

Practical applications of machine-learned flows on gauge fields

Ryan Abbott,^{b,c} Michael S. Albergo,^d Denis Boyda,^{b,c} Daniel C. Hackett,^{a,b,c,*} Gurtej Kanwar,^e Fernando Romero-López,^{b,c} Phiala E. Shanahan^{b,c} and Julian M. Urban^{b,c}

^a*Fermi National Accelerator Laboratory, Batavia, IL 60510, U.S.A.*

^b*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

^c*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*

^d*Center for Cosmology and Particle Physics, New York University, New York, NY 10003, USA*

^e*Albert Einstein Center, Institute for Theoretical Physics, University of Bern, 3012 Bern, Switzerland*

E-mail: dhackett@fnal.gov

Normalizing flows are machine-learned maps between different lattice theories which can be used as components in exact sampling and inference schemes. Ongoing work yields increasingly expressive flows on gauge fields, but it remains an open question how flows can improve lattice QCD at state-of-the-art scales. We discuss and demonstrate two applications of flows in replica exchange (parallel tempering) sampling, aimed at improving topological mixing, which are viable with iterative improvements upon presently available flows.

*The 40th International Symposium on Lattice Field Theory (Lattice 2023)
July 31st - August 4th, 2023
Fermi National Accelerator Laboratory*

*Speaker

1. Introduction

Numerical lattice quantum chromodynamics (QCD) is an integral part of the modern particle and nuclear theory toolkit [1–9]. In this framework, the discretized path integral is computed using Monte Carlo methods. Computationally, this is very expensive, and grows more so as physical limits of interest are approached [10–12]. Consequently, algorithmic developments are an important driver of progress. For example, resolving topological freezing [12–14]—an issue that arises in sampling gauge field configurations with state-of-the-art Markov chain Monte Carlo (MCMC) algorithms like heatbath [15–19] or Hybrid/Hamiltonian Monte Carlo (HMC) [20–22]—would provide access to finer lattice spacings than presently affordable. To such ends, recent work has explored how emerging machine learning (ML) techniques may be applied to lattice QCD [23, 24]. Of particular interest has been the possibility of accelerating gauge-field sampling [25–34] using normalizing flows [35–37], a class of generative statistical models with tractable density functions.

In this framework, a flow f is a learned, invertible (diffeomorphic) map between gauge fields. Abstractly, flows may be thought of as bridges between different distributions over gauge fields (or, equivalently, different theories or choices of action parameters). When applied to samples V from some chosen base distribution $r(V)$ as $U = f(V)$, the flow induces a model distribution $q(U)$, whose density may be computed exactly as

$$q(U) = r(V) / J_f(V) , \quad (1)$$

where $J_f(V) = |\det \partial f(V) / \partial V|$ is the Jacobian determinant of the flow transformation. The flow is an *exact* bridge between the base and model distributions: ensembles distributed per r may be transformed to ensembles distributed per q and vice versa. However, the form of the model distribution is a complicated function of the architecture and precise parameter values of the learned flow and does not typically correspond to a distribution of physical interest.

Instead, for sampling applications, one typically variationally optimizes the flow so that q approximates some target density of interest p . In this case, the flow serves as an *approximate* bridge between the base r and target p . Because the density of q can be evaluated, this approximation can be corrected to obtain exact results for p . Samples drawn from r and flowed are samples from q , which can be used to compute reweighted expectations under the target as $\langle O \rangle_p = \langle pO / q \rangle_q$ or used as proposals in independence Metropolis to construct a p -distributed Markov chain [38–42].

Most demonstrations thus far have been in the context of the simplest application of such an approximate bridge: full generative modeling or “(approximate) direct sampling”. In this case, the flow bridges between a Haar uniform base $r(U) = 1$ and the target theory of interest p , specified in terms of an action as $p(U) \propto \exp[-S_p(U)]$. If achieved for QCD at scale, this would offer significant advantages over the present state of the art, including embarrassingly parallel sampling and a natural resolution of topological freezing [24]. However, this requires higher-quality models than presently available.

This raises an important question: how can presently available flows be used to improve lattice QCD? The key idea is that we are free to use a flow to bridge between any r and p of our choosing. Neither distribution need be trivial, and choosing more similar distributions makes for a substantially easier modeling task. In this Proceedings, we explore two promising applications following this idea. After detailing our numerical setup in Sec. 2, we provide preliminary numerical demonstrations in

Secs. 3 and 4 of how flows may accelerate sampling of QCD field configurations using the parallel tempering / replica exchange algorithm. The first in Sec. 3 uses flows to bridge nearby physical theories to accelerate sampling in all of them. The second in Sec. 4 uses flows to repair unphysical action defects introduced to accelerate topological mixing. A recent publication, Ref. [43], presents a related but distinct application: flows can be employed to generate correlated ensembles for multiple different target theories, enabling correlated noise cancellation in computations of derivatives with respect to lattice action parameters.¹

2. Numerical setup

For the two numerical demonstrations below we consider pure-gauge SU(3) defined with the Wilson gauge action [44],

$$S_g(U) = -\frac{\beta}{N_c} \sum_x \sum_{\mu < \nu} \text{Re Tr } P_{\mu\nu}(x), \quad (2)$$

where $P_{\mu\nu}(x)$ is the plaquette with extent in the μ and ν directions. We emphasize that flowed approaches to fermions and the fermion determinant have already been explored [30, 32, 45], including in QCD [33]. These methods may be applied to generalize everything presented here to QCD. The choice to test in pure gauge theory is made only for simplicity and to reduce experimentation costs.

In both demonstrations, the objective is to use flows to improve topological mixing in sampling of pure-gauge targets. The baseline to be improved upon is alternating hits of (pseudo)heatbath (HB) and (pseudo)overrelaxation (OR), denoted as HB+OR [15–19]. We compute topological charges using the clover definition at Wilson flow time $t/a^2 = 2$ [46–49].

MCMC sampling efficiency is typically assessed using the integrated autocorrelation time. For some observable O , it is defined in terms of the autocorrelation function, $C_O(\delta t) = \langle O(\delta t)O(0) \rangle - \langle O(\delta t) \rangle \langle O(0) \rangle$, where equilibration is assumed and the expectations are over all possible Markov chains, as

$$\tau_{\text{int}}(O) = \frac{1}{2} + \sum_{\delta t=1}^{\infty} C_O(\delta t). \quad (3)$$

In practice, the sum must be truncated at some $t_{\text{max}} < \infty$ due to noise in C_O estimates at large δt . Autocorrelations increase the variance of estimators of O by a factor of $2\tau_{\text{int}}(O)$ versus independent samples with the minimal $\tau_{\text{int}} = 1/2$ [50]. Thus, computational cost is linear in τ_{int} .

The model architectures used in both demonstrations are variations on the residual flows [34] used in Ref. [43]. This “coupling layer” architecture uses variable partitioning to guarantee invertibility and a tractable Jacobian [36, 51]. Schematically, each layer updates a subset of “active links” conditioned on the other “frozen” ones by applying a step of gradient flow with respect to a learned potential (cf. the Wilson gauge action in Wilson flow). In each layer, active links are chosen as those in one direction and for one checkerboard parity over sites. The models here have 16 layers such that each link in the lattice is updated exactly twice. Our learned potential is parametrized by applying a learned equivariant smearing to the frozen link field, then summing over plaquettes

¹Preliminary results for this strategy were presented in the talk corresponding to this Proceedings.

each constructed out of one active link and a frozen smeared staple as in Ref. [43, Sec. 2C]. Certain modifications are required for the defect repair architecture in Sec. 4, as discussed therein.

We use reverse KL self-training [38, 52, 53] with path gradients [34, 54] to optimize the models. Note that the base distributions here are pure gauge theories at $\beta > 0$, rather than Haar uniform ($\beta = 0$). This does not modify the training procedure formally, but does require sampling the base distribution using HB+OR. In practice, we accomplish this by evolving B independent Markov chains alongside training, where B is the batch size. In some cases we also employed transfer learning between different volumes and action parameters [55]. Further details of the ML methodology are provided in Ref. [43].

3. Transformed Replica EXchange (T-REX)

The parallel tempering or Replica EXchange (REX) algorithm [56, 57] simultaneously samples parallel Markov chains for a sequence of different, overlapping target densities (p_0, p_1, \dots) on the same lattice geometry. The algorithm proceeds by evolving the chains independently with an appropriate MCMC update for each target (here, HB+OR for the β of that chain), and then occasionally proposing swaps of configurations between neighboring chains. Intuitively, this swapping allows mixing with faster-moving targets to accelerate evolution in slower ones. However, the swap acceptance rate (AR) falls as the lattice size grows, requiring increasingly many chains interpolating between any two target β s to maintain a finite swap AR.

As introduced in Ref. [58], a flow which approximately bridges two neighboring densities can be used to improve the swap AR between them, allowing more widely separated targets. For a flow f , a proposed swap of (U_0, U_1) between the chains for targets p_0 and p_1 proceeds as: flow each sample “across the bridge” as $(U'_1, U'_0) = (f^{-1}(U_1), f(U_0))$, then accept or reject the swap with probability²

$$p_{\text{acc}} = \min \left[1, \frac{p_0(U'_1)p_1(U'_0)}{p_0(U_0)p_1(U_1)} J_f(U_0) J_{f^{-1}}(U_1) \right]. \quad (4)$$

If accepted, the next state is (U'_1, U'_0) ; if rejected, the present state (U_0, U_1) is replicated. We call this algorithm “Transformed Replica EXchange” or T-REX. Several limiting cases are worth noting. For an exact flow between p_0 and p_1 , $p_{\text{acc}} = 1$. For an identity flow, $f(U) = U$, untransformed/standard REX is recovered. For two chains, one targeting Haar uniform and the other the theory of interest, and with no MCMC update for the target chain,³ we recover a sampling scheme equivalent to approximate direct sampling with independence Metropolis.

For a numerical test, we consider simultaneously sampling three chains with $\beta = (5.95, 6, 6.05)$ on a 12^4 lattice geometry. Applying untransformed REX, we find the swap AR is no more than 10^{-4} for both $5.95 \leftrightarrow 6$ and $6 \leftrightarrow 6.05$. To improve this, we train two independent flows using the setup of Sec. 2: one to bridge $5.95 \leftrightarrow 6$ and another for $6 \leftrightarrow 6.05$. Note that we train on 4^4 lattices before transferring to 12^4 lattices, and also transfer between target parameters rather than training each flow from scratch. Applying T-REX with these flows, we obtain swap ARs of $\approx 15\%$ and 20% , respectively.

²This is equivalent to the expression in Ref. [58].

³This amounts to the choice of the identity operation as the MCMC update for the target stream. Using a nontrivial update for the target stream is equivalent to alternating MCMC updates with flow model proposals as explored in Ref. [59].

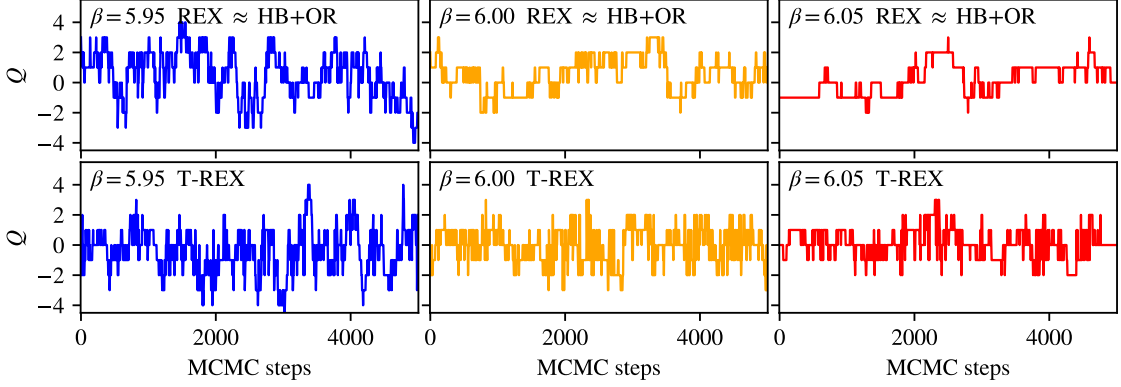


Figure 1: Evolution of the topological charge for a set of three different pure-gauge targets on a 12^4 volume, sampled simultaneously using either REX or T-REX. REX swaps are almost never accepted, so it is equivalent to independent HB+OR on each stream. One MCMC step is 5 HB hits followed by 2 OR hits. Swaps are proposed every 5 steps.

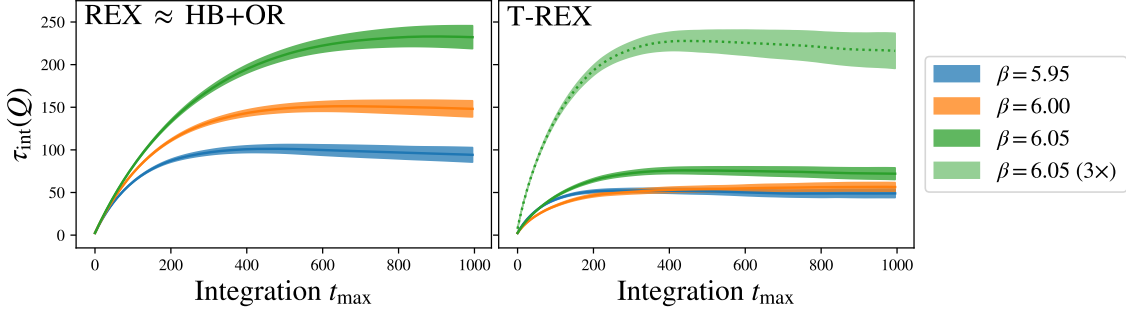


Figure 2: Running estimates of the integrated autocorrelation times of the topological charge for three pure-gauge actions, sampled simultaneously using either REX or T-REX as in Fig. 1. Units are MCMC steps as in Fig. 1. The curves show $1/2 + \sum_{\delta t=1}^{t_{\max}} C_Q(\delta t)$ and are expected to plateau at $\tau_{\text{int}}(Q)$. Uncertainties are evaluated as the standard error over 32 (40) independent streams for REX (T-REX), each of length 10000 (4500) steps. For T-REX, the $\beta = 6.05$ curve is reproduced with a factor of 3 for ease of cost comparison.

Fig. 1 compares MCMC histories of the topological charge on the stream for each β , as sampled using untransformed REX \approx HB+OR (equivalent here due to the \approx zero swap AR) and T-REX using the aforementioned flows. The onset of topological freezing is clearly visible in the untransformed REX streams. T-REX provides a clear improvement in the mixing rate in all three streams.

To be more quantitative, in Fig. 2 we estimate the integrated autocorrelation time τ_{AC} of the topological charge in each stream. Even the slowest-moving chain ($\beta = 6.05$) in T-REX is faster than the fastest chain ($\beta = 5.95$) with REX \approx HB+OR. Whether this represents an advantage depends on the computational goal. We consider two scenarios. First, in a multi-ensemble calculation this is an unambiguous speed-up if flow costs can be neglected: for the same number of HB+OR steps, there is a gain in effective statistics in every stream. In addition, due to the swapping, the three T-REX ensembles are correlated, which may be useful as discussed in Ref. [43]. Second, one might imagine only $\beta = 6.05$ is of interest, and the goal is to resolve topological freezing by bridging to

lower, unfrozen β s as in approximate direct sampling. In this case, the other β s are sampled only to improve mixing and discarded after as a side-product. We find that after paying the $3\times$ overhead factor of running HB+OR for three streams instead of one (and again neglecting flow costs), the cost of sampling $\beta = 6.05$ is the same with either HB+OR or T-REX. Apparently, it will be necessary to bridge more widely separated β s to obtain an advantage in this scenario.

4. Defect Repair Replica EXchange (DR-REX)

The different targets in replica exchange need not all be theories of physical interest. This allows for opportunistic choices of unphysical actions designed to accelerate topological charge mixing when sampled alongside a single physical target stream. An existing approach which exploits this is Parallel Tempering Boundary Conditions (PTBC) [60–62]. The idea is as follows. Open boundary conditions (OBCs) are known to alleviate topological freezing [63], but periodic boundary conditions (PBCs) are preferable otherwise. PTBC uses REX to sample a sequence of actions interpolating between OBCs and PBCs to accelerate topological mixing in the target PBC stream. In practice, rather than using OBCs across the full spatial extent of the lattice, introducing a localized OBC defect (“poking a hole in the boundary”) can be sufficient to speed up topological mixing while requiring fewer interpolating chains.

More precisely, these defected theories are defined here using a generalization of the Wilson pure gauge action with a separate β for each plaquette, i.e.,

$$S_g(U) = -\frac{1}{N_c} \sum_x \sum_{\mu < \nu} \beta_{\mu\nu}(x) \text{Re Tr } P_{\mu\nu}(x) . \quad (5)$$

An OBC-type defect is defined as one where $\beta_{\mu\nu}(x) = \beta_d < \beta$ for all defect plaquettes, as defined by those touching the timelike links in a $L_d^3 \times 1$ volume, and $\beta_{\mu\nu}(x) = \beta$ for all other plaquettes. True temporal OBCs are recovered for a full-lattice defect $L_d = L$ and $\beta_d = 0$. Other defect geometries can be considered and may work better in practice. HB+OR can be generalized straightforwardly to sample these actions.

The novel approach explored here is to apply T-REX to PTBC. In this case, the role of the flow is to (perhaps partially) repair the defect, hence we dub the algorithm Defect Repair Replica EXchange (DR-REX). The locality of the defect makes this a particularly appealing application. Because the physical effects of the defect are localized in its neighborhood, the defect repair flow needs only to act on a subvolume containing this neighborhood, formally performing the identity operation elsewhere. This confers two advantages. First, the swap AR given some subvolume flow is approximately the same for any target volume, with no exponential scaling [31, 55]. Second, the computational cost of applying the subvolume flow is independent of the target volume, such that applying the flow can be much less expensive than the MCMC updates for large-volume targets.

The geometry acted on by the flow is sketched in Fig. 3. A halo of links at the boundary of the subvolume is kept frozen throughout the flow. The absence of translational symmetry in the problem requires generalizing the architecture used in the previous application. We use the straightforward option of promoting every parameter in the model to a lattice of parameters, i.e. adding a site index to each parameter.⁴ We break reflection symmetries in the model in a similar manner. Although this

⁴This is analogous to generalizing convolutional layers to define “locally connected layers”.

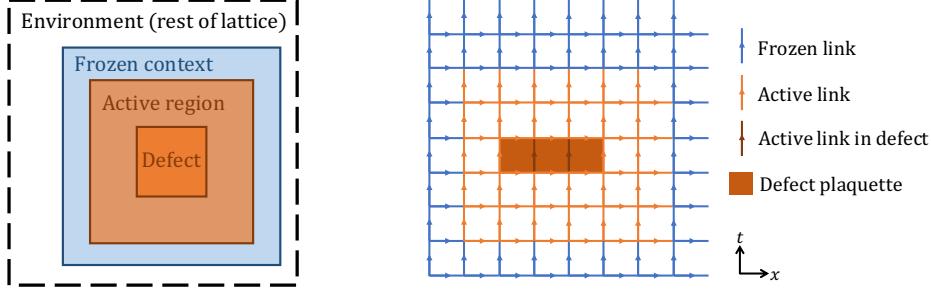


Figure 3: Sketches of the geometry for defect repair flows. At left, the flow performs the identity operation outside the orange region. The active region around the defect (orange) is acted upon, conditioned on its local context (blue). At right, the exact subvolume geometry used for the flows in this work. This geometry is symmetric in all spatial directions (y, z not shown).

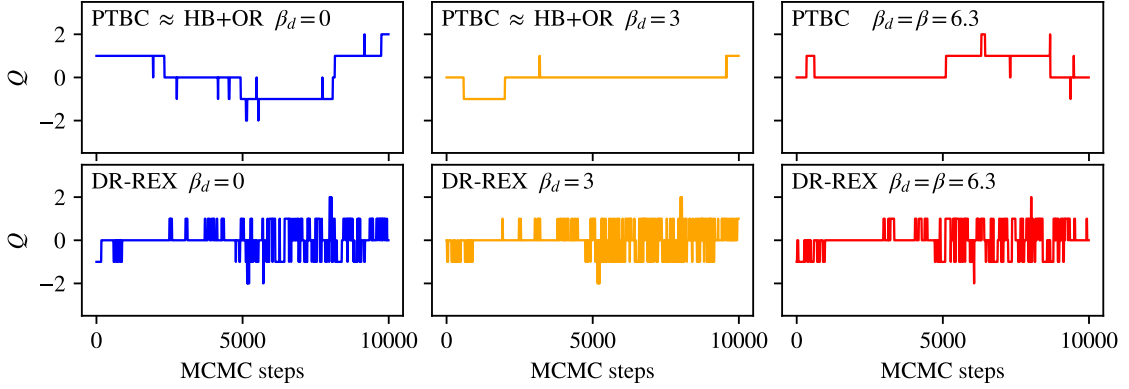


Figure 4: Evolution of the topological charge under PTBC or DR-REX for pure-gauge targets with $\beta = 6.3$ on a 16^4 volume with OBC defects of size 2^3 with $\beta_d = 0, 3, 6.3$, where $\beta_d = \beta = 6.3$ is the target stream with no defect. PTBC swaps are almost never accepted, so it is equivalent to independent HB+OR on each stream. One MCMC step is 1 HB hit followed by 5 OR hits. Swaps are proposed every 10 steps.

massively increases the parameter count, the computational cost of training or applying the flow is not substantially increased. Note that while this choice “hard codes” a flow to a specific subvolume geometry, it can still be applied to subvolumes embedded in different target volumes. Subvolume-transferrable models could be obtained by e.g. instead parametrizing this site dependence with neural networks, but we leave these complications for future work.

For our numerical test, we consider OBC defects of size 2^3 in a 16^4 target at $\beta = 6.3$. We train two defect repair flows which act on 8^4 subvolumes, one bridging between $\beta_d = 0$ and 3, and a second bridging 3 and 6.3, where $\beta_d = 6.3$ is the undefected target. Figure 3 shows the precise subvolume geometry. We train each flow independently, using smaller volumes (10^4 , then 12^4) than the target. Sampling these three theories with $\beta_d = 0, 3, 6.3$ with untransformed PTBC results in a very small swap rate; we observed no accepted swaps in 10^5 attempts. Adding the flows enables swap ARs of 23% and 28% in DR-REX, thus providing at least an $O(10^4)$ improvement. A more fair comparison may be that unflowed PTBC obtains comparable swap ARs with β_d interpolated over 7-8 chains, amounting to a $\sim 2\times$ improvement.

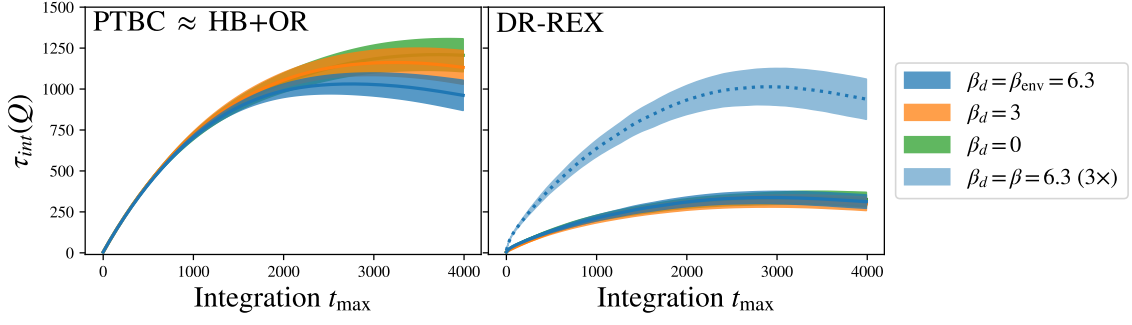


Figure 5: Running estimates of the integrated autocorrelation times of the topological charge for sampling using PTBC or DR-REX as in Fig. 4. Units are MCMC steps as in Fig. 4. The curves show $1/2 + \sum_{\delta t=1}^{t_{\max}} C_Q(\delta t)$ and are expected to plateau at $\tau_{\text{int}}(Q)$. Uncertainties are evaluated as the standard error over 24 independent streams, each of length 29000 steps. For DR-REX, the $\beta_d = 6.3$ curve is reproduced with a factor of 3 for ease of cost comparison.

Figure 4 compares topological charge histories for PTBC \approx HB+OR with those of DR-REX. Clearly, the topological charge is severely frozen for HB+OR, and changes more rapidly under DR-REX. However, much of this apparent mixing is due only to swapping, rather than tunneling events. To assess whether an advantage has been obtained, we estimate τ_Q^{int} in Fig. 5. Accounting for the $3\times$ overhead of running HB+OR for three chains, we find that DR-REX is only at the break-even point with HB+OR, neglecting flow costs. Evidently, the 2^3 OBC defect does not sufficiently increase the tunneling rate (note τ_Q^{int} is approximately the same across the different HB+OR streams). We conclude that it will be necessary to repair larger defects to obtain a true computational advantage.

5. Outlook

There are a number of promising avenues to employ machine-learned flows on gauge fields for computational advantages in lattice QCD calculations. Ref. [43] demonstrated uses of correlated ensembles generated by flows. In this Proceedings, we explore two different replica exchange methods, T-REX and DR-REX, to improve topological freezing. Computational advantage has yet to be shown when accounting for all costs associated with the learned components. However, if they may be neglected, there are already demonstrated advantages if several nearby ensembles are to be generated simultaneously. Moreover, there are structural reasons to believe these methods are especially promising.

Flowed replica exchange methods may be thought of as a generalization of approximate direct sampling with an adjustable trade-off between how much of the sampling task is accomplished by flows versus traditional MCMC updates. With two chains, T-REX provides an exact sampling scheme to use flows between a target theory and a nontrivial base distribution, even when exact sampling is not possible for the base distribution. This can enable useful flow-based sampling before full trivializing maps are available, and a smooth approach to the capabilities of approximate direct sampling as flow technology develops further. Using more chains to interpolate over a sequence of theories further eases requirements on flow model quality. We note that these same advantages also hold for the closely related class of algorithms including Continual Repeated

Annealed Flow Transport Monte Carlo (CRAFT) [64, 65] and various limiting cases thereof [66], including stochastic normalizing flows (SNFs) [67, 68].

The combination of flows with the PTBC algorithm, DR-REX, has important structural advantages that make it an especially promising application of flows. Flows can increase the swap AR between chains, potentially making larger defects practical than with untransformed replica exchange alone. Furthermore, defect repair flows which act on a fixed-size embedded subvolume can be applied to larger target volumes without increasing the application cost of the flow or the required model quality. The defects treated in the application here were too small to alleviate freezing, via either PTBC or DR-REX, but there is a clear path forward with well-defined learning tasks achievable with iterative improvements over what is done here.

With further model tuning (which was not extensively explored here), and continued development of flow technology, both T-REX and DR-REX present promising avenues to computational advantage in lattice QCD gauge field generation at scale. To fully exploit novel flow technology it will be important to further develop these methods and continue to explore other algorithmic approaches to accelerate sampling by incorporating flows.

Acknowledgements

We thank Aleksandar Botev, Kyle Cranmer, Alexander G. D. G. Matthews, Sébastien Racanière, Ali Razavi, and Danilo J. Rezende for useful discussions and valuable contributions to the early stages of this work. RA, DCH, FRL, PES, and JMU are supported in part by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under grant Contract Number DE-SC0011090. PES is additionally supported by the U.S. DOE Early Career Award DE-SC0021006, by a NEC research award, and by the Carl G and Shirley Sontheimer Research Fund. FRL acknowledges support by the Mauricio and Carlota Botton Fellowship. GK was supported by the Swiss National Science Foundation (SNSF) under grant 200020_200424. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. MSA is supported by the National Science Foundation under the award PHY-2141336. MSA thanks the Flatiron Institute for their hospitality. This work is supported by the U.S. National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, <http://iaifi.org/>) and is associated with an ALCF Aurora Early Science Program project, and used resources of the Argonne Leadership Computing Facility which is a DOE Office of Science User Facility supported under Contract DEAC02-06CH11357. The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center [69] for providing HPC resources that have contributed to the research results reported within this paper. Numerical experiments and data analysis used PyTorch [70], JAX [71], Haiku [72], Horovod [73], NumPy [74], and SciPy [75]. Figures were produced using matplotlib [76].

References

- [1] C. Morningstar, *The Monte Carlo method in quantum field theory*, [hep-lat/0702020](https://arxiv.org/abs/hep-lat/0702020).

- [2] USQCD collaboration, *Opportunities for Lattice QCD in Quark and Lepton Flavor Physics*, *Eur. Phys. J. A* **55** (2019) 195 [[1904.09479](#)].
- [3] USQCD collaboration, *Opportunities for Lattice QCD in Quark and Lepton Flavor Physics*, *Eur. Phys. J. A* **55** (2019) 195 [[1904.09479](#)].
- [4] USQCD collaboration, *Lattice QCD and Neutrino-Nucleus Scattering*, *Eur. Phys. J. A* **55** (2019) 196 [[1904.09931](#)].
- [5] USQCD collaboration, *The Role of Lattice QCD in Searches for Violations of Fundamental Symmetries and Signals for New Physics*, *Eur. Phys. J. A* **55** (2019) 197 [[1904.09704](#)].
- [6] USQCD collaboration, *Hadrons and Nuclei*, *Eur. Phys. J. A* **55** (2019) 193 [[1904.09512](#)].
- [7] USQCD collaboration, *Hot-dense Lattice QCD: USQCD whitepaper 2018*, *Eur. Phys. J. A* **55** (2019) 194 [[1904.09951](#)].
- [8] USQCD collaboration, *Status and Future Perspectives for Lattice Gauge Theory Calculations to the Exascale and Beyond*, *Eur. Phys. J. A* **55** (2019) 199 [[1904.09725](#)].
- [9] P.A. Boyle et al., *A lattice QCD perspective on weak decays of b and c quarks Snowmass 2022 White Paper*, in *2022 Snowmass Summer Study*, 5, 2022 [[2205.15373](#)].
- [10] U. Wolff, *Critical slowing down*, *Nucl. Phys. Proc. Suppl.* **17** (1990) 93.
- [11] S. Schaefer, R. Sommer and F. Virota, *Investigating the critical slowing down of QCD simulations*, *PoS LAT2009* (2009) 032 [[0910.1465](#)].
- [12] ALPHA collaboration, *Critical slowing down and error analysis in lattice QCD simulations*, *Nucl. Phys. B* **845** (2011) 93 [[1009.5228](#)].
- [13] B. Alles, G. Boyd, M. D’Elia, A. Di Giacomo and E. Vicari, *Hybrid Monte Carlo and topological modes of full QCD*, *Phys. Lett. B* **389** (1996) 107 [[hep-lat/9607049](#)].
- [14] L. Del Debbio, G.M. Manca and E. Vicari, *Critical slowing down of topological modes*, *Phys. Lett. B* **594** (2004) 315 [[hep-lat/0403001](#)].
- [15] M. Creutz, *Monte Carlo Study of Quantized SU(2) Gauge Theory*, *Phys. Rev. D* **21** (1980) 2308.
- [16] N. Cabibbo and E. Marinari, *A New Method for Updating SU(N) Matrices in Computer Simulations of Gauge Theories*, *Phys. Lett. B* **119** (1982) 387.
- [17] A.D. Kennedy and B.J. Pendleton, *Improved Heat Bath Method for Monte Carlo Calculations in Lattice Gauge Theories*, *Phys. Lett. B* **156** (1985) 393.
- [18] F.R. Brown and T.J. Woch, *Overrelaxed Heat Bath and Metropolis Algorithms for Accelerating Pure Gauge Monte Carlo Calculations*, *Phys. Rev. Lett.* **58** (1987) 2394.

- [19] S.L. Adler, *Overrelaxation Algorithms for Lattice Field Theories*, *Phys. Rev. D* **37** (1988) 458.
- [20] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.
- [21] R.M. Neal, *Probabilistic inference using Markov chain Monte Carlo methods*, Department of Computer Science, University of Toronto Toronto, ON, Canada (1993).
- [22] R.M. Neal, *Bayesian Learning for Neural Networks*, *Lecture Notes in Statistics* **118** (1996) .
- [23] D. Boyda et al., *Applications of Machine Learning to Lattice Quantum Field Theory*, in *2022 Snowmass Summer Study*, 2, 2022 [[2202.05838](#)].
- [24] K. Cranmer, G. Kanwar, S. Racanière, D.J. Rezende and P.E. Shanahan, *Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics*, *Nature Rev. Phys.* **5** (2023) 526 [[2309.01156](#)].
- [25] G. Kanwar, M.S. Albergo, D. Boyda, K. Cranmer, D.C. Hackett, S. Racanière et al., *Equivariant flow-based sampling for lattice gauge theory*, *Phys. Rev. Lett.* **125** (2020) 121601 [[2003.06413](#)].
- [26] D. Boyda, G. Kanwar, S. Racanière, D.J. Rezende, M.S. Albergo, K. Cranmer et al., *Sampling using $SU(N)$ gauge equivariant flows*, *Phys. Rev. D* **103** (2021) 074504 [[2008.05456](#)].
- [27] S. Foreman, X.-Y. Jin and J.C. Osborn, *Deep Learning Hamiltonian Monte Carlo*, in *9th International Conference on Learning Representations*, May, 2021 [[2105.03418](#)].
- [28] S. Foreman, T. Izubuchi, L. Jin, X.-Y. Jin, J.C. Osborn and A. Tomiya, *HMC with Normalizing Flows*, in *38th International Symposium on Lattice Field Theory*, Dec, 2021 [[2112.01586](#)].
- [29] S. Foreman, X.-Y. Jin and J.C. Osborn, *LeapfrogLayers: A Trainable Framework for Effective Topological Sampling*, in *38th International Symposium on Lattice Field Theory*, 12, 2021 [[2112.01582](#)].
- [30] M.S. Albergo, D. Boyda, K. Cranmer, D.C. Hackett, G. Kanwar, S. Racanière et al., *Flow-based sampling in the lattice Schwinger model at criticality*, [2202.11712](#).
- [31] J. Finkenrath, *Tackling critical slowing down using global correction steps with equivariant flows: the case of the Schwinger model*, [2201.02216](#).
- [32] R. Abbott, M.S. Albergo, D. Boyda, K. Cranmer, D.C. Hackett, G. Kanwar et al., *Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions*, *Phys. Rev. D* **106** (2022) 074506 [[2207.08945](#)].
- [33] R. Abbott, M.S. Albergo, A. Botev, D. Boyda, K. Cranmer, D.C. Hackett et al., *Sampling QCD field configurations with gauge-equivariant flow models*, in *39th International Symposium on Lattice Field Theory*, 8, 2022 [[2208.03832](#)].

- [34] R. Abbott, M.S. Albergo, A. Botev, D. Boyda, K. Cranmer, D.C. Hackett et al., *Normalizing flows for lattice gauge theory in arbitrary space-time dimension*, [2305.02402](#).
- [35] D.J. Rezende and S. Mohamed, *Variational inference with normalizing flows*, [1505.05770](#).
- [36] L. Dinh, J. Sohl-Dickstein and S. Bengio, *Density estimation using real nvp*, [1605.08803](#).
- [37] G. Papamakarios, E. Nalisnick, D.J. Rezende, S. Mohamed and B. Lakshminarayanan, *Normalizing flows for probabilistic modeling and inference*, *Journal of Machine Learning Research* **22** (2021) 1.
- [38] M.S. Albergo, G. Kanwar and P.E. Shanahan, *Flow-based generative models for Markov chain Monte Carlo in lattice field theory*, *Phys. Rev. D* **100** (2019) 034515 [[1904.12072](#)].
- [39] F. Noé, S. Olsson, J. Köhler and H. Wu, *Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning*, *Science* **365** (2019) eaaw1147 [[1812.01729](#)].
- [40] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *Equation of state calculations by fast computing machines*, *J. Chem. Phys.* **21** (1953) 1087.
- [41] W.K. Hastings, *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, *Biometrika* **57** (1970) 97.
- [42] L. Tierney, *Markov chains for exploring posterior distributions*, *the Annals of Statistics* (1994) 1701.
- [43] R. Abbott, A. Botev, D. Boyda, D.C. Hackett, G. Kanwar, S. Racanière et al., *Applications of flow models to the generation of correlated lattice QCD ensembles*, .
- [44] K.G. Wilson, *Confinement of Quarks*, *Phys. Rev. D* **10** (1974) 2445.
- [45] M.S. Albergo, G. Kanwar, S. Racanière, D.J. Rezende, J.M. Urban, D. Boyda et al., *Flow-based sampling for fermionic lattice field theories*, *Phys. Rev. D* **104** (2021) 114507 [[2106.05934](#)].
- [46] M. Lüscher, *Trivializing maps, the Wilson flow and the HMC algorithm*, *Commun. Math. Phys.* **293** (2010) 899 [[0907.5491](#)].
- [47] M. Lüscher, *Properties and uses of the Wilson flow in lattice QCD*, *JHEP* **08** (2010) 071 [[1006.4518](#)].
- [48] R. Narayanan and H. Neuberger, *Infinite N phase transitions in continuum Wilson loop operators*, *JHEP* **03** (2006) 064 [[hep-th/0601210](#)].
- [49] R. Lohmayer and H. Neuberger, *Continuous smearing of Wilson Loops.*, *Proc. Sci. Lattice* **2011** (2012) 249 [[1110.3522](#)].
- [50] ALPHA collaboration, *Monte Carlo errors with less errors*, *Comput. Phys. Commun.* **156** (2004) 143 [[hep-lat/0306017](#)].

- [51] L. Dinh, D. Krueger and Y. Bengio, *Nice: Non-linear independent components estimation*, [1410.8516](#).
- [52] S.-H. Li and L. Wang, *Neural network renormalization group*, *Phys. Rev. Lett.* **121** (2018) 260601 [[1802.02840](#)].
- [53] S. Kullback and R.A. Leibler, *On Information and Sufficiency*, *The Annals of Mathematical Statistics* **22** (1951) 79 .
- [54] L. Vaitl, K.A. Nicoli, S. Nakajima and P. Kessel, *Gradients should stay on path: Better estimators of the reverse- and forward kl divergence for normalizing flows*, 2022.
- [55] R. Abbott, M.S. Albergo, A. Botev, D. Boyda, K. Cranmer, D.C. Hackett et al., *Aspects of scaling and scalability for flow-based sampling of lattice QCD*, *Eur. Phys. J. A* **59** (2023) 257 [[2211.07541](#)].
- [56] R.H. Swendsen and J.-S. Wang, *Replica Monte Carlo Simulation of Spin-Glasses*, *Phys. Rev. Lett.* **57** (1986) 2607.
- [57] D.J. Earl and M.W. Deem, *Parallel tempering: Theory, applications, and new perspectives*, *Physical Chemistry Chemical Physics* **7** (2005) 3910.
- [58] M. Invernizzi, A. Krämer, C. Clementi and F. Noé, *Skipping the replica exchange ladder with normalizing flows*, *The Journal of Physical Chemistry Letters* **13** (2022) 11643.
- [59] D.C. Hackett, C.-C. Hsieh, M.S. Albergo, D. Boyda, J.-W. Chen, K.-F. Chen et al., *Flow-based sampling for multimodal distributions in lattice field theory*, [2107.00734](#).
- [60] M. Hasenbusch, *Fighting topological freezing in the two-dimensional CPN-1 model*, *Phys. Rev. D* **96** (2017) 054504 [[1706.04443](#)].
- [61] C. Bonanno, C. Bonati and M. D’Elia, *Large- N $SU(N)$ Yang-Mills theories with milder topological freezing*, *JHEP* **03** (2021) 111 [[2012.14000](#)].
- [62] C. Bonanno, M. D’Elia, B. Lucini and D. Vadacchino, *Towards glueball masses of large- N $SU(N)$ pure-gauge theories without topological freezing*, *Phys. Lett. B* **833** (2022) 137281 [[2205.06190](#)].
- [63] M. Luscher and S. Schaefer, *Lattice QCD without topology barriers*, *JHEP* **07** (2011) 036 [[1105.4749](#)].
- [64] A. Matthews, M. Arbel, D.J. Rezende and A. Doucet, *Continual repeated annealed flow transport Monte Carlo*, .
- [65] M. Arbel, A. Matthews and A. Doucet, *Annealed flow transport Monte Carlo*, *International Conference on Machine Learning* (2021) 318.
- [66] R.M. Neal, *Annealed importance sampling*, *Statistics and computing* **11** (2001) 125.

- [67] H. Wu, J. Köhler and F. Noé, *Stochastic normalizing flows*, *Advances in Neural Information Processing Systems* **33** (2020) 5933 [[2002.06707](#)].
- [68] M. Caselle, E. Cellini, A. Nada and M. Panero, *Stochastic normalizing flows as non-equilibrium transformations*, *JHEP* **07** (2022) 015 [[2201.08862](#)].
- [69] A. Reuther, J. Kepner, C. Byun, S. Samsi, W. Arcand, D. Bestor et al., *Interactive supercomputing on 40,000 cores for machine learning and data analysis*, *2018 IEEE High Performance extreme Computing Conference (HPEC)* (2018) 1 [[1807.07814](#)].
- [70] A. Paszke et al., *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds., pp. 8024–8035, Curran Associates, Inc. (2019), <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [71] J. Bradbury, R. Frostig, P. Hawkins, M.J. Johnson, C. Leary, D. Maclaurin et al., *JAX: composable transformations of Python+NumPy programs*, 2018.
- [72] T. Hennigan, T. Cai, T. Norman and I. Babuschkin, *Haiku: Sonnet for JAX*, 2020.
- [73] A. Sergeev and M. Del Balso, *Horovod: fast and easy distributed deep learning in TensorFlow*, [1802.05799](#).
- [74] C.R. Harris, K.J. Millman, S.J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau et al., *Array programming with numpy*, *Nature* **585** (2020) 357.
- [75] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau et al., *Scipy 1.0: fundamental algorithms for scientific computing in python*, *Nature methods* **17** (2020) 261.
- [76] J.D. Hunter, *Matplotlib: A 2d graphics environment*, *Computing in Science & Engineering* **9** (2007) 90.