

# Study of topological quantities of lattice QCD with a modified Wasserstein generative adversarial network

Lin Gao<sup>1,\*</sup>, Heping Ying<sup>2</sup>, and Jianbo Zhang<sup>2</sup>

<sup>1</sup>American Association for the Advancement of Science, Washington, DC 20005, USA

<sup>2</sup>School of Physics, Zhejiang University, Hangzhou 310027, China



(Received 8 October 2023; accepted 22 March 2024; published 22 April 2024)

We propose a modified Wasserstein generative adversarial network (M-WGAN) to study the distribution of the topological charge in lattice QCD based on Monte Carlo simulations. We construct new generator and discriminator in M-WGAN to support the generation of high-quality distribution. Our results show that the M-WGAN scheme of machine learning should be helpful for us to calculate efficiently the 1D distribution of topological charge compared with the method by the Monte Carlo simulation alone.

DOI: [10.1103/PhysRevD.109.074509](https://doi.org/10.1103/PhysRevD.109.074509)

## I. INTRODUCTION

Machine learning (ML) has improved the state-of-the-art technology in many domains [1–3]. Especially it has had many applications in lattice QCD. Kim A. Nicoli *et al.* exerted the deep generative models to estimate the absolute value of the free energy in lattice field theories [4]. Ankur Singha *et al.* developed a conditional normalizing flow-based sampling method for scalar lattice  $\phi^4$  theory to improve the problem of critical slowing down [5]. Yuki Nagai *et al.* developed the self-learning Monte Carlo (MC) simulations to resolve the autocorrelation problem [6]. Matteo Favoni and Andreas Ipp *et al.* propose lattice gauge equivariant convolutional neural networks to study applications of generic machine learning in lattice gauge theory [7]. These works have shown the great impact of ML in lattice QCD, including the calculation of physical quantities, the improvement of research methods, the reduction in autocorrelation, and the study of gauge invariant quantities, etc. The conventional computation of topological charge in lattice QCD based on the MC method consumes a lot of computation time [8,9], and ML could be explored as a potential method for generating topological quantities efficiently. Our research focuses on constructing new generator and discriminator of Wasserstein generative adversarial network (M-WGAN) to facilitate the exploration of the topological quantities in lattice QCD. It is committed to generating corresponding topological charge data based on a small amount of MC

data to calculate the topological susceptibility. In the following, we begin our studies with a brief introduction to the lattice QCD and ML.

In QCD, it is difficult to use perturbation methods like quantum electrodynamics in the low-energy region since the coupling constant of QCD becomes large [10]. Therefore, lattice QCD was introduced to study the QCD in the low-energy region [8], where Euclidean spacetime instead of Minkowski spacetime is used, and this simplifies the computations (as real numbers rather than complex numbers are involved). Furthermore, the action used here is that of the Wilson gauge action [9]

$$S_G[U] = \frac{\beta}{3} \sum_{n \in \Lambda} \sum_{\mu < \nu} \text{ReTr}[1 - U_{\mu\nu}(n)], \quad (1)$$

where  $\beta$  is the inverse coupling and  $U_{\mu\nu}(n)$  is the plaquette.

The topological charge density discussed in this article is defined by

$$q(n) = \frac{1}{32\pi^2} \epsilon_{\mu\nu\rho\sigma} \text{ReTr}[F_{\mu\nu}^{\text{clov}}(n) F_{\rho\sigma}^{\text{clov}}(n)], \quad (2)$$

related to the nonconserved axial vector current [11].  $F_{\mu\nu}^{\text{clov}}(n)$  is the clover improved lattice discretization of the field strength tensor  $F_{\mu\nu}(x)$  and can be noted as

$$F_{\mu\nu}^{\text{clov}}(n) = -\frac{i}{8a^2} \left[ (C_{\mu\nu}(n) - C_{\mu\nu}^\dagger(n)) - \frac{\text{Tr}(C_{\mu\nu}(n) - C_{\mu\nu}^\dagger(n))}{3} \right], \quad (3)$$

where the clover is given as

$$C_{\mu\nu}(n) = U_{\mu,\nu}(n) + U_{\nu,-\mu}(n) + U_{-\mu,-\nu}(n) + U_{-\nu,\mu}(n). \quad (4)$$

\*silvester\_gao@qq.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

The topological charge is further introduced as

$$Q_{\text{top}} = a^4 \sum_{n \in \Lambda} q(n), \quad (5)$$

which converges to an integer in the continuum limit [12]. The above nonconserved axial vector current is related to a symmetry breaking, which is called the  $U(1)_A$  anomaly. From a physical point of view, a possible explanation is that an instanton can be introduced to obtain the conserved charge [13]. Furthermore, the topological susceptibility  $\chi_t$  is defined as

$$\chi_t = \frac{1}{V} \langle Q_{\text{top}}^2 \rangle, \quad (6)$$

where  $V$  is the 4D volume. The  $\chi_t$  is related to the Witten-Veneziano relation, which indicates that  $\chi_t$  is proportional to the mass squared of  $\eta'$  meson for massless quarks [14]. In our investigations, when computing the topological susceptibility from the initial configurations, we incorporate the Wilson flow as a technique to smooth the configurations and alleviate ultraviolet divergences [15–17].

On the other hand, ML can be classified into supervised learning, unsupervised learning, semisupervised learning, transfer learning, and reinforcement learning. Supervised learning in machine learning typically demands millions of training examples to attain optimal results [3]. However, it proves impractical when dealing with scenarios involving just a few hundred training data points. Generative adversarial networks (GANs) were proposed in 2014 [2], which would gradually improve the capabilities of generator and discriminator through their mutual adversarial game. It is often necessary to know the probability distribution of a set of data, such as the age distribution of a biological population or the distribution of pixels in an image. Compared with the method of maximum likelihood estimation, which directly estimates the parameters of the probability density, GANs are implicit models that can infer the probability distribution  $p(x)$  without explicitly expressing the probability density function [3]. For simple distribution like the age distribution of a biological population, one can guess a probability density function and then use the data to estimate the parameters of the probability density function. However, the probability density function may have millions or even billions of parameters for complicated distribution, and it will be difficult to guess a probability density function in this case. In contrast, the implicit models GANs have a good performance in data generation for the complex distribution. The topological charge that we are going to discuss obeys a certain distribution, so GANs can be used to generate data of the topological charge. After GANs were proposed, there have been many variants, one of which is

Wasserstein generative adversarial network (WGAN) [18]. The WGAN introduces the Wasserstein distance

$$\mathbb{E}_{x \sim \mathbb{P}_r}[D(x)] - \mathbb{E}_{x \sim \mathbb{P}_G}[D(x)], \quad (7)$$

where  $\mathbb{P}_r$  is the distribution of real data,  $\mathbb{P}_G$  is the distribution generated by the generator and the discriminator  $D(x)$  is required to be a 1-Lipschitz function. As a result, WGAN greatly improves the stability of GAN training and the quality of results.

## II. MODEL AND DATA PREPARATIONS

For our purpose, the suitable ML model should be constructed to explore the characteristics of the topological charge in lattice QCD. We checked that the results will be poor if the neural networks of generator and discriminator are too simple in WGAN through testing. For example, if we use a simple neural network to generate the distribution of the topological charge, we will find that the symmetry of the distribution is not good or many bars of the distribution histogram stick together instead of being discrete. Therefore, the new generator and discriminator are constructed to study the distribution of the topological charge.

The overall structure of M-WGAN is shown in Fig. 1. The real distribution in Fig. 1 refers to the distribution generated by the MC method, and the fake distribution refers to the distribution generated by machine learning at a certain epoch in the training process. These two distribution histograms are only used to help illustrate the outline of M-WGAN in the training process. The training process of M-WGAN is similar to that of

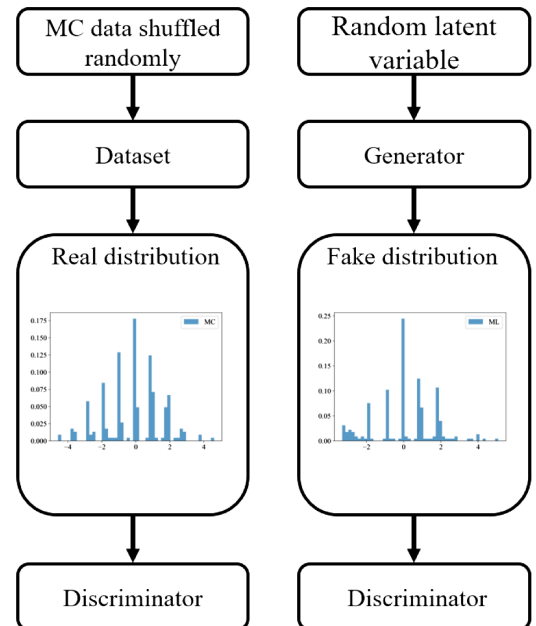


FIG. 1. The overall structure of M-WGAN.

TABLE I. The structure of the generator for M-WGAN.

Layer (type)	Output shape	Parameter number
Linear-1	$[-1, 6400]$	6,400
BatchNorm1d-2	$[-1, 6400]$	12,800
LeakyReLU-3	$[-1, 6400]$	0
Linear-4	$[-1, 100]$	640,000
BatchNorm1d-5	$[-1, 100]$	200
LeakyReLU-6	$[-1, 100]$	0
Linear-7	$[-1, 4]$	400
BatchNorm1d-8	$[-1, 4]$	8
LeakyReLU-9	$[-1, 4]$	0
Linear-10	$[-1, 100]$	400
BatchNorm1d-11	$[-1, 100]$	200
LeakyReLU-12	$[-1, 100]$	0
Linear-13	$[-1, 4]$	400
BatchNorm1d-14	$[-1, 4]$	8
LeakyReLU-15	$[-1, 4]$	0
Linear-16	$[-1, 100]$	400
BatchNorm1d-17	$[-1, 100]$	200
LeakyReLU-18	$[-1, 100]$	0
Linear-19	$[-1, 6400]$	640,000
BatchNorm1d-20	$[-1, 6400]$	12,800
LeakyReLU-21	$[-1, 6400]$	0
Linear-22	$[-1, 1]$	6,400

WGAN [3,18]. The differences are mainly in the sampling method of training data and the structure of the generator and the discriminator.

The structure of the generator for M-WGAN is explained in Table I. The fully connected layers are applied to reshape the input random latent variable with normal distribution. The batch normalization is exerted to improve generation. The structure of the discriminator for M-WGAN is described in Table II. Three fully connected layers are used. LeakyReLU activation is implemented in all layers except for the output layer.

The input of generator for M-WGAN is a random latent variable tensor with shape  $[-1, 1]$ , and the output is a tensor with shape  $[-1, 1]$ , where  $-1$  is an undetermined parameter. For example, we need to generate 1600 topological charge values, the input shape is  $[1600, 1]$ , and the output shape is also  $[1600, 1]$ . Finally, the output needs to be reshaped into an 1600-dimensional vector. These 1600 data can form a distribution.

TABLE II. The structure of the discriminator for M-WGAN.

Layer (type)	Output shape	Parameter number
Linear-1	$[-1, 64]$	128
LeakyReLU-2	$[-1, 64]$	0
Linear-3	$[-1, 3200]$	208,000
LeakyReLU-4	$[-1, 3200]$	0
Linear-5	$[-1, 1]$	3,201

We update the parameters to maximize the Wasserstein distance and the optimizer is RMSProp [18]. These settings are the same as those of the original WGAN.

As a result, the M-WGAN can generate the distribution of the topological charge directly to be applied to calculate the topological charge susceptibilities after training. The M-WGAN realize unsupervised generation without labels. In addition, some programs are based on Pytorch [19].

The next part is the preparation of original data. The software Chroma [20] is used to generate configurations of pure gauge field on individual workstation. In order to obtain the topological charge density data, the configurations are simulated first with the pseudo heat bath algorithm and smoothed by Wilson flow [9], then topological charge density data can be calculated from such configurations. In detail, the periodic boundary conditions and hot start have been applied and the updating steps are repeated 10 times for the visited link variable because the computation of sum of staples is costly.

The Wilson flow step time  $\varepsilon_f = 0.01$  and total number of steps  $N_{\text{flow}} = 600$  are chosen. The reason why we use these two parameters are as follows.  $\sqrt{8t_f}$  is used to characterize the smoothing range [16,21]. In the Sec. 6.2.1 of [21], the author says that when the flow smearing exceeds the lattice spacing, such that  $\sqrt{8t_f}/a \gg 1$ , the discretization effects will be less visible. In this paper, we use the parameters  $N_{\text{flow}} = 600$  and  $\varepsilon_f = 0.01$ , so that  $\sqrt{8t_f}/a \approx 6.93 \gg 1$ . Therefore, the degree of smearing is sufficient and  $N_{\text{flow}} = 600$  is enough to remove the ultraviolet divergences. In addition, referring to Appendix C of Ref. [16], we discover that the error of the Runge-Kutta scheme is small enough when  $\varepsilon_f = 0.01$ .

The demonstration that the sampled configurations has reached thermalization is as follows. To detect whether the system has reached thermalization, it is possible to check a physical quantity starting from cold start and hot start is consistent after a certain number of MC steps. The average of the plaquette with  $\beta = 6.0$  was chosen to study the thermalization and the mean value of plaquette is defined as  $\langle \frac{1}{3} \text{ReTr} U_{\text{plaq}} \rangle$ . It is found from Fig. 2 that the system has reached thermalization after about 200 sweeps because the mean value of plaquette has evolved to a similar value starting from either cold or hot start.

Furthermore, it is necessary to calculate the integrated autocorrelation time which is computed by the topological charge in this article. For a Markov sequence of field configurations generated by MC simulations,  $X_i$  is a random variable and we could introduce the autocorrelation function

$$\begin{aligned}
 C_X(X_i, X_{i+t}) &= \langle (X_i - \langle X_i \rangle)(X_{i+t} - \langle X_{i+t} \rangle) \rangle, \\
 &= \langle X_i X_{i+t} \rangle - \langle X_i \rangle \langle X_{i+t} \rangle.
 \end{aligned} \tag{8}$$

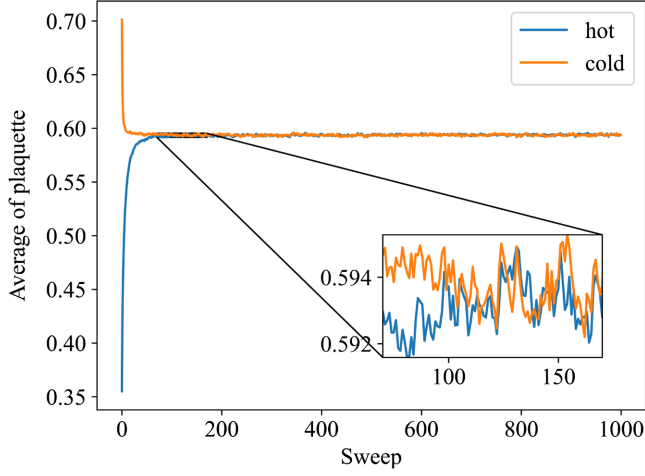


FIG. 2. The evolution of the average of plaquette under different initial conditions.

We define  $C_X(t) = C_X(X_i, X_{i+t})$  at thermalization and the normalized autocorrelation function  $\Gamma_X(t) = \frac{C_X(t)}{C_X(0)}$ . We introduce the integral autocorrelation time as

$$\tau_{X,\text{int}} = \frac{1}{2} + \sum_{t=1}^N \Gamma_X(t). \quad (9)$$

It is calculated that the integrated autocorrelation time is 0.416. This result indicates that each of data can be regarded as independent because of  $N/2\tau_{X,\text{int}} > N$  [9], where a total of 1000 configurations are sampled with intervals of 200 sweeps.

In addition, the static QCD potential is applied to set the scale and can be parametrized by [9]

$$V(r) = A + \frac{B}{r} + \sigma r. \quad (10)$$

The Sommer parameter  $r_0$  is defined as

$$\left( r^2 \frac{dV(r)}{dr} \right)_{r=r_0} = 1.65, \quad (11)$$

and the Sommer parameter  $r_0 = 0.49\text{fm}$  is used [22]. The scales are summarized in Table III.

TABLE III. The parameters used for the production of configurations and the associated value of  $r_0$ .

Volume	$\beta$	$a[\text{fm}]$	$N_{\text{cnfg}}$	$r_0/a$	$L[\text{fm}]$
$24 \times 12^3$	6.0	0.093(3)	1000	5.30(15)	1.11(3)

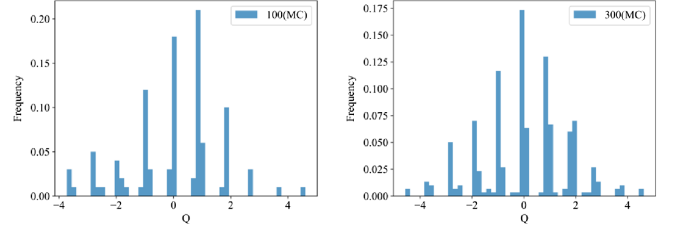


FIG. 3. The distribution of topological charge based on the MC simulations with Wilson flow. A total number of Wilson flow time steps  $N_{\text{flow}} = 600$  are chosen. The numbers of topological charge are 100 for the left plot and 300 for the right plot.

### III. NUMERICAL RESULTS

First, we show our results of 100 and 300 topological charge data simulated by the MC simulations with Wilson flow. It is found from Fig. 3 that the topological charge  $Q$  is approximately concentrated on the integer positions being consistent with the conclusion that the topological charge converges to an integer in the continuum limit mentioned above. It is calculated from Fig. 3 (right) that the fourth root of topological susceptibility  $\chi_t^{1/4} = 191.8 \pm 3.9 \text{ MeV}$  when  $N_{\text{flow}} = 600$ . Therefore,  $a\chi_t^{1/4}$  is approximately equal to 0.09, which is consistent with Ref. [23]. Moreover, the distribution of the topological charge should be symmetrical about the origin. However, we can find that the symmetry of distribution is poor in Fig. 3 (left) due to the poor statistics of data. Therefore, it is important to improve the distribution from the data of increased statistics. In MC simulations, increasing data may result in a rapid increase in computation time and storage usage, but these problems can be almost avoided with an appropriate ML model. Once the ML model is trained, it can immediately generate a corresponding data to improve the accuracy of the results. Next, we will discuss the details of two methods, the MC simulations with Wilson flow and ML with the M-WGAN scheme, as well as combine two methods to generate configurations more efficiently.

For the MC method, we have simulated a total of 1600 configurations and calculated the topological charge from the configurations. Then the fourth root of the topological susceptibility  $\chi_t^{1/4} = 190.9 \pm 1.7 \text{ MeV}$  was calculated from 1600 topological charge data when 25 CPU cores were used in our calculations.

For ML, we concern to apply M-WGAN to generate the topological charge data based on the original MC training data. At first, we need to determine how much training data

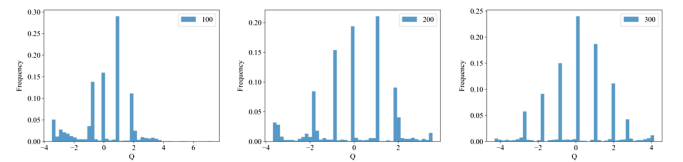


FIG. 4. Comparison of models with 100, 200, 300 training data.



TABLE IV. The fourth root of the topological susceptibility  $\chi_t^{1/4}$  for MC and ML.

Data volume	$\chi_t^{1/4}(\text{MC})/\text{MeV}$	$\chi_t^{1/4}(\text{ML})/\text{MeV}$
400	$191.2 \pm 3.5$	$191.7 \pm 3.3$
600	$189.4 \pm 2.9$	$190.7 \pm 2.6$
800	$190.6 \pm 2.4$	$191.5 \pm 2.3$
1000	$191.4 \pm 2.1$	$190.7 \pm 2.1$
1200	$191.2 \pm 2.0$	$191.7 \pm 1.9$
1400	$191.6 \pm 1.8$	$191.7 \pm 1.8$
1600	$190.9 \pm 1.7$	$191.0 \pm 1.6$

to be suitable. We tested the training processes with total training data volumes of 100, 200, and 300 and randomly select three-quarters of data volumes each iteration to train the model. The distribution of topological charge generated by the model trained using different data volumes is shown in the Fig. 4. It can be observed that the distribution of Fig. 4 (middle) is mainly concentrated at the integers compared with Fig. 4 (left), but its peak does not align with the origin. Furthermore, the distribution of Fig. 4 (right) is mainly concentrated at the integers and roughly symmetrical about the origin. Therefore, it is found that the model trains better as the amount of training data increases. By the test experiences, we choose to use 300 MC data as the ML training data sample and randomly pick out three-quarters of the 300 data in a training step to perform our ML train scheme. The total amount of training data we use is 300, and we randomly select three-quarters of 300 data, that is 225 data, to train the machine learning model each iteration. In other words, the batch size is 225. Compared with fixed selection, random selection can increase the diversity of data, which is more conducive to the training of machine learning model. In addition, three-quarters of the 300 data are selected each time because such a large amount of data is more beneficial in training a better distribution.

For our study, we should notice the accuracy of the physical results obtained with the MC and ML methods. It can be found from Table IV and Fig. 5 that the data error

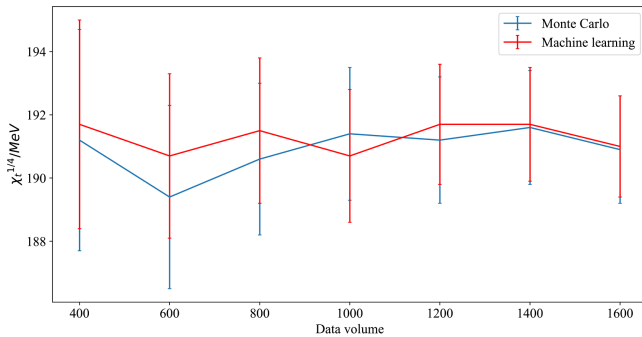
FIG. 5. The fourth root of the topological susceptibility  $\chi_t^{1/4}$  for MC and ML methods under different data size.

TABLE V. The time consumption and hard disk storage usage for MC and ML methods. The time and storage of M-WGAN incorporate the effects of training data.

Method	Time/h	Storage/MB	$\chi_t^{1/4}/\text{MeV}$
MC	136	18230	$190.9 \pm 1.7$
ML	26	3423	$191.0 \pm 1.6$

gradually decreases as size of data increases for both MC and ML, where 300 training data were used to train M-WGAN and then 1600 output data were obtained. We apply the jack knife to analysis the error of data in calculations [9].

As shown in Table V, the results from both methods are consistent within the statistical errors. But it is much faster to generate the same amount of data with the ML method, which also takes up less storage space than the MC method. Therefore, we can apply ML to generate suitable data based on MC data to deal with the error more efficiently.

Furthermore, the distributions of topological charge generated by the MC and ML methods are shown in Fig. 6. We can conclude that both the distributions have good symmetry, and their data are discretely distributed at integers. These characteristics are consistent with the nature of the topological charge. In addition, the integrated autocorrelation time of 1600 data for ML is 0.41, which indicates that these data are independent.

#### IV. CONCLUSION

In this paper, we have studied the topological quantities of lattice QCD by using MC simulations combined with the M-WGAN. Especially we have applied the M-WGAN to generate the distribution of topological charge to show the potential for applications of ML technique in a study of lattice QCD.

By our experience, the conclusions are as follows. Compared with the MC by the pseudo heat bath algorithm and the Wilson flow, the M-WGAN shows its efficiency for the MC simulations in terms of time cost and storage as shown in Tables IV and V. The data generated by the M-WGAN trained with 300 data are more accurate than the corresponding data by the MC simulation as the topological susceptibility is concerned.

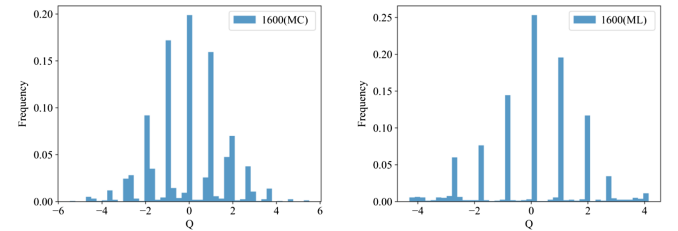


FIG. 6. The distributions of topological charge for MC and ML methods.

The pseudodistribution generated by this model after a necessary training can be applied to calculate the correct topological susceptibility in the SU(3) lattice QCD.

For future studies, we hope that the M-WGAN can be applied to tackle other physics problems in lattice QCD and provide an alternative approach to simulating some interesting and important quantities in a much more efficient way.

## ACKNOWLEDGMENTS

This work was done based on the chroma applied to simulate the configurations of the lattice gauge field. We are grateful to the relevant contributors to the chroma, and H. Y. would like to thank all his coworkers for lattice QCD field studies.

- 
- [1] J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
  - [2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, in *Advances in Neural Information Processing Systems* (2014), Vol. 27.
  - [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Commun. ACM* **63**, 139 (2020).
  - [4] K. A. Nicoli, C. J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel, S. Nakajima, and P. Stornati, *Phys. Rev. Lett.* **126**, 032001 (2021).
  - [5] A. Singha, D. Chakrabarti, and V. Arora, *Phys. Rev. D* **107**, 014512 (2023).
  - [6] Y. Nagai, A. Tanaka, and A. Tomiya, *Phys. Rev. D* **107**, 054501 (2023).
  - [7] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, *Phys. Rev. Lett.* **128**, 032003 (2022).
  - [8] K. G. Wilson, *Phys. Rev. D* **10**, 2445 (1974).
  - [9] C. Gattringer and C. B. Lang, *Quantum Chromodynamics on the Lattice* (Springer, Berlin, 2010), Vol. 788, 10.1007/978-3-642-01850-3.
  - [10] H. Georgi and H. D. Politzer, *Phys. Rev. D* **14**, 1829 (1976).
  - [11] G. 't Hooft, *Phys. Rep.* **142**, 357 (1986).
  - [12] M. F. Atiyah and I. M. Singer, *Ann. Math.* **93**, 139 (1971).
  - [13] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Y. S. Tyupkin, *Phys. Lett.* **59B**, 85 (1975).
  - [14] E. Witten, *Nucl. Phys.* **B156**, 269 (1979).
  - [15] M. Reuter, *Mod. Phys. Lett. A* **12**, 2777 (1997).
  - [16] M. Lüscher, *J. High Energy Phys.* **08** (2010) 071.
  - [17] P. J. Moran, D. B. Leinweber, and J. Zhang, *Phys. Lett. B* **695**, 337 (2011).
  - [18] M. Arjovsky, S. Chintala, and L. Bottou, [arXiv:1701.07875](https://arxiv.org/abs/1701.07875).
  - [19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga *et al.*, *Advances in Neural Information Processing Systems* (2019), Vol. 32.
  - [20] R. G. Edwards and B. Joó, *Nucl. Phys.* **B140**, 832 (2005).
  - [21] H. M. M. Vege, Solving SU(3) Yang-Mills theory on the lattice: A calculation of selected gauge observables with gradient flow, Master's thesis, University of Oslo (main), 2019.
  - [22] R. Sommer, *Proc. Sci. LATTICE2013* (**2014**) 015 [[arXiv:1401.3270](https://arxiv.org/abs/1401.3270)].
  - [23] A. Athenodorou and M. Teper, *J. High Energy Phys.* **11** (2020) 172.