)_y \right\rangle_c, \quad \tilde{G}(k) \equiv \int d^2x \; e^{ikx} G(x)$$

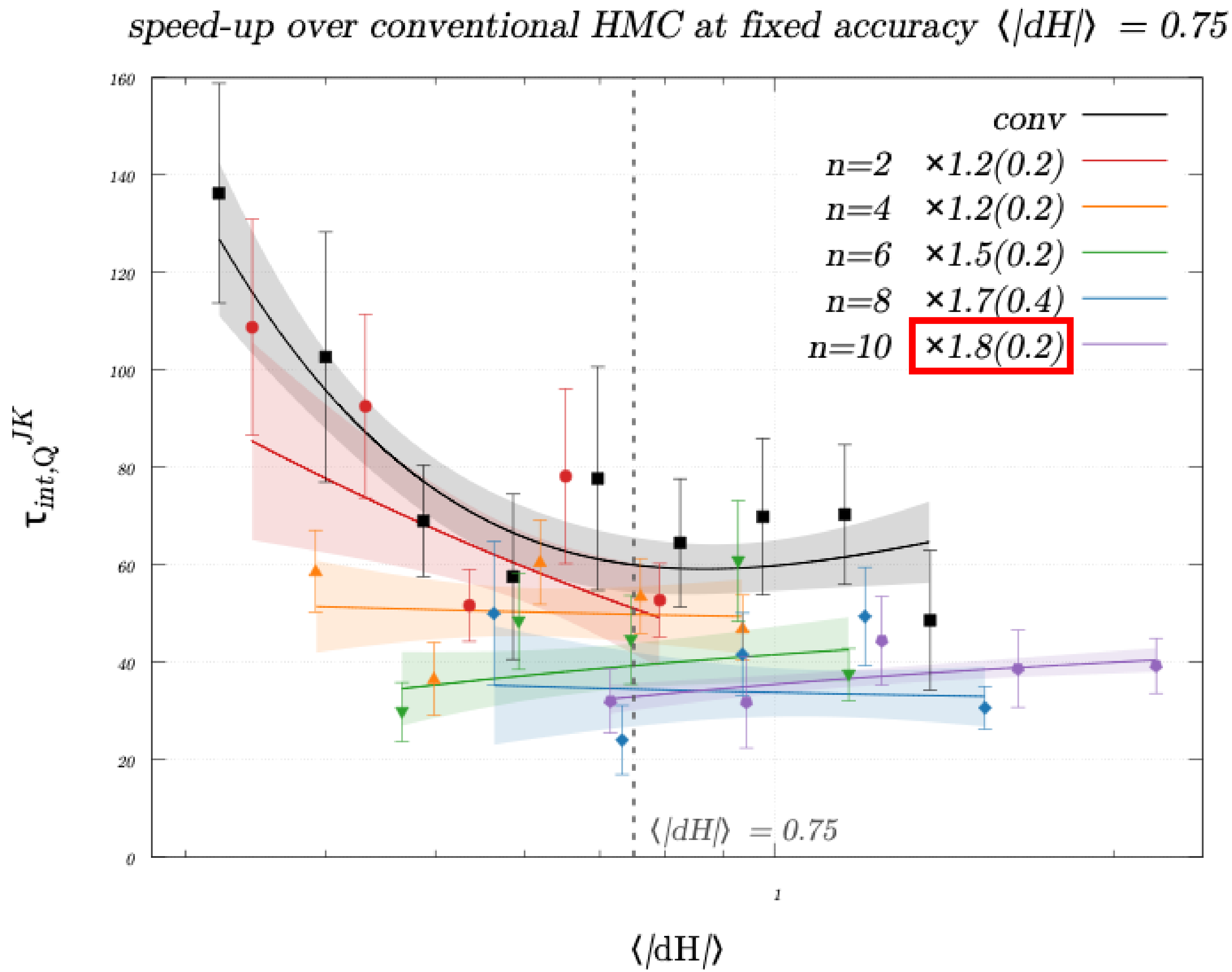
See, e.g., Flynn Jüttner Lawson Sanfilippo 2015



Reach the decorrelated regime with better acceptance

After fitting τ_{int} vs s_{tot} and τ_{int} vs $\langle |dH| \rangle$ with some power-law ansatze geometrical definition: **Berg-Lüscher 1981**

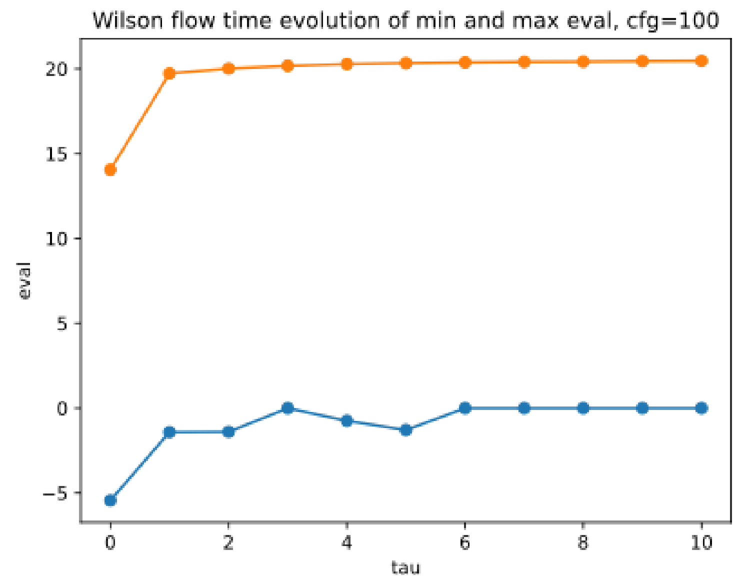
- Topological charge counts the $U(1)$ vortices, which are ultralocal (no dynamics in 2D pure Maxwell)



- Building neural network to build $K(U)$

- Currently exploring an ideal target function:
  Hessian, covariant Laplacian, …

—Ideas to extract the universal part:
Chebyshev, Lanczos/Arnoldi, …



Chris Kelly

Summary

- RMHMC: ideal application of the rapidly-developing AI/ML methodologies; we import the coupling layer in this study

- The result from our toy model study with the CPN model looks attractive
  https://github.com/nobuyukimatsumoto68/2024_latticeCPN_dev/tree/main
  (work-in-progress fork of https://github.com/jglaubitz/2024_latticeCPN_dev)

- An optimization with AI/ML in QCD study is under way
  GPT: Christoph https://github.com/lehner/gpt/tree/2024-10
  HP Cortex: Chris Kelly https://github.com/giltirn/HPCortex

- The diagonal part of the coupling layer gives room to tweak the effective action in the HMC analytically by the Jacobian determinant:

$$H_{\mathrm{HMC}} \equiv \frac{1}{2} p^T K(U) p + S(U) - \frac{1}{2} \log \det K(U)$$

- Trivializing-map-like choice → randomized coupling; did not accelerate

- Nonetheless, in cond-mat: *"superdiffusion"* = *changing the critical exponents;* would be interesting to find a perturbation, which can even be nonlocal, that changes the diffusion properties

Thanks!