nature reviews physics

https://doi.org/10.1038/s42254-023-00616-w

Perspective

Check for updates

Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics

Kyle Cranmer [1], Gurtej Kanwar [2], Sébastien Racanière [3], Danilo J. Rezende [3] & Phiala E. Shanahan [4,5]

Abstract

Sampling from known probability distributions is a ubiquitous task in computational science, underlying calculations in domains from linguistics to biology and physics. Generative machine-learning (ML) models have emerged as a promising tool in this space, building on the success of this approach in applications such as image, text and audio generation. Often, however, generative tasks in scientific domains have unique structures and features — such as complex symmetries and the requirement of exactness guarantees — that present both challenges and opportunities for ML. This Perspective outlines the advances in ML-based sampling motivated by lattice quantum field theory, in particular for the theory of quantum chromodynamics. Enabling calculations of the structure and interactions of matter from our most fundamental understanding of particle physics, lattice quantum chromodynamics is one of the main consumers of open-science supercomputing worldwide. The design of ML algorithms for this application faces profound challenges, including the necessity of scaling custom ML architectures to the largest supercomputers, but also promises immense benefits, and is spurring a wave of development in ML-based sampling more broadly. In lattice field theory, if this approach can realize its early promise it will be a transformative step towards first-principles physics calculations in particle, nuclear and condensed matter physics that are intractable with traditional approaches.

Sections

[1]Physics Department, University of Wisconsin-Madison, Madison, WI, USA. [2]Albert Einstein Center for Fundamental Physics, Institute for Theoretical Physics, University of Bern, Bern, Switzerland. [3]Google DeepMind, London, UK. [4]Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, MA, USA. e-mail: pshana@mit.edu

Introduction

Theoretical nuclear physics has the ironic feature that although the fundamental laws are well understood, the computations required to make quantitative, first-principles predictions are in many cases currently infeasible. The strong nuclear force is fundamentally described by the quantum field theory known as quantum chromodynamics (QCD), which details the dynamics of constituent particles — quarks and gluons — that arise as excitations of underlying quantum fields. This theory successfully predicts a wide range of phenomena that occur at different energy scales, ranging from the high-energy collisions at the Large Hadron Collider to the properties and interactions of composite particles such as the proton and neutron, as well as the nuclei they form. At high energies, the interactions between quarks and gluons are weak, and accurate QCD calculations can be made using a perturbative expansion, which is often represented with Feynman diagrams. At the lower energies relevant for much of nuclear physics, the interactions between quarks and gluons are strong, and the perturbative approach breaks down. In this regime, quantitative predictions can be achieved through a computational approach known as lattice QCD, in which the quark and gluon fields are represented on a discrete spacetime lattice. Many key aspects of nuclear physics can be computed precisely in this framework. For example, such calculations reveal how the masses of the proton and neutron arise from the fundamental quarks and gluons[1], and they have been used to make predictions of the masses of new composite particles later discovered by experiments at CERN[2–4]. However, the reach of this approach is limited by its computational cost, and controlled first-principles QCD calculations of nuclear structure and reactions, for example, would require a scale of computational resources that is currently infeasible[5]. Without breakthrough developments, many important studies will remain impossible even with the world's next generation of exascale supercomputers (quintillions (10^{18}) of operations per second, or the equivalent of 50 million laptops working in concert). If the computational cost of lattice field theory can be greatly reduced, fundamental questions in particle, nuclear and condensed matter physics will be answered. For example, first-principles calculations can probe the fine-tunings in nuclear physics that are deeply important for understanding our existence, by revealing how sensitive the production of carbon in the Universe via the triple-α process is to the free parameters of the theory, explaining why protons and neutrons cluster inside nuclei, and elucidating how the lightest elements formed in the first minutes of the Universe's existence via Big Bang nucleosynthesis[6].

Calculations in lattice QCD are cast in the form of statistical averages with respect to a distribution of quark and gluon field configurations. A major component of the computational cost of lattice QCD calculations is the estimation of these averages by Monte Carlo sampling techniques. (Sampling is one of several computationally intensive steps in lattice QCD calculations. Others, such as the inversion of Dirac operators for the calculation of physical observables, may also be accelerated using machine-learning (ML) approaches[7–9].) Sampling representative configurations of a system to quantitatively evaluate its properties is ubiquitous in physics, being used in fields spanning from ab initio molecular dynamics and statistical physics to astrophysics, and many others. However, sampling from the highly structured, high-dimensional and multimodal distribution of configurations in lattice QCD presents an extraordinarily difficult computational challenge. This problem has historically been the impetus for the development of what have become foundational techniques in computational statistics and high-performance computing, with far-reaching implications within and beyond physics. For example, both the classic Metropolis–Hastings Markov chain Monte Carlo algorithm[10] and Hamiltonian/hybrid Monte Carlo (HMC)[11] were first developed in the context of theoretical nuclear physics, with the latter conceived specifically for lattice QCD. Similarly, the IBM Blue Gene series of supercomputers trace their origins back to the QCDOC (quantum chromodynamics on a chip) computer built specifically for this particular application[12].

The rapid advance of ML over the past few years has spurred the emergence of a new class of algorithms that are revolutionizing computing for both science and industry applications. For example, the extraordinary success of the ML tool AlphaFold[13] in protein folding took the world of biology by surprise, redefining the pace of progress in a field where algorithmic developments had been slow for decades. For lattice QCD, which has historically driven a virtuous cycle of innovations in scientific computing, these advances promise a new chapter. In particular, the rise of generative modelling with ML[14,15] suggests the particular application of sampling algorithms for lattice QCD. The sampling problem in lattice QCD has several key features that present both challenges and opportunities to ML. On the one hand, any algorithm must be asymptotically exact, preventing the direct application of certain generative ML approaches such as generative adversarial networks or variational autoencoders (VAEs). A practical challenge is also presented by the extreme scale of lattice QCD samples used in state-of-the-art calculations, each of the order of several terabytes at the current time. On the other hand, the forms of the relevant probability distributions are exactly known, which can inform the design and training of sampling architectures. In particular, these distributions are invariant under complicated and high-dimensional symmetry groups which significantly reduce the dimensionality and complexity of the problem if they can be incorporated exactly. Although it has required considerable effort to develop ML models that incorporate the symmetries of lattice QCD into ML architectures, the investment has paid dividends in the efficacy of the resulting algorithms.

This Perspective reviews the unique requirements and features of a class of ML-based sampling strategies that have been recently developed for lattice QCD applications and places these developments in the broader context of ML for sampling in scientific domains. Although this endeavour remains in its early stages, it is already clear that it has considerable potential, not only to emulate the transformative impact that ML has had in applications such as AlphaFold[13], but also to spur the advancement of ML itself.

Lattice QCD and the sampling problem

The lattice method for computing physical observables in quantum field theories such as QCD proceeds by discretizing space and time onto a four-dimensional grid (or 'lattice'), with spacing a between neighbouring points and a finite volume V. In this framework, the fundamental particle degrees of freedom of the theory — quarks and gluons in QCD — are represented through 'quantum fields' that consist of complex numbers, vectors or matrices associated with the points and edges (or 'links') of the lattice. Quantities of physical interest are then defined by integrals over these field degrees of freedom, and the continuum, infinite-volume theory is recovered by taking the limit $a \to 0$, $V \to \infty$.

A general physical observable can be defined in terms of quantum 'operators' $\mathcal{O}$ and computed as a statistical expectation value[16]:

$$\langle \mathcal{O} \rangle = \int \mathcal{D}\Phi \ \mathcal{O}[\Phi] p[\Phi], \text{ where } p[\Phi] = \mathrm{e}^{-S[\Phi]}/Z. \quad (1)$$

Here the notation $\int \mathcal{D}\Phi$ schematically indicates integration over all configurations of the discretized quantum fields collectively denoted by Φ, and $Z = \int \mathcal{D}\Phi e^{-S[\Phi]}$ is a normalizing constant. The 'action' $S[\Phi]$ encodes the dynamics of the theory by defining the statistical distribution $p[\Phi]$; in QCD, it describes the fluctuations and interactions of the quark and gluon fields. The operator $\mathcal{O}$ can be chosen to study various physical properties of the theory; for example, the mass of the proton can be calculated using an operator that represents the interaction of two up quarks and one down quark.

In practice, the integral in equation (1) cannot be computed analytically and is instead evaluated by Monte Carlo integration, that is, using an ensemble of N field configurations $\{\Phi_1,\ldots,\Phi_N\}$ sampled from the distribution $p[\Phi]$. Physical quantities are then computed as $\langle\mathcal{O}\rangle \approx \frac{1}{N}\sum_{i=1}^{N}\mathcal{O}[\Phi_i]$ with an uncertainty that is systematically improvable by taking N large. The first step of any lattice field theory calculation is thus a sampling problem. Although the challenge of generating lattice field configurations is reminiscent of sampling problems in many other fields, the structure of the quantum fields, the complicated symmetries of the distribution $p[\Phi]$ and the sheer scale of the required calculations set this apart as a particularly difficult computational problem.

Structure and symmetries of field configurations

In typical lattice quantum field theories, the discretized quantum fields not only extend over the spacetime lattice, but also have 'internal' degrees of freedom represented mathematically by a vector or matrix structure at each point or edge of the lattice. In particular, in QCD the gluon field U is encoded by SU(3) variables — 3×3 complex unitary, unit-determinant matrices — on each edge of the lattice, whereas the quark fields Ψ are encoded by 4×3 complex matrices on each site of the lattice, as shown in Fig. 1. For QCD, the calculation of a physical observable via equation (1) can thus be expressed as

$$\begin{aligned}\langle\mathcal{O}\rangle &= \frac{1}{Z}\int \mathcal{D}U\mathcal{D}\overline{\Psi}\mathcal{D}\Psi\,\mathcal{O}[U,\overline{\Psi},\Psi]e^{-S[U,\overline{\Psi},\Psi]}\\ &= \frac{1}{Z}\int \mathcal{D}U\mathcal{O}'[U]e^{-S_{\text{eff}}[U]},\\ \text{where } Z &= \int \mathcal{D}U\mathcal{D}\overline{\Psi}\mathcal{D}\Psi e^{-S[U,\overline{\Psi},\Psi]}.\end{aligned} \tag{2}$$

Here the notation $\int \mathcal{D}U$ indicates integration over all values of the discretized gluon field U, whereas the integration $\int \mathcal{D}\overline{\Psi}\mathcal{D}\Psi$ over all values of the discretized quark fields are Gaussian integrals that are evaluated analytically, yielding a modified operator $\mathcal{O}'$ and the modified weight $p[U] = e^{-S_{\text{eff}}[U]}/Z$ over gluon field configurations. (In particular, the integral $\int \mathcal{D}\overline{\Psi}\mathcal{D}\Psi$ is a Berezin integral[17] over elements of a Grassmann algebra, which must be analytically treated to produce an integral amenable to numerical evaluation.) In practice, auxiliary degrees of freedom known as 'pseudo-fermions'[18] are also typically introduced as stochastic estimators for determinants appearing in $p[U] = \exp(-S_{\text{eff}}[U])/Z$. State-of-the-art lattice QCD calculations involve fields of size up to $256^3 \times 512 \approx 8.6$ billion lattice sites with quantum fields represented by roughly 50 degrees of freedom per lattice site (this counting includes four SU(3) matrices for each lattice site, yielding $4 \times 8 = 32$ degrees of freedom, as well as complex 4×3 matrices with $2 \times 3 \times 4 = 24$ degrees of freedom for each site, arising from the pseudo-fermion fields), meaning that, in practice, calculations involve Monte Carlo integration over as many as 10^{12} variables.

Symmetries in a lattice field theory manifest as transformations of field configurations that leave the probability density $p[U]$ and the integration measure invariant. The action, and hence p, is typically invariant under both discrete geometric symmetries of the hypercubic Euclidean spacetime, such as discrete translations, rotations and reflections, and under internal symmetry transformations. For example, one contribution to the lattice QCD action is given by

$$S_g[U] = -\frac{\beta}{6}\sum_{x}\sum_{\substack{\mu=1\\ \nu=1}}^{4}\text{ReTr}[U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^\dagger(x+\hat{\nu})U_\nu^\dagger(x)], \tag{3}$$

where β is a parameter of the theory that is related to the lattice spacing a, x is summed over the sites of the discretized lattice, and $\hat{\mu}$, $\hat{\nu}$ indicate vectors of length a in the μ and ν directions, respectively (see Fig. 1). From this expression, it can be seen how 'gauge' symmetry is manifest in QCD, as $p[U]$ is invariant under the transformation of the gauge field U according to

$$U_\mu(x) \to \Omega(x)U_\mu(x)\Omega^\dagger(x+\hat{\mu}) \tag{4}$$

for all possible choices of $\Omega(x) \in \text{SU}(3)$ over all lattice sites. Because this symmetry is specified by one SU(3)-valued matrix per lattice site (so eight degrees of freedom per site), the symmetry group may have a dimension as large as 10^{11} in state-of-the-art calculations.

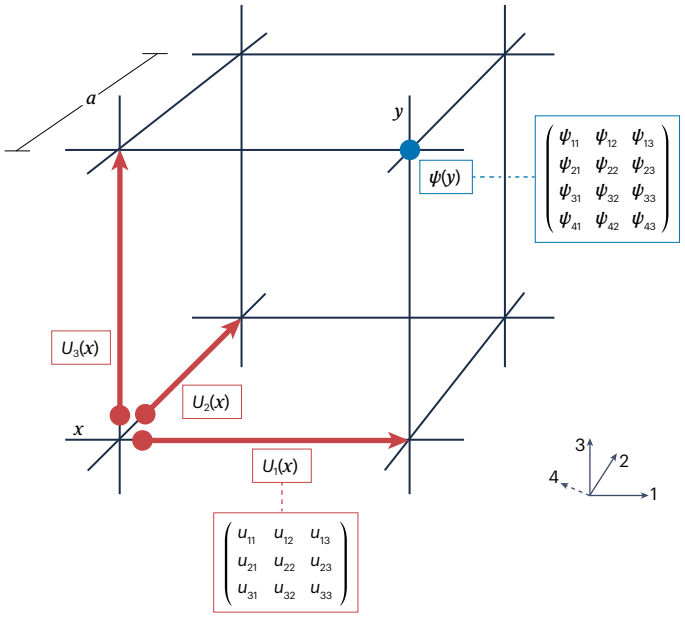

Fig. 1 | Depiction of a single cube within the spacetime lattice of a lattice QCD calculation. Shown are some elements $U_\mu(x)$ of the discretized gluon field (red), each associated with an edge $(x, x+\hat{\mu})$ from site x to the neighbouring site in direction $\mu \in \{1,2,3,4\}$, and an element $\Psi(y)$ of the discretized quark field (blue), associated with a site y. The value of each $U_\mu(x)$ is a complex unitary 3×3 matrix with determinant 1, that is, an SU(3) matrix, and each $\Psi(y)$ is a 4×3 complex matrix. a is the lattice spacing between neighbouring points. The fourth dimension of the lattice is suppressed for clarity.

Approaches and challenges to sampling field configurations

Conventionally, the generation of an ensemble of lattice fields distributed according to $p[\Phi]$ is performed iteratively using a Markov

process, in which a chain of configurations $\{\Phi_1,\Phi_2,\ldots\}$ is generated by a sequence of stochastic updates beginning from an initial configuration Φ_0. In particular, the HMC algorithm was first conceived of in the 1980s specifically for this application in lattice field theory[11] and has since become a mainstay of the computational science community. In this paradigm, the rapid exploration of the state space is achieved by a directed evolution from each configuration to a new proposed configuration, which avoids an inefficient random walk. Exactness of the distribution is guaranteed by applying the Metropolis–Hastings procedure to accept the proposed configuration with an appropriate probability[10,19] (see also the next section).

Despite the outstanding success of this approach — which remains the workhorse of lattice field theory — generating ensembles of field configurations is one of the notable computational costs of first-principles QCD calculations. In particular, because the approach evolves configurations via a local dynamical process, increasingly many updates are required to decorrelate samples on physical length scales as the continuum limit is approached ($a \to 0$). This is a manifestation of the phenomenon known as 'critical slowing-down' in this context[11,20]. Simultaneously, the distribution of QCD gauge fields spans topologically distinct sectors, and Markov-based sampling algorithms such as HMC can become 'trapped' or 'frozen' in sectors of fixed topology.

Any alternative approach to sampling lattice field configurations will need to satisfy several key requirements in order to be practically viable. Most importantly, it must be statistically improvable: that is, the true probability distribution $p[\Phi]$, including the various symmetries that this distribution respects, must be recovered in the limit of a large number of samples. Furthermore, the approach must be efficiently scalable to state-of-the-art lattice field theory studies, which involve field configurations as large as many terabytes of memory each, with as many as 10^{12} degrees of freedom. Finally, the approach must improve on the impressive success of the HMC framework and mitigate the challenges of critical slowing-down and topological freezing in some regime of physical interest.

ML for sampling lattice field configurations

An ML approach to sampling lattice field configurations is an appealing proposition: it offers a new paradigm of algorithms that are optimized specifically for the task at hand. Enabling radically different approaches to sampling, ML may mitigate critical slowing-down and other key challenges faced by traditional Markov-process algorithms such as HMC. Even for cases where HMC works well, ML may still provide advantages, for example by enabling embarrassingly parallel, rather than sequential, sampling, or by learning to approximate a large number of computational steps of traditional algorithms with fewer operations, as is observed in other fields such as ML approximations of partial differential equation solvers[21].

However, the application of ML to lattice field theory is not straightforward, given the previously highlighted features of the lattice field theory problem. In particular, for the theory of QCD, gauge field samples are collections of matrices that are constrained to be SU(3) matrices (see Fig. 1), whereas samples in typical ML domains such as images or natural language models are represented by vectors of unconstrained real numbers. When considering generative models based on diffeomorphisms, as discussed below, it is imperative to ensure that these constraints are satisfied by all transformations. The associated SU(3) gauge symmetry is also very atypical in comparison with usual ML applications, although other symmetries such as 4D translations may be handled by traditional ML methods such as convolutional neural networks. Moreover, generative models for traditional applications such as images and language do not require asymptotic guarantees of exactness in sampling, whereas these are critical in the lattice field theory context. Finally, the sheer scale of state-of-the-art lattice QCD calculations, both in terms of the scale of lattice samples and the computational cost required to manipulate them, presents a challenge to ML approaches. Figure 2 illustrates these stark contrasts between the lattice field generation problem and other sampling tasks that have been revolutionized through ML, such as image generation. Clearly, achieving state-of-the-art sampling performance with new ML algorithms in the context of lattice field theory will require the development of new algorithms and innovation in ML.

Classes of generative models for sampling in lattice field theory

ML models designed to (approximately) sample from a target density are known as generative or probabilistic models. A generative model typically consists of three components: a space of latent or hidden variables equipped with a density, a set of observed or target variables, and a parametric map that transforms points in the latent space into points in the target space. Optimization is performed on the parametric map so that the density it induces in the target space approximates the target density. A wide variety of ML-based generative architectures have been developed over the past decade, with transformative successes particularly evident in applications to sound/image data[22–25] and language data[14,15,26–30].

One notable difference between these applications and the challenge of sampling field configurations for lattice QCD is that the true distribution over the space of images, sounds or text is not known, so the model distribution is learnt from data samples. For lattice field configurations, not only is the unnormalized target distribution known, but it must be sampled from with asymptotic guarantees of exactness. This can be achieved with ML models if they feature tractable likelihoods (the model probability density can be computed for any given sample); in this case, they can be embedded inside sampling algorithms with asymptotic guarantees, such as a Markov chain, as discussed further below.

A tractable likelihood also allows one to optimize an ML model by minimizing a probability divergence $D(q_\theta; p)$ between the model probability density $q_\theta[U]$ parameterized by θ and the known target probability density $p[U] = e^{-S_{\text{eff}}[U]}/Z$. However, in stark contrast to typical ML sampling problems, training models for lattice QCD sampling requires estimating the gradients $\nabla_\theta D(q_\theta; p)$ using only samples from the model or perhaps only a small number of 'ground truth' data samples. This restricts the family of probability divergences that can be used; f-divergences such as the Kullback–Leibler divergence[31] are commonly used, as they can be expressed as an expectation value under q_θ, allowing the divergence and its gradient to be estimated from model samples alone.

Any ML approach to sampling for lattice QCD must also be 'scalable' as the number of lattice sites, $M = V/a^4$, is increased. Ideally, its computational and memory costs should scale linearly or sublinearly with the number of lattice sites, which we denote by $O(M)$ below. This applies to all aspects of the model: drawing samples, evaluating the likelihood q_θ, and evaluating the gradients $\nabla_\theta D(q_\theta; p)$. This consideration restricts or rules out certain classes of models, as discussed below.

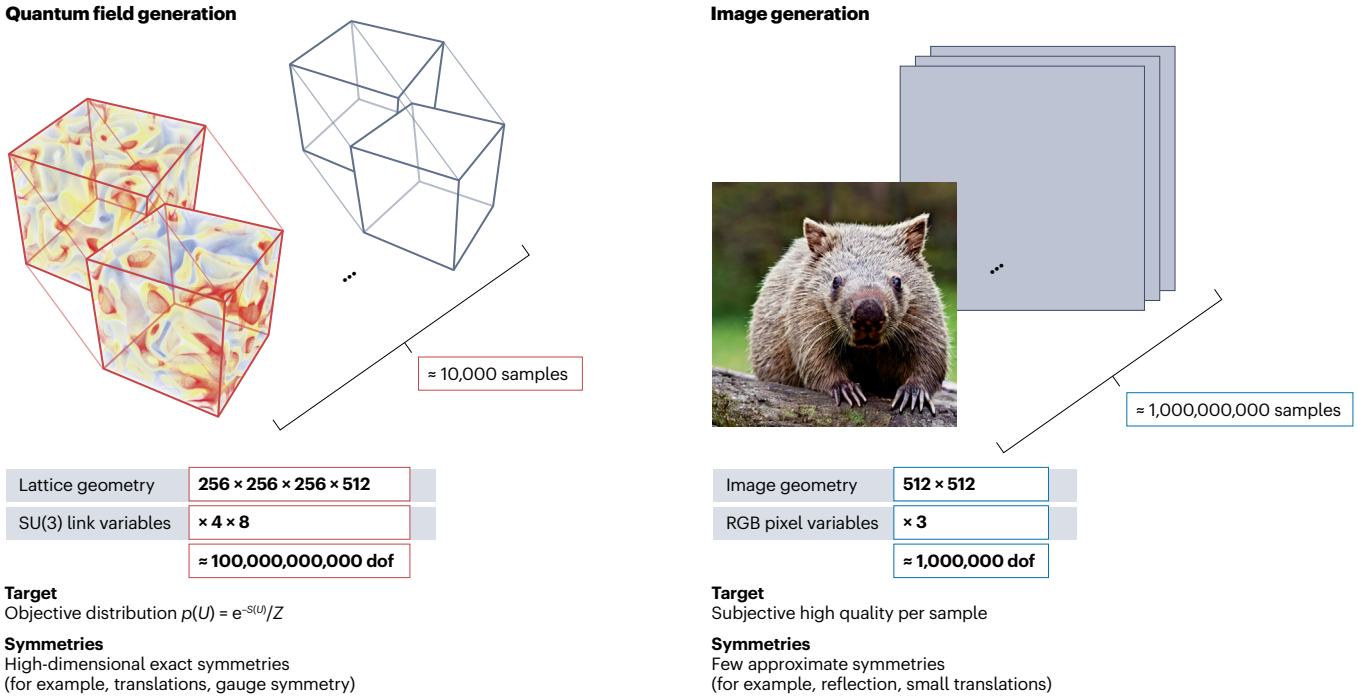

Fig. 2 | Comparison between the sampling tasks of quantum field generation for lattice quantum chromodynamics and image generation. In addition to differences in the target and symmetries of the problems, the hierarchy of degrees of freedom (dof) per sample to number of samples is inverted for quantum field generation as compared with image generation. The action S encodes the dynamics of the theory by defining the statistical distribution p, U is the gluon field, and Z is a normalizing constant. The image on the right side is reprinted from Kaggle (https://www.kaggle.com/datasets/vitaliykinakh/stable-imagenet1k) under a Creative Commons licence CC0 1.0).

The features of various generative modelling frameworks that could be considered for the lattice QCD sampling problem are outlined below.

- **Latent-variable models** such as generative adversarial networks[32] and VAEs[33,34] typically have efficient $O(M)$ sampling, but intractable likelihoods (involving marginalization of latent variables).
- **Autoregressive models**[14,15,22,26–30,35] typically have efficient $O(M)$ likelihood evaluation. Sampling can also be achieved with $O(M)$ cost in principle, but existing implementations are impractically slow.
- **Continuous-time models** include diffusion models[23,24] defined via stochastic differential equations and continuous-time normalizing flows[36] defined via ordinary differential equations. In these models, likelihood computation requires integrating a scalar ordinary differential equation defined by the divergence of a vector field (the marginal score function). This computation typically has computational cost $O(M^3)$ unless additional structure is forced on to the model[37].
- **Discrete-time normalizing flow models**[38–43] remain as good candidate models. In a discrete flow, the generative process maps a latent vector z (a field configuration in the lattice field theory context) sampled from a base density into the target density via the composition of a series of parametric diffeomorphisms $F_1, F_2, \ldots, F_n$. If z is sampled with density $r(z)$, then a flow sample $x = F(z)$ has known density $q(x) = r(z)|\det \partial F/\partial z|^{-1}$, where $F = F_1 \circ \ldots \circ F_n$ is the composed diffeomorphism. By restricting to F_i for which $\partial F_i/\partial z$ is a triangular matrix, the cost of $|\det \partial F/\partial z|$ is only $O(M)$ (ref. 38).

These models, however, also have intrinsic limitations that must be worked around, such as topology preservation of the diffeomorphism[44] and difficulty in modelling tail-behaviour of a target density if the tails are not already in the base density[45].

One can also consider sampling in an augmented space, where the data space is augmented with an additional set of auxiliary or latent variables. In this setting, it may be viable to reconsider VAEs[33,34] or VAE–flow hybrids[46], but currently there are no results demonstrating that these methods perform well compared to models working directly on the data space for lattice QCD.

Methods to guarantee asymptotic exactness

Several mechanisms have been proposed to combine generative models with Markov chain Monte Carlo and importance-sampling algorithms in order to inherit their asymptotic convergence guarantees. One of the simplest mechanisms is neural importance sampling, in which a model density $q_\theta(x) \approx p(x)$ is used to evaluate expectations under the target p via $\mathbb{E}_p[\mathcal{O}(x)] = \mathbb{E}_{q_\theta}[\frac{p(x)}{q_\theta(x)}\mathcal{O}(x)]$ (ref. 47). An appealing alternative is to incorporate generative models into an asymptotically exact Markov process, which allows existing analysis techniques to be used or existing Markov chain updates to be combined with the ML sampling approach. Generally, the Metropolis–Hastings algorithm uses an ergodic transition kernel $K(x'|x)$ to propose Markov chain updates $x \to x'$ which are accepted with probability

$$p_{\text{acc}}(x'|x) = \min\left(1, \frac{K(x|x')p(x')}{K(x'|x)p(x)}\right) \quad (5)$$

to ensure that the asymptotic equilibrium distribution is the desired target distribution[10,19]. Although this method is guaranteed to converge[48], the speed of convergence depends on the target density and the choice of transition kernel $K(x'|x)$.

ML models can be combined with the Metropolis–Hastings approach by using generative models to construct the kernel $K(x'|x)$. A direct approach is to use the trained model $q_\theta(x) \approx p(x)$ to produce independent and identically distributed (iid) proposal samples, $K(x'|x) = q_\theta(x')$, with the convergence rate determined by the quality of the model approximation to the target distribution. More advanced techniques include neural transport Monte Carlo[49,50], where Markov chain Monte Carlo is performed in the latent space; learned Monte Carlo proposals[51–64], where the goal is to directly learn the kernel $K(x'|x)$; learned sequential Monte Carlo sampling[46,65–67], which combines deterministic flows with sequential Monte Carlo annealing techniques; Monte Carlo variance minimization[47,68,69]; and stochastic normalizing flows[46], which interleave deterministic transforms with latent-variable VAE-like components and Markov chain Monte Carlo transforms. HMC is a particularly successful family of kernels and can also be combined with learnt components. In the presence of pseudo-fermions, simulating the Hamiltonian dynamics for HMC requires an expensive computation of 'force terms', which is another area where acceleration may be possible using ML[62,70–73].

Incorporating manifold constraints and gauge symmetry in ML models

When the target density for sampling features an exact or approximate symmetry, breaking that symmetry in a sampling algorithm will result in computational inefficiencies. In particular, continuous exact symmetries naturally reduce the effective dimensionality of the target distribution, such that incorporating them directly reduces the difficulty of modelling that target. In addition, training a model with the symmetry explicitly encoded modifies the structure of the loss landscape in a way that may make training feasible in cases where attempting to approximately 'learn' the symmetry may not be. Moreover, if guarantees of exactness are required, any remaining symmetry-breaking after training will result in additional costs to correct that breaking via the approaches discussed in the previous subsection. It is thus often advantageous, and in some cases critical, to incorporate symmetries into ML architectures and/or training approaches.

Modern ML offers many techniques that seek to take advantage of known symmetries. The simplest such method is data augmentation[74], where the available training set is 'augmented' with randomly transformed input/output pairs. Another technique consists of explicitly adding a term to the optimization target that encourages equivariance or invariance with respect to a group of transformations[75,76]. Both approaches only serve to assist the training, and symmetries still have to be learnt by the model architecture. Alternatively, it is possible to construct architectures such that they respect known symmetries by construction; standard convolutional neural networks, for example, are equivariant maps with respect to translations. The most commonly studied symmetries are finite groups[77] and SE(3), the group of isometries of 3D Euclidean space[78,79]. Naturally, the larger the symmetry group, the harder it is to learn the symmetries, either via data augmentation or via additional optimization targets. In the case of lattice QCD, the gauge symmetry has a prohibitively large dimension scaling with the number of lattice sites, hence it is likely essential to build gauge-equivariant and invariant neural networks for this application. For example, with current architectures and training approaches it has been demonstrated that it is essential to exactly incorporate gauge symmetry[80], but not translations and hypercubic transformations[81], for successful training of flow-based sampling algorithms for lattice field theory.

Building generative models that exactly incorporate the symmetry constraints of lattice QCD is a non-trivial task that has required the introduction of several new ML models to treat both gauge and pseudo-fermionic degrees of freedom[82–90]. This approach relies on the observation that starting from a base distribution that is gauge-invariant (such as the Haar measure on SU(N)) and applying a gauge-equivariant diffeomorphism to this base density yields a new density that is also gauge-invariant[38,76,84,91]. One can reduce the problem of building gauge-equivariant diffeomorphisms on the gauge degrees of freedom situated on the edges of a lattice to the problem of building matrix-conjugation-equivariant diffeomorphisms[80,81] on SU(3), which is a simpler problem. As described in equation (4), a gauge symmetry transformation T_Ω is parameterized by a field of SU(3) variables $\Omega(x)$ and acts on an edge, or 'link', variable as $T_\Omega U_\mu(x) = \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$. As a consequence, a product of link variables along a closed loop $\Lambda(x)$ starting and ending at a point x transforms as $\Lambda(x) \to \Omega(x)\Lambda(x)\Omega^\dagger(x)$. Mathematically this is referred to as a matrix-conjugation, or adjoint transformation. If $U_\mu(x)$ is a link, $\Gamma(x, x+\hat{\mu})$ is a product of links along an open path from x to $x+\hat{\mu}$ that does not contain $U_\mu(x)$, and g: SU(3) → SU(3) is a conjugation-equivariant diffeomorphism, then a gauge-equivariant diffeomorphism f can be constructed as

$$f(U_\mu(x)) = g(U_\mu(x)\Gamma^\dagger(x, x+\hat{\mu}))\Gamma(x, x+\hat{\mu}), \tag{6}$$

as shown in Fig. 3. The diffeomorphism f then manifestly satisfies gauge equivariance, $f(T_\Omega U) = T_\Omega f(U)$. Building gauge-equivariant diffeomorphisms for pseudo-fermion degrees of freedom is also possible using extensions of this approach based on parallel transport[92].

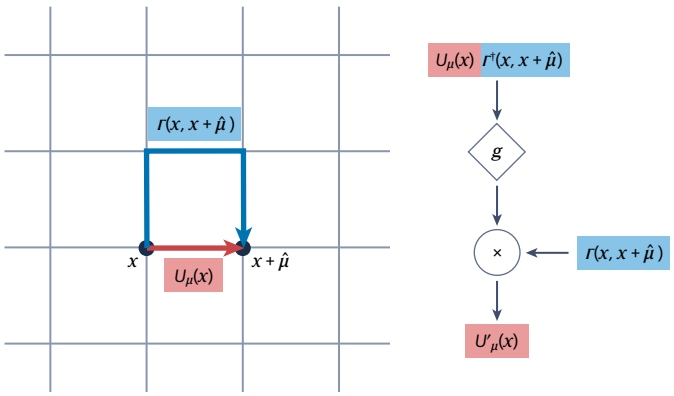

Fig. 3 | Illustration of a gauge-equivariant transformation layer[80,81]. Here, a parametric gauge-equivariant diffeomorphism is constructed from a diffeomorphism g satisfying $g(\Omega U \Omega^\dagger) = \Omega g(U) \Omega^\dagger$ (matrix-conjugation equivariance). To update the matrix-valued link $U_\mu(x)$ in red, this transformation first updates the plaquette $U_\mu(x)\Gamma^\dagger(x, x+\hat{\mu})$ containing that link, before 'pushing' that update onto the link by assigning $U'_\mu(x) = g(U_\mu(x)\Gamma^\dagger(x, x+\hat{\mu}))\Gamma(x, x+\hat{\mu})$. The output link $U'_\mu(x)$ transforms appropriately under gauge transformations when $U_\mu(x)$ and $\Gamma(x, x+\hat{\mu})$ are transformed. See the main text for further details.

A roadmap for ML-based sampling in lattice QCD

The great potential of ML-based sampling for lattice field theories has inspired rapid developments that have already demonstrated profound

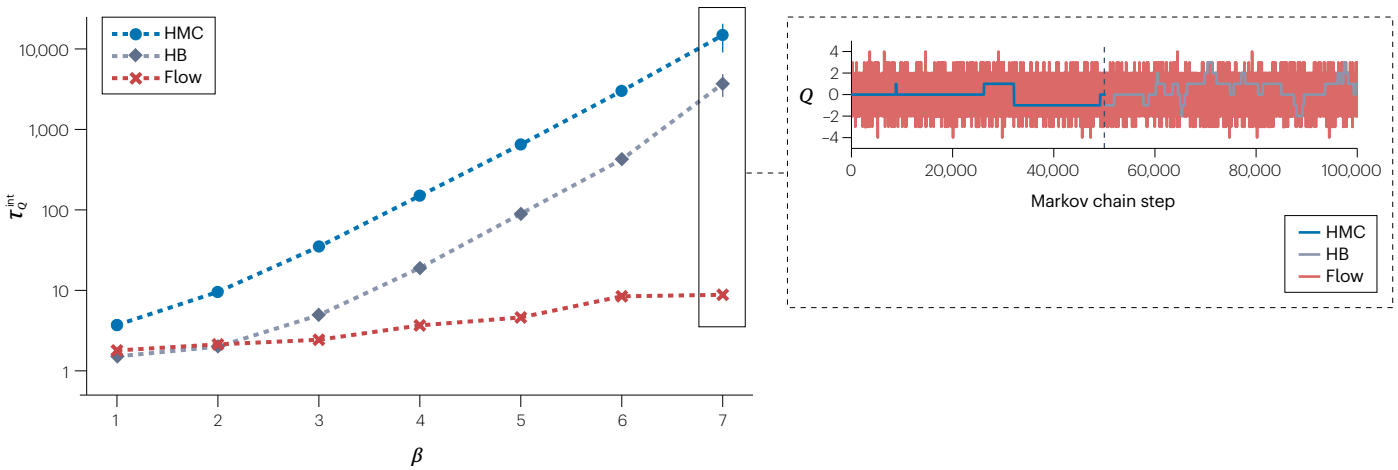

Fig. 4 | Demonstration of the advantages of flow-based sampling in a U(1) lattice gauge theory in two spacetime dimensions[83]. The inset shows the rapid mixing of topological charge Q when sampling with normalizing flows, compared with Hamiltonian/hybrid Monte Carlo (HMC) and heat bath (HB) algorithms for the action defined by $\beta = 7$ (see equation (3) for the definition of the analogous parameter in quantum chromodynamics). The main graph shows the asymptotically improved scaling of τ_Q^{int} towards the continuum limit $\beta \to \infty$, where τ_Q^{int} is the 'integrated autocorrelation time' of the topological charge, which is a measure of cost in Markov-process sampling and can be interpreted here as a metric for critical slowing-down. Reproduced with permission from ref. 83, APS.

successes. Following early applications of flow-based sampling to field theories other than QCD[80,89], the approach was developed for theories in two spacetime dimensions, specifically for SU(3) gauge fields without dynamical quark (fermionic) degrees of freedom[81] and for U(1) gauge fields with dynamical fermions[93]. Combining these advances enabled the first application of flow models to sampling QCD in 4D[94], albeit with small spacetime volumes. ML-accelerated updating schemes have been developed, again for small volumes and with the SU(2) gauge group instead of SU(3) (ref. 62), and continuous-time models inspired by previous work in the lattice field theory community[95,96] have been applied to simple lattice field theories[97] and both U(1) gauge theory[82] and SU(3) gauge theory[98] in 2D. These approaches have had astounding success; Fig. 4 illustrates the advantages of flow-based sampling in one particular toy theory, but the conclusion that ML-accelerated sampling schemes can overcome the critical slowing-down and topological freezing challenges faced by HMC has been clear and universal. It is important to emphasize that this success includes theories with fermionic degrees of freedom, where ML-accelerated sampling schemes have been developed to integrate with the usual approach of pseudo-fermions[92,94,99]. However, the crucial aspect missing in all applications so far is a demonstration of the effectiveness of ML techniques at the scale of state-of-the-art lattice QCD calculations in nuclear and particle physics.

We expect that not only will we soon see applications of ML-accelerated sampling to lattice field theory at scale, but that running at scale is a key ingredient necessary to realize the full potential of ML in this context. In particular, we anticipate that the first impact of ML for this application will be that, once the potentially high cost of training is paid, ML-based sampling will be orders of magnitudes faster than traditional HMC, mitigating critical slowing-down, overcoming topological freezing, and opening the door to a sampling regime where this training cost can be efficiently amortized, as depicted in Fig. 5. At precisely which scale this advantage will be reached is not yet clear; the computational cost of training ML models in this context may vary by orders of magnitude between different architectures and training approaches[100]. As the optimal approach to model parameterization and training can depend sensitively on the number of samples which are ultimately required, the balance of training and sampling costs is highly problem-dependent, and the regime in which flow-based sampling outperforms HMC for lattice QCD applications will depend on precisely how the flow models are used (and reused). It is already evident, however, that achieving this paradigm of efficient ML-accelerated sampling will require considerable investment; it is clear that in the field of generative ML as a whole, the substantial progress in text[14,15,26–30] and image modelling[22–25] has required pushing the boundaries of model size. So far, the generative ML experiments for lattice field theory are of comparatively small scale, despite the target scale of the problem itself being comparable to, or even larger than, applications in these domains. Success will thus probably require model scales, and corresponding investments in upfront training, that will constitute a change in paradigm of computational resource use for the theoretical physics community. Nevertheless, as a fundamentally structured problem, we anticipate that scaling the custom ML solutions developed for lattice field theory to large models will pay dividends and that lattice QCD will join the list of scientific problems that have seen important impacts from ML at state-of-the-art scale, such as low-dimension Bayesian parameter inference for astrophysics[101], quantum Monte Carlo[102] and protein folding[13].

Beyond the anticipated impacts of mitigating sampling challenges such as critical slowing-down and topological freezing, ML models have the potential to catalyse other paradigm shifts in lattice field theory. For example, they naturally offer new opportunities for community resource sharing. Ensembles of lattice field configurations are large enough in size (petabytes for state-of-the-art ensembles) that they cannot be easily shared, and massive investments in tape resources are made to store them. In contrast, even the largest ML models contain only a few terabytes of parameters. These can easily be shared, allowing research groups around the globe to efficiently generate their own configurations or reproduce ensembles from a known seed, in both cases capitalizing on community-owned pretrained models. Another important opportunity is that ML-based samplers can be conditioned

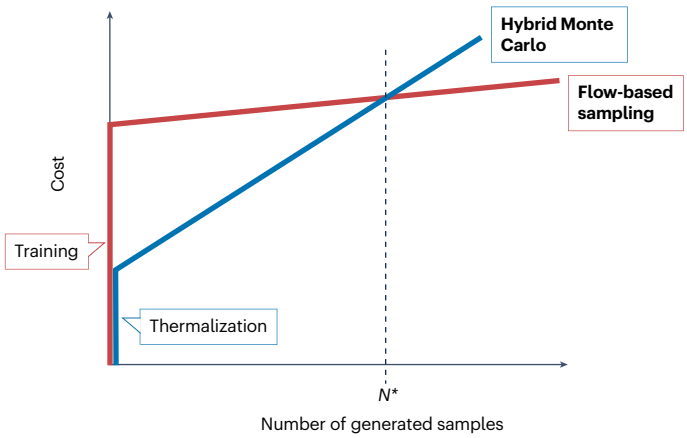

Fig. 5 | Sketch of the upfront and sampling costs of hybrid Monte Carlo compared with flow-based models. Hybrid Monte Carlo (HMC) shown in blue; flow-based sampling in red. Whereas machine-learning models require a potentially large upfront computational cost for training, they offer the hope of efficient sampling, such that this training cost can be amortized. That is, for sufficiently many samples, flow-based sampling that is more efficient than HMC (for example, as in Fig. 4) can outperform HMC in terms of total cost. The minimum number of samples for which flow-based sampling is more efficient than HMC is denoted by N^*.

on various parameters of the theory, from the lattice spacing and volume to physical parameters such as the strength of coupling of the fundamental particles of the theory. The potential to generate 'correlated' sets of samples at different parameters, interpolated, or even extrapolated[103,104], from the parameters used during training, is qualitatively distinct from what is possible using traditional sampling algorithms such as HMC (in principle this is also possible for pre-selected parameter sets, using approaches such as parallel tempering) and offers new parameter extrapolation methods. Ultimately, one could even imagine more general ways of conditioning these models (for example, on a symbolic description of the target action), enabling new approaches such as direct measurements of the effect of modifying the action on physical observables. As such, ML techniques hold the promise of redefining conventional wisdom in this field, with implications that have yet to be fully explored.

Outlook

As ML continues to evolve in scope and complexity, its applications in science are being driven into two broad categories: those that can adapt existing ML technologies (usually developed to model images, sound and text) and those that demand ground-up development and inspire innovation. Lattice field theory is becoming established as a prime example of the latter, being simultaneously an important science application in which algorithmic acceleration will have wide-reaching implications for fundamental physics and a massive computational challenge of a scope and scale that has driven advances in computation and algorithms for decades.

In particular, the challenge of sampling lattice field configurations for nuclear and particle physics calculations is a definitive proving-ground for generative ML models in science. With strict requirements of asymptotic exactness, and an ultimate scale at which each sample is several terabytes in size, it is clear that symmetries, structure and domain knowledge must be incorporated into ML architectures designed for this task. In that regard, the application of ML to sampling in lattice field theory is a key exemplar informing the debate that is currently underway in the ML community around incorporating domain or expert knowledge: whereas ardent supporters of modern deep learning often argue against the long-term value of incorporating such knowledge (the 'bitter lesson' theory[105], advocated by Richard Sutton), in the physical sciences, and in particular in theoretical physics calculations, there are often precise mathematical formulations of domain knowledge that not only make little sense to ignore, but are intractable to learn from data. Adding to the complexity of this engineering and design challenge, existing algorithmic benchmarks for sampling in lattice field theory are extremely high; new ML samplers must compete against well-established algorithms that have been optimized in co-design with high-performance computing systems for more than four decades.

As such, ML for lattice field theory is not only a benchmark for the application of ML in science, but it is also an endeavour with paradigm-shifting potential for physics. If the success already achieved in sampling field theories at toy scales can be mimicked at state-of-the-art scales, it will transform the computational landscape of a field that is one of the largest consumers of open-science supercomputing (computing available to public scientific applications) worldwide, with impacts across particle, nuclear and condensed matter physics and beyond. On the ML side, it will be a flagship example of the power of sophisticated domain-specific customization and engineering to achieve transformative impact in computational science.

Published online: 4 August 2023

References

1. Borsanyi, S. et al. Ab initio calculation of the neutron–proton mass difference. *Science* **347**, 1452–1455 (2015).
2. Brown, Z. S., Detmold, W., Meinel, S. & Orginos, K. Charmed bottom baryon spectroscopy from lattice QCD. *Phys. Rev. D* **90**, 094507 (2014).
3. Aaij, R. et al. Observation of two new Ξ_b^- baryon resonances. *Phys. Rev. Lett.* **114**, 062004 (2015).
4. Aaij, R. et al. Observation of the doubly charmed baryon Ξ_{cc}^{++}. *Phys. Rev. Lett.* **119**, 112001 (2017).
5. Joó, B. et al. Status and future perspectives for lattice gauge theory calculations to the exascale and beyond. *Eur. Phys. J. A* **55**, 199 (2019).
6. Detmold, W. et al. Hadrons and nuclei. *Eur. Phys. J. A* **55**, 193 (2019).
7. Calì, S., Hackett, D. C., Lin, Y., Shanahan, P. E. & Xiao, B. Neural-network preconditioners for solving the Dirac equation in lattice gauge theory. *Phys. Rev. D* **107**, 034508 (2023).
8. Lehner, C. & Wettig, T. Gauge-equivariant pooling layers for preconditioners in lattice QCD. Preprint at https://arxiv.org/abs/2304.10438 (2023).
9. Lehner, C. & Wettig, T. Gauge-equivariant neural networks as preconditioners in lattice QCD. Preprint at https://arxiv.org/abs/2302.05419 (2023).
10. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
11. Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **195**, 216–222 (1987).
12. Chen, D. et al. QCDOC: a 10-teraflops scale computer for lattice QCD. *Nucl. Phys. B Proc. Suppl.* **94**, 825–832 (2001).
13. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
14. Hoffmann, J. et al. Training compute-optimal large language models. Preprint at https://arxiv.org/abs/2203.15556 (2022).
15. Thoppilan, R. et al. Lamda: Language models for dialog applications. Preprint at https://arxiv.org/abs/2201.08239 (2022).
16. Peskin, M. E. & Schroeder, D. V. *An Introduction to Quantum Field Theory* (Addison-Wesley, 1995).
17. Berezin, F. A. The method of second quantization. *Pure Appl. Phys.* **24**, 1–228 (1966).
18. Gattringer, C. & Lang, C. B. *Quantum Chromodynamics on the Lattice* Vol. 788 (Springer, 2010).
19. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).
20. Schaefer, S., Sommer, R. & Virotta, F. Critical slowing down and error analysis in lattice QCD simulations. *Nucl. Phys. B* **845**, 93–119 (2011).
21. Beck, C., Hutzenthaler, M., Jentzen, A. & Kuckuck, B. An overview on deep learning-based approximation methods for partial differential equations. *Discrete Contin. Dyn. Syst. B* **28**, 3697–3746 (2023).

22. Oord, A. v. d. et al. Wavenet: a generative model for raw audio. Preprint at https://arxiv.org/abs/1609.03499 (2016).
23. Dhariwal, P. & Nichol, A. Diffusion models beat GANs on image synthesis. *Adv. Neural Inf. Process. Syst.* **34**, 8780–8794 (2021).
24. Saharia, C. et al. Photorealistic text-to-image diffusion models with deep language understanding. *Adv. Neural Inf. Process. Syst.* **35**, 36479–36494 (2022).
25. Child, R. Very deep VAEs generalize autoregressive models and can outperform them on images. Preprint at https://arxiv.org/abs/2011.10650 (2020).
26. Kaplan, J. et al. Scaling laws for neural language models. Preprint at https://arxiv.org/abs/2001.08361 (2020).
27. Brown, T. et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **33**, 1877–1901 (2020).
28. Lieber, O., Sharir, O., Lenz, B. & Shoham, Y. *Jurassic-1: Technical Details and Evaluation* White Paper (AI21 Labs, 2021).
29. Rae, J. W. et al. Scaling language models: methods, analysis & insights from training gopher. Preprint at https://arxiv.org/abs/2112.11446 (2021).
30. Smith, S. et al. Using DeepSpeed and Megatron to train Megatron-Turing NLG 530B, a large-scale generative language model. Preprint at https://arxiv.org/abs/2201.11990 (2022).
31. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).
32. Goodfellow, I. et al. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27**, 2672–2680 (2014).
33. Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *Proceedings of Machine Learning Research* ***32(2)***, 1278–1286 (2014).
34. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at https://arxiv.org/abs/1312.6114 (2014).
35. Van Oord, A., Kalchbrenner, N. & Kavukcuoglu, K. Pixel recurrent neural networks. *Proceedings of Machine Learning Research* **48**, 1747–1756 (2016).
36. Chen, C. et al. Continuous-time flows for efficient inference and density estimation. *Proceedings of Machine Learning Research* **80**, 824–833 (2018).
37. Chen, R. T. & Duvenaud, D. K. Neural networks with cheap differential operators. *Adv. Neural Inf. Process. Syst.* **32**, 9961–9971 (2019).
38. Papamakarios, G., Nalisnick, E. T., Rezende, D. J., Mohamed, S. & Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* **22**, 1–64 (2021).
39. Rezende, D. & Mohamed, S. Variational inference with normalizing flows. *Proceedings of Machine Learning Research* **37**, 1530–1538 (2015).
40. Tabak, E. G. & Turner, C. V. A family of nonparametric density estimation algorithms. *Commun. Pure Appl. Math.* **66**, 145–164 (2013).
41. Dinh, L., Sohl-Dickstein, J. & Bengio, S. Density estimation using real NVP. In *International Conference on Learning Representations* (ICLR, 2017).
42. Kingma, D. P. & Dhariwal, P. Glow: generative flow with invertible 1x1 convolutions. *Adv. Neural Inf. Process. Syst.* **31**, 10215–10224 (2018).
43. Papamakarios, G., Pavlakou, T. & Murray, I. Masked autoregressive flow for density estimation. *Adv. Neural Inf. Process. Syst.* **30**, 2338–2347 (2017).
44. Huang, C.-W., Dinh, L. & Courville, A. Augmented normalizing flows: bridging the gap between generative flows and latent variable models. Preprint at https://arxiv.org/abs/2002.07101 (2020).
45. Laszkiewicz, M., Lederer, J. & Fischer, A. Marginal tail-adaptive normalizing flows. *Proceedings of Machine Learning Research* **162**, 12020–12048 (2022).
46. Wu, H., Köhler, J. & Noé, F. Stochastic normalizing flows. *Adv. Neural Inf. Process. Syst.* **33**, 5933–5944 (2020).
47. Müller, T., McWilliams, B., Rousselle, F., Gross, M. & Novák, J. Neural importance sampling. *ACM Trans. Graph.* **38**, 1–19 (2019).
48. Robert, C. P., Casella, G. & Casella, G. *Monte Carlo Statistical Methods* Vol. 2 (Springer, 1999).
49. Hoffman, M. et al. Neutralizing bad geometry in Hamiltonian Monte Carlo using neural transport. Preprint at https://arxiv.org/abs/1903.03704 (2019).
50. Nijkamp, E. et al. Learning energy-based model with flow-based backbone by neural transport MCMC. Preprint at https://arxiv.org/abs/2006.06897 (2020).
51. Wang, T., Wu, Y., Moore, D. & Russell, S. J. Meta-learning MCMC proposals. *Adv. Neural Inf. Process. Syst.* **31**, 4146–4156 (2018).
52. Song, J., Zhao, S. & Ermon, S. A-NICE-MC: adversarial training for MCMC. *Adv. Neural Inf. Process. Syst.* **30**, 5140–5150 (2017).
53. Li, Z., Chen, Y. & Sommer, F. T. A neural network MCMC sampler that maximizes proposal entropy. *Entropy* **23**, 269 (2021).
54. Huang, L. & Wang, L. Accelerated Monte Carlo simulations with restricted Boltzmann machines. *Phys. Rev. B* **95**, 035105 (2017).
55. Liu, J., Qi, Y., Meng, Z. Y. & Fu, L. Self-learning Monte Carlo method. *Phys. Rev. B* **95**, 041101 (2017).
56. Liu, J., Shen, H., Qi, Y., Meng, Z. Y. & Fu, L. Self-learning Monte Carlo method and cumulative update in fermion systems. *Phys. Rev. B* **95**, 241104 (2017).
57. Nagai, Y., Shen, H., Qi, Y., Liu, J. & Fu, L. Self-learning Monte Carlo method: continuous-time algorithm. *Phys. Rev. B* **96**, 161102 (2017).
58. Shen, H., Liu, J. & Fu, L. Self-learning Monte Carlo with deep neural networks. *Phys. Rev. B* **97**, 205140 (2018).
59. Xu, X. Y., Qi, Y., Liu, J., Fu, L. & Meng, Z. Y. Self-learning quantum Monte Carlo method in interacting fermion systems. *Phys. Rev. B* **96**, 041119 (2017).
60. Chen, C. et al. Symmetry-enforced self-learning Monte Carlo method applied to the Holstein model. *Phys. Rev. B* **98**, 041102 (2018).
61. Nagai, Y., Okumura, M. & Tanaka, A. Self-learning Monte Carlo method with Behler–Parrinello neural networks. *Phys. Rev. B* **101**, 115111 (2020).
62. Nagai, Y., Tanaka, A. & Tomiya, A. Self-learning Monte Carlo for non-Abelian gauge theory with dynamical fermions. *Phys. Rev. D* **107**, 054501 (2023).
63. Pawlowski, J. M. & Urban, J. M. Reducing autocorrelation times in lattice simulations with generative adversarial networks. *Mach. Learn. Sci. Technol.* **1**, 045011 (2020).
64. Foreman, S. et al. HMC with normalizing flows. *PoS* **LATTICE2021**, 073 (2022).
65. Arbel, M., Matthews, A. & Doucet, A. Annealed flow transport Monte Carlo. *Proceedings of Machine Learning Research* **139**, 318–330 (2021).
66. Matthews, A. G. D. G., Arbel, M., Rezende, D. J. & Doucet, A. Continual repeated annealed flow transport Monte Carlo. *Proceedings of Machine Learning Research* **162**, 15196–15219 (2022).
67. Caselle, M., Cellini, E., Nada, A. & Panero, M. Stochastic normalizing flows as non-equilibrium transformations. *J. High Energy Phys.* **2022**, 1–31 (2022).
68. Veach, E. & Guibas, L. J. Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, 419–428 (1995).
69. Müller, T., Rousselle, F., Keller, A. & Novák, J. Neural control variates. *ACM Trans. Graph.* **39**, 1–19 (2020).
70. Li, S.-H. & Wang, L. Neural network renormalization group. *Phys. Rev. Lett.* **121**, 260601 (2018).
71. Li, S.-H., Dong, C.-X., Zhang, L. & Wang, L. Neural canonical transformation with symplectic flows. *Phys. Rev. X* **10**, 021020 (2020).
72. Tomiya, A. & Nagai, Y. Gauge covariant neural network for 4 dimensional non-Abelian gauge theory. Preprint at https://arxiv.org/abs/2103.11965 (2021).
73. Tanaka, A. & Tomiya, A. Towards reduction of autocorrelation in HMC by machine learning. Preprint at https://arxiv.org/abs/1712.03893 (2017).
74. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48 (2019).
75. Mitrovic, J., McWilliams, B., Walker, J. C., Buesing, L. H. & Blundell, C. Representation learning via invariant causal mechanisms. In *International Conference on Learning Representations* (2020).
76. Rezende, D. J., Racanière, S., Higgins, I. & Toth, P. Equivariant Hamiltonian flows. Preprint at https://arxiv.org/abs/1909.13739 (2019).
77. Cohen, T. & Welling, M. Group equivariant convolutional networks. *Proceedings of Machine Learning Research* **48**, 2990–2999 (2016).
78. Fuchs, F., Worrall, D., Fischer, V. & Welling, M. SE(3)-transformers: 3D roto-translation equivariant attention networks. *Adv. Neural Inf. Process. Syst.* **33**, 1970–1981 (2020).
79. Du, W. et al. SE(3) equivariant graph neural networks with complete local frames. *Proceedings of Machine Learning Research* **162**, 5583–5608 (2022).
80. Kanwar, G. et al. Equivariant flow-based sampling for lattice gauge theory. *Phys. Rev. Lett.* **125**, 121601 (2020).
81. Boyda, D. et al. Sampling using *SU*(*N*) gauge equivariant flows. *Phys. Rev. D* **103**, 074504 (2021).
82. Jin, X.-Y. Neural network field transformation and its application in HMC. In *The 38th International Symposium on Lattice Field Theory* Vol. 396, 600 (PoS, 2022).
83. Kanwar, G. et al. Equivariant flow-based sampling for lattice gauge theory. *Phys. Rev. Lett.* **125**, 121601 (2020).
84. Katsman, I. et al. Equivariant manifold flows. *Adv. Neural Inf. Process. Syst.* **34**, 10600–10612 (2021).
85. Finkenrath, J. Tackling critical slowing down using global correction steps with equivariant flows: the case of the Schwinger model. Preprint at https://arxiv.org/abs/2201.02216 (2022).
86. de Haan, P., Rainone, C., Cheng, M. & Bondesan, R. Scaling up machine learning for quantum field theory with equivariant continuous flows. Preprint at https://arxiv.org/abs/2110.02673 (2021).
87. Albergo, M. S. et al. Flow-based sampling for fermionic lattice field theories. *Phys. Rev. D* **104**, 114507 (2021).
88. Hackett, D. C. et al. Flow-based sampling for multimodal distributions in lattice field theory. Preprint at https://arxiv.org/abs/2107.00734 (2021).
89. Albergo, M. S., Kanwar, G. & Shanahan, P. E. Flow-based generative models for markov chain monte carlo in lattice field theory. *Phys. Rev. D* **100**, 034515 (2019).
90. Vaitl, L., Nicoli, K. A., Nakajima, S. & Kessel, P. Path-gradient estimators for continuous normalizing flows. *Proceedings of Machine Learning Research,* **162**, 21945–21959 (2022).
91. Köhler, J., Klein, L. & Noé, F. Equivariant flows: exact likelihood generative learning for symmetric densities. *Proceedings of Machine Learning Research* **119**, 5361–5370 (2020).
92. Abbott, R. et al. Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions. *Phys. Rev. D* **106**, 074506 (2022).
93. Albergo, M. S. et al. Flow-based sampling for fermionic lattice field theories. *Phys. Rev. D* **104**, 114507 (2021).
94. Abbott, R. et al. Sampling QCD field configurations with gauge-equivariant flow models. In *The 39th International Symposium on Lattice Field Theory* Vol. 430, 036 (PoS, 2023).
95. Lüscher, M. Trivializing maps, the Wilson flow and the HMC algorithm. *Commun. Math. Phys.* **293**, 899–919 (2010).
96. Lüscher, M. & Weisz, P. Perturbative analysis of the gradient flow in non-Abelian gauge theories. *J. High Energy Phys.* **2011**, 1–23 (2011).

97. Gerdes, M., de Haan, P., Rainone, C., Bondesan, R. & Cheng, M. C. N. Learning lattice quantum field theories with equivariant continuous flows. Preprint at https://arxiv.org/abs/2207.00283 (2022).
98. Bacchio, S., Kessel, P., Schaefer, S. & Vaitl, L. Learning trivializing gradient flows for lattice gauge theories. *Phys. Rev. D* **107**, L051504 (2023).
99. Albergo, M. S. et al. Flow-based sampling in the lattice Schwinger model at criticality. *Phys. Rev. D* **106**, 014514 (2022).
100. Abbott, R. et al. Aspects of scaling and scalability for flow-based sampling of lattice QCD. Preprint at https://arxiv.org/abs/2211.07541 (2022).
101. Gabbard, H., Messenger, C., Heng, I. S., Tonolini, F. & Murray-Smith, R. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. *Nat. Phys.* **18**, 112–117 (2022).
102. Musaelian, A. et al. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **14**, 579 (2023).
103. Singha, A., Chakrabarti, D. & Arora, V. Conditional normalizing flow for Markov chain Monte Carlo sampling in the critical region of lattice field theory. *Phys. Rev. D* **107**, 014512 (2023).
104. Lehner, C. & Wettig, T. Gauge-equivariant neural networks as preconditioners in lattice QCD. Preprint at https://arxiv.org/abs/2302.05419 (2023).
105. Sutton, R. *The Bitter Lesson* (2019); https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf.

Acknowledgements

We thank W. Detmold and R. D. Young for comments on the manuscript. P.E.S. was supported in part by the US Department of Energy, Office of Science, Office of Nuclear Physics, under grant contract number DE-SC0011090, and by Early Career Award DE-SC0021006, by a NEC research award, and by the Carl G. and Shirley Sontheimer Research Fund. G.K. was supported by funding from the Schweizerischer Nationalfonds (grant agreement no. 200020_200424).

Author contributions

The authors contributed equally to all aspects of the article.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Physics* thanks Tanmoy Bhattacharya and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.