

Rediscovery of numerical Lüscher's formula from the neural network*

Yu Lu (陆宇)^{1†} Yi-Jia Wang (王一佳)^{1‡} Ying Chen (陈莹)^{1,2§} Jia-Jun Wu (吴佳俊)^{1‡}

¹School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

²Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Abstract: We present that by predicting the spectrum in discrete space from the phase shift in continuous space, the neural network can remarkably reproduce the numerical Lüscher's formula to a high precision. The model-independent property of the Lüscher's formula is naturally realized by the generalizability of the neural network. This exhibits the great potential of the neural network to extract model-independent relation between model-dependent quantities, and this data-driven approach could greatly facilitate the discovery of the physical principles underneath the intricate data.

Keywords: Luescher's formula, neural network, finite volume effects, LQCD

DOI: 10.1088/1674-1137/ad3b9c

I. INTRODUCTION

Physicists are always going after a concise description of data. Generally, this concise description boils down to analytic expressions or conserved quantities, which are usually dodging and hiding and cannot be trapped easily. Nowadays the rapid progress of machine learning (ML) techniques are helping physicists to meet their goals, as manifested by the applications, such as AI Feynman [1, 2] and AI Poincaré [3, 4]. For a review of ML techniques in physics, see [4] and several applications in hadron physics in Refs. [5–11] and references therein.

In most cases of modern physics, a concise description is generally realized at more abstract levels, such as analytic differential or integral equations whose solutions are supposed to explain the data. If these equations are explicitly known but cannot be solved easily even through numerical methods, ML may help to work out the solutions through Physics-informed-neural-network (PINN) approach [12]. In a more challenging case that there are conceptually links between physical principles and realistic phenomena but we cannot write down the exact expressions, maybe we can also resort to the data driven ML for uncovering the underneath connections.

A typical example is the study of the strong interaction in the low energy regime. It is known that the properties of hadrons are necessarily dictated by quantum chro-

modynamics (QCD), the fundamental theory of the strong interactions. However, due to the unique self-interacting properties of gluons, the strong coupling constant is large at the low energy regime and makes the standard perturbation theory inapplicable. Up to now, lattice QCD (LQCD) is the most important *ab initio* non-perturbative method for investigating the low energy properties of the strong interactions. LQCD is defined on the discretized Euclidean spacetime lattice and adopts the numerical simulation as its major approach. The major observables of LQCD are energies and matrix elements of hadron systems. However, except for the properties of ground state hadrons without strong decays, it is usually non-trivial for lattice results (on the Euclidean spacetime lattice) to be connected with experimental observables in the continuum Minkowski spacetime. For example, most hadrons are resonances observed in the invariant mass spectrum of multi-hadron system in decay or scattering processes, while what the lattice QCD can calculate are the discretized energy levels of related hadron systems on finite lattices. Therefore, the connection must be established.

One successful approach to address this issue is called Lüscher's formula [13–15], developed by Lüscher and collaborators more than 30 years ago. By making use of the finite volume effects, Lüscher's formula describes relation of the spectrum $E(L)$ of a two-body system on the finite lattice of size L with the scattering phase shift $\delta(E)$

Received 3 March 2024; Accepted 7 April 2024; Published online 8 April 2024

* Partly supported by the National Natural Science Foundation of China (12221005, 11935017, 12070131001, 12175239), (CRC 110 by DFG and NNSFC) and the National Key R&D Program of China (2020YFA0406400)

[†] E-mail: ylu@ucas.ac.cn

[‡] E-mail: wangyijia18@mails.ucas.ac.cn

[§] E-mail: chenying@ihep.ac.cn

[‡] E-mail: wujiajun@ucas.ac.cn

©2024 Chinese Physical Society and the Institute of High Energy Physics of the Chinese Academy of Sciences and the Institute of Modern Physics of the Chinese Academy of Sciences and IOP Publishing Ltd

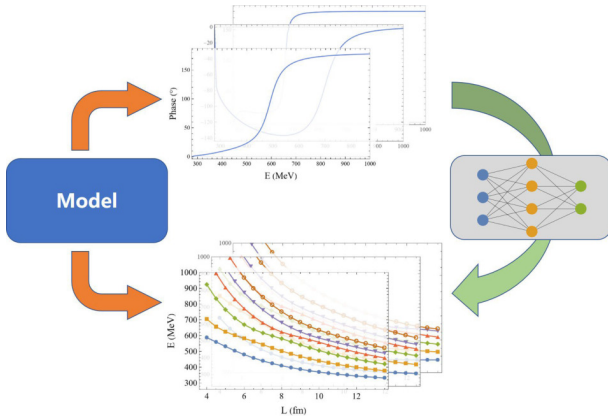


Fig. 1. (color online) The workflow of this work.

of this system in the continuum Minkowski space. The extension of Lüscher's formula to three-body systems is still undergoing [16–26]. Lüscher's formula and its extension are not only practically useful, but also is invaluablely model-independent on the theoretical side. Deriving these model-independent theoretical approaches are very challenging and require a lot of wisdom and insight.

For the multi-channel case, more than one free parameters in the scattering amplitude will show up in the infinite volume, while Lüscher's formula offers only one constrain to connect the volume size and scattering amplitude at the discrete energy levels. To extract the information of scattering amplitude in the infinite volume, we will need several different finite volume sizes which share the same energy level. However, one cannot know a priori which volume size will produce the desired spectra without doing the expensive lattice calculations. Therefore, one practical way in the multi-channel process is to build a model to relate the scattering amplitude at different energy levels in order to use the Lüscher's formula to translate the lattice spectrum into phase shifts or other information. As a consequence, model dependence will inevitably enter into such calculation. In contrast, since the neural network is trained via the data-driven way, it is naturally model independent, or at least its model dependence can be safely ignored. To this end, as a first step, we should answer whether the neural network can rediscover the numerical Lüscher's formula in single channel case.

Another challenge comes from the Lüscher's formula itself. Unlike extracting the analytic expression of the conserved quantities from the trajectory by ML approach, the Lüscher's formula is beyond the elementary function, therefore it can only be evaluated numerically. This property also leaves a great challenge for neural network to discovery it.

Encouraged by the achievement of machine learning in various areas, it is intriguing to ask if the neural network is able to discover the Lüscher's formula and its variance after fed by plenty data of spectra on lattice and the corresponding phase shifts. If a model-independence

link does exist, in principle, a highly generalizable neural network will be a decent approximation of this link, because of the universal approximation theorem [27–29]. In this paper, we will show that neural network is able to rediscover the numerical Lüscher's formula to a high precision.

This paper is organized as follows. Section II is devoted to build the theoretical formalism to generate the data of energy levels in the finite volume and phase shifts in the infinite space. Then we will elaborate on the construction of neural network and its training setup in Sec. III. In Sec. IV, we will analyse the result in detail, and provide the evidence to show the numerical form of Lüscher's formula is generated from neural network. Finally, we give a brief conclusion in Sec. V.

II. THEORETICAL FORMALISM

It is known that Lüscher's formula connects the finite volume energy level E and the S -wave phase shift $\delta(E)$ as [15]

$$\delta(E) = \arctan\left(\frac{q\pi^{3/2}}{\mathcal{Z}_{00}(1; q^2)}\right) + n\pi, \quad (1)$$

where $q = \frac{k_0 L}{2\pi}$ is defined with k_0 being the on-shell momentum of energy E , and the generalized zeta function $\mathcal{Z}_{00}(1; q^2)$ is defined as

$$\mathcal{Z}_{00}(1; q^2) := \frac{1}{\sqrt{4\pi}} \sum_{\vec{n} \in \mathbb{Z}^3} (\vec{n}^2 - q^2)^{-1}. \quad (2)$$

The system which we check against Lüscher's formula is the elastic $\pi\pi$ S -wave scattering process. In order to generate the training and test set, which consists of the phase shift $\delta(E)$ and the finite volume spectrum $E(L)$ for a given lattice size L , we model this scattering process by Hamiltonian Effective Field Theory (HEFT) [30].

Following Refs. [31, 32], we assume that $\pi\pi$ scattering can be described by vertex interactions and two-body potentials. In the rest frame, the Hamiltonian of a meson-meson system takes the *energy-independent* form as follows,

$$H = H_0 + H_I. \quad (3)$$

The non-interacting part is

$$H_0 = |\sigma\rangle m_\sigma \langle \sigma| + 2 \int d\vec{k} |\vec{k}\rangle \omega(|\vec{k}|) \langle \vec{k}|, \quad (4)$$

where $|\sigma\rangle$ is the bare state with mass m_σ , and $|\vec{k}\rangle$ is for the $\pi\pi$ channel state with relative momentum $2\vec{k}$ in the rest

frame of σ , and $\omega(k) = \sqrt{m_\pi^2 + k^2}$.

The interaction Hamiltonian is

$$H_I = \tilde{g} + \tilde{v}, \quad (5)$$

where \tilde{g} is a vertex interaction describing the decays of the bare state into two-pion channel,

$$\tilde{g} = \int d\vec{k} \{ |\vec{k}\rangle \tilde{g}^*(k) \langle \sigma | + h.c. \}, \quad (6)$$

and the direct $\pi\pi \rightarrow \pi\pi$ interaction (only S -wave) is defined by

$$\tilde{v} = \int d\vec{k} d\vec{k}' |\vec{k}\rangle v(k, k') \langle \vec{k}'|. \quad (7)$$

For the S -wave, the $\pi\pi$ scattering amplitude is then defined by the following coupled-channel equation,

$$t(k, k'; E) = V(k, k') + \int_0^\infty \tilde{k}^2 d\tilde{k} \frac{V(k, \tilde{k}) t(\tilde{k}, k'; E)}{E - 2\omega(\tilde{k}) + i\epsilon}, \quad (8)$$

where the coupled-channel potential is

$$V(k, k') = \frac{g^*(k)g(k')}{E - m_\sigma} + v(k, k'). \quad (9)$$

We choose the normalization $\langle \vec{k} | \vec{k}' \rangle = \delta(\vec{k} - \vec{k}')$, such that the S -matrix (and thereby the phase shift $\delta(E)$) in each partial-wave is related to the T -matrix by

$$S(E) \equiv e^{i2\delta(E)} = 1 + 2iT(k_{on}, k_{on}; E) \quad (10)$$

with

$$T(k_{on}, k_{on}; E) = -\pi \frac{k_{on} E}{4} t(k_{on}, k_{on}; E), \quad (11)$$

and $2\omega(k_{on}) = E$.

On the other hand, the HEFT provides direct access to the multi-particle energy eigenstates in a periodic volume characterized by the size length L . The quantized three momenta of the π meson is $k_n = \sqrt{n} \frac{2\pi}{L}$ for $n = n_x^2 + n_y^2 + n_z^2$ where $n_x, n_y, n_z = 0, \pm 1, \pm 2, \dots$. Then the Hamiltonian matrices with discrete momenta are,

$$[H_0] = \begin{pmatrix} m_\sigma & 0 & 0 & \cdots \\ 0 & 2\omega(k_0) & 0 & \cdots \\ 0 & 0 & 2\omega(k_1) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (12)$$

$$[H_I] = \begin{pmatrix} 0 & \tilde{g}(k_0) & \tilde{g}(k_1) & \cdots \\ \tilde{g}(k_0) & \tilde{v}(k_0, k_0) & \tilde{v}(k_0, k_1) & \cdots \\ \tilde{g}(k_1) & \tilde{v}(k_1, k_0) & \tilde{v}(k_1, k_1) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (13)$$

The corresponding finite-volume matrix elements are given by

$$\tilde{g}(k_n) = \sqrt{\frac{C_3(n)}{4\pi}} \left(\frac{2\pi}{L} \right)^{3/2} g(k_n), \quad (14)$$

$$\tilde{v}(k_i, k_j) = \frac{\sqrt{C_3(i)C_3(j)}}{4\pi} \left(\frac{2\pi}{L} \right)^3 v(k_i, k_j), \quad (15)$$

where the factor $C_3(n)$ is the degeneracy of (n_x, n_y, n_z) that gives the same n . The factor $\sqrt{\frac{C_3(n)}{4\pi}} \left(\frac{2\pi}{L} \right)^{3/2}$ follows from the quantization conditions in a finite box of a size L , where only S -wave contribution is included. With this Hamiltonian matrix, the spectra in the finite volume are the eigenvalues $E(L)$ of H satisfying $H|\Psi_E\rangle = E(L)|\Psi_E\rangle$.

For $g(k)$ and $v(k, k')$, we let them to be

$$g(k) = \frac{g_\sigma}{\sqrt{m_\pi}} f(c; k), \quad (16)$$

$$v(k, k') = \frac{g_{\pi\pi}}{m_\pi^2} u(d; k) u(d; k') \quad (17)$$

and in order to explore different types of data, three different forms of the $f(a; k)$ and $u(a; k)$ is assumed,

$$f_A(a; k) = \sqrt{u_A(a; k)} = \frac{1}{(1 + (ak)^2)}, \quad (18)$$

$$f_B(a; k) = \sqrt{u_B(a; k)} = \frac{1}{(1 + (ak)^2)^2}, \quad (19)$$

$$f_C(a; k) = u_C(a; k) = e^{-(ak)^2}, \quad (20)$$

which are model A, B and C, respectively.

Note that the shapes of the potentials in the momentum space become sharper and sharper from model A to model C. Since a sharper potential in momentum space has a larger effective range in coordinate space, and therefore has a more prominent finite volume effect, which is an artifact of a finite lattice. This artifact can be attributed to the deviation of the discrete momentum summation from the continuous momentum integration of

the kernel function of the model. It is proved that the finite volume correction to Lüscher's formula behaves as e^{-mL} with m being the typical energy scale of the model and a sharper potential will suffer larger corrections in general.

III. NEURAL NETWORK AND TRAINING SETUP

To ensure the diversity of the data, the training set should cover both the broadest and the sharpest potentials. In practice, the data from model A and C are used as the training set, while the data from model B serve as the test set. Each data set includes 100 evenly-sampled phase shift $\delta(E)$ values from $2m_\pi \approx 277\text{MeV}$ to 1 GeV, and the lowest 10 energy levels of $E(L)$ for different lattice sizes $L \in [10, 13]$ fm with a step size 0.5 fm. The parameter space is spanned by $m_\sigma(\text{MeV})$, g_σ , $c(\text{fm})$, $g_{\pi\pi}$ and $d(\text{fm})$, ranging from [350, 700], [0.5, 5], [0.5, 2], [0.1, 1], [0.5, 2], respectively. The space is randomly sampled by 2500 points for each model A, B and C. Once the parameters are fixed, we calculate the phase shift $\delta(E)$ in continuum space and $E(L)$ for lattice size L in [10, 13] fm with step size 0.5 fm. The phase shift $\delta(E)$ is evenly sampled by 100 points from $2m_\pi \approx 277$ MeV to 1 GeV, and the lowest 10 energy levels of $E(L)$ were kept for training and testing.

We summarize the workflow and the structure of the neural network in Fig. 1 and Fig. 2 respectively. Since the phase shift $\delta(E)$ contains the full information of a scattering process, it is natural to expect that the finite volume energy $E(L)$ can be predicted from $\delta(E)$. This is treated as a feature extraction task. To be precise, for a given phase shift $\delta(E)$ and a lattice size L , the neural network is designed to predict the *lowest* 10 energy levels $E(L)$ above the $\pi\pi$ threshold. With some trial-and-error, we construct a small feed-forward fully-connected neural network with "SoftPlus" activation function. The neural network is trained by ADAM method, with learning rate 10^{-3} and batch size 10^4 . 10% of the training set is kept for validation and training process finishes after 4×10^4 epoch.

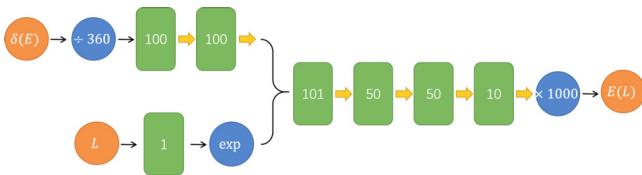


Fig. 2. (color online) The structure of our neural network. Green round rectangles with integer n represent the linear layer with size n , which consists of all the learning parameters. Orange circles denote the input and output nodes and blue circles are layers with operations marked in the middle. The yellow thick arrow marks the "SoftPlus" activation function and the right brace is a conjunction of the corresponding layers.

In Fig. 2, L and $\delta(E)$ are fed separately into different ports, since they are different physical quantities ($\delta(E)$ has nothing to do with L). In order to speed up the training, we also normalize the input $\delta(E)$ by dividing 360 and output $E(L)$ by multiplying 1000. Instead of the widely used rectified linear unit, we also find our task prefers more smooth activation functions, such as "SoftPlus". Although a strict proof is still missing, we speculate this preference of smooth activation functions originates from the regularity or even analyticity of the formulas in physics. Compared with the tiny neural network such as LeNet-5 in computer vision [33], our network is even smaller. However, it turns out that such a simple network is already adequate to make notable predictions.

IV. RESULTS ANALYSIS

For a regression task, one natural test is to calculate the deviation $\Delta(E) = E_{\text{model}} - E_{NN}$ of the neural network prediction E_{NN} from the ground truth values E_{model} from models.

As shown by the histograms in Fig. 3, $\Delta(E)$ s of all the three models cluster around zero which ensures the precision of the neural network. It is also reasonable to see the precision on the test set (model B), is slightly worse than that of the training set (model A and C).

For the test set, there is an additional feature in Fig. 3: The distribution of $\Delta(E)$ has a slightly heavier tail on the right. This implies that E_{NN} is generally smaller than E_{model} . It turns out, this systematic underestimation of the spectrum is not a flaw of neural network, on the contrary, it reveals that neural network is successfully trained as a decent model-independent feature extractor which essentially approximates the Lüscher's formula.

To see this, we plot the Lüscher's formula along with the model predictions in Fig. 4, where we insert the spectrum back to the phase shift and make a scattering plot of $[E_L, \delta(E_L)]$.

By definition, these points should agree with the Lüscher's formula up to a *model-dependent* correction term e^{-mL} . Theoretically, it is difficult to foresee the magnitude or even the sign of this correction term.

As shown in Fig. 4, data from model A are nearly identical to what Lüscher's formula predicts. This agrees with what we have anticipated, since the potentials in model A are generally narrow in coordinate space, the correction terms are small. Compared with this ideal-matching case, data from model C becomes much noisier. However, the mainstream of it still agrees considerably with Lüscher's formula, and the blue points are evenly scattered along two sides of red curves. It can also be seen in Fig. 5 that, with the increase of the volume size, the model becomes closer to the Lüscher's formula predictions.

Compared with model A and C, a new feature from

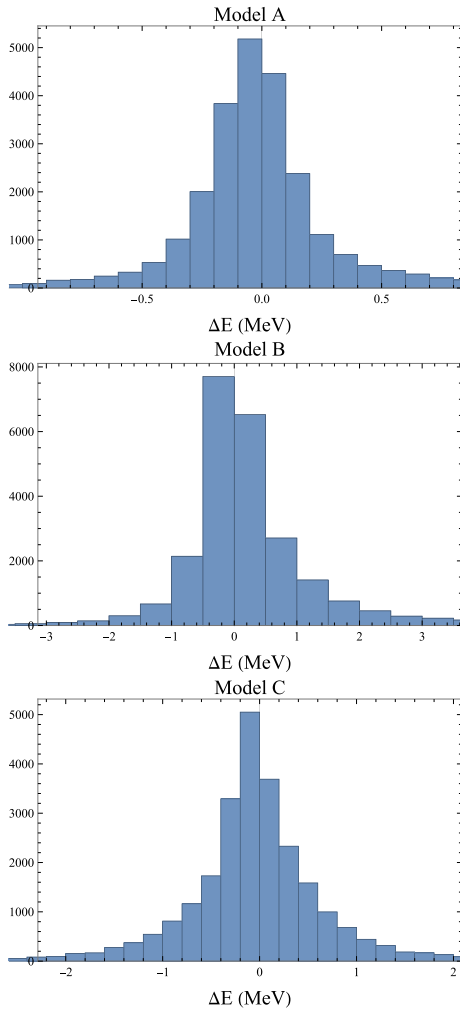


Fig. 3. (color online) The histogram of $\Delta(E) \equiv E_{\text{model}} - E_{NN}$ at $L = 10$ fm, where E_{model} , E_{NN} represent the predictions from the neural network and the model, respectively. The neural network is trained on the data from model A and C, and the data from model B serves as test set.

the model B is that the spectra from the model are systematically larger than what Lüscher's formula have predicted. Thus, if the neural network learns the Lüscher's formula well, spectra from model B will be naturally larger than neural network predictions.

The above statement can also be confirmed by comparing the two plots of the second column in Fig. 4. After training, the neural network suppresses the energy levels towards what Lüscher's formula predicts when it applies on model B, thus leading to a less accurate results and notable non-central distribution in Fig. 3. Thus, the deviation on model B signifies that neural network successfully captures the model-independent ingredients in the process $\delta(E) \rightarrow E(L)$ and effectively treats the model-dependent features as noise.

We speculate that this may partially due to the small size of the neural network (28362 parameters V.S.

3.5×10^5 energy points + 5×10^5 phase shift points in training set), which keeps the neural network from learning or even memorizing the highly model-dependent feature (see. e.g. Ref. [34] for the risk when the number of parameters exceeds the number of data points). Since Lüscher's formula is the only model-independent approach, this lead to our central conclusion that we get a neural network reprint of the numerical Lüscher's formula.

To make a stronger evidence that the numerical Lüscher's formula is learned by neural network, it is necessary to expand the test set and explore the generalizability further, i.e., challenge the neural network by more different types of phase shifts. This will not only reveal more interesting structure of the neural network, but also guide us to spot a subtle deficiency in the above treatment.

One typical pattern of the phase shift $\delta(E)$ in our training and test set is that, with the increase of energy, $\delta(E)$ will departure from zero at $2m_\pi$ threshold, gain a sharp or broad resonance structure in the middle steps and end up to be 0° or $\pm 180^\circ$. Here, we will challenge the neural network by feeding a *constant* phase shift $\delta(E) = \delta_0$, where δ_0 ranges in $[-180^\circ, +180^\circ]$. Since this constant phase shift is far beyond our training set, it would be impossible to pass the test if the neural network were doing nothing but a trivial memorization.

In Fig. 6, the agreement between Lüscher's formula and neural network is even more fascinating except an unexpected twist around $\delta = 0^\circ$. To be precise, if we track the lowest level of the spectrum E_1 , the neural network concludes from the data that E_1 should generally increase with the decrease of δ_0 . However, once δ_0 crosses the zero from above, another lower energy level will emerge. Thus, as a function of δ_0 , E_1 is not a continuous function at zero. This discontinuity is essentially caused by the periodicity of the phase shift: $\delta(E)$ and $\delta(E) + n\pi$ corresponds to the same physics. On the other side, since the neural network is designed to predict the lowest 10 energy levels above the threshold, and the activation functions are continuous in order to do back-propagation in the training process, the best neural network can achieve is to make a soft transition between the neighbor red curves around $\delta_0 = 0^\circ$, resulting several zigzag tracks in Fig. 6. It is also worthy to find that this twist structure does not manifest itself in Fig. 4, which makes this constant-phase-shift-test valuable.

We circumvent this twist issue by the following approach. The energy level E is marked as E_1 only when $\delta(E)$ is negative and $E < 2\sqrt{(\frac{2\pi}{L})^2 + m_\pi^2}$, otherwise, the valid energy levels starts from E_2 . Noting that this does not request any pre-knowledge of the Lüscher's formula. It is essentially a convention that $\delta(E)$ is zero at the following free energies

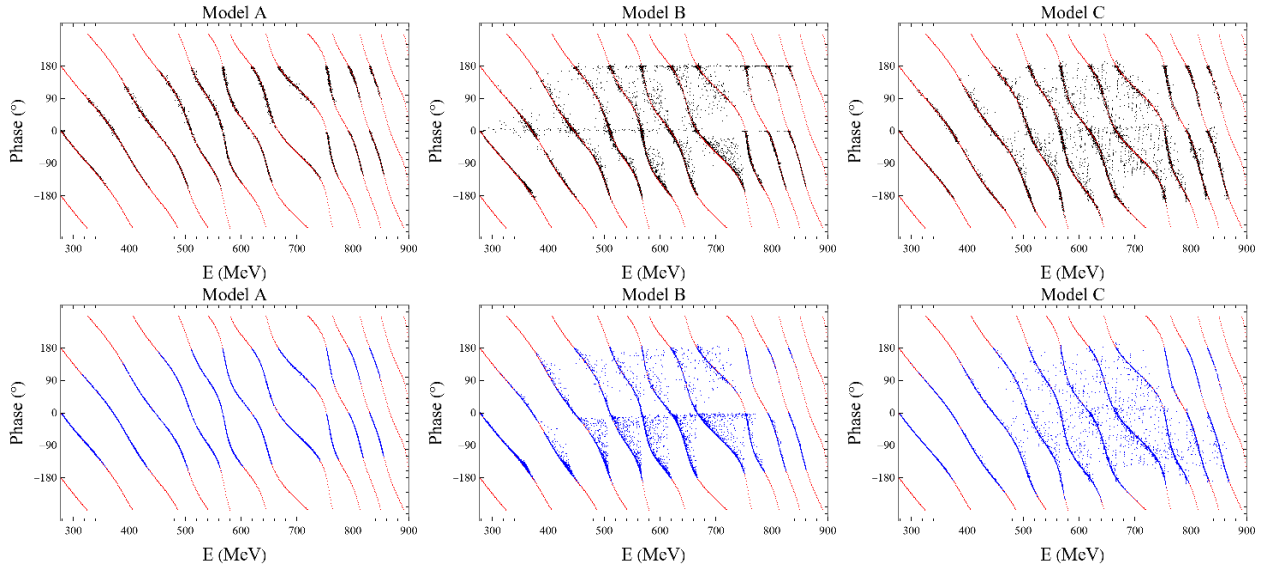


Fig. 4. (color online) Comparison of the Lüscher's formula (red), predictions from the neural network (black) and models (blue), where lattice size is 10 fm.

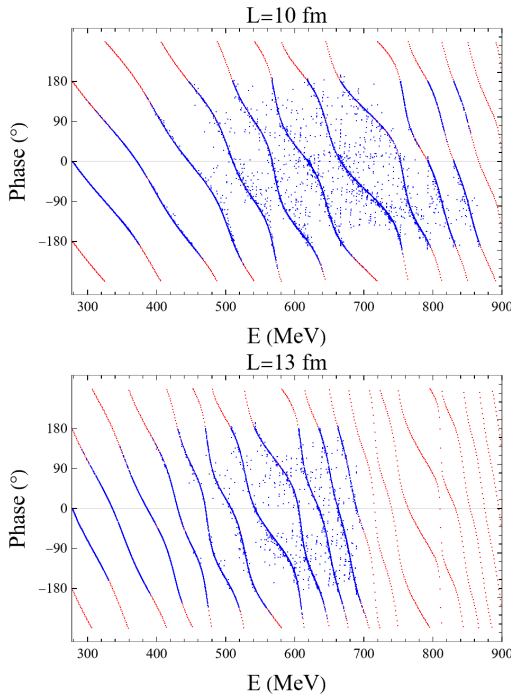


Fig. 5. (color online) Comparison of model C (blue points) with the Lüscher's formula in different volume sizes.

$$E_{\text{free}} := 2 \sqrt{\vec{n}^2 \left(\frac{2\pi}{L} \right)^2 + m_\pi^2}, \quad (21)$$

where $\vec{n} = (n_x, n_y, n_z)$, $n_{x,y,z} = 0, \pm 1, \pm 2, \dots$. Retraining the neural network with the above modification results in a superb agreement with the Lüscher's formula, which is shown in Fig. 7. The slightly worse precision around $\pm 180^\circ$ can be improved by either increasing the size of the neural network or we can simply ignore the neural

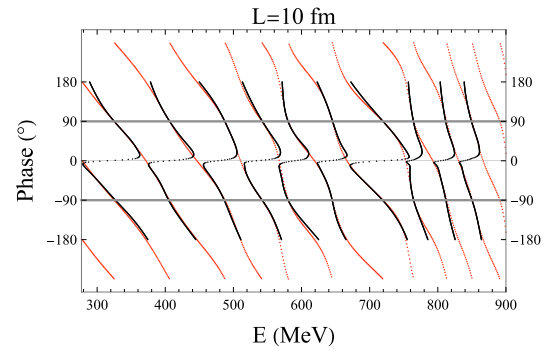


Fig. 6. (color online) Prediction (black dots) from the neural network when phase shift is constant $\delta(E) = \delta_0$, $\delta_0 \in [-180^\circ, 180^\circ]$. The precise Lüscher's formula curve is marked as red dots. One period boundary $\pm 90^\circ$ is marked by gray horizontal line for comparison.

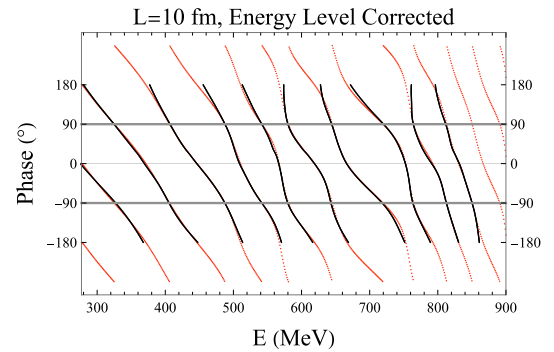


Fig. 7. (color online) Same as Fig. 6, with energy level issue explicitly addressed.

network predictions by constraining it within a period, such as $[-90^\circ, +90^\circ]$ and extrapolate the results to other regions by periodicity. After addressing this twist issue,

we finally strengthen the previous conclusion that the numerical form of Lüscher's formula is learned by the neural network.

V. SUMMARY AND OUTLOOK

In this paper, we have shown that the numerical form of Lüscher's formula can be rediscovered by the neural network when it is trained to predict the spectrum on lattice from the phase shift in continuous space. From the perspective of pragmatism, the neural network is able to exploit the sophisticated data and extract valuable inform-

ation. In a broad sense, our work is an concrete example to demonstrate how to extract model-independent link between model-dependent quantities in a data-driven approach. Surprised by the capability of the neural network, we believe its potential is still waiting for physicists to explore.

VI. ACKNOWLEDGMENT

We are grateful to Yan Li, Qian Wang, Ross D. Young and James M. Zanotti for useful discussions.

References

- [1] S.-M. Udrescu and M. Tegmark, *Sci. Adv.* **6**(16), eaay2631 (2020)
- [2] S.-M. Udrescu, A. Tan, J. Feng *et al.*, arXiv:2006.10782 [cs.LG]
- [3] Z. Liu, M. Tegmark, *Phys. Rev. Lett.* **126**(18), 180604 (2021), arXiv:2011.04698
- [4] G. Carleo, I. Cirac, K. Cranmer *et al.*, *Rev. Mod. Phys.* **91**(4), 045002 (2019), arXiv:1903.10563
- [5] P. E. Shanahan, A. Trewartha, and W. Detmold, *Phys. Rev. D* **97**(9), 094506 (2018), arXiv:1801.05784
- [6] S. Y. Chen, H. T. Ding, F. Y. Liu *et al.*, arXiv: 2110.13521
- [7] D. L. B. Sombillo, Y. Ikeda, and T. Sato, *Phys. Rev. D* **104**(3), 036001 (2021), arXiv:2105.04898
- [8] D. Liu, C. Sun, and J. Gao, *JHEP* **08**, 088 (2022), arXiv:2201.06586
- [9] Z. Zhang, R. Ma, J. Hu *et al.*, arXiv: 2208.03165
- [10] H. Chen, W.-Q. Niu, H.-Q. Zheng, arXiv: 2205.03572
- [11] J. Liu, Z. Zhang, J. Hu, Q. Wang, *Phys. Rev. D* **105**(7), 076013 (2022), arXiv:2202.04929
- [12] M. Raissi, P. Perdikaris, and G. Karniadakis, *J. Comput. Phys.* **378**, 686 (2019)
- [13] M. Luscher, *Commun. Math. Phys.* **104**, 177 (1986)
- [14] M. Luscher, *Commun. Math. Phys.* **105**, 153 (1986)
- [15] M. Luscher, *Nucl. Phys. B* **354**, 531 (1991)
- [16] M. T. Hansen and S. R. Sharpe, *Phys. Rev. D* **90**(11), 116003 (2014), arXiv:1408.5933
- [17] A. W. Jackura, S. M. Dawid, C. Fernández-Ramírez *et al.*, *Phys. Rev. D* **100**(3), 034508 (2019), arXiv:1905.12007
- [18] M. T. Hansen and S. R. Sharpe, *Ann. Rev. Nucl. Part. Sci.* **69**, 65 (2019), arXiv:1901.00483
- [19] T. D. Blanton and S. R. Sharpe, *Phys. Rev. D* **103**(5), 054503 (2021), arXiv:2011.05520
- [20] H.-W. Hammer and J.-Y. Pang, *JHEP* **09**, 109 (2017), arXiv:1706.07700
- [21] F. Müller and A. Rusetsky, *JHEP* **03**, 152 (2021), arXiv:2012.13957
- [22] F. Müller, J.-Y. Pang, A. Rusetsky *et al.*, *JHEP* **02**, 158 (2022), arXiv:2110.09351
- [23] M. Mai and M. Döring, *Eur. Phys. J. A* **53**(12), 240 (2017), arXiv:1709.08222
- [24] M. Mai and M. Döring, *Phys. Rev. Lett.* **122**(6), 062503 (2019), arXiv:1807.04746
- [25] R. Brett, C. Culver, M. Mai *et al.*, *Phys. Rev. D* **104**(1), 014501 (2021), arXiv:2101.06144
- [26] M. Mai, M. Döring, and A. Rusetsky, *Eur. Phys. J. ST* **230**(6), 1623 (2021), arXiv:2103.00577
- [27] K. Hornik, M. Stinchcombe, and H. White, *Neural Networks* **2**(5), 359 (1989)
- [28] K. Hornik, *Neural Networks* **4**(2), 251 (1991)
- [29] M. Leshno, V. Y. Lin, A. Pinkus *et al.*, *Neural Networks* **6**(6), 861 (1993)
- [30] J.-J. Wu, T. S. H. Lee, A. W. Thomas *et al.*, *Phys. Rev. C* **90**(5), 055206 (2014), arXiv:1402.4868
- [31] A. Matsuyama, T. Sato, and T. S. H. Lee, *Phys. Rept.* **439**, 193 (2007), arXiv:nucl-th/0608051
- [32] H. Kamano, S. X. Nakamura, T. S. H. Lee *et al.*, *Phys. Rev. D* **84**, 114019 (2011), arXiv:1106.4523
- [33] Y. LeCun, B. Boser, J. S. Denker *et al.*, *Neural Computation* **1**(4), 541 (1989)
- [34] C. Zhang, S. Bengio, M. Hardt *et al.*, arXiv: 1611.03530