

Unveiling the pole structure of S-matrix using deep learning

D. L. B. Sombillo^{a,b,*}, Y. Ikeda^{c,*}, T. Sato^{b,†}, and A. Hosaka^{a,d,•}
^a*National Institute of Physics, University of the Philippines Diliman,
 Quezon City 1101, Philippines.*
^{*}*e-mail: sombillo@rcnp.osaka-u.ac.jp; dbsombillo@up.edu.ph*
<https://orcid.org/0000-0001-9357-7236>
^b*Research Center for Nuclear Physics (RCNP),
 Osaka University, Ibaraki, Osaka 567-0047, Japan.*
[†]<https://orcid.org/0000-0001-5216-5657>
^c*Department of Physics, Kyushu University,
 Fukuoka 819-0395, Japan.*
^{*}<https://orcid.org/0000-0002-2235-1464>
^d*Advanced Science Research Center, Japan Atomic Energy Agency,
 Tokai, Ibaraki 319-1195, Japan.*
[•]<https://orcid.org/0000-0003-3623-6667>

Received 23 December 2021; accepted 1 March 2022

Particle scattering is a powerful tool to unveil the nature of various subatomic phenomena. The key quantity is the scattering amplitude whose analytic structure carries the information of the quantum states. In this work, we demonstrate our first step attempt to extract the pole configuration of inelastic scattering using the deep learning method. Among various problems, motivated by the recent new hadron phenomena, we develop a curriculum learning method of deep neural network to analyze coupled channel scattering problems. We show how effectively the method works to extract the pole configuration associated with resonances in the πN scattering.

Keywords: πN scattering.

DOI: <https://doi.org/10.31349/SuplRevMexFis.3.0308067>

1. Introduction

The past decades saw the proliferation of new peak structures in hadronic scattering cross-sections with properties eluding the conventional quark model [1–6]. Specifically, most of these peaks appear close to some two-hadron threshold, leading to a variety of possible interpretations of its nature. Some enhancements are claimed to be candidates of the conventional qqq or $q\bar{q}$ resonant excitation [7,8], or one of the novels multi-quark states like pentaquarks or tetraquarks [9,10]. The proximity of peaks to the two-hadron threshold suggests that the coupled channel interaction in the S -wave of two hadrons plays a significant role [11]. For instance, the observed peaks may be generated by the two-hadron interaction, which may lead to either a hadronic molecule or a virtual state [12–14]. On the other hand, peaks can also be interpreted as kinematical threshold effects where the presence of a dynamical physical state is not relevant. These kinematical effects can be attributed to either the two-body threshold cusp or the three-body triangle singularity [15–17]. Identifying which of the observed near-threshold peaks correspond to resonances is crucial to our ongoing quest to understand the mechanism of color confinement.

For the model-dependent treatment, the pole position is obtained by fitting the parameters of a model to the experimental data. In a coupled-channel analysis, the model contains some channel coupling parameters that can be switched

off. This model's feature allows us to determine the nature of enhancement by tracing the origin of the pole in the zero-coupling limit [18–21]. Alternatively, a model-independent analysis is possible by using the effective range expansion around some threshold of interest. Specifically, a pole-counting method was proposed in Refs. [22, 23] which relates the number of near-threshold poles on different unphysical sheets to the nature of enhancement. It was later shown in Ref. [24] that the pole-counting method is consistent with Weinberg's compositeness criterion [25]. It follows that the identification of pole configuration, which we define to be the number of poles in each sheet associated with the enhancement, is a crucial starting point in a model-independent analysis of near-threshold phenomena. The task is now reduced to a classification problem. For this purpose, the deep learning approach is one of the best ways to be employed. Here, we show how to extract the pole configuration of the S-matrix using deep learning.

Deep learning is a versatile tool used in the physical sciences [26]. The availability of a large dataset is essential to improve the performance of a deep neural network (DNN) model. Unlike in other fields, experimental data in hadron physics is limited and with statistical uncertainties. It is, therefore, imperative to introduce a general method to generate the teaching dataset containing a large number of simulated amplitudes with known pole configurations. Here, the simulated amplitudes are mock data generated from the gen-

eral properties of the S-matrix. The generation of teaching dataset is followed by the construction and improvement of the DNN model, which is then applied to the experimental data. We have demonstrated in Ref. [27] the feasibility of using deep learning in the analysis of single-channel scattering where a generic S-matrix is used to generate the teaching dataset. Here, we extend our method to the coupled-channel case.

In this paper, we present a deep learning approach to study the coupled-channel scattering with near-threshold enhancement. We design a classification DNN that determines the S-matrix pole configuration when an input partial-wave amplitude is given. In doing so, we introduce a general S-matrix, with controlled Riemann sheet poles, to generate the teaching dataset. Our present method achieves two significant improvements - the inclusion of energy resolution in the DNN design and the utilization of amplitude's error bars in the final DNN inference. The simulated energy resolution, which is done by using random energy points, introduces noise in the teaching dataset. We find that the presence of noise requires the curriculum method to initiate the learning process. As a concrete example, we treat the GW-SAID elastic πN scattering analysis in Ref. [28–30] as the experimental data. It must be emphasized that our method can only determine the Riemann sheets where the poles are located, but the method does not tell their positions. Our approach can be extended to general resonance analysis, not only in hadron physics but also in nuclear or atomic physics, where low-energy scattering reveals interesting universal features.

2. S-matrix poles

Let us consider the two-hadron scattering with two channels where the relative momentum of the i th channel ($i = 1, 2$) is p_i , with reduced mass μ_i , and threshold at T_i . Specifically, we set the πN system as channel 1 and the ηN as channel 2. Aside from the πN and ηN thresholds, there are also the $K\Lambda$ and $K\Sigma$ as shown in Fig. 1. We justify the two-channel analysis as follows. On the one hand, there is no significant threshold effect with the opening of the $K\Lambda$ channel and, thus, can be ignored [31]. On the other hand, we can avoid the possible effect of $K\Sigma$ channel [32] by not going beyond the $K\Sigma$ threshold. Hence, it suffices to consider two channels for the present study.

Now, the non-relativistic relation between the scattering energy E and channel momentum p_i is

$$E = \frac{p_i^2}{2\mu_i} + T_i. \quad (1)$$

The non-relativistic treatment will suffice in our present purpose since we restrict our analysis around the second threshold. The smooth variation of the lower channel can be approximated by Eq. (1) where p_1 serves as a label to E for the lower channel. Using analyticity, unitarity, and hermiticity in the relevant energy region [33, 34], the lower channel S-matrix takes the form

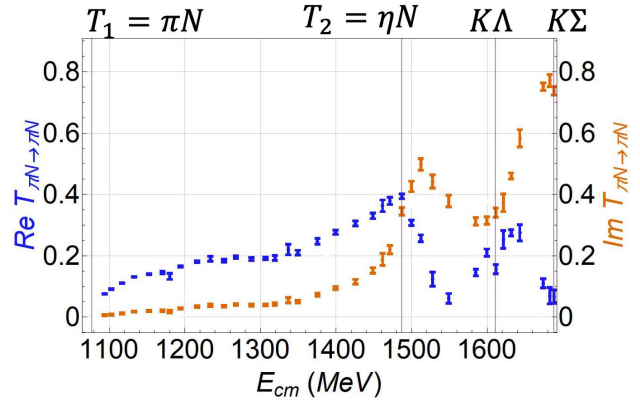


FIGURE 1. The elastic πN amplitude of the GW-SAID in Ref. [28]. The two-hadron thresholds are shown as vertical thin lines. Only the πN and ηN channels are considered in the present study. The $K\Lambda$ and $K\Sigma$ thresholds are shown for reference purposes only.

$$S_{11}(p_1, p_2) = \frac{D(-p_1, p_2)}{D(p_1, p_2)}, \quad (2)$$

with the scattering amplitude $T_{11}(p_1, p_2)$ obtained via $S_{11} = 1 + 2iT_{11}$. The function $D(p_1, p_2)$ contains all the pole singularities of $S_{11}(p_1, p_2)$.

In the generation of simulated amplitudes, we have to consider the general case of more than one pole. To do this, we express $D(p_1, p_2)$ as a product of independent $D_j(p_1, p_2)$ such that $D_j(p_1, p_2) = 0$ is satisfied by one assigned j th pole, $E = E_{\text{pole}}^{(j)}$, on a chosen Riemann sheet. Here, the assigned pole $E_{\text{pole}}^{(j)}$ can be one of the poles related to the observed peaks in the scattering region. The form of $D_j(p_1, p_2)$ can be deduced by hermiticity, where the conjugate pair of energy poles is implied. That is, if $E = E_{\text{pole}}^{(j)}$ is a solution to $D_j(p_1, p_2) = 0$ on one Riemann sheet, so is its complex conjugate $E = E_{\text{pole}}^{(j)*}$ on the same sheet. Thus, we can write $D_j(p_1, p_2)$ in a form where the conjugate partner is explicitly included, and the Riemann sheet is transparently shown:

$$D_j(p_1, p_2) = \left[(p_1 - i\beta_1^{(j)})^2 - \alpha_1^{(j)2} \right] + \lambda \left[(p_2 - i\beta_2^{(j)})^2 - \alpha_2^{(j)2} \right], \quad (3)$$

where the absolute values of $\alpha_i^{(j)}$ and $\beta_i^{(j)}$ for $i = 1, 2$ are determined by the assigned j th energy pole $E = E_{\text{pole}}^{(j)}$ through Eq. (1). The Riemann sheet of $E_{\text{pole}}^{(j)}$ can be chosen arbitrarily by choosing the signs of $\beta_1^{(j)}$ and $\beta_2^{(j)}$. The parameter λ is chosen to control the other set of conjugate solutions to $D_j(p_1, p_2) = 0$ (called the shadow pole) so that they will not appear on the physical sheet and will remain far from the relevant scattering region. This ensures that $E_{\text{pole}}^{(j)}$ is the only nearby pole produced by $D_j(p_1, p_2)$. For a more detailed discussion of shadow pole related to Eq. (3), we refer the reader to Ref. [35].

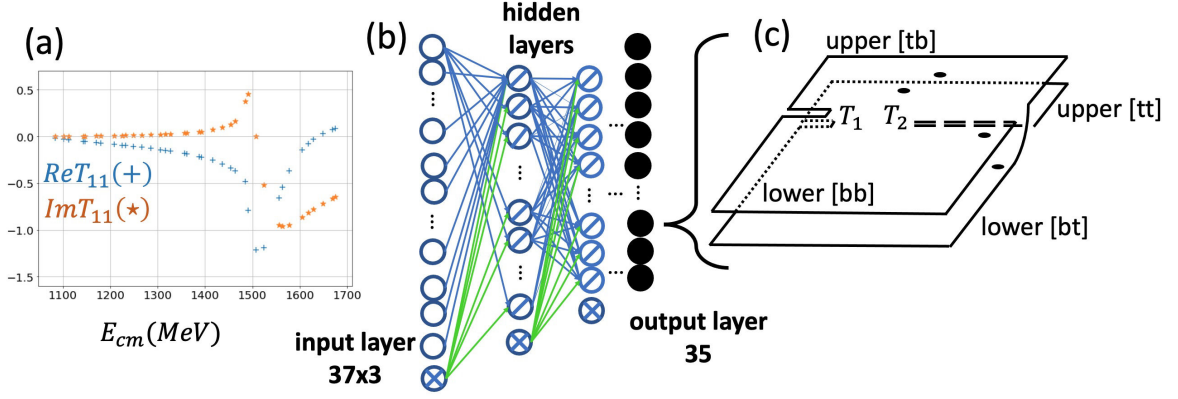


FIGURE 2. Schematic operation of our DNN model where (a) is one of the simulated amplitudes (b) is the DNN model and (c) is one of the representative output pole configuration. For (b), \circ s are for input nodes, \otimes s are for bias nodes \circ s are for hidden layer nodes and \bullet s are for output nodes. The dashed lines in (c) are connected to each other.

3. Development of DNN

The parametrization that we have introduced allows us to generate simulated amplitudes on arbitrary energy points. However, it is always the case that the experimental data has energy resolution and error bars for the amplitude. The resolution can be accommodated in the design of DNN as follows. We start by determining the region of interest, which in this study is from πN to the $K\Sigma$ thresholds. The GW-SAID data has 37 points within the mentioned region, so we split the πN - $K\Sigma$ into 37 bins. Anticipating that the energy points may not be equally spaced, we randomly select a representative energy point in each bin using a uniform probability distribution. Then, we calculate the amplitude on the randomly picked energy points. Proper labels can be assigned to each simulated amplitude because we have complete control of the poles. The amplitude's error bars are utilized in the last stage of deep learning analysis.

The relevant structures in the GW-SAID data are seen only on some finite interval of the scattering region, as shown in Fig. 1. It is, therefore, practical to count only the poles that are within the range $T_2 - 50 \leq \text{Re}E_{\text{pole}} \leq T_2 + 200$, below the real axis ($-200 \leq \text{Im}E_{\text{pole}} < 0$) of the $[bt]$ or $[bb]$ and poles above real axis ($0 < \text{Im}E_{\text{pole}} \leq 200$) of the $[tb]$. Here, the energy units are in MeV. We do not have to count the conjugate poles because they do not correspond to different independent states. Also, remote background poles are randomly added, by inserting extra $D_{\ell \neq j}(p_1, p_2)$ factor in $D(p_1, p_2)$, but are also not counted. In addition, we restrict the maximum number of nearby poles in any unphysical sheet to four since there are only two prominent structures in the πN scattering amplitude. The constraint gives us a total of 35 possible configurations ranging from no nearby poles to at least one pole in each unphysical Riemann sheet. We randomly generate poles inside the counting region to produce an equal number of simulated amplitudes per configuration, giving us a total of around 1.8×10^6 simulated amplitudes for the teaching dataset. An independent 3.5×10^4 set of simu-

lated amplitudes is produced to measure the performance of the DNN and to check for possible overfitting.

Figure 2 shows the schematic diagram of our DNN model's operation. Given an input amplitude in Fig. 2a), our DNN will take the random energy points and the amplitude's real and imaginary parts as the 37×3 input-node values. The output nodes are configured to match the total number of possible pole configurations. If the weights and biases (lines in Fig. 2b)) are already optimized, then the DNN can give its output, which is schematically represented in Fig. 2c). It must be emphasized that our DNN can only count the number of poles in each Riemann sheet and not give the exact pole positions. Now, the performance of our DNN model can be improved in a training loop. Here, all the simulated amplitudes are fed in a forward pass to estimate the cost function, and some variant of stochastic gradient descent is used to do the backpropagation. We trained six different architectures and found that none of them learn the classification problem. In particular, after around 500 epochs, the training accuracies are only around 2.86%, which is the accuracy if one makes a random guess out of the 35 possibilities. Our preliminary numerical experiment suggests that our classification problem is too complex for our DNN models, perhaps due to the noise present in the dataset. Thus, to initiate the learning process, we use the curriculum method.

The basic idea of the curriculum method is to recognize the easy examples of a complex dataset and use them for initial training of the model [36]. For most classification pro-

TABLE I. Chosen DNN Architecture.

Layer	Number of nodes	Activation Function
Input	111+1	
1st	200+1	ReLU
2nd	200+1	ReLU
3rd	200+1	ReLU
Output	35	Softmax

blems, an additional procedure is expected since one has to identify which part of the dataset is less complicated than the other [37–39]. For our case, an additional procedure is needed. We choose the at-most-1-pole configuration as the easy dataset and slowly add multi-pole classifications until all the datasets are introduced. We measure our models’ performance on the easy classification set and then devote our computing resources to the best-performing model for the complete curriculum learning [35]. We have chosen the DNN model with specifications given in Table I.

Except for the easy dataset, one new classification is added every after 100 epochs. We also vary the mini-batch size in each epoch as we saw fit. The smaller mini-batch size is optimal in the early part of the new-classification-epoch while the large ones are better for the later part. After 3,200 epochs, we now have a pre-trained DNN model that can detect up to four nearby poles on any Riemann sheet with training accuracy of 63.5% and testing accuracy of 68.3%. We further train the model up to 31,050 epochs, obtaining a final performance of 76.5% for the training and 80.4% for the testing.

4. Results and discussion

The GW-SAID amplitude in Fig. 1 is used as the experimental data in the following discussions. Note that the proper description of experimental data must include the uncertainties, and the deep learning approach is no exception here. In addition, we should be able to estimate the confidence of the trained DNN’s prediction once the experimental data is fed. To accomplish these, we first interpret the error bars of the amplitude as a collection of points weighted by some probability distribution. The most natural choice is to treat each error bar as one standard deviation of a Gaussian distribution. We produce a collection of 10^6 amplitudes by combining the points in each error bar. Finally, the amplitudes are fed directly to the trained DNN models, and the outputs are counted. The output with the highest count is the possible pole configuration of the experimental data.

Out of the 35 possibilities, our trained DNN linked only four-pole configurations to the πN scattering amplitude. As shown in Table II, the structures in Fig. 1 are due to one pole

TABLE II. Result of the DNN inferences on the GW-SAID πN scattering amplitude. The error bar is interpreted as one standard deviation to generate 10^6 amplitudes from the experimental data.

Percentage	bt	bb	tb
44.6%	1	1	2
34.1%	1	1	1
16.4%	0	1	3
4.9%	0	1	2

in $[bt]$, one pole in $[bb]$ and at most two poles in $[tb]$. Note that no assumptions are made on any of the detected poles since they are produced independently in the training dataset. It is now up to some dynamical model to interpret their origin and explain how they relate to each other.

Our approach can go beyond the conventional model-fitting scheme, where the error bar is typically interpreted as one standard deviation of a Gaussian distribution. In the Ref. [35], we showed that the choice of a probability distribution for the description of error bars does not affect the outcome of DNN inference. In particular, the same set of configurations is obtained, and the configuration with the highest count remains no matter which error bar interpretation is used. This observation demonstrates that our deep learning approach gives a statistically robust interpretation of experimental data.

5. Conclusion and Outlook

In this paper, we have shown for the first time how to implement deep learning in the study of coupled-channel scattering with near-threshold enhancement. The DNN model is designed to accommodate the energy resolution of the experimental data and the error bars, which are utilized in the final DNN inference. As a result, the confidence of DNN’s inference can be estimated by counting the output, with the highest count being the best description of experimental data.

The deep learning approach used in this paper can be extended to any peak analysis of the experimental data. In the present study, we focused only on identifying the pole configuration that best describes a near-threshold enhancement. No *a priori* assumption is made on the detected poles; they are independently produced in the generation of the teaching dataset. However, to further identify the nature of near-threshold phenomena, we need to employ a suitable dynamical model to reproduce the detected poles. We can then determine the origin of enhancements by using the obtained pole configuration and tracing their trajectories through the coupling mechanism of a dynamical model.

Acknowledgment

This study was supported in part by MEXT as “Program for Promoting Researches on the Supercomputer Fugaku” (Simulation for basic science: from fundamental laws of particles to creation of nuclei). DLBS is supported in part by the DOST-SEI ASTHRDP postdoctoral research fellowship. YI is partly supported by JSPS KAKENHI Nos. JP17K14287 (B) and 21K03555 (C). AH is supported in part by JSPS KAKENHI No. JP17K05441 (C) and Grants-in-Aid for Scientific Research on Innovative Areas, No. 18H05407 and 19H05104.

1. S. L. Olsen, T. Skwarnicki, and D. Zieminska, Nonstandard heavy mesons and baryons: Experimental evidence, *Rev. Mod. Phys.* **90**, (2018) 015003, arXiv:1708.04012 [hep-ph].
2. R. Aaij *et al.* (LHCb Collaboration), Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$, *Phys. Rev. Lett.* **122**, (2019) 222001, arXiv:1904.03947 [hep-ex].
3. R. Aaij *et al.* (LHCb Collaboration), Observation of structure in the J/ψ -pair mass spectrum, *Sci. Bull.* **65** (2020) 1983, arXiv:2006.16957 [hep-ex].
4. X.-K. Dong, V. Baru, F.-K. Guo, C. Hanhart, and A. Nefediev, Coupled-Channel Interpretation of the LHCb Double- J/ψ Spectrum and Hints of a New State Near the $J/\psi J/\psi$ Threshold, *Phys. Rev. Lett.* **126**, 132001 (2021), arXiv:2009.07795 [hep-ph].
5. E. Wang, W.-H. Liang, and E. Oset, Analysis of the $e^+e^- \rightarrow J/\psi DD$ reaction close to the threshold concerning claims of a $\chi_{cc0}(2P)$ state, *Eur. Phys. J. A* **57**, (2021) 38, arXiv:1902.06461 [hep-ph].
6. J. Haidenbauer and U. G. Meißner, On the structure in the ΛN cross section at the ΣN threshold, (2021), arXiv:2105.00836 [nucl-th].
7. S. Capstick and W. Roberts, Quark models of baryon masses and decays, *Prog. Part. Nucl. Phys.* **45**, (2000) S241, arXiv:nucl-th/0008028.
8. R. L. Jaffe, Ordinary and extraordinary hadrons, *AIP Conf. Proc.* **964**, (2007) 1, arXiv:hep-ph/0701038.
9. A. Ali, J. S. Lange, and S. Stone, Exotics: Heavy Pentaquarks and Tetraquarks, *Prog. Part. Nucl. Phys.* **97**, (2017) 123, arXiv:1706.00610 [hep-ph].
10. Y.-R. Liu, H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, Pentaquark and Tetraquark states, *Prog. Part. Nucl. Phys.* **107**, (2019) 237, arXiv:1903.11976 [hep-ph].
11. X.-K. Dong, F.-K. Guo, and B.-S. Zou, Explaining the Many Threshold Structures in the Heavy-Quark Hadron Spectrum, *Phys. Rev. Lett.* **126** (2021) 152001, arXiv:2011.14517 [hep-ph].
12. F.-K. Guo, C. Hanhart, U.-G. Meißner, Q. Wang, Q. Zhao, and B.-S. Zou, Hadronic molecules, *Rev. Mod. Phys.* **90** (2018) 015004, arXiv:1705.00141 [hep-ph].
13. Y. Yamaguchi, A. Hosaka, S. Takeuchi, and M. Takizawa, Heavy hadronic molecules with pion exchange and quark core couplings: a guide for practitioners, *J. Phys. G* **47** (2020) 053001, arXiv:1908.08790 [hep-ph].
14. T. Hyodo, Structure and compositeness of hadron resonances, *Int. J. Mod. Phys. A* **28** (2013) 1330045, arXiv:1310.1176 [hep-ph].
15. M. Bayar, F. Aceti, F.-K. Guo, and E. Oset, A Discussion on Triangle Singularities in the $\Lambda_b \rightarrow J/\psi K^- p$ Reaction, *Phys. Rev. D* **94**, 074039 (2016), arXiv:1609.04133 [hep-ph].
16. F.-K. Guo, X.-H. Liu, and S. Sakai, Threshold cusps and triangle singularities in hadronic reactions, *Prog. Part. Nucl. Phys.* **112** (2020) 103757, arXiv:1912.07030 [hep-ph].
17. S. X. Nakamura, Triangle singularity appearing as an X(3872)-like peak in $B \rightarrow (J/\psi \pi^+ \pi^-) K \pi$ *Phys. Rev. D* **102** (2020) 074004, arXiv:1912.11830 [hep-ph].
18. W. R. Frazer and A. W. Hendry, S-Matrix Poles Close to Threshold, *Phys. Rev.* **134** (1964) B1307.
19. B. C. Pearce and B. F. Gibson, Observable effects of poles and shadow poles in coupled-channel systems, *Phys. Rev. C* **40** (1989) 902.
20. A. M. Badalyan, L. P. Kok, M. I. Polikarpov, and Y. A. Simonov, Resonances in coupled channels in nuclear and particle physics, *Phys. Rep.* **82** (1982) 31.
21. C. Hanhart, J. R. Pelaez, and G. Rios, Remarks on pole trajectories for resonances, *Phys. Lett. B* **739** (2014) 375, arXiv:1407.7452 [hep-ph].
22. D. Morgan and M. Pennington, $f_0(S^*)$: molecule or quark state?, *Phys. Lett. B* **258** (1991) 444.
23. D. Morgan, Pole counting and resonance classification, *Nucl. Phys. A* **543** (1992) 632.
24. V. Baru, J. Haidenbauer, C. Hanhart, Y. Kalashnikova, and A. E. Kudryavtsev, Evidence that the $a(0)(980)$ and $f(0)(980)$ are not elementary particles, *Phys. Lett. B* **586** (2004) 53, arXiv:hep-ph/0308129.
25. S. Weinberg, Evidence that the deuteron is not an elementary particle, *Phys. Rev.* **137** (1965) B672.
26. G. Carleo *et al.*, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91** (2019) 045002, arXiv:1903.10563 [physics.comp-ph].
27. D. L. B. Sombillo, Y. Ikeda, T. Sato, and A. Hosaka, Classifying the pole of an amplitude using a deep neural network, *Phys. Rev. D* **102** (2020) 016024, arXiv:2003.10770 [hep-ph].
28. http://gwdac.phys.gwu.edu/analysis/pin_analysis.html.
29. R. L. Workman, R. A. Arndt, W. J. Briscoe, M. W. Paris, and I. I. Strakovsky, Parameterization dependence of T matrix poles and eigenphases from a fit to πN elastic scattering data, *Phys. Rev. C* **86**, 035202 (2012), arXiv:1204.2277 [hep-ph].
30. R. A. Arndt, W. J. Briscoe, I. I. Strakovsky, and R. L. Workman, Extended partial-wave analysis of πN scattering data, *Phys. Rev. C* **74** (2006) 045205, arXiv:nucl-th/0605082.
31. H. Kamano, S. X. Nakamura, T. S. H. Lee, and T. Sato, Nucleon resonances within a dynamical coupled-channels model of πN and γN reactions, *Phys. Rev. C* **88** (2013) 035209, arXiv:1305.4351 [nucl-th].
32. N. Kaiser, P. B. Siegel, and W. Weise, Chiral dynamics and the S11 (1535) nucleon resonance, *Phys. Lett. B* **362**, (1995) 23, arXiv:nucl-th/9507036.
33. R. G. Newton, Structure of the Many-Channel S-matrix, *J. Math. Phys.* **2** (1961) 188.
34. K. J. L. Couteur and R. E. Peierls, The structure of a non-relativistic S-matrix, *Proc. R. Soc. Lon. A.* **256** (1960) 115.
35. D. L. B. Sombillo, Y. Ikeda, T. Sato, and A. Hosaka, Model independent analysis of coupled-channel scattering: A deep learning approach, *Phys. Rev. D* **104** (2021) 036001, arXiv:2105.04898 [hep-ph].

36. J. L. Elman, Learning and development in neural networks: the importance of starting small, *Cognition* **48** (1993) 71.
37. Y. Bengio, J. Louradour, R. Collobert, and J. Weston, Curriculum learning, in Proceedings of the 26th Annual International Conference on Machine Learning, ICML'09 (Association for Computing Machinery, New York, NY, USA, 2009) p. 41-48.
38. A. Graves, M. G. Bellemare, J. Menick, R. Munos, and K. Kavukcuoglu, Automated curriculum learning for neural networks, in Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research, Vol. 70 (PMLR, International Convention Centre, Sydney, Australia, 2017) pp. 1311-1320.
39. G. Hacohen and D. Weinshall, On The Power of Curriculum Learning in Training Deep Networks, (2019), arXiv:1904.03626 [cs.LG].