Signal-to-noise improvement through neural network contour deformations for 3D $SU(2)$ lattice gauge theory

William Detmold,[a,b] **Gurtej Kanwar,**[c] **Yin Lin,**[a,b,*] **Phiala E. Shanahan**[a,b] **and Michael L. Wagman**[d]

[a]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b]*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[c]*Albert Einstein Center, Institute for Theoretical Physics, University of Bern, 3012 Bern, Switzerland*
[d]*Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*

E-mail: yin01@mit.edu

Complex contour deformations of the path integral have been demonstrated to significantly improve the signal-to-noise ratio of observables in previous studies of two-dimensional gauge theories with open boundary conditions. In this work, new developments based on gauge fixing and a neural network definition of the deformation are introduced, which enable an effective application to theories in higher dimensions and with generic boundary conditions. Improvements of the signal-to-noise ratio by up to three orders of magnitude for Wilson loop measurements are shown in $SU(2)$ lattice gauge theory in three spacetime dimensions.

*Speaker

https://pos.sissa.it/

1. Introduction

The problem of statistical noise is ubiquitous in lattice field theory calculations, which employ Monte Carlo sampling to estimate field theory expectation values. Euclidean correlation functions $C(\tau)$ parameterized by imaginary-time separation τ are generically affected by an exponential decay of the signal-to-noise (StN) ratio with increasing τ [1, 2],

$$\text{StN}[C(\tau)] \equiv |\langle C(\tau)\rangle| / \sqrt{\text{Var}[C(\tau)]} \sim \exp(-\text{const.} \times \tau). \tag{1}$$

This severe *signal-to-noise problem* crucially limits the values of τ which provide meaningful information at a given level of precision. The signal-to-noise problem also affects correlation functions parameterized by multiple distance scales. In lattice gauge theories, the expectation values of rectangular Wilson loops $W_{r\times\tau}$ of dimension $r \times \tau$ can be interpreted as correlation functions of a static particle/anti-particle pair separated by a spatial distance r and Euclidean time τ. A hallmark of confinement in pure gauge theories is the area-law scaling of Wilson loop expectation values, $\langle W_{r\times\tau}\rangle \sim e^{-\sigma A}$, where $A = r \times \tau$ and σ is the confining string tension [3]. The signal-to-noise problem can be shown to be maximally severe in this case, with $\text{StN} \sim e^{-\sigma A}$, making this an interesting test bed for approaches to mitigate the signal-to-noise problem in gauge theories.

Recently, [4, 5] introduced an approach to signal-to-noise improvement based on *complex contour deformations* of the path integral, which can be used to define families of observables with identical expectation values but generally distinct variance properties. Similar approaches have been applied to solving sign problems introduced by a complex action [6–16], for example arising from a non-zero chemical potential or a real-time path integral; see [15] for a comprehensive review. Optimizing the choice of deformation was shown in [4, 5] to exponentially reduce the noise of the Euclidean-time correlation function of Wilson loops in 2D $U(1)$ and $SU(N)$ lattice gauge theories. However, a key simplification in these results was to first change integration variables in the path integral from gauge links to (untraced) plaquettes as allowed by the 2D geometry with open boundary conditions.

In this work, we introduce a new approach to defining contour deformations of the path integral for $SU(N)$ lattice gauge theory which is applicable to gauge theories defined in arbitrary spacetime dimensions. In contrast to previous work, these contour deformations are defined to act directly on link degrees of freedom and are therefore applicable to theories defined with generic boundary conditions, including those with periodic boundaries. We introduce and demonstrate the approach in the context of an $SU(2)$ pure gauge theory in 3D, where it is possible to exponentially improve the signal-to-noise ratio of Wilson loops over a wide range of areas. We also investigate the lattice volume dependence of these results. Finally, we conclude by discussing the straightforward generalization to $SU(N)$ and higher dimensions and giving an outlook for this work.

2. Path integral deformations for signal-to-noise improvement

We start from the Euclidean path integral definition of the expectation value of a quantum operator O in a lattice gauge theory,

$$\langle O\rangle \equiv \frac{1}{Z} \int d[U]O[U]e^{-S[U]}, \quad Z \equiv \int d[U]e^{-S[U]}, \tag{2}$$

where $\int d[U]$ indicates integration with respect to the Haar measure of gauge field configurations, $S[U]$ is a discretized action, and $O[U]$ is the representation of the operator acting on the field configurations. In this work we assume $S[U] \in \mathbb{R}$, in which case equation (2) is typically estimated by a statistical expectation value, $\langle O \rangle = \mathbb{E}_S[O[U]]$, where $\mathbb{E}_S[\cdot]$ expresses the statistical mean over field configurations U sampled according to the probability measure $d[U]e^{-S[U]}$.

When the Boltzmann weight $\exp(-S[U])$ and operator $O[U]$ are complex analytic ("holomorphic"), the path integral can be *deformed* continuously into the complexified domain of gauge field configurations without modifying its value, as a result of Cauchy's theorem [6, 17]. For continuous gauge groups, the complexified space can be straightforwardly identified by complexifying the associated Lie algebra; for example, $SU(N)$ is complexified to $SL(N, \mathbb{C})$, the group of $N \times N$ matrices with unit determinant. Typically, the weight $\exp(-S[U])$ and operators $O[U]$ are initially defined only on the original (real) domain of integration, but these can be analytically continued to the complexified domain by replacing instances of the Hermitian conjugate links $U_\mu^\dagger(x)$ with inverses $U_\mu^{-1}(x)$.

Following [4, 5], we consider a deformation of the original domain of the integral in equation (2) to a new domain $\widetilde{\mathcal{M}}$, abstractly written as

$$\langle O \rangle = \frac{1}{Z} \int_{\widetilde{\mathcal{M}}} d[\tilde{U}] O[\tilde{U}] e^{-S[\tilde{U}]}, \tag{3}$$

where we leave the evaluation of Z unchanged. In practice, the integration over the new manifold $\widetilde{\mathcal{M}}$ is defined by an injective map $\tilde{U}(U)$, such that we can write the result of the deformation as

$$\begin{aligned}
\langle O \rangle &= \frac{1}{Z} \int d[U] J[U] O[\tilde{U}(U)] e^{-S[\tilde{U}(U)]} \\
&= \mathbb{E}_S \left[J[U] O[\tilde{U}(U)] e^{-S[\tilde{U}(U)]+S[U]} \right] \equiv \mathbb{E}_S \left[Q[U] \right],
\end{aligned} \tag{4}$$

where $J[U] = \det(\partial \tilde{U}/\partial U)$ is the Jacobian of the map. Because the new observable $Q[U] = J[U]O[\tilde{U}(U)]e^{-S[\tilde{U}(U)]+S[U]}$ has identical expectation value to O, one can choose to estimate it instead of $O[U]$. It will generally have different variance properties from $O[U]$, and by optimizing the choice of deformation $\tilde{U}(U)$ this variance can be minimized.

3. Contour deformation of $SU(2)$ Wilson loops

We next focus on the signal-to-noise problem of Wilson loops of various sizes in $SU(2)$ pure gauge theory in three dimensions, introducing our choice of deformations $\tilde{U}(U)$ in this context.

3.1 Complexification and analytic continuation

We consider the Wilson gauge action, which is given by

$$S[U] = -\frac{\beta}{2N_c} \sum_{x,\mu<\nu} \mathrm{Tr}\left(P_{\mu\nu}(x) + P_{\mu\nu}^{-1}(x)\right),$$
$$P_{\mu\nu}(x) = U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^{-1}(x+\hat{\nu})U_\nu^{-1}(x). \tag{5}$$

Here, β is the inverse coupling constant and $P_{\mu\nu}$ is the oriented plaquette located at position x. For the $SU(2)$ gauge group, $N_c = 2$ and $U_\mu(x), P_{\mu\nu}(x) \in SU(2)$. We have analytically continued the action in the usual way by using inverses $P_{\mu\nu}^{-1}(x)$ and $U_\mu^{-1}(x)$ in equation (5).

In this work, we adopt the Bronzan parametrization of $SU(2)$ matrices [18]

$$U_\mu(x) = U\big(\theta_\mu(x), \phi_{1,\mu}(x), \phi_{2,\mu}(x)\big) \equiv \begin{pmatrix} \sin\big(\theta_\mu(x)\big)e^{i\phi_{1,\mu}(x)} & \cos\big(\theta_\mu(x)\big)e^{i\phi_{2,\mu}(x)} \\ -\cos\big(\theta_\mu(x)\big)e^{-i\phi_{2,\mu}(x)} & \sin\big(\theta_\mu(x)\big)e^{-i\phi_{1,\mu}(x)} \end{pmatrix}, \quad (6)$$

where $\theta_\mu(x) \in [0, \pi/2]$ is the polar angle, and $\phi_{1,\mu}(x), \phi_{2,\mu}(x) \in [0, 2\pi)$ are periodic azimuthal angles with periods of 2π. The complexified domain $SL(2, \mathbb{C})$ is given by considering $\theta_\mu(x), \phi_{i,\mu}(x) \in \mathbb{C}$. For general $U_\mu(x) \in SL(2, \mathbb{C})$, the inverse is equal to $U_\mu^{-1}(x) = U(\theta_\mu(x), -\phi_{1,\mu}(x), -\phi_{2,\mu}(x))^T$. In terms of these parameters, the path integral in equation (2) can be written as

$$\langle O \rangle = \frac{1}{Z} \int \Big(\prod_{x,\mu} d\theta_\mu(x) d\phi_{1,\mu}(x) d\phi_{2,\mu}(x) h(\theta_\mu(x)) \Big) e^{-S[U(\Theta)]} O[U(\Theta)], \quad (7)$$

where Θ indicates the set of all angular variables, $U(\Theta)$ is the set of gauge links parametrized by those variables, and $h_\mu(x) = 1/(4\pi^2)\sin(2\theta_\mu(x))$ is the normalized $SU(2)$ Haar measure.

In [5], an equivalent parametrization of $SU(2)$ matrices based on Euler angles was also considered and could serve as another starting point for deformation. Though equivalent, the choice of parameterization can significantly affect whether specific deformations are simple or complicated to write. For the simple class of deformations discussed next, this class defined using the Bronzan parametrization contains all analogous deformations defined using Euler angles, motivating the use of the Bronzan parametrization here.

3.2 Constant deformations

To feasibly explore the effects of contour deformations on the signal-to-noise problem, we drastically reduce the space of possible deformations $\tilde{U}(U)$ by considering only variable-independent deformations, or *constant deformations*, which have been used successfully in [4, 5] to reduce Wilson loops variances. In the Bronzan parametrization, this is defined by deforming each $\theta_\mu(x)$, $\phi_{1,\mu}(x)$, and $\phi_{2,\mu}(x)$ by constant shifts into the imaginary direction. Because the $\theta_\mu(x)$ are not periodic, deformations must fix the endpoints at 0 and $\pi/2$, which leaves no non-trivial constant shifts for these variables. On the other hand, $\phi_{1,\mu}(x)$ and $\phi_{2,\mu}(x)$ are periodic variables and their contour of integration can safely be shifted, giving in general the deformation map

$$\tilde{U}\big(\theta_\mu(x), \phi_{1,\mu}(x), \phi_{2,\mu}(x)\big) = U\big(\theta_\mu(x), \tilde{\phi}_{1,\mu}(x), \tilde{\phi}_{2,\mu}(x)\big), \quad (8)$$

where

$$\tilde{\phi}_{i,\mu}(x) = \phi_{i,\mu}(x) + i\Delta_{i,\mu}(x), \ (i = 1, 2), \quad (9)$$

in terms of the shift fields, $\Delta_{i,\mu}(x) \in \mathbb{R}$.

Because $\Delta_{i,\mu}(x)$ does not depend on the explicit values of $\theta_\mu(x)$ and $\phi_{i,\mu}(x)$, the Jacobian $J[U] = \det(\partial\tilde{U}/\partial U)$ of this transformation is trivial. The normalized Haar measure $h_\mu(x)$ is also not changed by this deformation. We can thus write the new observable

$$Q[\Theta] \equiv O[U(\tilde{\Theta})]e^{-S[U(\tilde{\Theta})]+S[U(\Theta)]}, \quad (10)$$

where $\tilde{\Theta}$ is the complexified set of all angles, and as before we have $\langle O \rangle = \mathbb{E}_S[Q[\Theta]]$.

3.3 Maximal tree gauge fixing

Unsurprisingly, minimizing the variance of Q by optimizing the deformation parameters $\Delta_{i,\mu}(x)$ leads to negligible improvement to the signal-to-noise ratio. Intuitively, constant deformations are restricted in this case to independently deform gauge links, which do not independently encode gauge invariant physical quantities. Instead, inspired by the previous construction of constant deformations with open boundary conditions in [5], we first restrict the path integral by fixing to a maximal tree gauge and then deform the remaining degrees of freedom in this formulation.

Even with a tree-based gauge fixing scheme such as a maximal tree gauge, there is significant choice as to how the fixed links are arranged. We have found that these choices can drastically affect the signal-to-noise improvements which we have been able to achieve. After some exploration, the most significant reduction in variance was observed by choosing the maximal tree gauge defined by

$$
\begin{aligned}
U_0(x_0 < L_0 - 1, x_1 = 0, x_2 = 0) &= \mathbb{I}, \\
U_1(x_0, x_1 < L_1 - 1, x_2 = 0) &= \mathbb{I}, \\
U_2(x_0, x_1, x_2 < L_2 - 1) &= \mathbb{I}.
\end{aligned}
\tag{11}
$$

Here (L_0, L_1, L_2) are the lattice dimensions in lattice units and $x = (x_0, x_1, x_2)$ is the lattice coordinate with $x_i \in \{0, 1, \cdots, L_i - 1\}$ $(i = 0, 1, 2)$. To obtain the most significant reduction in variance, we also chose the arrangement of the Wilson loop observable to lie in the $(0, 1)$-plane with the lower left corner located at the origin $x = (x_0 = 0, x_1 = 0, x_2 = 0)$.
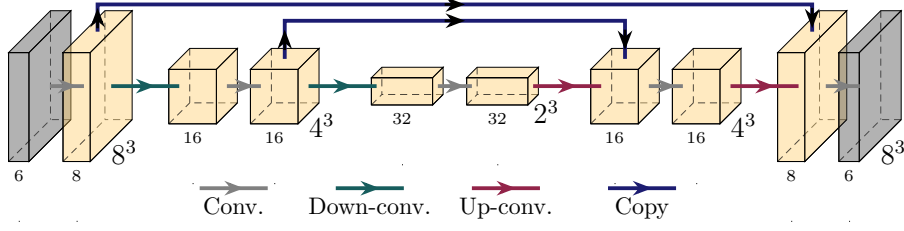
3.4 Deformation parameters from U-nets

For small lattices, for example with dimensions 8^3, optimizing $\Delta_{i,\mu}(x)$ directly on gauge-fixed configurations straightforwardly provides orders-of-magnitude improvement of Wilson loop signal-to-noise ratios. However, if the lattice dimension is taken large while statistics are held fixed, the abundance of parameters $\Delta_{i,\mu}(x)$ quickly leads to an overfitting regime during optimization. We expect on physical grounds that the reduction in variance achieved for Wilson loops of identical size should be nearly equivalent across lattices with dimensions sufficiently larger than the size of the loop (when β is held fixed), but these expected variance improvements are not reached when overfitting is encountered on larger lattices.

To overcome this training issue, we reparametrize the shift field, $\Delta_{i,\mu}(x)$, as

$$
\Delta_{i,\mu}(x) = f_{i,\mu}(x; w),
\tag{12}
$$

where $f_{i,\mu}(x; w)$ is the output of a neural network and w is the set of weights defining the neural network, which implicitly parameterize this output. This is a new parameterization of approximately the same space of deformations, albeit one where the parameters w at some finite neural network complexity are less prone to overfitting. To produce $f_{i,\mu}(x; w)$, we use the *U-net architecture* [19] applied to a constant input given by a collection of Boolean fields specifying the gauge-fixed links and the links making up the Wilson loop. Figure 1 shows an example of the U-net for 8^3 lattices.

For a generic lattice geometry, the U-net is constructed so that each coarsening step ("down-convolution") reduces the lattice dimensions by a factor of two along each axis while increasing the number of channel dimensions by a factor of two; on the other hand, the prolongation steps ("up-convolution") double the lattice dimensions along each axis while reducing the channel dimension

Figure 1: U-net architecture used to compute constant shift deformation parameters (gray output) from the gauge fixing and Wilson loop masks (gray input). The architecture is defined using a combination of convolutions (gray arrows), down-convolutions (green arrows), up-convolutions (red arrows), and copy-and-add operations (blue arrows). Each box depicts the number of channels (width) and the 3D lattice geometry (height and depth, third dimension suppressed) at each stage of the network. The 6 input channels correspond to gauge-fixing masks and the Wilson loop masks for the three orientations of links, while the six output channels correspond to the components of $\Delta_{i,\mu}(x)$.
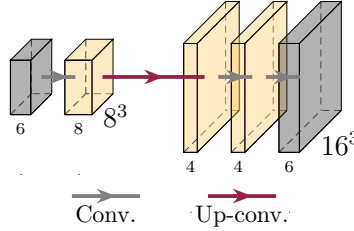
Figure 2: The decoder architecture that takes $\Delta_{i,\mu}(x)$ for a Wilson loop of size $r \times \tau$ in an 8^3 lattice geometry as the input and outputs $\Delta_{i,\mu}(x)$ for a Wilson loop of size $(2r) \times (2\tau)$ in a 16^3 lattice geometry. The meanings of the operations and boxes are the same as in figure 1.

by a factor of two. In this work, we always coarsen until the lattice dimensions are 2^3 at the coarsest level before prolongation. Because the gauge-fixed links cannot be deformed, we multiply by a mask as a final step to enforce $f_{i,\mu}(x; w) = 0$ for all μ, x where $U_\mu(x)$ are gauge-fixed.

The use of the U-net in training not only allows us to reach much smaller variances on lattice sizes $\gtrsim 16^3$, but also enables much faster training on smaller lattices by reducing the total number of updating steps to the weights. For these reasons, all results in this work are based on this U-net parametrization of the shift field.

3.5 Decoder network and transfer learning

The U-net parametrization by itself is sufficient to parameterize the shift field at any given lattice volume and Wilson loop geometry. However, as a practical improvement, we additionally introduce and study the use of a *decoder network* that uses convolutions to transform a shift field from a smaller volume to a larger volume for lattices with identical β and gauge fixing schemes. Figure 2 shows an example of the convolutional decoder architecture that we use to transform shift fields between 8^3 and 16^3 lattice volumes.

To train the network to transform between a smaller and larger lattice geometry differing by a factor of two in each dimension, we first train a U-net to produce a shift field $\Delta_{i,\mu}(x)$ for a Wilson loop size of $r \times \tau$ on the smaller lattice volume. The decoder takes this shift field as the input and outputs a new shift field which is used to deform the path integral for a Wilson loop size of $(2r) \times (2\tau)$, with the variance of this larger loop used as a target for optimization.

Comparing deformations produced from a U-net targeting the final lattice geometry versus using the U-net output at smaller volume together with a decoder network, we find that both reach the same variance improvement once properly trained. However, once the decoder network is trained to transform from 8^3 to 16^3 lattice volumes, we can fine-tune the same network by a smaller number of optimization steps to transform the shift field on the 16^3 lattices to a shift field suitable for 32^3 lattices. This transfer learning scheme is crucial for training on lattice volumes $\gtrsim 32^3$ where the training costs needed for directly training a U-net become large.

4. Training details and results

In this work, we present results for lattice theories with fixed $\beta = 3.75$, periodic boundary conditions, and three choices of lattice sizes, 8^3, 16^3, and 32^3. The bare coupling corresponds to a string tension of $a^2\sigma \approx (0.4)^2$ in lattice units [20]. For the three volumes, we respectively generate 1.5×10^5, 1.1×10^5, and 2.4×10^4 configurations for training and 500 configurations for testing. No significant autocorrelations of plaquette values are observed within these configurations.

For the 8^3 ensemble, we find the optimal constant deformation by directly training U-nets as described in section 3.4; for the 16^3 ensemble, we use the decoder architecture shown in figure 1. We also compared directly training a U-net on the 16^3 ensemble, with those two methods giving the same performance. For the 32^3 ensemble, we reuse the decoder network previously trained for the 16^3 ensemble and fine-tune it with fewer than 10^3 steps of weight updates, a much lesser amount of additional training than directly training the shift fields.

To avoid numerical stability issues, we train in all cases by minimizing the logarithm of the variance of the targeted Wilson loop for each geometry using the numerically stable LogSumExp function. We also use the trick introduced in [5] of measuring and optimizing only the $(1,1)$ component of the untraced Wilson loop. All training is done on a single compute node with eight NVIDIA A100 GPUs. We use mini-batch sizes of 32 configurations per GPU with the default settings of the Adam optimizer [21]. Training for each Wilson loop size takes at most a few node-hours.

The resulting improvement factors in the Wilson loop variance are shown in figure 3. Up to three orders of magnitude of variance reduction is obtained at the largest Wilson loop studied. As the area is increased, we find increasing improvements. The reduction in variance obtained using the techniques discussed above is also nearly independent of the total lattice volume, depending much more strongly on the Wilson loop geometry.

5. Outlook and conclusion

In this work, we showed for the first time how to use constant contour deformations to mitigate the signal-to-noise problem of Wilson loops for three dimensional $SU(2)$ gauge theory with periodic
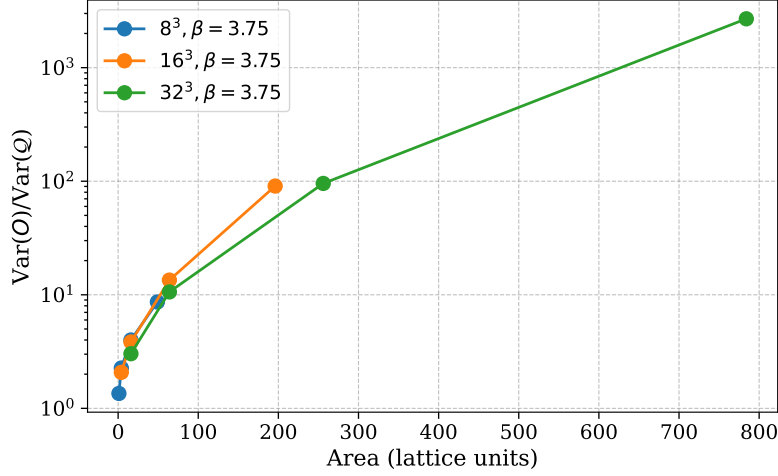
Figure 3: Ratios of variances between the original (O) and improved (Q) Wilson loop observables of various areas for 8^3, 16^3, and 32^3 ensembles.

boundary conditions. To achieve this result, we used a maximal tree gauge fixing scheme (equation (11)), a U-net parametrization of the deformation (equation (12) and figure 1), and decoder networks with transfer learning (section 3.5 and figure 2).

Our approach of constant contour deformations parametrized by a U-net can easily be generalized to 4D by using four-dimensional convolutions and a four-dimensional maximal tree gauge. These deformations are also applicable to other choices of N_c, including the $SU(3)$ gauge group, by working with the corresponding Bronzan angular parametrizations [18]. We are currently working to apply the method to four dimensional $SU(2)$ and $SU(3)$ gauge theories.

Within the framework of the constant deformation presented in this work, many avenues are left to be explored that may lead to better performance. Most importantly, we have observed that the choice of the gauge fixing scheme heavily influences the final variances of improved observables. In addition, the maximal tree gauge fixing scheme we use here completely obscures the translation symmetry of the lattice, as it must be combined with gauge transformations to maintain the gauge fixing scheme. This means that performing volume averaging that is commonly used for variance reduction is expensive since each translation requires additional gauge fixing. Other gauge fixing schemes, such as partial subsets of a maximal tree gauge, could simplify the process of (partial) volume averaging and may yield similar or better variance improvements. Finally, field-independent deformations represent a small class of all possible deformations. Exploring contour deformation beyond this simple choice, such as the Fourier series expansion presented in [5], may further boost the performance and eliminate the needs of gauge fixing completely.

References

[1] G. Parisi *Phys. Rept.* **103** (1984) 203–211.

[2] G. P. Lepage in *Theoretical Advanced Study Institute in Elementary Particle Physics*. 6, 1989.

[3] E. H. Fradkin and L. Susskind *Phys. Rev. D* **17** (1978) 2637.

[4] W. Detmold, G. Kanwar, M. L. Wagman, and N. C. Warrington *Phys. Rev. D* **102** no. 1, (2020) 014514, `arXiv:2003.05914 [hep-lat]`.

[5] W. Detmold, G. Kanwar, H. Lamm, M. L. Wagman, and N. C. Warrington *Phys. Rev. D* **103** no. 9, (2021) 094517, `arXiv:2101.12668 [hep-lat]`.

[6] AuroraScience Collaboration, M. Cristoforetti, F. Di Renzo, and L. Scorzato *Phys. Rev. D* **86** (2012) 074506, `arXiv:1205.3996 [hep-lat]`.

[7] M. Cristoforetti, F. Di Renzo, A. Mukherjee, and L. Scorzato *Phys. Rev. D* **88** no. 5, (2013) 051501, `arXiv:1303.7204 [hep-lat]`.

[8] G. Aarts *Phys. Rev. D* **88** no. 9, (2013) 094501, `arXiv:1308.4811 [hep-lat]`.

[9] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington *JHEP* **05** (2016) 053, `arXiv:1512.08764 [hep-lat]`.

[10] A. Alexandru, G. Basar, and P. Bedaque *Phys. Rev. D* **93** no. 1, (2016) 014504, `arXiv:1510.03258 [hep-lat]`.

[11] A. Alexandru, G. Basar, P. F. Bedaque, S. Vartak, and N. C. Warrington *Phys. Rev. Lett.* **117** no. 8, (2016) 081602, `arXiv:1605.08040 [hep-lat]`.

[12] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence *Phys. Rev. D* **96** no. 9, (2017) 094505, `arXiv:1709.01971 [hep-lat]`.

[13] A. Alexandru, G. Basar, P. F. Bedaque, and G. W. Ridgway *Phys. Rev. D* **95** no. 11, (2017) 114501, `arXiv:1704.06404 [hep-lat]`.

[14] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence *Phys. Rev. D* **97** no. 9, (2018) 094510, `arXiv:1804.00697 [hep-lat]`.

[15] A. Alexandru, G. Basar, P. F. Bedaque, and N. C. Warrington *Rev. Mod. Phys.* **94** no. 1, (2022) 015006, `arXiv:2007.05436 [hep-lat]`.

[16] Y. Mori, K. Kashiwa, and A. Ohnishi *PTEP* **2018** no. 2, (2018) 023B04, `arXiv:1709.03208 [hep-lat]`.

[17] E. Witten *AMS/IP Stud. Adv. Math.* **50** (2011) 347–446, `arXiv:1001.2933 [hep-th]`.

[18] J. B. Bronzan *Phys. Rev. D* **38** (1988) 1994.

[19] O. Ronneberger, P. Fischer, and T. Brox in *MICCAI 2015: 18th International Conference*, pp. 234–241, Springer. 2015. `arXiv:1505.04597 [cs.CV]`.

[20] A. Athenodorou and M. Teper *JHEP* **02** (2017) 015, `arXiv:1609.03873 [hep-lat]`.

[21] D. P. Kingma and J. Ba in *ICLR 2015*, Y. Bengio and Y. LeCun, eds. 2015. `http://arxiv.org/abs/1412.6980`.