Journal of the Physical Society of Japan **94**, 031006 (2025)

https://doi.org/10.7566/JPSJ.94.031006

Special Topics

Machine Learning Physics

Machine Learning for Lattice QCD

Akio Tomiya*

Tokyo Woman's Christian University, Suginami, Tokyo 167-8585, Japan

(Received October 15, 2024; accepted January 6, 2025; published online February 21, 2025)

In this review, we explore the application of machine learning (ML) to lattice quantum chromodynamics (QCD), a key tool in studying nonperturbative phenomena in particle physics. By integrating ML techniques such as neural networks, lattice QCD simulations are significantly enhanced, enabling challenges like critical slowing down and topological charge to be addressed. These methods reduce computational costs and improve accuracy in configuration generation and physical measurements. Despite concerns over the black-box nature of ML, its application shows great promise in advancing lattice QCD research beyond traditional methods.

1. Introduction

Particle theory aims to understand the physical phenomena at scales smaller than atomic nuclei, driven by four fundamental forces.[1] Three of these forces — the electromagnetic, weak, and strong forces — are described by gauge theories within quantum field theory. The electroweak force, combining the electromagnetic and weak forces, is well understood through the $SU(2) \times U(1)$ gauge theory and the Higgs mechanism. The strong force described by $SU(3)$ gauge theory, has a large coupling constant and perturbative calculations are difficult.

In collider experiments such as those at the Large Hadron Collider (LHC), protons, composed of quarks and gluons, are accelerated and collided, producing new particles via relativistic effects ($E = mc^2$). The resulting particles, their trajectories, and momentum distributions are predicted by the Standard Model, which uses the Lagrangian or Hamiltonian of the $SU(3)$ gauge theory.

Relativistic quantum theory is naturally calculated through quantum field theory. In conventional quantum theory for many-body systems, the number of particles is fixed, making it challenging to handle processes involving particle creation or annihilation. This approach corresponds to the canonical ensemble in statistical mechanics. Quantum field theory, on the other hand, aligns with the grand canonical ensemble, where the particle number can vary, facilitating the description of such processes. The chemical potential can be either zero or non-zero in these systems. In condensed matter physics, quantum field theory is often referred to as the method of second quantization method.[2]

The gauge theory for the strong force is called quantum chromodynamics (QCD). QCD involves two types of fields: one corresponding to fermions called quarks and the other to vector bosons called gluons. Let us look at them one by one. The quark field transforms as a fermion (spin 1/2) under Lorentz transformations and corresponds to the electron in condensed matter physics. It has three internal degrees of freedom, called color degrees of freedom. The quark field is divided into up-type ($Q_e = 2/3$) and down-type ($Q_e = -1/3$) depending on their electric charge. Throughout this review, we mostly use the natural unit $c = \hbar = 1$.

The quark field behaves like a vector concerning the color degrees of freedom (i.e., in the fundamental representation), and it is written as $\psi = [\psi_\mathrm{r}\ \psi_\mathrm{g}\ \psi_\mathrm{b}]^\top$, where the subscripts denote the colors, and each $\psi_c = \psi_c(x)$ is a four-dimensional fermion field with $x = (\vec{r}, t)$. In the absence of interactions, it follows the Dirac equation (precisely, it depends on spin and additional conditions, which are omitted here). The corresponding Lagrangian (density) is

$$\mathcal{L} = \bar{\psi} \sum_{\mu=0}^{3} \gamma^\mu \partial_\mu \psi, \tag{1}$$

where $\bar{\psi} = \psi^\dagger \gamma^0$ and γ^μ are the four-dimensional gamma matrices that satisfy the Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = -\eta^{\mu\nu}$. The derivative $\partial^\mu = \partial/\partial_\mu$ is shorthand for $\partial/\partial_0, \partial/\partial_1, \partial/\partial_2, \partial/\partial_3$ with the 0th component being the time derivative, and the others being spatial derivatives.

The gluon field transforms as a vector (spin 1) under Lorentz transformations. It does not carry a charge but has eight internal degrees of freedom corresponding to color. These are akin to phonons in condensed matter physics and mediate interactions between quarks. The corresponding Lagrangian is

$$\mathcal{L} = -\frac{1}{4} \sum_{a=1}^{8} \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} F^a_{\mu\nu} F^{a\mu\nu}, \tag{2}$$

where $F^a_{\mu\nu}$ is a tensor corresponding to the electromagnetic field strength $f_{\mu\nu}$, but the explicit form is not necessary here and is omitted. Like the electromagnetic field, the gluon field has only transverse waves with two polarizations.

For later explanations, let us also introduce fields that do not appear in QCD. These are used for the explanation of subsequent methods and for checking the actual procedures when lattice QCD is implemented. This is the relativistic scalar field. The scalar field corresponds to a real number at each spacetime point and can be thought of as, for example, the density of a substance. It is defined as $\phi = \phi(x) \in \mathbb{R}$. In relativistic quantum field theory, it obeys the Klein–Gordon equation $(\partial_\mu \partial^\mu - m^2)\phi = 0$. It is important to note that ϕ corresponds to the density of matter and is not a wave function (see a textbook on quantum field theory for more details[3]). The Lagrangian of the scalar field, with the metric $\eta^{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$ used to change the sign of the spatial part, is

$$\begin{aligned} \mathcal{L} &= \frac{1}{2} \sum_{\mu=0}^{3} \partial_\mu \phi \partial^\mu \phi - \frac{m^2}{2} \phi^2 \\ &= \frac{1}{2} (\partial_0 \phi)^2 - \frac{1}{2} \sum_{\mu=1}^{3} (\partial_\mu \phi)^2 - \frac{m^2}{2} \phi^2. \end{aligned} \tag{3}$$

Looking at the first line, it corresponds exactly to the formula for the Lagrangian in nonrelativistic analytical mechanics, $L =$ kinetic energy $-$ potential energy. The second line has been rewritten for later use. Integrating the Lagrangian density over spacetime yields the action where by, after integration by parts (assuming the field vanishes at infinity) in the first expression,

$$S = \int \mathrm{d}^4x\,\mathcal{L} = \int \mathrm{d}^4x\left(-\frac{1}{2}\sum_{\mu=0}^{3}\phi\partial_\mu\partial^\mu\phi - \frac{m^2}{2}\phi^2\right), \tag{4}$$

is obtained. To quantize the theory, one can define the Hamiltonian via a Legendre transformation and impose canonical quantization conditions on the fields. In this case, the momentum also becomes a field. Specifically, using the Lagrangian before integration by parts, we obtain $\pi(x) = \delta\mathcal{L}/\delta(\partial_0\phi) = \partial_0\phi$.

The Hamiltonian is obtained by the Legendre transformation and spatial integration,

$$H = \int \mathrm{d}^3x\left(\frac{1}{2}\pi^2(x) + \frac{1}{2}\sum_{\mu=1}^{3}(\partial_\mu\phi)^2 + \frac{m^2}{2}\phi^2\right). \tag{5}$$

Note that the sign of the potential term has changed, and information for the time direction has disappeared. Imposing canonical quantization conditions on the field $\phi(\vec{x}, t)$ and momentum $\pi(\vec{x}, t)$,

$$[\phi(\vec{x}, t), \pi(\vec{y}, t)] = \mathrm{i}\delta^{(3)}(\vec{x} - \vec{y}), \tag{6}$$

solving the Schrödinger equation gives the quantum state of the field. The quantum state at time t' is related to the state at $t = 0$ as

$$|\Psi(t')\rangle = \mathrm{e}^{-\mathrm{i}Ht'}|\Psi(0)\rangle. \tag{7}$$

In the operator formalism, the expectation value is given by

$$\langle O\rangle = \langle\Psi|O|\Psi\rangle. \tag{8}$$

Calculations in lattice QCD are often performed using the path integral formalism. In the path integral formalism, the expectation value is given by

$$\langle O\rangle = \frac{1}{Z}\int\mathcal{D}\phi\,\mathrm{e}^{\mathrm{i}S[\phi]}O[\phi], \tag{9}$$

where $O[\phi]$ is a physical quantity and a function of $\phi(x)$. Z is a constant called the partition function, corresponding to the integral of the right-hand side without $O[\phi]$. The reason it is called the partition function will become clear later. Here, $\mathcal{D}\phi = \prod_x \mathrm{d}\phi(x)$, representing an infinite-dimensional continuous integral over the entire spacetime. If the terms include only up to second derivatives and the potential term is quadratic, the integral can be evaluated using Gaussian integration. However, if there is a term corresponding to the interaction between particles, such as ϕ^4, the integral cannot be performed exactly. Those concerned with mathematical rigor may have noticed the peculiar nature of this integral. There are three main approaches taken.

1. Expand around free field configurations (perturbation theory)
2. Discretize spacetime (lattice field theory) and calculate on computers
3. Other topic-dependent methods

Here, we take the second approach. This method is typically used with the Euclidean time formalism. The Euclidean time formalism is a method of calculation where the time variable is formally replaced as $t \to -\mathrm{i}\tau$, $\tau \in \mathbb{R}$. This is also refereed as the imaginary time formalism. Thus, the time evolution of the state is modified to

$$|\Psi(\tau)\rangle = \mathrm{e}^{-H\tau}|\Psi(0)\rangle. \tag{10}$$

Here, the Hamiltonian is not changed, so physical quantities like energy are unaffected. However, calculating phenomena that involve real-time evolution requires an analytic continuation of the results. The path integral formalism can also be applied in the Euclidean time formalism,

$$\langle O\rangle = \frac{1}{Z}\int\mathcal{D}\phi\,\mathrm{e}^{-S^{(\mathrm{E})}[\phi]}O[\phi], \tag{11}$$

where the action $S^{(\mathrm{E})}$ is modified and given by

$$S^{(\mathrm{E})} = \int \mathrm{d}^3x\,d\tau\,\mathcal{L}^{(E)}, \tag{12}$$

with $\partial_\mu \equiv (\partial_1, \partial_2, \partial_3, \partial_\tau)$ and

$$\mathcal{L}^{(E)} = \frac{1}{2}\sum_{\mu=1}^{4}\partial_\mu\phi\partial_\mu\phi + \frac{m^2}{2}\phi^2 = -\frac{1}{2}\sum_{\mu=1}^{4}\phi\partial_\mu^2\phi + \frac{m^2}{2}\phi^2, \tag{13}$$

where the last equality uses integration by parts for $S^{(\mathrm{E})}$. $S^{(\mathrm{E})}$ is called the Euclidean action. One can see that the sign of the potential term matches that of the Hamiltonian in classical mechanics. Even when including an interaction term $\lambda\phi^4$, a similar expression can be obtained. From the equations, it is clear that the path integral in Euclidean time is very similar to the expression for the expectation value in classical statistical mechanics,

$$\langle O\rangle = \frac{1}{Z}\int\prod_i \mathrm{d}q_i\,\mathrm{d}p_i\,\mathrm{e}^{-H^{(\mathrm{cl})}[q,p]}O[p, q], \tag{14}$$

although the mathematical meaning differs because of the continuous infinite-dimensional integration measure $\mathcal{D}\phi$. At this point, we perform discretization, i.e., we define $\phi(an) = \phi(x) \equiv \phi_n$ only on discrete *lattice* points,[4] with a being the lattice spacing. This is reminiscent of phonon calculations in condensed matter theory. This is the lattice field theory. Now we have

$$S^{(\mathrm{E,lat})} = a^4\sum_{n\in\mathbb{L}^4}\left(-\frac{1}{2}\phi_n(\partial^2\phi)_n^{(\mathrm{lat})} + \frac{m^2}{2}\phi_n^2 + \frac{\lambda}{4!}\phi_n^4\right), \tag{15}$$

where $\mathbb{L}^4$ denotes the set of points in a four-dimensional hypercubic lattice and n represents the lattice coordinates, i.e., a tuple of four integers. Moreover,

$$(\partial^2\phi)_n^{(\mathrm{lat})} = \frac{1}{a^2}\sum_{\mu=1}^{4}(\phi_{n+\hat{\mu}} + \phi_{n-\hat{\mu}} - 2\phi_n) \tag{16}$$

with $\hat{\mu}$ as the unit vector in the μ direction. The kinetic term can be written explicitly as

$$S^{(\mathrm{E,lat})} = a^4\sum_{n\in\mathbb{L}^4}\left(-\frac{1}{2}\frac{1}{a^2}\sum_{\mu=1}^{4}(\phi_n\phi_{n+\hat{\mu}} + \phi_n\phi_{n-\hat{\mu}} - 2\phi_n^2)\right.$$
$$\left. + \frac{m^2}{2}\phi_n^2 + \frac{\lambda}{4!}\phi_n^4\right). \tag{17}$$

Although this action represents a quantum field system, it has become a "classical Hamiltonian" of a many-body system in (four-dimensional) space. The kinetic term has been transformed into an interaction via springs between neighboring points. In this manner, the expectation value in the path integral formalism is given by

$$\langle O\rangle^{(\mathrm{lat})} = \frac{1}{Z}\int \mathcal{D}\phi\, \mathrm{e}^{-S^{(\mathrm{E,lat})}[\phi]} O^{(\mathrm{lat})}[\phi]. \tag{18}$$

When there are no fermions in the system, $S^{(\mathrm{E,lat})}[\phi] \in \mathbb{R}$, so $\mathrm{e}^{-S^{(\mathrm{E,lat})}[\phi]} > 0$. Although this quantity deviates from that in continuous spacetime when $a \neq 0$, it matches in the limit $a \to 0$ (along with the adjustment of coupling constants, i.e., renormalization). This is called the continuum limit.[5] While the discussion of the continuum limit here has been rough, a precise discussion can be conducted using the renormalization group. Although detailed discussions are avoided here, in typical calculations, the four-dimensional volume is taken to be finite. Boundary conditions for spatial directions are typically taken as periodic. By using boundary conditions cleverly in the time direction, the Matsubara formalism can be employed.

The path integral measure is now fixed and

$$\mathcal{D}\phi = \prod_{n\in\mathbb{L}} \mathrm{d}\phi_n. \tag{19}$$

This is perfectly mathematically defined. Moreover, the degrees of freedom in the lattice formulation become the same as those in statistical mechanics.

Finite degrees of freedom allow us to use the methods of statistical mechanics, for example, the renormalization group. We will focus on the Markov chain Monte Carlo (MCMC) method. In this method, $\mathrm{e}^{-S^{(\mathrm{E,lat})}[\phi]}/Z$ is treated as a probability distribution, and the field configuration $\{\phi_n\}_{n\in\mathbb{L}^4}$ is sampled using importance sampling. This is the same as the MC calculation of the Ising model. Specifically, when $S^{(\mathrm{E,lat})}[\phi]$ has only nearest-neighbor interactions and is thus local, the heatbath method can be used as in the Ising model. In other words, when focusing on ϕ_n at a lattice point, $\phi_{n\pm\hat{\mu}}$ surrounding it can be treated as a heat bath and ϕ_n can be sampled.

Up to this point, gauge fields have been ignored, but let us now explain how they are treated in field theory. For a while, let us return to real-time and continuous spacetime. In field theory, gauge fields such as gluons and photons interact through derivatives. The covariant derivative for the scalar field is obtained by replacing $\partial_\mu\phi \to \partial_\mu\phi + \mathrm{i}Q_e A_\mu \phi$, where Q_e is the electric charge and the four-dimensional $A_\mu = (A_0, \vec{A})$ is the gauge field. A_0 corresponds to the scalar potential and the three-dimensional components correspond to the vector potential. This replacement may seem artificial, but it allows the imposition of gauge invariance on the action. Gauge invariance means invariance under the transformation

$$\phi(x) \to \mathrm{e}^{\mathrm{i}Q_e\omega(x)}\phi(x), \quad A_\mu(x) \to A_\mu(x) - \partial_\mu\omega(x), \tag{20}$$

where $\omega(x) \in \mathbb{R}$ is a function on four-dimensional spacetime. This transformation is a symmetry involving a coordinate-dependent phase, particularly called $U(1)$ gauge symmetry. The kinetic term for the gauge field A_μ is taken as $\sum_{\mu\nu} f_{\mu\nu} f^{\mu\nu}$, with $f_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, preserving gauge symmetry. Moreover, by choosing it this way, the Maxwell equations can be derived using the Euler–Lagrange equation. A_μ is a function taking values in the tangent plane of $U(1)$.

We have so far introduced the gauge field into the scalar field, but it can also be introduced for fermions. Introducing fermions and the $U(1)$ gauge field yields what is called quantum electrodynamics (QED). In this case, the fermion corresponds to a relativistic electron. This theory has been shown to be accurate through calculations such as those of the anomalous magnetic moment. Both the scalar field case and the fermion field case are referred to as gauge theories. Why is gauge theory special? There are various ways to say this, but in gauge theory, a renormalizable theory can be constructed, and by imposing gauge symmetry (and several symmetries), experimental results can be reproduced, and various predictions can be made. On the other hand, if gauge symmetry is lost, quantum corrections cannot be controlled and divergences appear everywhere (non-renormalizable).

The present subject, QCD, can also be constructed similarly. It is pointed out by Yang and Mills, Weyl, and Uchiyama that the $U(1)$ gauge symmetry of QED can be extended to general compact non-Abelian continuous groups. Namely, the formally replaced version to $SU(3)$ is nothing but QCD, mathematically. Here, we use the fact that the gauge group is $SU(3)$ based on experimental results. The qualitative property is not markedly different from $SU(2) = \{\mathrm{e}^{\mathrm{i}\sum_a \theta^a\sigma^a}\}$, where σ^a is the Pauli matrix. In most cases, one can consider $SU(2)$ instead, if one is not familiar with $SU(3)$.

In QCD, $\Omega(x) \in SU(3)$, and $A_\mu(x) = \sum_{c=1}^{8} A_\mu^c T_c$ is formulated as a function taking values in traceless Hermitian matrices in $su(3)$ Lie algebra and T_c is a matrix that is the basis of it. Lagrangian (2) is invariant under the transformation

$$\begin{aligned} &\psi(x) \to \Omega(x)\psi(x), \\ &A_\mu(x) \to \Omega(x)A_\mu(x)\Omega^{-1}(x) + i\Omega(x)\partial_\mu\Omega^{-1}(x). \end{aligned} \tag{21}$$

Now, A_μ is a function taking values in the tangent plane of the $SU(3)$ group.

1.1 *Spin to lattice gauge theory*

Spin systems and gauge theories are tightly connected.[6] To see this, let us look at the work of Wegner from 1971.[7] He proposed a dual model of the two-dimensional Ising model as an extension of classical statistical mechanics. The ferromagnetic Ising model on a square lattice assigns a spin variable $s_n \in \{-1, 1\} = \mathbb{Z}_2$ to each lattice point, described by the Hamiltonian,

$$H = -\beta\sum_{\langle n,n'\rangle} s_n s_{n'} \simeq \beta \sum_{\langle n,n'\rangle}(1 - s_n s_{n'}), \tag{22}$$

where β is the inverse temperature with coupling (a positive parameter) and is part of the Boltzmann weight, but it is included in the Hamiltonian. The Boltzmann weight is e^{-H}. The notation $\langle n, n'\rangle$ in the summation symbol indicates that only nearest-neighbor lattice points are summed. The second equality holds since a constant can be added to the Hamiltonian. Statistical expectation values can be obtained by summing all possible configurations of s_n. This model is invariant under $s_n \to -s_n$, which is referred as global $\mathbb{Z}_2$ invariance.

Wegner considered placing spin variables $s_\mu(n) \in \{-1, 1\}$ on the bonds of a two-dimensional lattice. For example, $s_\mu(n)$ is a spin on a bond at site n for direction $\mu = 1, 2$. The *Hamiltonian* S is given by

$$S = \beta \sum_n \sum_{\mu,\nu} (1 - s_{\mu\nu}(n)), \tag{23}$$

where $s_{\mu\nu}(n) = s_\mu(n) s_\nu(n+\hat{\mu}) s_\mu^{-1}(n+\hat{\nu}) s_\nu^{-1}(n)$. This $s_{\mu\nu}(n)$ is called a plaquette that takes the value ±1. Directions $\mu, \nu = 1, 2$ in this case. n is a lattice coordinate in two dimensions. Although the inverse is meaningless in this special model, it is added for later convenience. Geometrically, plaquettes are defined on a square surface of each minimal square on the lattice. Statistical expectation values can be obtained by summing all possible configurations of $s_\mu(n)$. The Hamiltonian is deliberately denoted as S for later correspondence as well. This model does not change the value of S under the transformation

$$s_\mu(n) \to w_n s_\mu(n) w_{n+\mu}^{-1} \tag{24}$$

using a site-dependent transformation with variable $w_n \in \{-1, 1\} = \mathbb{Z}_2$. This is a symmetry that exists because the w_n corresponding to the same point is cancelled out. This is called $\mathbb{Z}_2$ gauge theory or Ising gauge theory.

1.2 Lattice gauge theory

Having come this far, it is easy to transition from Wegner's discrete gauge theory to Wilson's lattice gauge theory for a continuous group.[4] Consider a $U(1)$ gauge field $A_\mu(n)$ and set the lattice gauge field as $u_\mu(n) = \mathrm{e}^{\mathrm{i}aA_\mu(n)}$. The Euclidean lattice action is

$$S = \beta \sum_n \sum_{\mu,\nu} \mathrm{Re}[1 - p_{\mu\nu}], \tag{25}$$

where the plaquette is defined as $p_{\mu\nu}(n) = u_\mu(n) u_\nu(n+\hat{\mu}) u_\mu^*(n+\hat{\nu}) u_\nu^*(n)$. β is a parameter corresponding to the inverse temperature, which will be determined. $u_\mu(n)$ is called a link variable and integrated over $U(1)$ in the path integral. This ansatz, while somewhat ad hoc (for justification, see textbooks[8,9]), shows that S reproduces the action of the continuum theory (a Maxwell action) under the identification $\beta = 1/g^2$ if we employ the ansatz[10] $p_{\mu\nu} \approx \mathrm{e}^{\mathrm{i}a^2 f_{\mu\nu}}$ with $f_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Gauge invariance is guaranteed by a transformation similar to $\mathbb{Z}_2$ gauge theory, $u_\mu(n) \to w_n u_\mu(n) w_{n+\mu}^*$ but with $w_n \in U(1)$. This reproduces Maxwell equations.

Similarly, in $SU(N_c)$ lattice gauge theory, setting $U_\mu(n) = \mathrm{e}^{\mathrm{i}aA_\mu(n)}$ with $A_\mu(n) \in su(N_c)$, the Lie algebra and a traceless Hermitian matrix of $N_c \times N_c$, the link $U_\mu(n)$ becomes an element of the Lie group $SU(N_c)$, and the action can be written as

$$S = \beta \sum_n \sum_{\mu,\nu} \mathrm{Re}\,\mathrm{tr}[\mathbf{1} - P_{\mu\nu}], \tag{26}$$

where

$$P_{\mu\nu}(n) = U_\mu(n) U_\nu(n+\hat{\mu}) U_\mu^\dagger(n+\hat{\nu}) U_\nu^\dagger(n). \tag{27}$$

This $P_{\mu\nu}$ is also called a plaquette, which corresponds to energy density. Gauge invariance is guaranteed by a similar transformation to other gauge theories,

$$U_\mu(n) \to \Omega_n U_\mu(n) \Omega_{n+\mu}^\dagger \tag{28}$$

but with $\Omega_n \in SU(N_c)$. If we use the ansatz $P_{\mu\nu} \approx \mathrm{e}^{\mathrm{i}a^2 F_{\mu\nu}}$, which can be obtained from $U_\mu = \mathrm{e}^{\mathrm{i}aA_\mu}$, we obtain

$$P_{\mu\nu} \approx \mathrm{e}^{\mathrm{i}a^2 F_{\mu\nu}} = \mathbf{1} + \mathrm{i}a^2 F_{\mu\nu} - \frac{1}{2} a^4 F_{\mu\nu} F_{\mu\nu} + O(a^6). \tag{29}$$

This leads to

$$S \approx \frac{\beta a^4}{2} \sum_n \sum_{\mu,\nu} \mathrm{tr}\, F_{\mu\nu} F_{\mu\nu} \xrightarrow{a\to 0} \frac{\beta}{2} \int \mathrm{d}^4 x \sum_{\mu,\nu} \mathrm{tr}\, F_{\mu\nu} F_{\mu\nu}. \tag{30}$$

The position of the Lorentz indices μ, ν is different from one in the Minkowski spacetime but this is because we are using imaginary time formulation. Namely, it reproduces action of the continuum theory (2) in the limit $a \to 0$, with $\beta = 2N_c/g^2$ and $F_{\mu\nu} = \sum_{c=1}^{8} F_{\mu\nu}^c T^c$. Although the degrees of freedom on the lattice are continuous, they can still be thought of similarly to spin models, i.e., spins on bonds. In total, quantum expectation values are given by

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}U \mathrm{e}^{-S[U]}\, O[U], \tag{31}$$

where $Z = \int \mathcal{D}U\, \mathrm{e}^{-S[U]}$ and $O[U]$ is a functional of configuration U. This integration corresponds to the sum over all possible spin configurations. Namely, it is integrated over all possible lattice gauge configurations. Here, the path integral measure $\mathcal{D}U$ is the Haar measure invariant under $SU(N_c)$ transformations.

The naive limit $a \to 0$, where the action reproduces the continuum theory, is called the classical continuum limit. In other words, it is not enough to just ensure this limit in quantization via path integrals. To make this argument precise, analysis using the renormalization group (RG) is necessary. The conclusion gained from the analysis with RG is that the continuum limit corresponds to a continuous phase transition[11] (of a four-dimensional statistical system). It is guaranteed that the continuum limit can be taken even when quantum effects are included for two-dimensional $U(1)$ and four-dimensional $SU(N)$ systems.

If a is not small enough, higher order effects of a (commonly referred to as lattice artifacts, corresponding to higher dimensional operators in the terminology in the phenomenology of particle physics. It is also referred to as irrelevant operators in the renormalization group) become visible. Thus, a should be taken to be as small as possible. On the other hand, approaching the continuum limit negatively impacts sampling because the system is close to the critical point. This trade-off is a major issue.

1.3 Procedure of lattice QCD

The calculation of quantum expectations via path integrals (equivalent to the partition function in condensed matter physics) involves sampling using the MCMC method. The process can be divided into three main steps.

1. Generation of gauge field configurations. The standard method is the hybrid Monte Carlo (HMC) method, described below. Large matrices must be treated, requiring supercomputers.
2. Calculation of physical quantities (referred to as measurement). Physical quantities are calculated for each configuration, such as energy density and a quark potential. The complexity varies depending on the quantity being measured. For pure gauge fields, the calculation is typically straightforward, but when quarks are involved, solving linear equations for large sparse

matrices is again necessary. Depending on the quantity being calculated, a supercomputer may be required.

3. Statistical processing and extrapolation to the continuum limit. This procedure is referred to as analysis. The physical quantities obtained for each configuration are averaged, and standard errors are added. By averaging physical quantities that vary owing to quantum fluctuations for each configuration, quantum expectations with a cutoff are obtained. The extrapolation to the continuum limit (lattice spacing $a \to 0$) is then performed. This is usually calculated on a laptop computer. This review will not treat analysis.

Most of the computation time is spent on steps 1 and 2, and the following sections will discuss how machine learning can be applied to these processes.

1.4 Lattice QCD with quarks

So far, the discussion has been limited to bosons, but when fermions (quarks) are included, sampling becomes problematic. While explaining the path integral formalism for fermions is not the purpose of this review, it is briefly explained here as it will be used in the following sections. The path integral for fermions, widely discussed in textbooks,[8,9] is given by

$$\int \mathcal{D}\bar{\psi}\mathcal{D}\psi \mathrm{e}^{-2\int \bar{\psi}(D+m)\psi} = \det(D+m)^2, \tag{32}$$

where D is the covariant derivative on the lattice (i.e., a huge sparse matrix). This does not cause the sign problem for zero chemical potential.[12] Moreover,

$$\det(D+m)^2 = \int \mathcal{D}\phi^\dagger \mathcal{D}\phi \mathrm{e}^{-\int \phi^\dagger (D^\dagger D + m^2)^{-1}\phi} \tag{33}$$

holds if D has the chiral symmetry. Here, ϕ is a complex scalar field and uses γ_5 hermiticity. In other words, the contribution to the quark expectation can be written as the nonlocal boson integral (the pseudofermion method). The right-hand side of the integral can be constructed using a Gaussian distribution variable ξ with $\mathrm{e}^{-\xi^\dagger \xi}$ by the transformation $\phi = (D+m)\xi$. Here, D is a huge sparse matrix on the lattice (almost diagonal band matrix with a size of about 10^5 rows and columns).

In steps 1 and 2 of lattice calculations, solvers related to this $D+m$ are called. Specifically, calculations related to $(D+m)^{-1}$ are performed. This is very troublesome. It involves inverting a huge sparse matrix and is usually solved using methods like the conjugate gradient. If one side length is L and the size of the gauge group is N_c, then the size of the rows and columns is $L^4 N_c N_{\mathrm{sp}}$ (with N_{sp} being the number of spinor elements, typically 4 in four dimensions), so with $L = 10$ and $N_c = 3$, the matrix size is about 10^5. Thus, solvers are a target of applications of new techniques with machine learning.

1.5 Neural nets for lattice QCD

We have explained the basics of lattice field theory and lattice QCD. From here, we will discuss topics related to machine learning, including those up to the Lattice 2024 conference held in the summer of 2024 and the preceding Workshop in Swansea. However, this review is not comprehensive, and some topics may have been omitted. Refer to the original papers for details, as this review is biased by the author's perspective.

1.6 Basic neural nets

Readers of this review are assumed to have some knowledge of neural nets, but we provide a brief explanation (See Ref. 13 in detail.). A neural net is a variational function with a lot of parameters. Typically,

$$\vec{y} = f_\theta(\vec{x}), \tag{34}$$

where vector $\vec{x}$ is input and vector $\vec{y}$ is output. Here, θ is a set of parameters. f_θ is constructed by repeatedly applying affine transformation and non-linear functions (activation functions). Since the output depends on the parameter values, adjusting the parameters is crucial, and this tuning is called learning. It is important to note that neural nets are differentiable with respect to the parameters, allowing parameter tuning (learning) using derivatives. The desired processing is specified through an error function, similar to fitting. The internal design of neural nets involves various insights, which are the main focus here. It is worth mentioning that neural nets are universal approximaters at the infinite parameter limit. Namely, they can mimic any function.

1.7 Convolution and equivariance

There is a type of neural net called a convolutional neural net. This can be thought of as a net that takes an image as input and outputs a processed image. In other words, it acts like an image filter.

Consider inputting an image $\vec{x}$ into an image filter f. When the image is translated (ignoring boundary effects) and then input, the output matches the translated version of the original output. This means that the translation operation $\hat{T}$ and the filter processing are cummutative, i.e., $f(\hat{T}\vec{x}) = \hat{T}f(\vec{x})$. This property is called *equivariance*.[14–16]

Gauge-equivariant nets have been proposed for gauge symmetry (28) as well. Below, we will explain a gauge symmetric neural net that typically takes a gauge field configuration as input, processes it, and outputs a gauge field configuration (or a variable of the same shape).

1.8 Spectral flow

The first gauge-symmetric neural net developed is known as spectral flow.[17] The spectral flow can be constructed as follows. $U(1)$ gauge fields can be written as $U_\mu(n) = \mathrm{e}^{\mathrm{i}\theta_\mu(n)}$. $\theta_\mu(n)$ is a real number, but it takes values within $-\pi < \theta_\mu(n) \leq \pi$. Applying a non-compact projection (NCP) transformation[18] makes it take values over the entire real line, and it can then be input into a (conventional real-valued) neural net. The same operation can be performed even if all links are input. $U(1)$ gauge symmetry can also be preserved equivariantly.

For $SU(N_c)$, while some additional operations are necessary,[19] a similar construction can be achieved. The concept of a maximal torus allows the diagonarization of the $SU(N_c)$ matrix, isolating the $U(1)$ component. For instance, in the case of $SU(3)$, given an element U, it can be diagonalized as $P^{-1}UP = \mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1}, \mathrm{e}^{\mathrm{i}\theta_2}, \mathrm{e}^{-\mathrm{i}\theta_1 - \mathrm{i}\theta_2})$. Repeating this process for all links enables the application of a neural net designed for the $U(1)$ case.[20] In this manner, $SU(N_c)$ gauge (28) symmetry can be preserved equivariantly.

1.9 Gauge-equivariant neural nets

Gauge-equivariant neural nets have been proposed as neural nets that extract gauge-invariant features.[21] These nets calculate quantities that behave equivariantly under gauge transformation (28) by combining links. In the paper, calculations of Wilson loops and other quantities are performed, demonstrating better performance than conventional neural nets that do not consider gauge symmetry.

1.10 Gauge-covariant neural nets

Gauge-covariant neural nets[22] are extended versions of the traditional method in lattice QCD, known as smearing.[23–25]

Smearing is a technique to smooth out fluctuations at the smallest scales covariantly with (28). It involves summing the product of nearby links that transform in the same way under gauge transformations, with appropriate coefficients. This process reduces noise and helps extract the long-range structures in the gauge field.

Gauge-covariant neural nets treat the weights of this addition as trainable weights. Since the gauge covariant nets are almost the same as the conventional method, implementation is extremely simple. Gauge-covariant neural nets can be constructed using APE-type smearing or stout-type smearing, but here we focus on the stout type. This is called the residual flow sometimes, but it is the same as the gauge-covariant neural net.

The original stout smearing is given by

$$U_\mu^{(l+1)}(n) = \mathrm{e}^{\mathrm{i}\sum_f w^{(l,f)} Q_\mu^{(l,f)}(n)} U_\mu^{(l)}(n), \tag{35}$$

where l is the smearing level corresponding to the number of layers. f indicates a type of loop. $w^{(l,f)} \in \mathbb{R}$ is a trainable weight. $Q_\mu^{(l,f)}(n)$ takes values in the $su(N_c)$ Lie algebra, constructed as

$$Q_\mu(n) = \frac{\mathrm{i}}{2}(\Omega_\mu^\dagger(n) - \Omega_\mu(n)) - \frac{\mathrm{i}}{2N_c}\mathrm{Tr}(\Omega_\mu^\dagger(n) - \Omega_\mu(n)). \tag{36}$$

Here, $\Omega_\mu(n)$ is a (untraced) closed loop consisting of links associated with $U_\mu(n)$. The minimal example is plaquette (27). $\Omega_\mu(n)$ takes values in the $SU(N_c)$ Lie group. In the covariant neural net, coefficients are promoted to trainable parameters.

It is interesting that in this neural net, operations such as projection and normalization to ensure $SU(N_c)$ symmetry (28) act as activation functions.

As pointed out by Lüscher,[26] the continuum limit of the smearing step of stout smearing gives the gradient flow. This corresponds precisely to neural ODE[27] in neural nets, meaning that the gradient flow can be designed to produce the desired output. This has been developed as the continuous normalizing flow.[28]

1.11 Gauge-covariant transformer: CASK

Transformers currently boast state-of-the-art performance in neural net architectures and are already applied in web services like ChatGPT.[29] The core technology of transformers, the attention mechanism,[30] allows the consideration of contributions from words far apart, enabling powerful reasoning. We develop transformers that preserve global symmetries of physical systems to simulations of spin systems.[31,32] This transformer has an attention mechanism that preserves global symmetry. The construction method involves first performing two types of block spin transformations (corresponding to Key and Query) using trainable weights, with the attention matrix being the rotationally invariant correlation function of those block spins. Multiplying the attention matrix by the spins corresponding to value gives the attention-weighted spins. Adding these spins to the original spins and normalizing them give the output. This structure can be systematically repeated to construct the net. This transformer is used to construct effective models in the self-learning Monte Carlo method, resulting in performance improvements and scaling laws.[31,32]

A gauge-covariant transformer (CASK, covariant attention with stout kernel) is a transformer with global symmetry extended to gauge symmetry (28).[33] The block spin transformation is replaced with stout smearing (35) (APE smearing is also acceptable) using a gauge-invariant attention matrix. Preliminary results suggest that it performs better than conventional gauge-covariant neural nets.

1.12 Parallel transport convolution layer

The action of the Dirac operator (the covariant derivative of quarks on the lattice) is centered on the covariant parallel transport operator. It is natural to parametrize the parallel transporter and attempt to mimic the action of quarks in the presence of a gauge field. The parallel transport convolution (PTC) layer[34] is designed as a gauge-equivariant convolutional net for fermions in lattice field theory. In other words, the PTC layer is introduced in the context of flow-based sampling to incorporate the effects of pseudofermions. By cleverly summing covariantly translated pseudofermion fields, it mimics the probability distribution of pseudofermions under gauge fields. The application of the PTC layer is discussed in the following section.

2. Production

Configuration generation is computationally expensive. Therefore, it is natural to try to reduce the computational cost using machine learning.

As mentioned above, the continuum limit is achieved by adjusting the parameters in the action of field theory to approach the critical point as a four-dimensional statistical system. However, as one approaches the critical point as a statistical system, one encounters a phenomenon known as critical slowing down. Simply put, critical slowing down occurs when the generated samples become similar to each other, reducing the efficiency of the Monte Carlo method. While it is wellknown that calculations near the critical point are difficult in condensed matter physics, in lattice QCD, it is troublesome because it is related to reducing discretization errors. Below, we introduce efforts to reduce critical slowing down.

2.1 Self-learning Monte Carlo method

The MCMC method is a technique for generating sequences of samples according to a certain probability distribution and using these samples to calculate expectations.[35] In the MCMC method, detailed balance is often emphasized. The detailed balance condition is crucial because it guarantees that the samples generated by the Markov chain will converge to the exact expectation value, provided that the condition is met. The typical examples of algorithms that satisfy this condition are the Metropolis method and the Metropolis–Hastings method. Both methods consist of two steps: generating a candidate and performing the Metropolis or Metropolis–Hastings test. The specifics of the algorithm are determined by how the candidate is generated. The Metropolis–Hastings method falls back to the Metropolis method when a reversible conditional probability is used to generate the candidate. The hybrid Monte Carlo or Hamiltonian Monte Carlo,[36] the de facto standard in lattice QCD, is also a type of Metropolis method.

The self-learning Monte Carlo method uses an effective Hamiltonian learned through training. If the effective Hamiltonian allows a global update, one can use the global update. For example, in Ref. 35, the target theory is an Ising model with a four-point interaction, and the effective theory is a conventional Ising model used for the global update. The exact distribution is obtained using the Metropolis–Hastings test with these two Hamiltonians.

The interesting aspect of the self-learning Monte Carlo method is its ability to correct itself. It can autonomously generate samples, allowing the model to learn and improve during the sample generation process. Even when using an untrained model, the generated samples are still distributed according to the target system (with a high rejection rate).

The self-learning Monte Carlo method has been applied to lattice gauge theory,[37] where its implementation utilized an effective action parameterized in a gauge-invariant/covariant manner. In this calculation, the coupling constants of the effective action are determined using linear regression, and exact calculations are performed for the four-dimensional system.

The self-learning HMC[38] uses the parametrized action in the molecular dynamics part of the HMC. Since molecular dynamics is reversible in the fictitious time, only the Metropolis test is needed to ensure the convergence. A concrete example with gauge-covariant neural nets has been studied, leading to the world's first calculation of a four-dimensional system with dynamical fermions using machine learning.[22]

2.2 *Flow-based sampling*

The flow-based sampling is a method that is considered efficient for dealing with critical slowing down and the sampling of topological sectors in gauge theory.[17,19,39–42] This method is based on change-of-variables integration in path integrals. The approach leverages invertible neural nets to parameterize physical probability distributions, enabling the generation of configurations that are distributed according to the target distribution.

Recent developments show the effectiveness of flow-based models in lattice QCD, particularly through the implementation of gauge-equivariant flows that respect the gauge symmetries of the theory. These models have been successfully applied in four dimensional QCD simulations with pseudofermions,[42,43] showcasing their potential to handle the computational challenges posed by fermionic degrees of freedom. Moreover, the scalability of these models has been a subject of recent studies[44–46] highlighting their ability to maintain efficiency even as the lattice size increases. See[47] regarding the scale separation.

Here, we explain the algorithm with a concrete example. Consider a system with a scalar field ϕ,

$$\int \mathcal{D}\phi\, \mathrm{e}^{-S[\phi]} = \int \mathcal{D}\varphi\, \mathrm{e}^{-S[\varphi]}\mathcal{J}[\varphi] \equiv \int \mathcal{D}\varphi\, \mathrm{e}^{-S'[\varphi]}, \tag{37}$$

where $\mathcal{J}$ is the Jacobian of the transformation from ϕ to φ. If $S' = S[\varphi] - \log \mathcal{J}[\varphi]$ does not contain kinetic terms, then the integral can be performed independently at each point. Such a change of variables $\phi = \phi[\varphi]$ is called a trivializing map. In the case of supersymmetry, an explicit example of such a map has been constructed, known as the Nicolai map.[48] Lüscher also showed that, in the absence of supersymmetry,[49] the gradient flow is an example of a trivializing map in gauge theory in the strong coupling expansion.

The flow-based sampling starts from a configuration generated by a trivial Boltzmann weight with only potential terms, $r[\varphi] \propto \exp(-\sum_n \varphi^2(n))$. This can be done at relatively negligible cost and the generated configurations are rough, similar to Ising spins at the high-temperature limit. The trivial configurations are then deformed by a bijective neural net $\phi = f_\theta^{-1}[\varphi]$ (corresponding to constructing an un-trivializing map). The specific form of the trivializing map $f_\theta[\phi]$ can be designed in various ways, but typically, the framework of normalizing flows, which gives a properly normalized neural net, is used (originally an architecture for image generation). This works as a cooling process intuitively. The problem with this naive method is the computation of the Jacobian. However, the design of the net allows the Jacobian to be computed easily using the idea of coupling layers,[39,50] which is indeed employed. The next question is how to choose the error function for the neural net to achieve the un-trivializing map. For this purpose, the reverse Kullback–Leibler (KL) divergence is used. Suppose the normalized Boltzmann weight appearing in the path integral is $P[\phi] = \mathrm{e}^{-S[\phi]}/Z$, and the effective Boltzmann weight given by the neural net is $Q[\varphi] \propto r[\varphi]\mathcal{J}^{-1}[\varphi]$. Then,

$$D_{\mathrm{KL}}[Q\|P] = \int \mathcal{D}\varphi\, Q[\varphi] \log \frac{Q[\varphi]}{P[f_\theta^{-1}[\varphi]]} \tag{38}$$

can be used as the error function to adjust the parameters by minimizing it. In other words, while the KL divergence is an asymmetric function measuring the difference between probability distributions, it is zero only when they match. Therefore, if $P \approx Q$, the expectation value is evaluated as

$$\langle O \rangle \propto \int \mathcal{D}\phi O[\phi]P[\phi], \tag{39}$$

$$\approx \int \mathcal{D}\phi O[\phi]\mathcal{J}^{-1} r[f_\theta[\phi]], \tag{40}$$

$$= \int \mathcal{D}\varphi O[f_\theta^{-1}[\varphi]]r[\varphi]. \tag{41}$$

Furthermore, the evaluation of $D_{\mathrm{KL}}[Q\|P]$ in the path integral can be replaced by the random sampling of φ at each point, making it efficient. Note that the partition function does not contribute to this optimization as it is a constant with respect to the parameters of the neural net.

Once learning is complete (or self-learning is used), the approximate configuration generated can be considered a sample from the exact distribution by applying the Metropolis–Hastings method. Namely, the flow-based sampling algorithm is an exact algorithm. Although a higher acceptance ratio is desirable, it is worth noting that an embarrassingly parallel strategy can be employed (except for the Metropolis–Hastings step), making it attractive even if the acceptance rate is low.

The example here is calculating scalar fields, which has already been extended to and performed in gauge theories. In fact, such a calculation has been performed in two-dimensional pure $U(1)$ gauge theory,[17] two-dimensional pure $SU(N_c)$ gauge theory,[19] the two-dimensional Schwinger model (including the full quantum effects of quarks),[41] and preliminary tests in four-dimensional QCD with dynamical fermions.[43] As an exact algorithm, it gives results consistent with the de facto standard algorithm HMC (for four dimensions) and the exact solution (for two dimensions).

In lattice gauge theory, it is well known that sampling topological sectors is problematic, especially in setups close to the continuum limit, where the topological charge is known to get stuck in the 0 sector and also topology tunneling tends to be frozen. Because there are high barriers between these sectors close to the continuum limit, this problem is challenging for the HMC. Note that, from the beginning, the topological sector of lattice gauge theory is defined by the admissibility condition.[51] Namely, it allows the system to be divided into sectors where the topological charge is well defined only for sufficiently smooth gauge fields. In other words, rough gauge field configurations are not separated topological sectors in the configuration space. In flow-based sampling, admissible configurations are constructed from sampling at parameters where admissibility is not met at all, making it strong against this kind of problem.[17,41] Therefore, it seems to be a very promising method for sampling topological charges.

While there are reports of significant successes, there have also been comments from other research groups pointing out issues such as mode collapse in the scalar field theory[52] and the lack of mitigation of critical slowing down,[53–57] making this a very active research topic.

There have also been reports of using continuous flow-type neural nets,[28,58] similar to Neural ODE, to improve efficiency in this flow.

The Schwinger–Dyson equation can be used to write the trivializing map as a set of gauge-covariant loops.[59] It is concluded that although the trivialization of realistic beta would require too many loops, making exact calculations costly, this method has potential.

2.3 *Transformed replica exchange (T-REX)*

A notable advantage of flow-based sampling is the ability to generate correlated ensembles, which can reduce the uncertainty in observables that are sensitive to changes in theory parameters.

T-REX is an improved on the replica exchange (REX) algorithm,[60,61] which samples multiple Markov chains simultaneously. In this algorithm, proposals are made to exchange configurations between adjacent target densities, and the evolution of slower targets is accelerated using faster-mixing targets. In standard REX, the acceptance rate of exchanges decreases as the lattice size increases. T-REX uses a flow that bridges target densities, improving the acceptance rate of these exchanges and enabling the simultaneous sampling of a wider range of target densities.

DR-REX is a method that simultaneously samples not only physically meaningful target densities but also nonphysical actions with intentionally introduced defects to accelerate the freezing of topology.[61] In this method, a flow is applied to repair local open boundary condition defects, mitigating topology freezing while sampling the target physical action. This allows the flow to be applied only to the repaired part, keeping the flow's application cost constant regardless of the target volume.

2.4 *Flowed HMC*

In flow-based sampling, normalizing flows act as a change of variables connecting the coarse lattice and the target lattice. This can also be used within the HMC, leading to the development of the flowed HMC.[62] In the HMC, molecular dynamics are used internally, and the idea is to temporarily transform the gauge field to a coarse lattice, solve the virtual equations of motion on the coarse lattice, and return to the target system via flow (i.e., change of variables). This method is tested in four-dimensional $SU(3)$ calculations and is found to improve efficiency.

2.5 *L2HMC*

The L2HMC is an application of the generalized HMC, originally developed in the context of Bayesian inference.[63] This algorithm also modifies the molecular dynamics method within the HMC,[64–66] which is compatible with $SU(3)$ gauge dynamics. In molecular dynamics, symplectic numerical integration, which preserves energy, is used. Symplectic numerical integration, being a type of integrator that preserves energy, can be systematically constructed to be time-reversible. The L2HMC uses a neural net for the HMC force in the molecular dynamics method to enhance the tunneling of topological charge by altering the dynamics (i.e., tuning the force). On the other hand, since time-reversal symmetry is preserved, it converges to the exact expectation value. It is confirmed that during molecular dynamics, the probability distribution is altered [Fig. 6(b) in Ref. 65 for example], effectively enhancing tunneling. Intuitively, this can be seen as virtually increasing the temperature.

2.6 *Stochastic normalizing flow*

More aggressive ideas are aimed at accelerating sampling by using nonequilibrium physics.[67–71] This is known as the nonequilibrium Monte Carlo approach. A specific example is stochastic normalizing flow, which has been researched primarily by a group in INFN and has already been applied to scalar fields and the CP^N model. The idea is to connect the open boundary conditions and the periodic boundary conditions nonequilibrium-wise during the simulation. The innovative aspect is the use of Jarzynski's equation to calculate the expected values of observables obtained through nonequilibrium evolution. As a result, the autocorrelation time of topological charge is reduced in the model, which has been a long-standing goal in lattice QCD. There was also a presentation of preliminary results in $SU(3)$ gauge field calculations at the conference.[71,72]

2.7 *Diffusion model*

The diffusion model has been widely successful in the field of image generation.[73,74] It can be described as a physicsinspired method that uses Langevin dynamics. The main idea is to apply the concepts used in diffusion models for image generation to create configurations. While flow-based sampling uses (in some sense) only gradients, the diffusion model for image generation mimics the target distribution using noise and gradients (scores) to generate configurations. Interestingly, the diffusion model is being discussed as equivalent to the stochastic quantization process. The generated distribution can also be exactified using the Metropolis test.

2.8 *Perfect action*

The fixed point of the RG is scale-invariant and defines the continuum limit of lattice quantum field theory. Therefore, the lattice action that is slightly away from this continuum limit is expected to be very close to the continuum limit. Additionally, a lattice action along the RG flow gives the same expectation value even if the cutoff differs. It is expected that actions near the fixed point on the RG flow will have significantly smaller lattice artifacts. The idea that realizes this is called the perfect action.[75] It has been determined by traditional methods in lattice QCD, but there is research on writing the perfect action using neural nets and determining it through learning.[76] Preliminary results suggest that using the perfect action for HMC and gradient flow may allow scale-setting in lattice QCD with reduced discretization errors.

3. Measurement

Measurement can also be time-consuming, particularly for physical quantities involving quarks, such as correlation functions and hadron masses. Here, we introduce research on reducing computational costs using machine learning. Statistical errors account for a significant portion of the errors in lattice QCD calculations using MC methods. Reducing these errors can yield more precise results. Combined with ensemble generation, it can lead to overall reductions in computational cost.

3.1 *Bias-corrected approximation*

Some physical quantities are numerically expensive. For example, three-point functions of nucleons, various correlation functions appearing in the calculation of the anomalous magnetic moment, and $\mathrm{tr}(D+m)^{-n}$ are used in several calculations. Naturally, the question arises whether these can be calculated more cost-effectively using machine learning. On the other hand, using machine learning introduces approximations, making error evaluation challenging. The solution to this is the bias-corrected approximation method.[77–79)]

Suppose O is a physical quantity with high computational cost, and let O^{app} be the approximation constructed using a machine learning architecture. Then, the identity

$$\langle O \rangle = \langle O^{\mathrm{app}} \rangle + \langle (O - O^{\mathrm{app}}) \rangle \tag{42}$$

holds. If the expectation value is evaluated strictly, there is no gain. Using MC methods, the expectation value on the right-hand side can be evaluated as

$$\frac{1}{N_{\mathrm{conf}}} \sum_{c=1}^{N_{\mathrm{conf}}} O^{\mathrm{app}}[U_c] + \frac{1}{N_{\mathrm{bc}}} \sum_{c'=1}^{N_{\mathrm{bc}}} (O[U_{c'}] - O^{\mathrm{app}}[U_{c'}]), \tag{43}$$

where if $N_{\mathrm{conf}} \gg N_{\mathrm{bc}}$, the computational cost can be reduced. The second term is called the bias correction term. This method originated from all mode averaging (AMA), a technique developed in a context unrelated to machine learning. In the first study, a gradient-boosted decision tree is used to construct O^{app}. This part can be replaced by various machine learning architectures. Since the bias from machine learning can be canceled out while keeping the computational cost low, further applications are conceivable.

3.2 *Control variates*

The idea behind control variates is to reduce statistical errors by cleverly subtracting variables from the operator defining the physical quantity.[80,81] Since statistical errors are proportional to the variance, reducing the variance of the physical quantity is beneficial. This can be achieved as follows. Let O be the physical quantity and f be a quantity that satisfies $\langle f \rangle = 0$. Then, from this assumption, $\langle O \rangle = \langle O - f \rangle$. In other words, O and $O - f$ yield the same expectation value, but their variances differ. Specifically,

$$\langle (O - f)^2 \rangle = \langle O^2 \rangle + \langle f^2 \rangle - 2\langle Of \rangle \tag{44}$$

holds. If we can find an f that correlates strongly with O, i.e., for which $\langle Of \rangle$ is positive and large, we can reduce the variance. While f can be determined from the Schwinger–Dyson equation, a study has been presented where f is expressed using machine learning. A neural net is designed with constraints from the perspective of symmetry, and it is found that noise reduction is achieved in low-dimensional scalar fields and gauge theories. Like the bias-corrected approximation, this has many potential applications.

3.3 *Contour deformation for noise reduction*

Fermions can cause problems even when there is no sign problem. That is, measurement of observables involving fermions tends to be noisy. Therefore, research has been conducted to reduce noise by deforming the contour in path integrals.[82,83] By representing and parametrizing the contour in path integrals using neural nets, it is possible to reduce the measurement error by optimizing the contour.

3.4 *Multigrid method*

In lattice QCD, the generation of configurations and the calculation of quark propagators involve numerous D^{-1} calculations. Thus, accelerating the conjugate gradient method for D^{-1} calculations is crucial. The multigrid method mitigates the numerical costs, which effectively serves as a preconditioner.[84,85] In the multigrid method, a coarse grid is created from the original grid, and solving the problem on the coarse grid helps improve the solution on the original grid. The relationship between the original and coarse grids acts as the preconditioning step.

Before the rise of machine learning, multigrid methods were extensively studied. However, significant gains can now be achieved using machine learning techniques. Specifically, the use of the PTC layer enables the construction of an approximate Dirac operator on the coarse grid, effectively transmitting local information and efficiently handling long-range interactions. This approach significantly reduces the number of iterations required to solve the Dirac equation, leading to substantial reductions in computation time, especially in large-scale lattice QCD calculations.

3.5 Spectral functions

Since lattice QCD MC calculations use the imaginary-time formalism, accessing real-time information is extremely difficult. Spectral functions are functions common to both imaginary-time Green's functions (referred to as correlators in lattice QCD) and real-time Green's functions.[86] Once we obtain the spectral function, we can extract information for real time. Since imaginary-time Green's functions can be calculated using MC methods, once the spectral function is known, real-time physical quantities can be investigated.

This involves what is known as an ill-posed problem. It can be explained as follows. Let $\vec{G}$ be the imaginary-time Green's function that can be measured in lattice QCD. Originally, it is a function of imaginary time τ, but since only about $O(10)$ points can be taken, it is written as a $O(10)$-dimensional vector. The spectral function is also written as a vector $\vec{\rho}$; although it is a continuous function, we consider it to be a 1000-dimensional (approximately) vector for convenience. The kernel, a rectangular matrix, is denoted by K. $\vec{G}$ and $\vec{\rho}$ are related by

$$\vec{G} = K\vec{\rho}. \tag{45}$$

Since $\vec{G}$ and K are known, it seems that solving this system of equations will determine $\vec{\rho}$, but because of a lack of information, it cannot be solved. Thus, additional information is needed to solve it, which leads to the method of sparse modeling using L_1 regularization. By performing singular-value decomposition on K, $K = USV^\top$, where S is a rectangular matrix with singular values on the diagonal, and U and V are unitary matrices, the basis can be transformed as follows:

$$\underbrace{U^\top\vec{G}}_{=\vec{G}_{\mathrm{IR}}} = S\,\underbrace{V^\top\vec{\rho}}_{=\vec{\rho}_{\mathrm{IR}}}. \tag{46}$$

This is called the intermediate representation (IR) basis. In this basis, $\vec{\rho}_{\mathrm{IR}}$ is reconstructed with fewer "modes". Symbolically, the loss function is written as

$$L = |\vec{G}_{\mathrm{IR}} - S\vec{\rho}_{\mathrm{IR}}|^2 + \lambda|\vec{\rho}_{\mathrm{IR}}|, \tag{47}$$

where the second term is a regularization term that works to reduce the number of elements of $\vec{\rho}_{\mathrm{IR}}$ in the IR basis. By using this approach, one can extract the information of the spectral functions.[87–90]

3.6 Related topics

The parton distribution function (PDF), which represents the distribution of momentum within nucleons, is a physical quantity that has recently attracted attention. It is difficult to determine experimentally and is often determined through fitting. Therefore, a neural net PDF,[91] which parametrizes the PDF using neural nets, has been proposed for some time.

4. Sign Problem

In the Euclidean path integral formulation of QCD, a nonzero chemical potential for the quark number adds a complex phase to $\det(D + m)$, which renders the Monte Carlo method infeasible for large chemical potentials. This difficulty is known as the "sign problem",[92] because the Boltzmann weight — required to be real and positive for the Markov chain Monte Carlo — becomes complex. If the quark chemical potential is zero, the weight remains real and positive, permitting standard Monte Carlo simulations, but once it turns complex, it can no longer be interpreted as a probability distribution. Approaches to address this include using complexified fields, tensor networks, or quantum computers. Complexified fields extend the fields into the complex plane, tensor networks simplify quantum states by discarding less relevant information, and quantum computers can directly simulate quantum states without encountering the sign problem.[93] In this review, we adopt the first approach in conjunction with machine learning. For latter applications, refer to Ref. 94 and references therein.

4.1 Path optimization method and related topics

The path optimization method is a type of technique based on variable transformations. The sign problem depends on the representation of the system, and in the path optimization method, variables are transformed to suppress fluctuations in the phase along the integration path. This variable transformation is expressed using a neural network in the path optimization method.[95–101] Moreover, many similar approaches have been investigated.[102,103]

5. Application of Field Theory to Machine Learning

In this section, we discuss the use of field theory techniques to understand machine learning, the opposite of what has been discussed so far.

The restricted Boltzmann machine (RBM) is a tool widely used in machine learning to learn probability distributions from data.[104] They analyze RBMs with scalar fields as nodes from the perspective of lattice field theory. Starting with the simplest Gaussian field, they show that the RBM acts as an ultraviolet regulator with a cutoff determined by either the number of hidden nodes or the model's mass parameter. These ideas are tested in the case of scalar fields, where the target distribution is known, and the implications for cases where the distribution is not known are explored using the MNIST dataset. Interestingly, it is shown that the infrared modes are learned first.[105]

6. Discussion

As we have seen, the application of machine learning to lattice QCD has garnered much anticipation. However, it has also faced critical opinions. Concerns have been raised about the potential loss of rigor in physics, the black-box nature of machine learning models, and biases within the models. These criticisms are valid. Nonetheless, there is also a risk in not adopting machine learning: being left behind in the advancement of computational science and losing competitiveness.

Machine learning is certainly not a panacea. However, when properly understood and integrated with existing physical theories and numerical calculations, it can become an extremely powerful tool. Machine learning provides efficient and effective solutions for specific tasks, enabling analyses that were previously impossible with traditional methods. As discussed, incorporating machine learning into well-known algorithms can mitigate and resolve biases, thereby enhancing the reliability of the results.

Ultimately, the decision to apply machine learning in lattice QCD must be carefully weighed. Machine learning serves as a complementary tool, and when used with an understanding of its limitations, it can significantly bolster traditional methods and provide new physical insights. By appropriately integrating machine learning, research in lattice QCD can advance further than ever before, leading to deeper understanding and discoveries.

Readers familiar with machine learning are likely to have encountered Rich Sutton's "The Bitter Lesson".[106] Sutton's argument emphasizes that while approaches leveraging human knowledge may be effective in the short term, in the long run, methods that fully exploit computational power prove to be the most successful. In the basics of this assertion, it is reasonable to conclude that, in lattice QCD, general-purpose methods that harness computational resources will eventually outperform more specialized approaches. However, lattice QCD operates within a framework defined by specific physical constraints and theoretical foundations. Therefore, incorporating bias correction and inductive biases into machine learning models is not only conducive to short-term success but also to long-term gains. Since lattice QCD is not concerned with solving generic tasks but rather focused on simulations constrained by known physical laws, there is no contradiction in applying these methods alongside Sutton's insights.

7. Summary

In this review, we have discussed the application of machine learning in lattice QCD. Although lattice QCD is the only method that can quantitatively elucidate nonperturbative phenomena, various methods have been proposed to keep pace with the increasing precision of experiments and observations. Among these, various approaches using neural nets for generating configurations and measurements have been explored, as introduced in this review. It bears repeating that many works not introduced in this review exist. If the reader finds this review interesting, they are encouraged to refer to the original papers.

Acknowledgment The work of A.T. was partially supported by JSPS KAKENHI Grant Numbers 20K14479, 22K03539, 22H05112, and 22H05111, and MEXT as "Program for Promoting Researches on the Supercomputer Fugaku" (Simulation for basic science: approaching the new quantum era; Grant Number JPMXP1020230411, and search for physics beyond the standard model using large-scale lattice QCD simulation and development of AI technology toward next-generation lattice QCD; Grant Number JPMXP1020230409).

*akio@yukawa.kyoto-u.ac.jp

1) R. L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. **2022**, 083C01 (2022).
2) This name is just a remnant of historical origin.
3) M. Srednicki, *Quantum Field Theory* (Cambridge University Press, Cambridge, U.K., 2007). p. 1.
4) K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
5) K. G. Wilson and J. B. Kogut, Phys. Rep. **12**, 75 (1974).
6) J. B. Kogut, Rev. Mod. Phys. **51**, 659 (1979).
7) F. J. Wegner, J. Math. Phys. **12**, 2259 (1971).
8) H. J. Rothe, *Lattice Gauge Theories: An Introduction* (World Scientific, Singapore, 2012) 4th ed., Vol. 43.
9) C. Gattringer and C. B. Lang, *Quantum Chromodynamics on the Lattice* (Springer, Berlin, 2010) Vol. 788.
10) This Anzatz can be derived from $u_\mu = e^{aiA_\mu}$.
11) The continuum limit corresponds to $\beta \to \infty$ limit because this theory is asymptotically free.
12) QCD vacuum corresponds to half-filled in the condensed matter terminology.
13) A. Tanaka, A. Tomiya, and K. Hashimoto, *Deep Learning and Physics* (Springer, 2021) Mathematical Physics Studies.
14) T. S. Cohen and M. Welling, arXiv:1602.07576.
15) M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, arXiv:2104.13478.
16) D. Marcos, M. Volpi, N. Komodakis, and D. Tuia, IEEE International Conference on Computer Vision (ICCV), 2017.
17) G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, Phys. Rev. Lett. **125**, 121601 (2020).
18) D. J. Rezende, G. Papamakarios, S. Racanière, M. S. Albergo, G. Kanwar, P. E. Shanahan, and K. Cranmer, arXiv:2002.02428.
19) D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shanahan, Phys. Rev. D **103**, 074504 (2021).
20) Precisely speaking, additional certain operations are necessary to prevent bias, such as shuffling within each Weyl chamber.[19]
21) M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, Phys. Rev. Lett. **128**, 032003 (2022).
22) Y. Nagai and A. Tomiya, arXiv:2103.11965.
23) M. Albanese et al. (APE Collaboration), Phys. Lett. B **192**, 163 (1987).
24) C. Morningstar and M. J. Peardon, Phys. Rev. D **69**, 054501 (2004).
25) S. Dürr, Comput. Phys. Commun. **172**, 163 (2005).
26) M. Lüscher, J. High Energy Phys. **2010** [08], 071 (2010) [Erratum: J. High Energy Phys. **03**, 092 (2014)].
27) R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, arXiv:1806.07366.
28) P. de Haan, C. Rainone, M. C. N. Cheng, and R. Bondesan, arXiv:2110.02673.
29) J. Achiam et al. (OpenAI), Gpt-4 Technical Report (2024).
30) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, arXiv:1706.03762.
31) Y. Nagai and A. Tomiya, J. Phys. Soc. Jpn. **93**, 114007 (2024).
32) A. Tomiya and Y. Nagai, PoS, LATTICE2023, 2024, 001.
33) A. Tomiya and Y. Nagai, Proceedings of Lattice 2024 at Liverpool, to appear. Jul. 29, 2024.
34) R. Abbott, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, B. Tian, and J. M. Urban, Phys. Rev. D **106**, 074506 (2022).
35) J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, Phys. Rev. B **95**, 041101(R) (2017).
36) S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
37) Y. Nagai, A. Tanaka, and A. Tomiya, Phys. Rev. D **107**, 054501 (2023).
38) Y. Nagai, M. Okumura, K. Kobayashi, and M. Shiga, Phys. Rev. B **102**, 041124(R) (2020).
39) M. S. Albergo, G. Kanwar, and P. E. Shanahan, Phys. Rev. D **100**, 034515 (2019).
40) D. C. Hackett, C.-C. Hsieh, M. S. Albergo, D. Boyda, J.-W. Chen, K.-F. Chen, K. Cranmer, G. Kanwar, and P. E. Shanahan, arXiv:2107.00734.
41) M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, and J. M. Urban, Phys. Rev. D **106**, 014514 (2022).
42) R. Abbott, M. S. Albergo, A. Botev, D. Boyda, K. Cranmer, D. C. Hackett, A. G. D. G. Matthews, S. Racanière, A. Razavi, D. J. Rezende, F. Romero-López, P. E. Shanahan, and J. M. Urban, Eur. Phys. J. A **59**, 257 (2023).
43) R. Abbott, M. S. Albergo, A. Botev, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, A. G. D. G. Matthews, S. Racanière, A. Razavi, D. J. Rezende, F. Romero-López, P. E. Shanahan, and J. M. Urban, PoS, LATTICE2022, 2023, 036.
44) J. Komijani and M. K. Marinkovic, PoS, LATTICE2022, 2023, 019.
45) L. Del Debbio, J. M. Rossney, and M. Wilson, Phys. Rev. D **104**, 094507 (2021).
46) A. Singha, D. Chakrabarti, and V. Arora, Phys. Rev. D **108**, 074518

(2023).

47) R. Abbott, M. S. Albergo, D. Boyda, D. C. Hackett, G. Kanwar, F. Romero-López, P. E. Shanahan, and J. M. Urban, PoS, LATTICE2023, 2024, 035.

48) H. Nicolai, Phys. Lett. B **89**, 341 (1980).

49) M. Lüscher, Commun. Math. Phys. **293**, 899 (2010).

50) L. Dinh, J. Sohl-Dickstein, and S. Bengio, arXiv:1605.08803.

51) M. Lüscher, Commun. Math. Phys. **85**, 39 (1982).

52) K. A. Nicoli, C. J. Anders, T. Hartung, K. Jansen, P. Kessel, and S. Nakajima, Phys. Rev. D **108**, 114501 (2023).

53) R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

54) U. Wolff, Nucl. Phys. B (Proc. Suppl.) **17**, 93 (1990).

55) H. B. Meyer, H. Simma, R. Sommer, M. D. Morte, O. Witzel, and U. Wolff, Comput. Phys. Commun. **176**, 91 (2007).

56) L. Del Debbio, G. M. Manca, and E. Vicari, Phys. Lett. B **594**, 315 (2004).

57) S. Schaefer, R. Sommer, and F. Virotta, Nucl. Phys. B **845**, 93 (2011).

58) M. Gerdes, P. de Haan, C. Rainone, R. Bondesan, and M. C. N. Cheng, SciPost Phys. **15**, 238 (2023).

59) P. Boyle, T. Izubuchi, L. Jin, C. Jung, C. Lehner, N. Matsumoto, and A. Tomiya, PoS, LATTICE2022, 2023, 229.

60) R. Abbott, A. Botev, D. Boyda, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, and J. M. Urban, Phys. Rev. D **109**, 094514 (2024).

61) R. Abbott, D. Boyda, D. C. Hackett, G. Kanwar, F. Romero-López, P. E. Shanahan, J. M. Urban, and M. S. Albergo, PoS, LATTICE2023, 2024, 011.

62) S. Foreman, T. Izubuchi, L. Jin, X.-Y. Jin, J. C. Osborn, and A. Tomiya, PoS, LATTICE2021, 2022, 073.

63) D. Levy, M. D. Hoffman, and J. Sohl-Dickstein, arXiv:1711.09268.

64) S. Foreman, X.-Y. Jin, and J. C. Osborn, 9th International Conference on Learning Representations, 2021, 5.

65) S. Foreman, X.-Y. Jin, and J. C. Osborn, PoS, LATTICE2021, 2022, 508.

66) S. Foreman, X.-Y. Jin, and J. C. Osborn, MLMC: Machine Learning Monte Carlo for Lattice Gauge Theory, 2023, 12.

67) H. Wu, J. Köhler, and F. Noé, arXiv:2002.06707.

68) M. Caselle, E. Cellini, A. Nada, and M. Panero, J. High Energy Phys. **2022** [07], 015 (2022).

69) C. Bonanno, A. Nada, and D. Vadacchino, J. High Energy Phys. **2024** [04], 126 (2024).

70) M. Caselle, E. Cellini, and A. Nada, arXiv:2409.15937.

71) A. Bulgarelli, E. Cellini, and A. Nada, 41st International Symposium on Lattice Field Theory, 2024, 9.

72) A. Bulgarelli, E. Cellini, and A. Nada, arXiv:2412.00200.

73) L. Wang, G. Aarts, and K. Zhou, 37th Conference on Neural Information Processing Systems, 2023, 11.

74) S. Soma, L. Wang, S. Shi, H. Stöcker, and K. Zhou, PoS, FAIRness2022, 2023, 055.

75) P. Hasenfratz and F. Niedermayer, Nucl. Phys. B **414**, 785 (1994).

76) K. Holland, A. Ipp, D. I. Müller, and U. Wenger, Phys. Rev. D **110**, 074502 (2024).

77) B. Yoon, T. Bhattacharya, and R. Gupta, Phys. Rev. D **100**, 014504 (2019).

78) B. J. Choi, H. Ohno, T. Sumimoto, and A. Tomiya, Proceedings of Lattice 2024 at Liverpool, to appear. Jul. 29, 2024.

79) H. Wittig, A. Conigli, A. Segner, L. Geyer, S. Kuberski, and T. Blum, Proceedings of Lattice 2024 at Liverpool, to appear. Jul. 31, 2024.

80) T. Bhattacharya, S. Lawrence, and J.-S. Yoo, Phys. Rev. D **109**, L031505 (2024).

81) P. F. Bedaque and H. Oh, Phys. Rev. D **109**, 094519 (2024).

82) W. Detmold, G. Kanwar, H. Lamm, M. L. Wagman, and N. C. Warrington, Phys. Rev. D **103**, 094517 (2021).

83) W. Detmold, G. Kanwar, M. L. Wagman, and N. C. Warrington, Phys. Rev. D **102**, 014514 (2020).

84) C. Lehner and T. Wettig, Phys. Rev. D **110**, 034517 (2024).

85) C. Lehner and T. Wettig, Phys. Rev. D **108**, 034503 (2023).

86) M. Asakawa, T. Hatsuda, and Y. Nakahara, Prog. Part. Nucl. Phys. **46**, 459 (2001).

87) H. Shinaoka, J. Otsuki, M. Ohzeki, and K. Yoshimi, Phys. Rev. B **96**, 035147 (2017).

88) J. Otsuki, M. Ohzeki, H. Shinaoka, and K. Yoshimi, Phys. Rev. E **95**, 061302(R) (2017).

89) E. Itou and Y. Nagai, J. High Energy Phys. **2020** [07], 007 (2020).

90) J. Takahashi, H. Ohno, and A. Tomiya, PoS, LATTICE2023, 2024, 028.

91) J. Rojo, S. Forte, G. Ridolfi, R. D. Ball, L. Del Debbio, M. Ubiali, V. Bertone, A. Guffanti, F. Cerutti, and J. I. Latorre, PoS, DIS2010, 2010, 244.

92) It is often stated that finding a general solution to the sign problem is known to be NP-hard.[107]

93) In this case, thermal state is not efficiently simulated.

94) L. Funcke, T. Hartung, K. Jansen, and S. Kühn, PoS, LATTICE2022, 2023, 228.

95) Y. Mori, K. Kashiwa, and A. Ohnishi, Phys. Rev. D **96**, 111501 (2017).

96) Y. Mori, K. Kashiwa, and A. Ohnishi, Prog. Theor. Exp. Phys. **2018**, 023B04 (2018).

97) K. Kashiwa, Y. Mori, and A. Ohnishi, Phys. Rev. D **99**, 014033 (2019).

98) K. Kashiwa and Y. Mori, Phys. Rev. D **102**, 054519 (2020).

99) Y. Namekawa, K. Kashiwa, A. Ohnishi, and H. Takase, Phys. Rev. D **105**, 034502 (2022).

100) Y. Namekawa, K. Kashiwa, H. Matsuda, A. Ohnishi, and H. Takase, Phys. Rev. D **107**, 034509 (2023).

101) K. Kashiwa, Y. Namekawa, A. Ohnishi, and H. Takase, Phys. Rev. D **108**, 094504 (2023).

102) A. Alexandru, G. Basar, P. F. Bedaque, and N. C. Warrington, Rev. Mod. Phys. **94**, 015006 (2022).

103) M. Rodekamp, E. Berkowitz, C. Gäntgen, S. Krieg, T. Luu, and J. Ostmeyer, Phys. Rev. B **106**, 125139 (2022).

104) P. Smolensky, Chapter 6: Information Processing in Dynamical Systems: Foundations of Harmony Theory, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, ed. D. E. Rumelhart and J. L. McClelland (MIT Press, Cambridge, MA, 1986) Foundations, Vol. 1, p. 194.

105) G. Aarts, B. Lucini, and C. Park, Phys. Rev. D **109**, 034521 (2024).

106) R. Sutton, The Bitter Lesson, March 2019. Accessed: 2024-10-14.

107) M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005).

Akio Tomiya was born in Hyogo, Japan in 1987. He studied physics at the Graduate School of Science, Osaka University, obtaining a Ph.D. in 2015. In late 2015, he was appointed as a postdoctoral researcher at Central China Normal University (CCNU) in Wuhan, China, where he remained until early 2018. In late 2018, he joined the RIKEN BNL Research Center as a Postdoctoral Researcher (SPDR), a position he held until early 2021. In late 2021, he moved to International Professional University of Technology in Osaka, Faculty of Engineering, Department of Information Engineering, where he served until early 2024. He has been the recipient of the 29th JPS Paper Award in 2024 and the 14th Particle Physics Medal: Young Scientist Award in Theoretical Particle Physics in 2019. Since 2024, he has held a full-time lectureship at Tokyo Woman's Christian University, School of Arts and Sciences, Division of Mathematical Sciences.