

A Machine Learning Approach for Lattice Gauge Fixing

Ho Hsiao ^{a,*} Benjamin J. Choi ^a, Hiroshi Ohno ^a and Akio Tomiya ^{b,c,d}

^aCenter for Computational Sciences, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan

^bDepartment of Information and Mathematical Sciences, Tokyo Woman's Christian University,
2-6-1 Zempukuji, Suginami-ku, Tokyo 167-8585, Japan

^cRIKEN Center for Computational Science,
7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe 650-0047, Japan

^dDepartment of Physics, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan

E-mail: hohsiao@ccs.tsukuba.ac.jp

Gauge fixing is an essential step in lattice QCD calculations, particularly for studying gauge-dependent observables. Traditional iterative algorithms are computationally expensive and often suffer from critical slowing down and scaling bottlenecks on large lattices. We present a novel machine learning framework for lattice gauge fixing, where Wilson lines are utilized to construct gauge transformation matrices within a convolutional neural network. The model parameters are optimized via backpropagation, and we introduce a hybrid strategy that combines a neural-network-based transformation with subsequent iterative methods. Preliminary tests on SU(3) gauge theory ensembles for Coulomb gauge demonstrate the potential of this approach to improve the efficiency of lattice gauge fixing. Furthermore, we show that the model exhibits lattice size transferability, where parameters optimized on smaller lattices remain effective for larger volumes without additional training. This framework provides a scalable path toward mitigating critical slowing down in high-precision gauge fixing.

The 42nd International Symposium on Lattice Field Theory (LATTICE2025)
2-8 November 2025
Tata Institute of Fundamental Research, Mumbai, India

*Speaker

1. Introduction

Gauge fixing, applying gauge transformations until a given condition, is a crucial but computationally demanding process in lattice quantum chromodynamics (QCD) calculations, particularly for large lattices where critical slowing down becomes a significant challenge. While gauge invariance is a fundamental feature of lattice formulations, gauge fixing is often necessary for studying gauge-dependent quantities, such as quark and gluon propagators [1], defining renormalization schemes [2], or extracting fundamental QCD Green's functions. Furthermore, smearing with a gauge-fixed configuration is shown to further reduce the noise in mass extractions [3]. Traditional iterative algorithms for gauge fixing, such as the Los Alamos (LA) method [4] and Cornell method [5], are known to suffer from inefficiencies as lattice sizes increase. In both cases, the updating procedure is purely local, so that information propagates only between neighboring lattice sites at each iteration. Motivated by these limitations of purely iterative algorithms with local updates, we introduce a novel machine-learning framework for lattice gauge fixing in SU(N) gauge theories to Landau (Coulomb) gauge. Our approach leverages Wilson lines to construct gauge transformation matrices and convolute the link variables through a neural network to efficiently acquire long distance information. We employ backpropagation to compute gradients to update the model parameters. For machine learning, mini-batch training is applied to learn optimal parameters. The optimal goal of our study is to replace the iterative procedure by a single gauge transformation. We utilize the software package, `Gaugefield.jl` [6], to compute gradient from the neural network and implement the traditional gauge fixing methods, while `Lux.jl` [7] is employed for machine learning.

In our initial study, we focus on the Coulomb gauge. It can be easily extended to Landau gauge, as formulated in the following sections. We investigate the performance of our approach with two models built by different neural network depths and lengths of Wilson lines. Various training schemes are tested to realize the training stability. Preliminary results in SU(3) gauge theory demonstrate the potential of this method to improve both the accuracy and computational efficiency of gauge fixing. Future work will explore the performance of the model across varying lattice sizes and its adaptability to different gauge configurations, including topologically frozen conditions, paving the way for more robust and scalable lattice QCD simulations.

2. Iterative Approaches

We briefly outline the iterative gauge-fixing procedures employed in this study. Comprehensive technical details are available in the original publications [4, 5] and in representative applications [3, 8]. In lattice calculations, gauge fixing is achieved by satisfying the following condition under certain gauge transformation g :

$$F[g] = \frac{1}{d_{\text{fix}} N_c N_{\text{vol}}} \text{Re} \sum_n \sum_{\mu}^{d_{\text{fix}}} \text{Tr} U_{\mu}^g(n), \quad \text{with} \quad \frac{\delta F}{\delta g} = 0. \quad (1)$$

The gauge-fixing functional, $F[g]$, computes the sum of traces of g -transformed link variables, $U_{\mu}^g(n)$, over the fixing direction, d_{fix} . In this work, we are interested in Landau (Coulomb) gauge

fixing, for which $d_{\text{fix}} = 4$ ($d_{\text{fix}} = 3$). For the LA method, one defines

$$w(n) = \sum_{\mu}^{d_{\text{fix}}} U_{\mu}(n) + U_{\mu}^{\dagger}(n - \hat{\mu}). \quad (2)$$

The gauge transformation is conducted on even or odd sites by $w(n) \rightarrow w^g(n) = g(n)w(n)$, which project $w(n)$ along its SU(2) group manifold that satisfy $\text{ReTr } g(n)w(n) \geq \text{ReTr } w(n)$. This maximization is known as the Cabibbo-Marinari method [9], which is widely applied in the heat bath algorithm. The Cornell method essentially bases on the steepest-descent (SD) approach, where the gauge transformation is given by

$$g(n) = \exp \alpha \Delta(n), \text{ where } \Delta(n) = \sum_{\mu} [U_{\mu}^{\dagger}(n - \mu) - U_{\mu}(n)]_{\text{TA}}. \quad (3)$$

The traceless anti-Hermitian (TA) operation follows the definition in Ref. [6]. The parameter α , acting as the step size in SD, should be properly chosen, either globally or locally, to guarantee efficient convergence. Hence, we further adopt the Mino method (see Appendix A in Ref. [3]), that computes the local α as

$$\alpha(n) = \frac{\text{Re} \sum_{\mu} \text{Tr} [\Delta(n) (U_{\mu}(n - \hat{\mu}) - U_{\mu}(n))]}{\text{Re} \sum_{\mu} \text{Tr} [\Delta^2(n) (U_{\mu}(n - \hat{\mu}) + U_{\mu}(n))]} . \quad (4)$$

An over-relaxation (OR) is also applied with a parameter $1 < \omega < 2$ in the gauge transformation by taking $g(n) = \exp [\omega \alpha(n) \Delta(n)]$.¹ In practice, we combine these two methods by first applying the LA method with a given number of iterations to make sure monotonically approaching gauge-fixed point, the maximum of Eq. (1). Subsequently, we switch to SD, using the Mino method and an OR parameter $\omega = 1.99$, until reaching desired tolerance with respect to $\Delta F[g]$, the change of $F[g]$.

3. Model for Machine Learning

Our approach aims to bypass traditional iterative procedures by, first, generalizing from local link variables to a sum of multi-length Wilson lines, to capture long-range correlations. Furthermore, we employ a convolutional neural network (CNN) to effectively construct complicated Wilson lines in deep layers. We then regard the SU(3) group projection as a non-linear activation function within a machine learning architecture. Lastly, machine learning optimization is applied to refine the parameters of each length of Wilson lines at each layer.

For each layer, ℓ , of the CNN, we define the Wilson line bundle (WLB) gauge transform matrix at the lattice position n as

$$g^{(\ell)}(n) \equiv \exp [\Omega^{(\ell)}(n)]_{\text{TA}}, \text{ where } \Omega^{(\ell)}(n) = \sum_{r=1}^s \theta_r^{(\ell)} L_r \left(n; \left\{ U_{\mu}^{(\ell-1)} \right\} \right). \quad (5)$$

¹It is also possible to improve the LA method by introducing OR [10]. We only apply OR in the SD method for simplicity. Furthermore, advanced techniques, such as Fourier acceleration [5, 11], conjugate gradient [12], and global optimization [13], are beyond the scope of this study.

In this expression, $L_r(n; \{U_\mu^{(\ell-1)}\})$ denotes the combination of length- r Wilson lines with one end fixed to reference point n , which is constructed by multiplying the gauge links of the previous layer, $U_\mu^{(\ell-1)}$. The training parameters, $\theta_r^{(\ell)}$, control the weight of such an object that contributes to the sum in $\Omega^{(\ell)}(n)$ up to a desired maximum length S . Since $\Omega^{(\ell)}(n)$ is a sum of SU(3) matrices, we take the exponential function with traceless anti-Hermitian (TA) operation to project it back to the SU(3) group for $g^{(\ell)}(n)$. In the CNN construction, we assign the first layer as the original link variables, namely, $U_\mu^{(1)}(n) \equiv U_\mu(n)$. The link variables of the next layer is obtained by applying the WLB gauge transformation as $U_\mu^{(\ell)} = g^{(\ell)}(n)U_\mu^{(\ell-1)}(n)g^{(\ell)\dagger}(n + \hat{\mu})$. In this way, we evolve the CNN to a number of layers L . Notice that the effective number of training parameters $N_{\text{param.}} = (L - 1) \times S$, excluding the first layer with the original gauge links.

In this model, the link variables at the final layer, $U_\mu^{(L)}(n)$, is the output of the CNN, which optimally is considered as a gauge-fixed configuration. Hence, we compute the objective function defined by

$$\mathcal{F}(\theta) = \frac{1}{d_{\text{fix}} N_c N_{\text{vol}}} \text{Re} \sum_n \sum_{\mu=1}^{d_{\text{fix}}} \text{Tr} U_\mu^{(L)}(n), \quad (6)$$

where d_{fix} denotes the number of directions depending the gauge to fix, N_c is the number of colors, and N_{vol} is the lattice volume. The training procedure is formulated as the maximization of the objective function. We compute the gradient of Eq. (6) with respect to the trainable parameters $\theta_r^{(\ell)}$ through backpropagation, which takes the form:

$$\frac{d\mathcal{F}(\theta)}{d\theta_r^{(\ell)}} = 2 \sum_{\mu=1}^{d_{\text{fix}}} \sum_n \text{Re} \text{Tr} \left[\Lambda_\mu^{H^{(\ell)}}(n) L_r(n; \{U^{(\ell-1)}\}) + \Lambda_\mu^{S^{(\ell)}}(n + \hat{\mu}) L_r(n + \hat{\mu}; \{U^{(\ell-1)}\}) \right], \quad (7)$$

where

$$\Lambda_\nu^{H^{(\ell+1)}}(m) = \left[\left\{ U_\nu^{(\ell)}(m) g^{(\ell+1)\dagger}(m + \hat{\nu}) \delta_\nu^{(\ell+1)}(m) \right\} \star \frac{\partial g^{(\ell+1)}(m)}{\partial Q^{(\ell+1)}(m)} \right]_{\text{TA}}, \quad (8)$$

$$\Lambda_\nu^{S^{(\ell+1)}}(m + \hat{\nu}) = \left[\left\{ U_\nu^{(\ell)\dagger}(m) g^{(\ell+1)\dagger}(m) \delta_\nu^{(\ell+1)\dagger}(m) \right\} \star \frac{\partial g^{(\ell+1)}(m + \hat{\nu})}{\partial Q^{(\ell+1)}(m + \hat{\nu})} \right]_{\text{TA}}. \quad (9)$$

The back-propagated contributions from the final layer, $\delta_\mu^{(\ell)}(n)$ and $\delta_\nu^{(\ell)\dagger}(n)$, are obtained by

$$\delta_\mu^{(\ell)}(n) = \frac{\partial \mathcal{F}(\theta)}{\partial U_\nu^{(\ell)}(n)} \quad \text{and} \quad \delta_\mu^{(\ell)\dagger}(n) = \frac{\partial \mathcal{F}(\theta)}{\partial U_\nu^{(\ell)\dagger}(n)}. \quad (10)$$

We define the star product for the rank-2 and rank-4 tensors as $[A \star T]^i_j \equiv \sum_{kl} A^l_k T^k_j^i$. The derivation of this backpropagation formula closely follows the case of stout smearing in Ref. [14], and the detailed derivation is deferred to a future publication.

4. Training schemes

To train our models, we make use of the JLDG public ensemble [15], RC32x48, listed in Table 1, while the ensemble RC48x48 plays the role of a validation set. As this is our first attempt,

Table 1: JLDG public ensembles used in this work. The ensembles were generated by the PACS-CS collaboration with the Iwasaki gauge action and the (2+1)-flavor non-perturbatively $O(a)$ -improved Wilson quarks [15]. For each ensemble, the spacial extent, N_s , the temporal extent, N_t , the lattice bare coupling, β , the clover coefficient c_{SW} , the hopping parameter for u and d quarks, κ_{ud} , the hopping parameter for s quark, κ_s , the lattice spacing, a , and the number of the available configurations, N_{conf} . are listed.

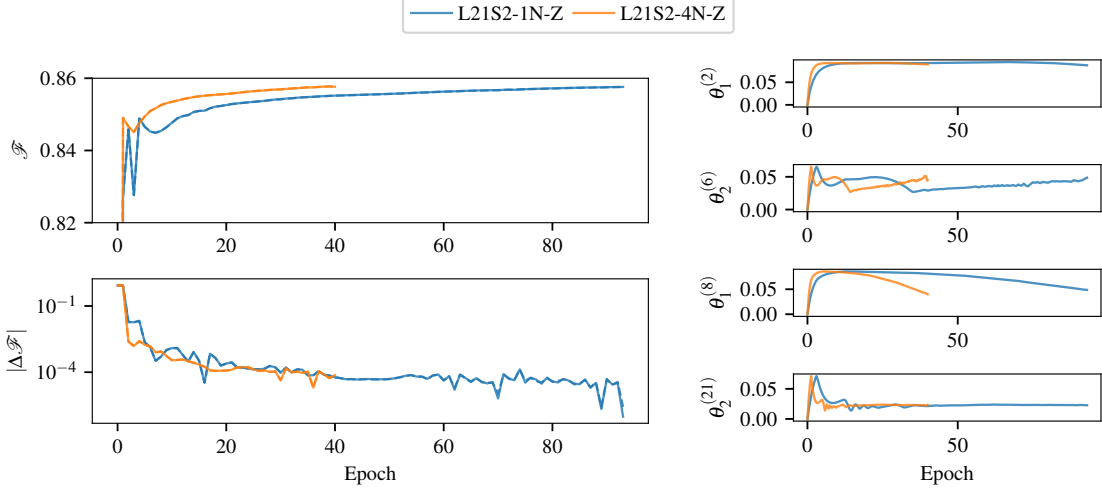
Ensemble	N_s	N_t	β	c_{SW}	κ_{ud}	κ_s	$a[\text{fm}]$	N_{conf}
RC32x48	32	48	1.9	1.715	0.1373316	0.1367526	0.0907(13)	402
RC48x48	48	48	1.9	1.715	0.1373316	0.1367526	0.0907(13)	200

we restrict ourselves to ensembles with identical lattice parameters, differing only in the lattice spacial extent. We construct two CNN architectures, denoted as L21S2 and L12S3. The former consists of 21 layers and incorporates Wilson-line combinations of length up to two, and the latter has 12 layers and extends the construction to including length-three Wilson lines. We apply the mini-batch training, which is a widely used strategy that strikes a balance between efficiency and stability during optimization in machine learning. Instead of updating model parameters using the entire training set (full-batch training) or a single data point at a time (stochastic training), mini-batch training processes small subsets of the data in each iteration. The training parameters are updated by the average of the gradients computed from the configurations within a mini batch. We employ adaptive moment estimation (Adam) optimizer to update the model parameters with a learning rate α . A single training epoch is completed once all mini-batches within the training set have been processed sequentially.

In this work, we consider two sizes for the training set: a baseline set of 40 configurations processed with a mini-batch size of 10, denoted by the suffix "-1N", and a larger training set of 160 configurations, denoted by "-4N", where the mini-batch size is 16. Most schemes are initialized from scratch with zero parameters (indicated by the suffix "-Z"), where the gauge transformation matrix is an identity matrix. It allows us to investigate the effect on the size of the training set. Alternatively, we also perform an incremental training test specifically with the L12S3 architecture. In the scheme L12S3-4N-W, the training begins with the parameters obtained at the 14th epoch of the smaller training set. The model then switches to the larger training set size and continues for additional 34 epochs. This two-stage process is denoted by the epoch count 14 + 34, and the suffix, "-W", stands for a warm start. The hyperparameters for each training scheme are summarized in

Table 2: Hyperparameters and the final results for each training scheme with ensemble RC32x48. The table details the training set size, mini batch size, CNN layer depth (L), maximum Wilson line length (S), the initialization strategy, and learning rate (α). The final training objective value is reported alongside the total epochs. In the case of L12S3-4N-W, the notation "14+34" indicates that training is initialized from the 14th epoch of L12S3-1N-Z and continue for an additional 34 epochs.

Scheme	Training set	Batch size	L	S	Initial param.	α	Epoch	\mathcal{F}
L21S2-1N-Z	40	10	21	2	zeros	0.01	93	0.8576
L21S2-4N-Z	160	16	21	2	zeros	0.01	40	0.8576
L12S3-1N-Z	40	10	12	3	zeros	0.005	32	0.8498
L12S3-4N-Z	160	16	12	3	zeros	0.005	39	0.8538
L12S3-4N-W	160	16	12	3	L12S3-1N-Z	0.005	14+34	0.8538

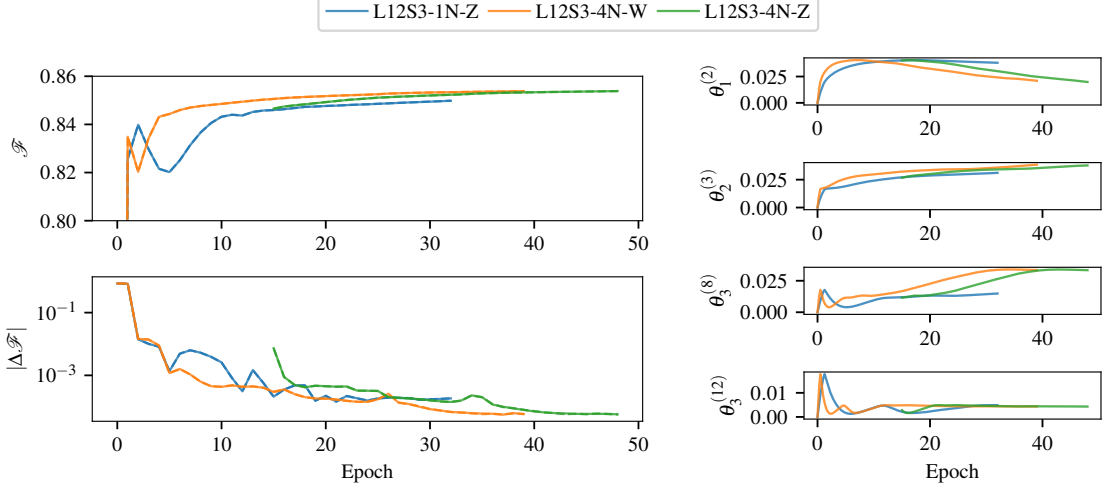


(a) The objective function and its absolute difference between two epochs. (b) Evolution of selected training parameters. **Figure 1:** Training history of the L21S2 CNN, consisting of 21 layers with Wilson lines of lengths one and two. The blue curves represent the L21S2-1N-Z scheme (small dataset), and the orange curves denote the L21S2-4N-Z scheme (large dataset).

Table 2, including the training set size, batch size, the CNN architecture, initial training parameters, learning rate α , the number of conducted epoch, and the final value of the objective function.

Figures 1 and 2 show the training history with respect to the epoch for CNN architectures L21S2, and L12S3, respectively, where the legend denotes the training scheme. In the left panels, we present the objective functions, \mathcal{F} , defined in Eq. (6) (top) and their relative change comparing to that of the previous epoch (bottom), denoted as $|\Delta\mathcal{F}|$. In the right panels, selective training parameters, $\theta_r^{(\ell)}$, for Wilson line with a length r at layer ℓ , are displayed. On one hand, as shown in Figure 1, the training for the L21S2 architecture reaches to the same final result for both the small (blue curves) and large (orange curves) datasets, in terms of the objective function and parameter values. On the other hand, the L12S3 architecture appears more sensitive to dataset size. As illustrated in Figure 2, $\Delta\mathcal{F}$ for the smaller dataset (blue) significantly plateaus after approximately 25 epochs. A comparison between L12S3-1N-Z (blue) and L12S3-4N-Z (orange), both starting from zero parameters, reveals significant discrepancies not only in the final objective values but also in the parameters, especially $\theta_1^{(2)}$ and $\theta_3^{(8)}$. This suggests that the smaller dataset lacks the necessary information to capture the complexity of the model.² To explore potential optimizations for future large-scale runs, we introduced an incremental training strategy in scheme L12S3-4N-W (green) by resuming training from the 14th epoch of the small-set run using the larger dataset. This transition causes a momentary increase in the objective function, followed by a correction that aligns the parameters and objective value with the L12S3-4N-Z results. This confirms that even when a model begins to converge toward a suboptimal steady-state due to limited data, an incremental shift to a larger dataset can successfully steer the parameters toward the behavior observed in the training with a larger set.

²Although the number of training parameters for L21S2 ($N_{\text{param.}} = 40$) is larger than that for L12S3 ($N_{\text{param.}} = 33$), the complexity of a model is driven by the combinations of involving Wilson lines. In the case of $d_{\text{fix}} = 3$, the number of combinations with the length-2 Wilson lines is 30, while that of the length-3 lines is 150. This necessitates a larger training set for proper generalization.



(a) The objective function and its absolute difference between two epochs. (b) Evolution of selected training parameters. **Figure 2:** Training history of the L12S3 CNN, featuring 12 layers and Wilson lines up to length three. The blue curves denote the L12S3-1N-Z scheme (small dataset), while the orange curves represent L12S3-4N-Z (large dataset). The green curves illustrate the incremental training scheme (L12S3-4N-W), where training resumed from the 14th epoch of the small-set run using the larger dataset. Its large value of $|\Delta\mathcal{F}|$ at the beginning reflects the reconfiguration of the model as it adapts to the richer dataset, eventually aligning with L12S3-4N-Z.

5. Preliminary results for gauge fixing

In this section, we present preliminary results for the gauge-fixing performance of our trained models, using outcomes from traditional methods as a baseline for comparison. As shown in Figures 1 and 2, the objective functions increase during the trainings, which are expected to approach the value evaluated from $F[g]$ using the traditional iterative approach. Given the limited training duration of these exploratory runs, the final objective values have not yet reached the theoretical optimal value of 0.86965(8); see Table 2 for final values. However, the primary goal of this contribution is not to achieve immediate convergence, but rather to explore the potential of the model to accelerate the gauge-fixing process. Specifically, we investigate whether the hybrid procedure, applying first WLB gauge transformation then the standard iterative approach, can significantly reduce the total number of iterations required to reach the target precision. To ensure a fair comparison, the number of iterations in the LA method is chosen to yield an $F[g]$ value equivalent to that of the pure iterative approach. The convergence tolerance for the subsequent SD method is set to $\Delta F[g] < 10^{-12}$. We select two representative sets of training parameters from L21S2-1N-Z and L12S3-4N-W to conduct such validation tests.

Figure 3 demonstrates the results of gauge fixing ensemble RC32x48, with 100 configurations that are not involved in trainings. The blue curves denote the results of the pure iterative approach, with first 250 iterations in the LA method. The orange and green curves represent the outcomes from L21S2-1N-Z and L12S4-4N-W, respectively, with only 200 LA iterations (50 iterations reduced). The top panel display the gauge-fixing functional, $F[g]$, with respect to the iteration step in SD method, where solid lines are the mean values, and the width of bands correspond to standard deviations. In the middle panel, we illustrate the relative difference between two iterations. The

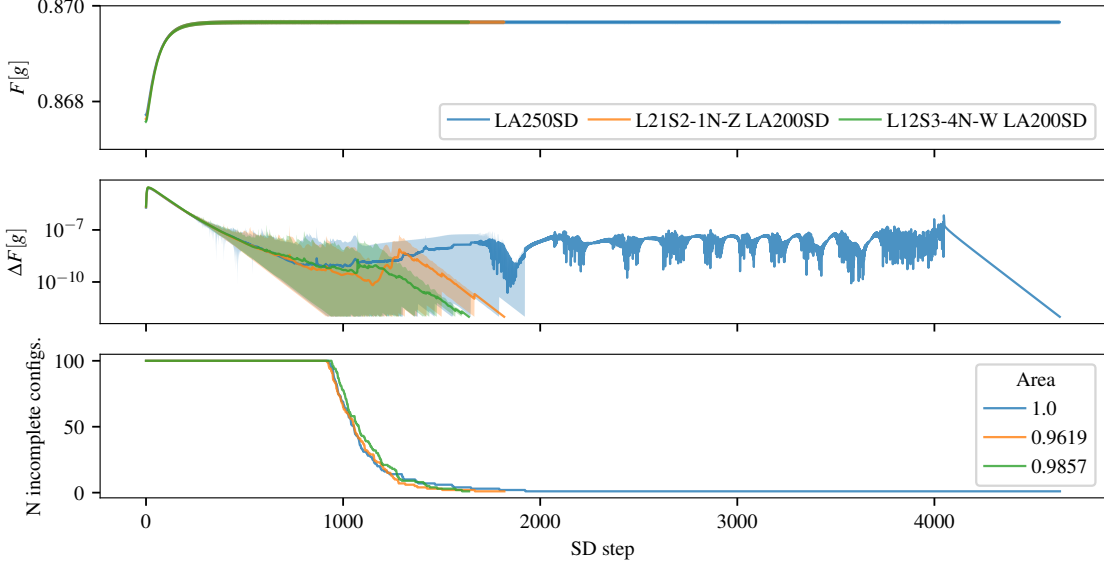


Figure 3: Gauge-fixing performance for the RC32x48 ensemble across 100 test configurations. Comparisons are shown between the pure iterative baseline (blue, beginning with 250 LA steps) and hybrid schemes using L21S2-1N-Z (orange) and L12S4-4N-W (green) trained parameter. The panels display: (Top) Evolution of the gauge-fixing functional $F[g]$, with solid lines indicating mean values and bands representing standard deviations. (Middle) Relative difference $\Delta F[g]$ between successive iterations; bands denote the range between maximum and minimum values across incomplete configurations. (Bottom) The number of incomplete configurations remaining throughout the SD iterations. Normalized total computational costs are provided in the legend, representing the area under each curve relative to the pure iterative baseline.

solid lines are the average, while the upper (lower) bounds of the bands are the maximum (minimum) over the incomplete configurations. We present the number of the incomplete configurations across the SD iterations in the bottom panel. The area below each curve corresponds to the total computational cost to fix all the configurations, which is normalized by such quantity obtained from the pure iterative approach as shown in the legend.

Although the neural network models were trained for a limited duration, the hybrid strategy demonstrates a clear advantage over the pure iterative baseline. As shown in the top panel, $F[g]$ for all schemes converges toward the same asymptotic limit. The improvement is evidenced in the middle panel, where the relative difference $\Delta F[g]$ in both orange and green decreases consistently, indicating that the initial WLB gauge transformation provides a configuration that is closer to the gauge fixed point. This smooth convergence avoids the critical slowing down seen in the pure iterative approach (blue). The total computational effort (bottom panel) yields efficiency factors of 0.9619 for the L21S2-1N-Z scheme (orange) and 0.9857 for the L12S3-4N-W scheme (green). These reductions also reflect the fact that L21S2-1N-Z acquires a larger objective value, compared to L12S3-4N-W; see Table 2. Based on the superior performance and higher final functional value achieved by the L21S2-1N-Z scheme, we selected this specific model to investigate the lattice-size transferability of our trained parameters to a larger lattice volume, ensemble RC48x48, without further training. In the top panel of Figure 4, the hybrid approach (orange) reaches to the same value as the traditional one (blue). The middle and bottom panels confirm that the relative difference $\Delta F[g]$ decreases consistently and the count of incomplete configurations drops steadily, resulting in

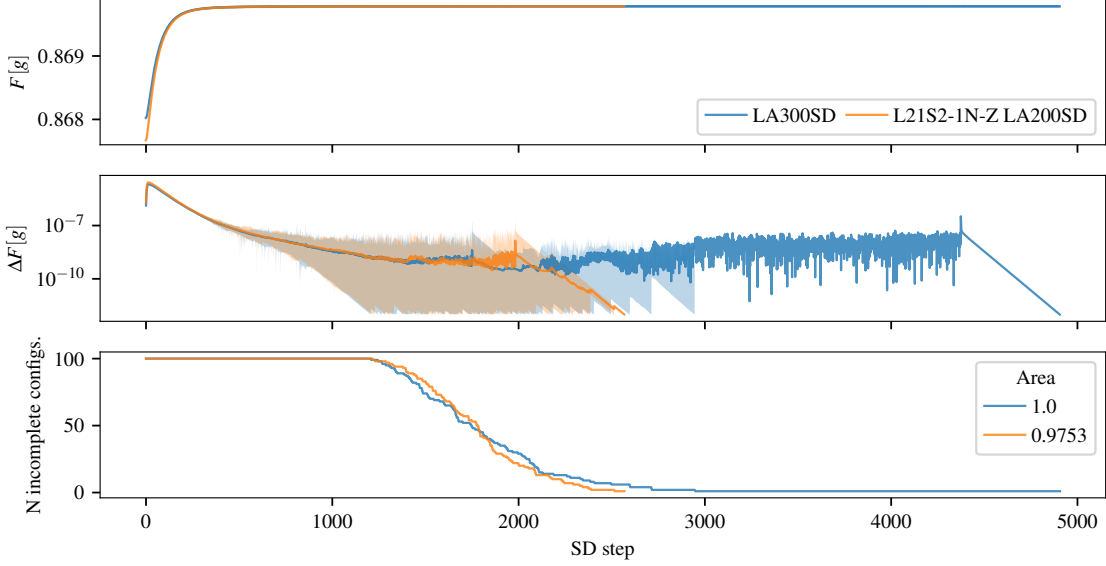


Figure 4: Gauge-fixing performance for the RC48x48 ensemble across 100 test configurations. The plot compares the pure iterative baseline (blue, starting with 300 LA steps) against the hybrid approach with trained parameters from L21S2-1N-Z scheme (orange). The panels illustrate: (Top) The evolution of the gauge-fixing functional $F[g]$, where solid lines represent mean values and shaded bands indicate standard deviations. (Middle) The relative difference $\Delta F[g]$ between successive iterations; the bands capture the range between the maximum and minimum values among incomplete configurations. (Bottom) The count of remaining incomplete configurations as a function of SD iterations. As indicated in the legend, the hybrid approach achieves a normalized total computational cost of 0.9753 relative to the pure iterative baseline.

a normalized computational cost of 0.9753. This demonstrates that the local gauge features learned by the L21S2 architecture are not only effective but also highly scalable, allowing the model to be trained on smaller, less expensive volumes while remaining valid for large-scale production runs.

6. Summary and Outlook

In this work, we introduced a machine learning framework for lattice gauge fixing designed to mitigate the computational bottlenecks and critical slowing down characteristic of traditional iterative methods. By utilizing Wilson lines within a CNN to construct gauge transformation matrices, we successfully implemented a model capable of maximizing the gauge-fixing functional for Coulomb gauge. Our hybrid strategy demonstrates a clear improvement in efficiency over the pure iterative approach, providing smooth and stable convergence throughout the process.

A significant finding of this study is the volume transferability of the trained parameters. Our model, processing translational invariant property, trained on a smaller lattice can apply successfully to a larger volume without further training. This demonstrates that the local gauge structures captured by the CNN are volume-independent, providing a scalable principle for future optimizations. Furthermore, we verified that an incremental training approach, moving from small to large datasets, allows the model to reconfigure its parameters to adapt to richer information density while reducing overall training cost.

In the ongoing research, we plan to evaluate the performance of this framework across even larger lattice sizes and investigate its adaptability to topologically frozen configurations. These

results provide more robust and scalable applications, with the ultimate goal of replacing iterative procedures with a single, highly efficient gauge transformation.

Acknowledgments

The authors thank the members of the project "Search for physics beyond the standard model using large-scale lattice QCD simulation and development of AI technology toward next-generation lattice QCD", especially Takeshi Yamazaki, for useful discussions. H. H., H. O., and A. T., were supported by JSPS KAKENHI Grant No. 22H05112. The work of A. T. was partially supported by JSPS KAKENHI Grants No. 20K14479, No. 22H05111, No. 22K03539 and JST BOOST, Japan Grant No. JPMJBY24F1. B. J. C. and part of this work were supported by MEXT as "Program for Promoting Researches on the Supercomputer Fugaku" (Grant Number JPMXP1020230411, JPMXP1020230409). This research used computational resources of Pegasus and Miyabi provided by Multidisciplinary Cooperative Research Program in Center for Computational Sciences, University of Tsukuba.

References

- [1] J.E. Mandula and M. Ogilvie *Phys. Lett. B* **185** (1987) 127.
- [2] L. Maiani, G. Martinelli and C.T. Sachrajda *Nucl. Phys. B* **368** (1992) 281.
- [3] CP-PACS collaboration *Phys. Rev. D* **67** (2003) 034503 [[hep-lat/0206009](#)].
- [4] R. Gupta, G. Guralnik, G. Kilcup, A. Patel, S.R. Sharpe and T. Warnock *Phys. Rev. D* **36** (1987) 2813.
- [5] C.T.H. Davies, G.G. Batrouni, G.R. Katz, A.S. Kronfeld, G.P. Lepage, K.G. Wilson *et al.* *Phys. Rev. D* **37** (1988) 1581.
- [6] Y. Nagai and A. Tomiya, [2409.03030](#).
- [7] A. Pal, *Lux: Explicit Parameterization of Deep Neural Networks in Julia*, Apr., 2023. [10.5281/zenodo.7808904](#).
- [8] H. Suman and K. Schilling, [hep-lat/9306018](#).
- [9] N. Cabibbo and E. Marinari *Phys. Lett. B* **119** (1982) 387.
- [10] J.E. Mandula and M. Ogilvie *Phys. Lett. B* **248** (1990) 156.
- [11] A. Cucchieri and T. Mendes *Phys. Rev. D* **57** (1998) 3822 [[hep-lat/9711047](#)].
- [12] RBC, UKQCD collaboration *Comput. Phys. Commun.* **187** (2015) 115 [[1405.5812](#)].
- [13] O. Oliveira and P.J. Silva *Comput. Phys. Commun.* **158** (2004) 73 [[hep-lat/0309184](#)].
- [14] Y. Nagai and A. Tomiya *Phys. Rev. D* **111** (2025) 074501 [[2103.11965](#)].
- [15] PACS-CS collaboration *Phys. Rev. D* **79** (2009) 034503 [[0807.1661](#)].