Field Transformation HMC

Xiao-Yong Jin for Lattice QCD ECP

May 12, 2020

Contents

Versions

- April 28, 2020, version 0
- April 28, 2020, version 1
- May 12, 2020, version 2

1 Introduction

Following the arguments of "Trivializing maps" [1], to evaluate,

$$\langle O \rangle = 1/Z \int \mathrm{d}x\, O(x) e^{-S(x)}, \tag{1}$$

we perform a change of variable,

$$x = F(y) \tag{2}$$

with a vector functions F, we have

$$\langle O \rangle = \frac{1}{Z} \int \mathrm{d}y\, |\det[J(y)]|\, O\left(F(y)\right) e^{-S(F(y))}, \tag{3}$$

where the Jacobian matrix,

$$J(y) = \frac{\partial F(y)}{\partial y}. \tag{4}$$

F has to satisfy,

- Injective (1 to 1), from the new integration domain to the old.

- Continuously differentiable (or differentiable and have continuous inverse).

Rewrite the integral as,

$$\langle O \rangle = \frac{1}{Z} \int dy\, O\left(F(y)\right) e^{-S(F(y)) + \ln|\det[J(y)]|}. \tag{5}$$

F is a trivializing map, when

$$S(F(y)) - \ln|\det[J(y)]| = \text{constant} \tag{6}$$

and our expectation value simplifies to

$$\langle O \rangle = \frac{1}{Z} \int dy\, O\left(F(y)\right). \tag{7}$$

In terms of HMC, we add the conjugate momenta, π, and use the equations of motion derived from the Hamiltonian,

$$\mathcal{H}(y,\pi) = \tfrac{1}{2}\pi^2 + S\left(F(y)\right) - \ln|\det[J(y)]|, \tag{8}$$

as

$$\frac{d}{dt}\pi = -\frac{\partial}{\partial y}\mathcal{H} = -J(y)S'\left(F(y)\right) + \text{tr}\left[J^{-1}\frac{d}{dy}J\right], \tag{9}$$

$$\frac{d}{dt}y = \frac{\partial}{\partial y}\mathcal{H} = \pi. \tag{10}$$

This is separable and can use the usual explicit, symplectic and symmetric discrete integrators.

Consider a change of variable in π,

$$\pi = J(y)p = J\left(F^{-1}(x)\right)p \tag{11}$$

we get a new Hamiltonian,

$$H(x,p) = \tfrac{1}{2}p^\dagger M p + S(x) - \ln|\det[J]| \tag{12}$$

where the positive definite M,

$$M(x) = J^\dagger\left(F^{-1}(x)\right)J\left(F^{-1}(x)\right) \tag{13}$$

is the kernel of the kinetic term considered Duane et al [2, 3].

2

2 Test in one dimensional model

We consider a one dimensional spin chain from complete fixing of the temporal links of a two dimensional gauge theory with action,

$$S(U) = -\sum_x \beta \operatorname{Re} \operatorname{tr}[U_x^\dagger U_{x+1}]. \tag{14}$$

With $U(1)$ gauge group, we use phase angle, $\theta \in [-\pi, \pi)$, and

$$U = e^{i\theta}, \tag{15}$$

$$S(U) = -\sum_x \beta \cos[\theta_{x+1} - \theta_x], \tag{16}$$

which is a one dimensional $O(2)$ model. With open boundary conditions, we can perform a change of variable for a chain of N spins,

$$\varphi_x = \theta_{x+1} - \theta_x, \tag{17}$$

$$Z = \int d\theta \prod_x e^{\beta \cos[\theta_{x+1} - \theta_x]} \tag{18}$$

$$= \int d\varphi \prod_x e^{\beta \cos[\varphi_x]} \tag{19}$$

$$= \prod_x \int_{-\pi}^{\pi} d\varphi e^{\beta \cos \varphi}, \tag{20}$$

where all the spins decouple. This gives us the expectation value of the plaquette, exactly the same as for the 2D $U(1)$ gauge theory in the limit,

$$\langle P \rangle = \langle \cos[\theta_{x+1} - \theta_x] \rangle = \langle \cos \varphi \rangle = \frac{I_1(\beta)}{I_0(\beta)}, \tag{21}$$

where I is the modified Bessel function of the first kind,

$$I_n(z) = \frac{1}{2\pi i} \oint e^{(t+1/t)z/2} t^{-n-1} dt \tag{22}$$

$$= \frac{1}{\pi} \int_0^{\pi} e^{z \cos \theta} \cos(n\theta) d\theta \quad \text{for } n \in \mathbb{Z}. \tag{23}$$

We can define the topological charge the same as in the 2D U(1) model,

$$Q = \frac{1}{2\pi} \sum_x \operatorname{Arg}[\theta_{x+1} - \theta_x]. \tag{24}$$

For the field transformation, we use "stout smearing" [4] style update, which is also used in the perturbative treatment in reference [1]. Given the 1D analogy of n cycle Wilson loops of length d,

$$W_{ndx}(\theta) = -\frac{\gamma_{nd}}{n^2} \cos[n(\theta_{x+d} - \theta_x)], \tag{25}$$

3

our smearing update from θ to θ' is

$$\theta'_x = F^{\text{even/odd}}_{ndx}(\theta) = \begin{cases} \theta_x + \frac{\partial}{\partial \theta_x} W_{ndy}(\theta) & \text{for "even" ("odd") sites,} \\ \theta_x & \text{for "odd" ("even") sites,} \end{cases} \tag{26}$$

where, nn order to have a tractable Jacobian determinant, we update "even" and "odd" sites separately,

$$\text{even: } x \bmod 2d < d, \tag{27}$$
$$\text{odd: } x \bmod 2d \geq d. \tag{28}$$

This defines a series of diffeomorphism with positive definite Jacobian determinant if

$$-0.5 < \gamma < 0.5. \tag{29}$$

We optimize the coefficients, γ, during HMC, by minimizing the loss function defined as,

$$l(\theta', \pi' | \theta, \pi) = -\frac{1}{N_{\text{batch}}} \sum_{\text{batch}} \max \left\{ 1, e^{H(\theta,\pi) - H(\theta',\pi')} \right\}$$
$$\left(\lambda \frac{1}{N} \sum_x \left[1 - \cos \left((\theta'_{x+1} - \theta'_x) - (\theta_{x+1} - \theta_x) \right) \right] \right. \tag{30}$$
$$\left. + \mu \left[\sum_x \sin(\theta'_{x+1} - \theta'_x) - \sum_x \sin(\theta_{x+1} - \theta_x) \right]^2 \right),$$

where (θ', π') and (θ, π) are respectively the proposed and initial configuration in the MD evolution, and the sum of sin is an approximate of the topological charge.

In the following test, we repeatedly apply a chain of 48 transformations twice, constructing the transformation, F_x, with a total of 96 applications of equation (26),

$$F_x = F^{\text{odd}}_{ndx} \circ F^{\text{even}}_{ndx} \circ \cdots \circ F^{\text{odd}}_{1,2,x} \circ F^{\text{even}}_{1,2,x} \circ F^{\text{odd}}_{1,1,x} \circ F^{\text{even}}_{1,1,x}, \tag{31}$$

with $n \in [1, 2, 3, 4]$ and $d \in [1, 2, 4, 8, 16, 32]$. There are 96 free parameters, γ, in total. In addition we also train the leapfrog step size. We train these 97 parameters during HMC with changing β from small to large in steps, everytime allow thermolizating after changing β before starting optimization. We train the model on a system with the number of sites, $N = 64$, and apply the trained model to a system with the number of sites, $N = 256$.

Figure 1 and 2 shows the evolution during the training, using 2048 parallel Markov chains. Each training step runs one FTHMC trajectory, and calculates the averaged loss function, equation (30), with its derivatives with respect to the training parameters, from the begins and ends of the trajectories of all the Markov chains, and then updates the training parameter towards the direction
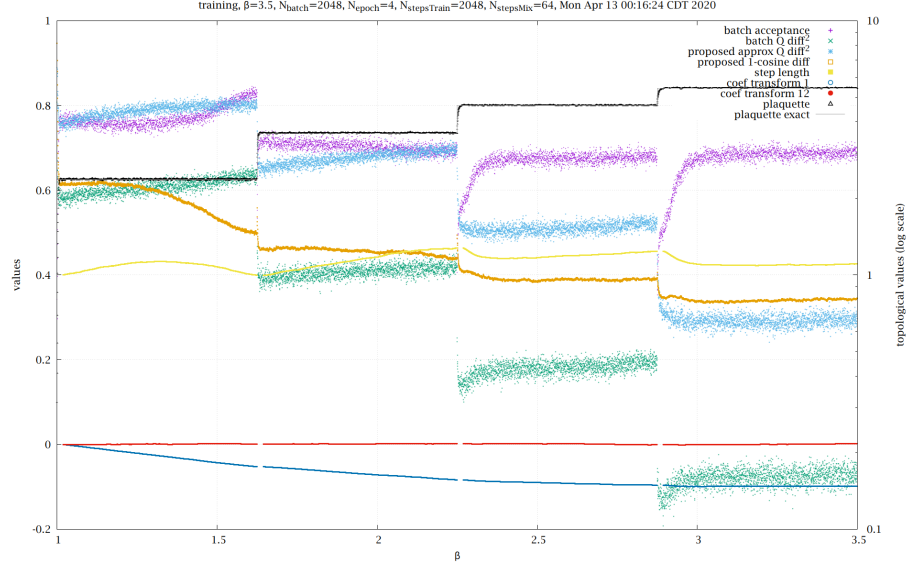
4

Figure 1: Annealed training steps for $N = 64$, at $\beta = 1.625$, 2.25, 2.875, and 3.5.

of reducing the loss function, using the Adam optimization algorithm [5]. Each FTHMC trajectory consists of 10 leapfrog steps, which is fixed during the training. We change the value of β in steps, at $\beta = 1.625$, 2.25, 2.875, 3.5, 4.125, 4.75, 5.375, 5.375, and 6.0. At each β value, we thermalize the Markov chain by running the FTHMC trajectory without computing the loss function and updating the parameters, for 64 trajectories for smaller β values and 256 trajectories for larger β values. We run 2048 training steps at that β value, and change β afterwards.

Figure 3 and 4 show the root-mean-squared difference of topological charges of consecutive Markov states over 1024 parallel Markov chains started from random, with thermalizations removed. The system is of size $N = 64$, corresponding to the training system size, and $\beta = 1.625$, 2.25, 2.875, 3.5, and 6.0.

We use the same parameters trained from $N = 64$ to apply to a system of size $N = 256$. We reduce the leapfrog step size to be half of the trained value to get a good acceptance rate. The resulting root-mean-squared difference of topological charges of consecutive Markov states over 256 parallel Markov chains are in figure 5 for $\beta = 3.5$ and figure 6 for $\beta = 6.0$. We change the number of leapfrog steps from 10 (used in training) to 20 and 30. Figure 7 shows the evolutions of topological charges in a few individual Markov chains at $\beta = 3.5$ using 10 leapfrog steps per trajectory. Figure 8 shows the same for $\beta = 6.0$ using 10 leapfrog steps per trajectory, while figure 9 is for $\beta = 6.0$
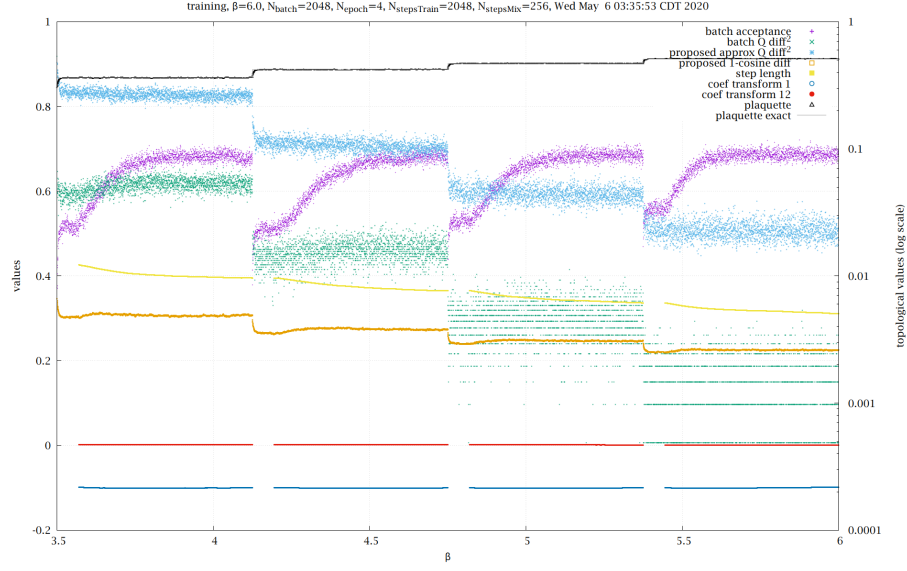
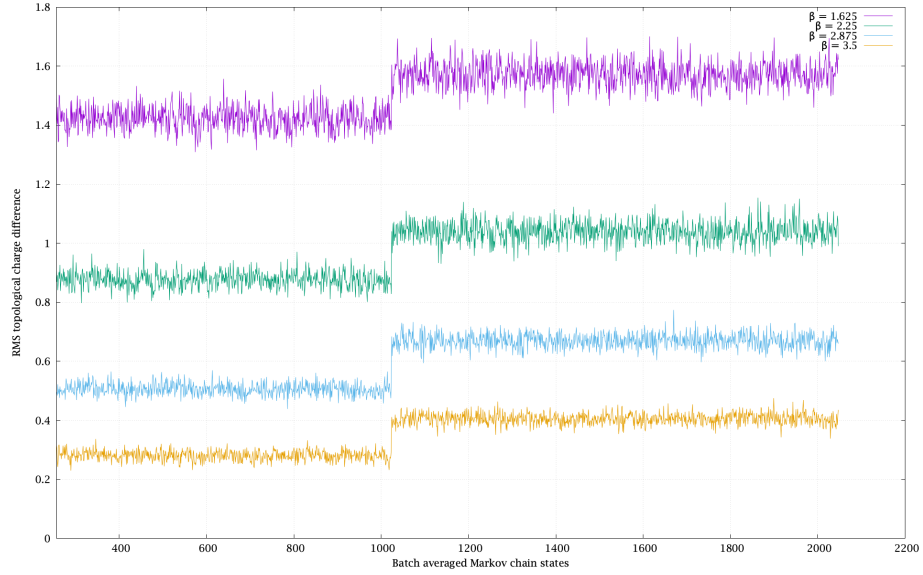Figure 2: Continued annealed training steps for $N = 64$, at $\beta = 4.125$, 4.75, 5.375, 5.375, and 6.0.

Figure 3: Topological charge change rate at $N = 64$, at $\beta = 1.625$, 2.25, 2.875, and 3.5, with traditional HMC (trajectories before 1024) and field transformation HMC (after 1024).
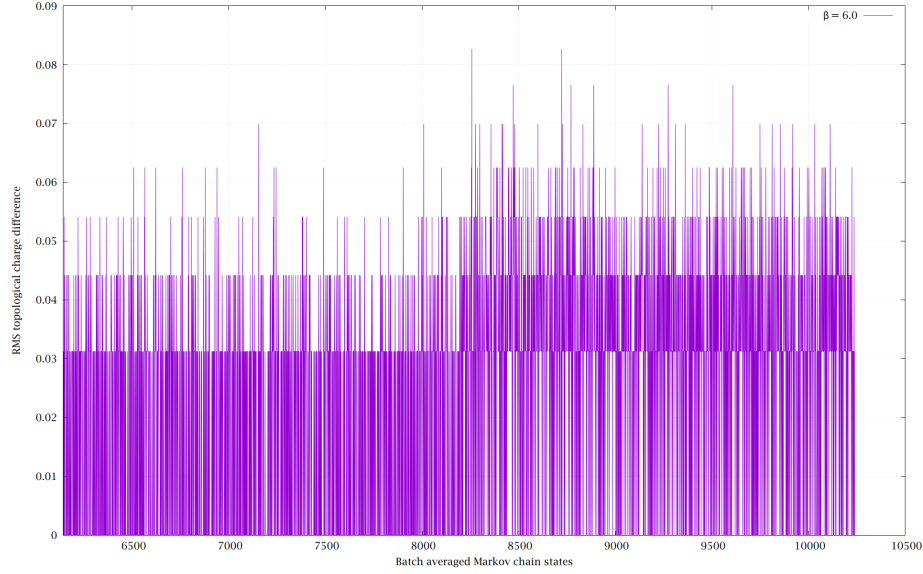
Figure 4: Topological charge change rate at $N = 64$, $\beta = 6.0$, with traditional HMC (trajectories before 8192) and field transformation HMC (after 8192).

using 30 leapfrog steps per trajectory.

3   Discussions and plans

A chain of "stout smearing" with optimized coefficients improves topoloical tunneling in our test one dimensional system. Parameters trained for a small system of size $N = 64$ works for a larger system of size $N = 256$.

Plans: (a) Try optimize the trivializing map directly, by reducing the force or the action difference. (b) Test with 2D U(1) model. (c) Try neural network parametrized smearing.

References

[1]   Martin Luscher. "Trivializing maps, the Wilson flow and the HMC algorithm". In: *Commun. Math. Phys.* 293 (2010), pp. 899-919. DOI: `10.1007/s00220-009-0953-7`. arXiv: `0907.5491` [`hep-lat`] (cit. on pp. 1, 3).

[2]   Simon Duane et al. "Acceleration of Gauge Field Dynamics". In: *Phys. Lett.* B176 (1986), p. 143. DOI: `10.1016/0370-2693(86)90940-8` (cit. on p. 2).
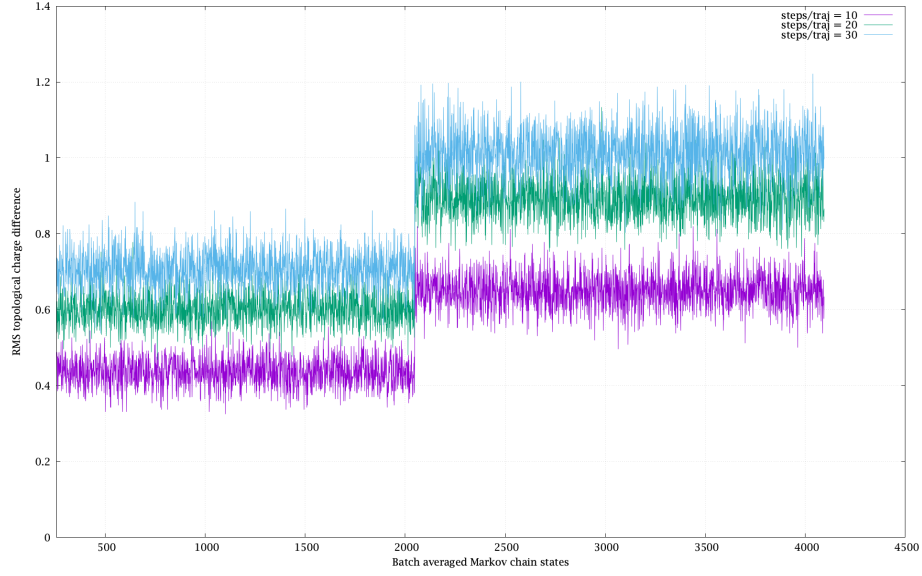
Figure 5: Apply trained parameters from $N = 64$ to $N = 256$, $\beta = 3.5$, with traditional HMC (trajectories before 2048) and field transformation HMC (after 2048).
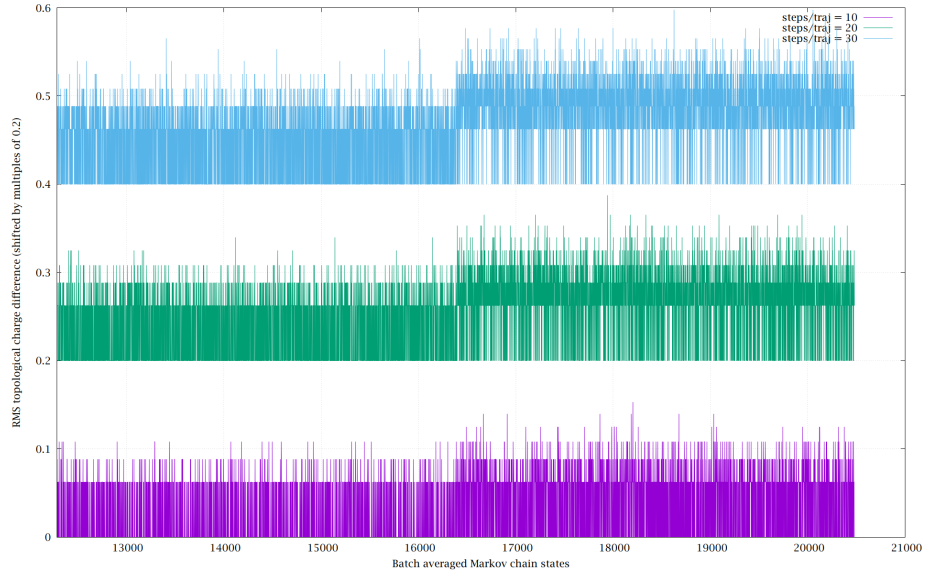

Figure 6: Apply trained parameters from $N = 64$ to $N = 256$, $\beta = 6.0$, with traditional HMC (trajectories before 16384) and field transformation HMC (after 16384).
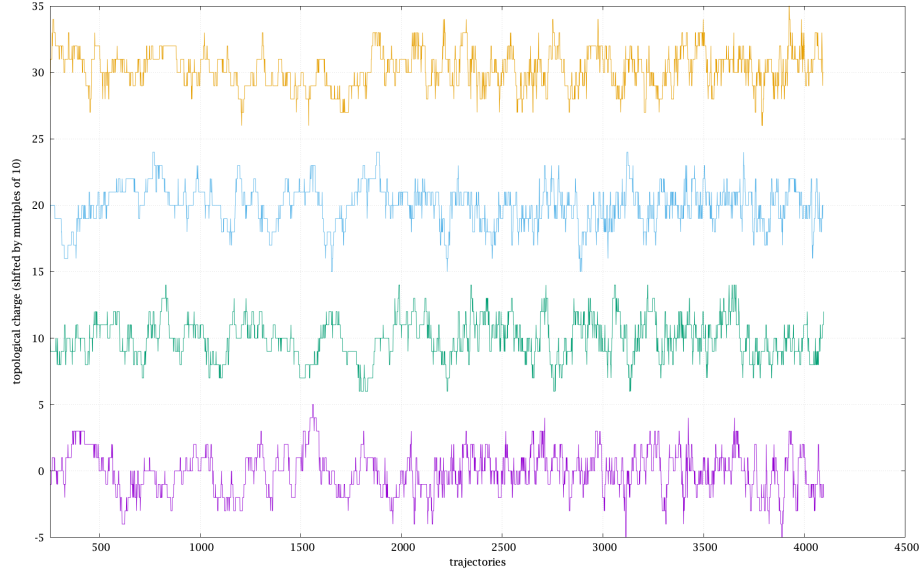
Figure 7: Individual MC evolution after applying trained parameters from $N = 64$ to $N = 256$, $\beta = 3.5$, with traditional HMC (trajectories before 2048) and field transformation HMC (after 2048).
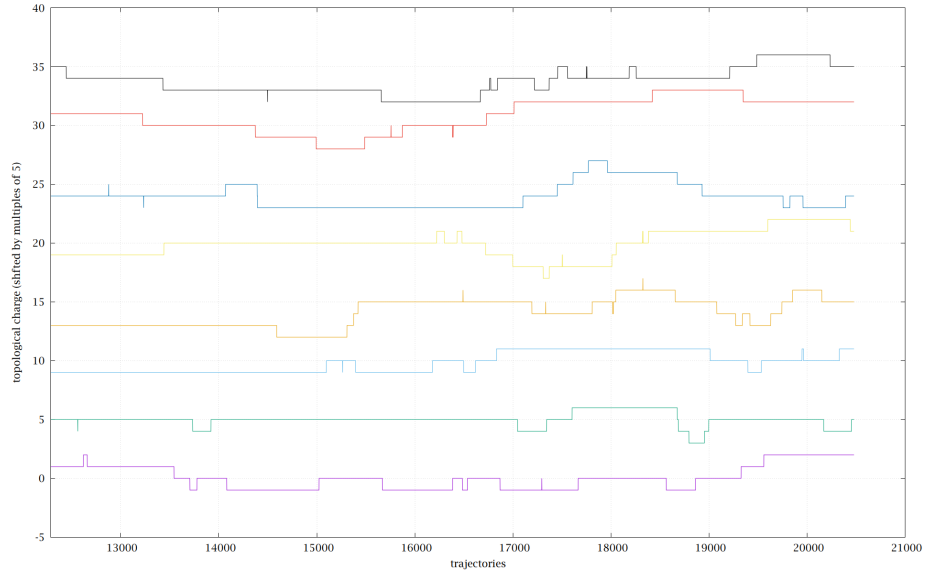

Figure 8: Individual MC evolution after applying trained parameters from $N = 64$ to $N = 256$, $\beta = 6.0$, with traditional HMC (trajectories before 16384) and field transformation HMC (after 16384).
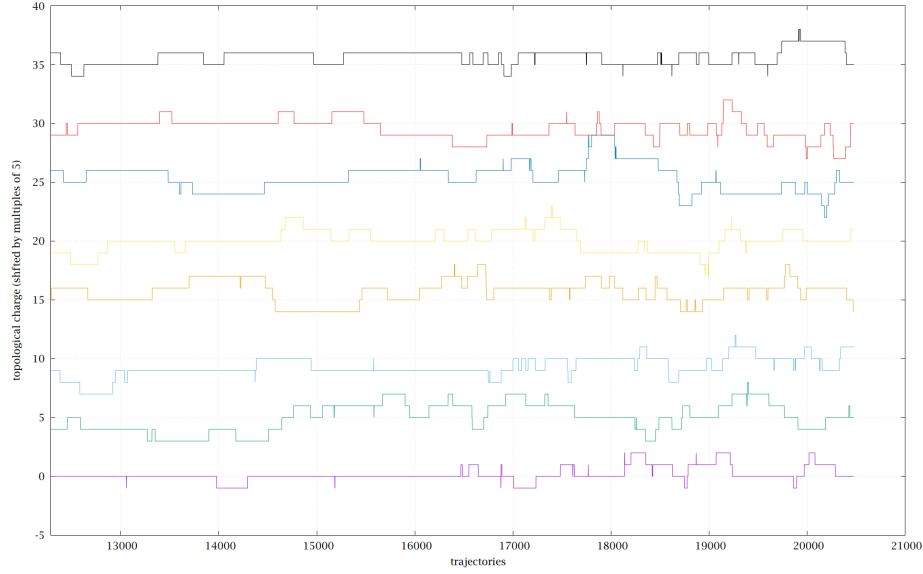
9

Figure 9: Individual MC evolution after applying trained parameters from $N = 64$ to $N = 256$, $\beta = 6.0$, using 30 steps per trajectory, with traditional HMC (trajectories before 16384) and field transformation HMC (after 16384).

[3]   Simon Duane and Brian J. Pendleton. "GAUGE INVARIANT FOURIER ACCEL-ERATION". In: *Phys. Lett.* B206 (1988), pp. 101–106. DOI: 10.1016/0370-2693(88)91270-1 (cit. on p. 2).

[4]   Colin Morningstar and Mike J. Peardon. "Analytic smearing of SU(3) link variables in lattice QCD". In: *Phys. Rev.* D69 (2004), p. 054501. DOI: 10.1103/PhysRevD.69.054501. arXiv: hep-lat/0311018 [hep-lat] (cit. on p. 3).

[5]   Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2014). arXiv: 1412.6980. URL: http://arxiv.org/abs/1412.6980 (cit. on p. 5).