PROCEEDINGS OF SCIENCE

Machine-learning techniques as noise reduction strategies in lattice calculations of the muon $g-2$

Thomas Blum,[a] **Alessandro Conigli,**[b,c] **Lukas Geyer,**[d] **Simon Kuberski,**[e] **Alexander Segner**[b,c,d] **and Hartmut Wittig**[d,b,c,*]

[a] *Department of Physics, University of Connecticut, Storrs, CT 06269, USA*
[b] *Helmholtz-Institut Mainz, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany*
[c] *GSI Helmholtz Centre for Heavy Ion Research, 64291 Darmstadt, Germany*
[d] *PRISMA$^+$ Cluster of Excellence and Institut für Kernphysik, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany*
[e] *Theoretical Physics Department, CERN, 1211 Geneva 23, Switzerland*

E-mail: hartmut.wittig@uni-mainz.de

Lattice calculations of the hadronic contributions to the muon anomalous magnetic moment are numerically highly demanding due to the necessity of reaching total errors at the sub-percent level. Noise-reduction techniques such as low-mode averaging have been applied successfully to determine the vector-vector correlator with high statistical precision in the long-distance regime, but display an unfavourable scaling in terms of numerical cost. This is particularly true for the mixed contribution in which one of the two quark propagators is described in terms of low modes. Here we report on an ongoing project aimed at investigating the potential of machine learning as a cost-effective tool to produce approximate estimates of the mixed contribution, which are then bias-corrected to produce an exact result. A second example concerns the determination of electromagnetic isospin-breaking corrections by combining the predictions from a trained model with a bias correction.

The 41st International Symposium on Lattice Field Theory (LATTICE2024)
28 July - 3 August 2024
Liverpool, UK

*Speaker

PoS(LATTICE2024)270

1. Motivation

The hadronic vacuum polarization (HVP) contribution, $a_\mu^{\rm hvp}$, to the muon anomalous magnetic has been a major focus of lattice calculations in recent years. The commonly used approach to compute $a_\mu^{\rm hvp}$ is based on the time-momentum representation [1] which expresses $a_\mu^{\rm hvp}$ as a convolution integral of the vector-vector correlation function $G(t)$ multiplied by an analytically known kernel function $\tilde{K}(t)$, i.e.

$$a_\mu^{\rm hvp} = \left(\frac{\alpha}{\pi}\right)^2 \int_0^\infty dt\, \tilde{K}(t) G(t)\,. \tag{1}$$

However, computing $a_\mu^{\rm hvp}$ at the required level of precision is numerically very costly due to the well-known "noise problem" which manifests itself in the exponential growth of statistical errors in $G(t)$ for large Euclidean times t [2, 3]. Furthermore, in order to match or exceed the precision of the traditional data-driven dispersive approach, it is necessary to include isospin-breaking corrections arising from QED and unequal up- and down-quark masses. While these contributions are small, they are very expensive to calculate.

In this contribution, we explore the potential of machine-learning methods to reduce the cost of evaluating computationally complex and/or noisy correlation functions. The idea is to use a trained neural net to produce approximate estimates for correlation functions at low numerical cost, thereby predicting an "expensive" observable through its correlation with a quantity that can be computed "cheaply". We report on first applications of this idea to compute the vector two-point correlation function entering the TMR definition of $a_\mu^{\rm hvp}$, as well as the calculation of electromagnetic isospin-breaking corrections to octet and decuplet baryon masses.

2. Low-mode averaging and machine-learning strategy

Deflation techniques such as low-mode averaging (LMA) [4, 5] have been applied successfully to mitigate the noise problem, e.g. in calculations of the HVP contribution using Wilson fermions [6]. LMA is based on splitting the quark propagator into a low-mode contribution and a part defined in the orthogonal complement of the space spanned by the chosen number of low eigenmodes of the system, according to

$$S(y,x) = S_{\rm eigen}(y,x) + S_{\rm rest}(y,x)\,, \tag{2}$$

where the subscripts "eigen" and "rest" denote the low- and high-mode parts, respectively. If we denote the chosen number of eigenmodes by $N_{\rm low}$, then the low-mode part of the propagator can be written in terms of the spectral decomposition as

$$S_{\rm eigen}(y,x) = \sum_{i=1}^{N_{\rm low}} \frac{v_i(y) \otimes (\gamma_5 v_i(x))^\dagger}{\lambda_i}, \quad (\gamma_5 D_{\rm w})\, v_i(x) = \lambda_i v_i(x)\,, \tag{3}$$

where $v_i(x)$ is the eigenmode of the hermitian Wilson-Dirac operator $\gamma_5 D_{\rm w}$ with eigenvalue λ_i. When computing two-point correlation functions using the decomposition of eq. (3), one obtains the "eigen-eigen", "rest-rest" and "rest-eigen" contributions schematically shown in Fig. 1. In the case of the vector-vector correlator $G(t)$ which is relevant for determining the HVP contribution, one

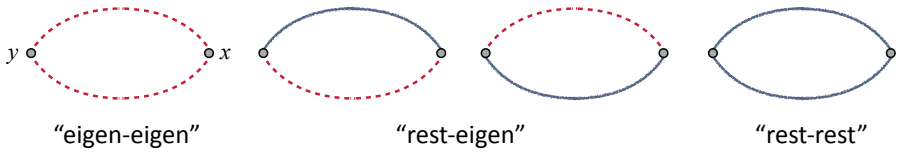

Figure 1: Sketch of the resulting two-point correlation function after applying the low-mode decomposition of eq. (3). Red dashed lines denote the low-mode part of the quark propagator expressed in terms of the spectral representation, while blue solid lines denote the high-mode part. The grey circles at points x and y are associated with a Dirac matrix.

Id	L/a	T/a	a [fm]	m_π	$N_{\rm cfg}$
A654	24	48	0.097	338	2500
D450	64	128	0.075	219	500
N451	48	96	0.075	291	1011

Table 1: CLS ensembles used in our project. Ensembles A654 and D450 were used to predict the rest-eigen contribution to pseudoscalar and vector two-point functions. Electromagnetic corrections to baryon masses were computed on ensemble N451.

finds that the "eigen-eigen" part dominates the long-distance regime, as expected (see Appendix C of Ref. [6] for a detailed account). While the mixed rest-eigen contribution is sub-dominant in the long-distance regime, its contribution to the variance can be sizable. Its evaluation scales with the number of eigenmodes, which substantially contributes to the cost, since typically $N_{\rm low} = \mathrm{O}(1000)$. Hence, if the rest-eigen part were accurately predicted through a machine-learning approach, it would potentially reduce the computational cost by a big margin.

Our machine-learning approach is inspired by the well-known all-mode-averaging (AMA) technique [7]. The goal is to compute many approximate estimates, $O_{\rm appx}$, at low numerical cost and obtain an exact result for the quantity of interest after applying a bias correction, i.e.

$$\langle O \rangle = \left\langle O_{\rm appx} \right\rangle + \left\langle (O - O_{\rm appx}) \right\rangle \ . \tag{4}$$

While the calculation of the correction $(O - O_{\rm appx})$ involves the exact (and hence more expensive) evaluation of the observable O, it is performed much less frequently than the computation of $O_{\rm appx}$. Provided that the approximate estimate can be computed with a smaller variance than the exact evaluation and that the error of the bias correction is sub-dominant, the procedure will be numerically more efficient. In the familiar case of AMA, the approximate solution is provided by the truncated solver technique; here we will replace it by a trained model. The idea to use trained networks for producing approximate estimates for an observable followed by a bias correction has been tried in the context of baryonic two- and three-point functions several years ago [8]. Here we will explore the potential of this method for obtaining statistically accurate results for the vector correlator $G(t)$ at large times. Specifically, we will train a model to predict the numerically expensive sub-leading rest-eigen contribution, given the eigen-eigen and rest-rest contributions as input.

3. Pseudoscalar and vector two-point correlation functions

Our pilot study was performed on two CLS ensembles, A654 and D450, of O(a)-improved Wilson fermions, both with periodic boundary conditions in time. Some of their features are listed in Table 1 together with those of a third ensemble (N451) used in our second test on predicting electromagnetic corrections to baryon masses. For each ensemble we divided the set of configurations into subsets for training, testing and bias correction. While the sets for testing and bias correction may share configurations, the training set is completely disjoint.

We conducted extensive tests on multiple common approaches for regression tasks, spanning from simple linear regression models to deep neural networks. This resulted in a large set of models with increasing complexity, where the number of trainable parameters varied from O(100) to O(10^5). We systematically explored the hyper-parameter space of our models using a grid-search algorithm, assessing the performance of each model through cross-validation on the training set. Our analysis suggests that architectures with a larger number of parameters might suffer from overfitting, particularly when the number of training epochs is increased. On the other hand, simple linear regression models typically struggle to capture the underlying correlation in our data, resulting in poorer predictions. In addition, we explored two different approaches: one where a single model predicts all timeslices simultaneously and another where separate models are trained for each individual time slice. Our analysis revealed comparable performance between these two strategies. For the rest of this study, we adopt the full timeslice prediction approach. Considering the results of our grid-search and the relatively small size of our training set compared with typical machine-learning applications, we employ a fully connected network with one hidden layer. The hidden layer employs ReLU activation functions to overcome the vanishing gradient problem, while the output layer uses a linear activation function to maintain unbounded predictions. To mitigate overfitting, we include dropout layers [9], which randomly deactivate a subset of parameters during training, thereby enhancing model generalization. To evaluate the performance of regression models, we used the mean squared error as our loss function. We found that normalizing the training set to have mean zero and standard deviation can significantly improve the network performance, leading to faster convergence [10]. Moreover, we explored variations of the training set by incorporating either the eigen-eigen data, the rest-rest data, or a combination of both into our models to evaluate their impact on performance. We find that a larger training set containing both the rest-rest and eigen-eigen data leads to improved network performance.

Before studying the more relevant case of the vector correlator, we applied our model to predict the rest-eigen contribution of the pseudoscalar correlation function which does not suffer from the noise problem. In Fig. 2 we show the rest-eigen contribution in the central region. The left panel compares the prediction from the model with the exact evaluation of the correlator on an identical set of configurations. For this comparison, 200 configuration were used in the training step. The right panel shows the quality of the prediction as a function of the size of the training set, N_{train}. The quality is expressed in terms of the absolute value of the quantity

$$A(t) = \left\langle C^{\text{exact}}(t) \right\rangle_{\text{test}} - \left\langle C^{\text{pred}}(t) \right\rangle_{\text{test}} , \tag{5}$$

which is a measure of the deviation between model prediction and exact calculation at timeslice t, both evaluated on the entire test set of configurations. While there is a tendency that larger training

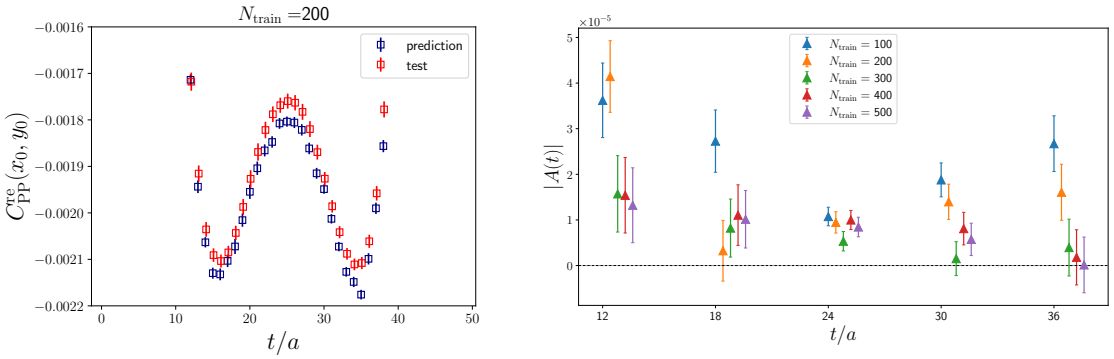

Figure 2: *Left:* The rest-eigen part of the pseudoscalar correlator on ensemble A654 as predicted by the model compared with the exact calculation for $N_{\text{train}} = 200$. *Right:* Deviation between prediction and exact calculation for different sizes of the training set.

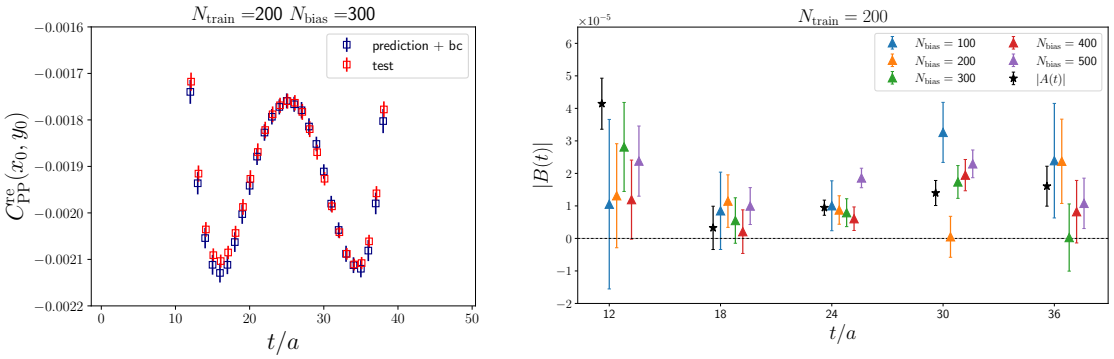

Figure 3: *Left:* The bias-corrected rest-eigen part of the pseudoscalar correlator on ensemble A654 with the exact calculation for $N_{\text{train}} = 200$, $N_{\text{bias}} = 300$. *Right:* Bias correction $B(t)$ for different values of N_{bias}. The deviation $A(t)$ between prediction and exact evaluation is shown as the black star.

sets yield a better prediction, the values of $A(t)$ are quite stable when $N_{\text{train}} \gtrsim 200$. The left panel of Fig. 3 shows that the bias-corrected prediction agrees very well with the exact calculation. In the right panel the absolute value of the bias correction, defined by

$$B(t) = \left\langle C^{\text{exact}}(t) - C^{\text{pred}}(t) \right\rangle_{\text{bias}} , \tag{6}$$

is shown for different values of N_{bias} and compared against the mismatch $A(t)$ between prediction and exact calculation. The fact that the magnitudes of $A(t)$ and $B(t)$ agree well within errors further illustrates the ability of our machine-learning model to produce accurate results for the rest-eigen part of the pseudoscalar correlator.

Moving to the case of the (light-quark connected) vector-vector correlator, we plot in Fig. 4 the deviation between prediction and exact calculation of the rest-eigen part, with and without bias correction, in units of the statistical error of the exact result. While the quality of the prediction for the vector correlator is similar to that of the pseudoscalar one (left panel), one finds that the statistical error of the bias-corrected result is about twice as large as for the exact calculation at large Euclidean times (right panel). This much more unfavourable behaviour compared to the pseudoscalar case is also illustrated by Fig. 5 where we compare the fraction of the variance contributed by the eigen-eigen, rest-eigen and rest-rest to the total uncertainty of the correlator. The right panel of the figure

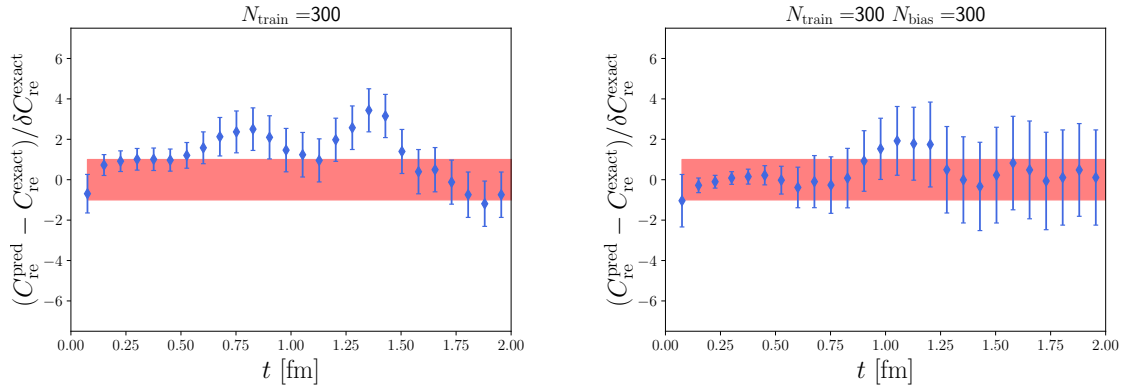

Figure 4: *Left:* Deviation between the predicted and exact rest-eigen part of the vector-vector correlator $G(t)$ in units of the statistical error of the exact result, plotted versus Euclidean time. The red horizontal band denotes the one-sigma error. *Right:* Bias-corrected prediction for the vector-vector correlator.

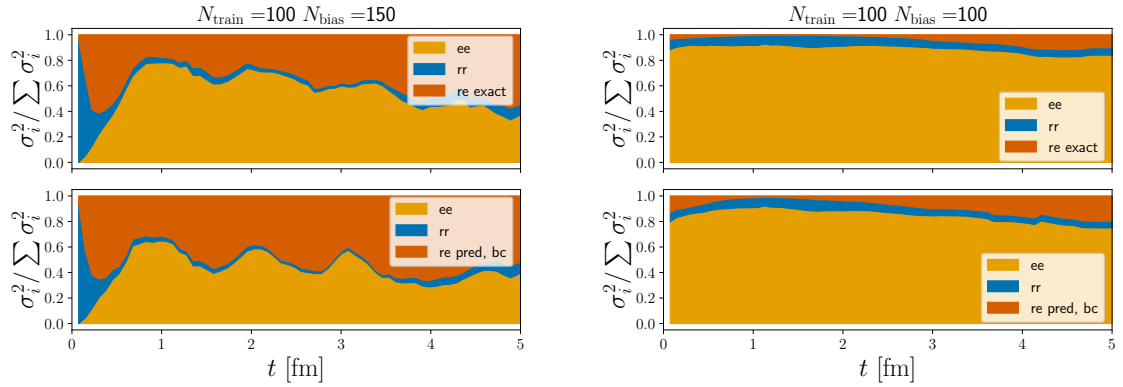

Figure 5: Fractional contributions of the eigen-eigen (ee), rest-eigen (re) and rest-rest (rr) parts to the total error of the correlator as a function of Euclidean time for ensemble D450. The upper plot in each panel shows the exact calculation with the machine-learning model shown in the lower plots. The vector correlator is shown in the two panels on the left, while the pseudoscalar case is shown on the right.

shows that the eigen-eigen contribution dominates the error in the long-distance regime of the pseudoscalar correlator, whereas the largest contribution to the uncertainty of the vector correlator comes from the rest-eigen part. A reduction of the total error in the bias-corrected prediction of the vector correlator is possible but at the expense of increasing the number of configurations used in the bias correction to a level where the numerical gain is all but gone.

4. Isospin-breaking corrections to baryon masses

Lattice calculations of the HVP contribution to the muon $g-2$ with sub-percent precision require that the uncertainty in the lattice scale be about two times smaller than the target precision, in order for the scale setting error to be sub-dominant [11]. The masses of the lowest-lying octet and decuplet baryons have emerged as reliable quantities for scale setting, since these states are stable in QCD and can be computed with small statistical errors. For the long-term goal of determining the Standard Model prediction to the muon $g-2$ with a precision that can rival the experimental measurement, it is necessary to include electromagnetic corrections as well as corrections due to the mass splitting

between up and down quarks. In Ref. [12] we have presented first results from an ongoing project to compute octet and decuplet baryons including strong and electromagnetic isospin-breaking effects. To this end we employ the approach pioneered by the RM123 collaboration [13, 14] which relies on an expansion about iso-symmetric QCD. Full details about the implementation of this formalism in the context of baryonic two-point correlation functions can be found in [12]. The numerical evaluation of baryon masses in QCD+QED via this approach is numerically very demanding, with 50% of the computer time spent on computing the relevant correlators including photon lines. However, as can be seen from Fig. 6, one finds that electromagnetic isospin-breaking corrections are strongly correlated with the quark mass detuning (i.e. the difference between the quark masses defining QCD+QED and iso-symmetric QCD, respectively). This suggests a machine-learning strategy that is based on training a model on the correlation between strong and electromagnetic isospin-breaking corrections to predict the latter and apply a bias correction to produce an exact result. We consider a simple linear model, i.e.

$$M(t) = \alpha(t)\, C^{(0)}(t) + \beta(t)\, C^{(1)}_{\Delta m_u}(t) + \gamma(t)\, C^{(1)}_{\Delta m_d}(t) + \delta(t)\, C^{(1)}_{\Delta m_s}(t) + \epsilon(t)\,, \tag{7}$$

where $C^{(0)}(t)$ denotes the baryon correlation function in iso-symmetric QCD, while $C^{(1)}_{\Delta m_q}(t)$, $q = u,\, d,\, s$ denote the contributions arising from the quark mass detuning. The parameters $\alpha,\, \beta, \ldots,\, \epsilon$ are determined in the training step such that the model predicts the QED contribution $C^{(1)}_{e^2}(t)$ given $C^{(1)}_{\Delta m_q}(t)$ as input. To this end we divided ensemble N451 into a training set of $N_{\text{train}} = 20$ configurations and a test set of size $N_{\text{test}} = 991$. For each configuration from the training set we performed the exact calculation of the QED contribution for all of the 32 quark sources used in our setup. For each timeslice t we then applied the linear regression of the QED contribution according to eq. (7). The saved parameters $\alpha(t), \ldots, \epsilon(t)$ from the training step were then used to evaluate the QED contribution via eq. (7) for each configuration of the test set and for all of the 32 quark sources. For the bias correction, we evaluated the exact QED contribution on all configurations of the test set, but only for a single quark source, thereby reducing the computational effort for evaluating electromagnetic corrections by a factor 32. Given that QED corrections account for about 50% of the total time, this produces a numerical gain by a factor 2. We found that the combined statistical errors of the prediction and the bias correction are competitive with the exact calculation for all 32 quark sources. This is seen clearly in Fig. 7 which shows the electromagnetic corrections to the Ω^- and Ξ^- effective masses. For both states, the bias-corrected predictions agree with the exact calculation within errors which are also of comparable size. We have studied the effect of increasing the number of sources in the bias correction and found that the errors stayed approximately constant.

5. Conclusions

Machine-learning techniques offer a viable alternative to conventional methods for calculating correlation functions in lattice QCD, although their efficiency and practicality depend on the specific application. In our study of isospin-breaking corrections to baryon masses, we observed a strong correlation between the electromagnetic corrections and quark mass detuning. These correlations enabled us to construct a model that achieved substantial resource savings in determining electromagnetic corrections, reducing the computer time required to reach our targeted precision by approximately 50%. In contrast, the analysis of the rest-eigen contribution to the vector correlator

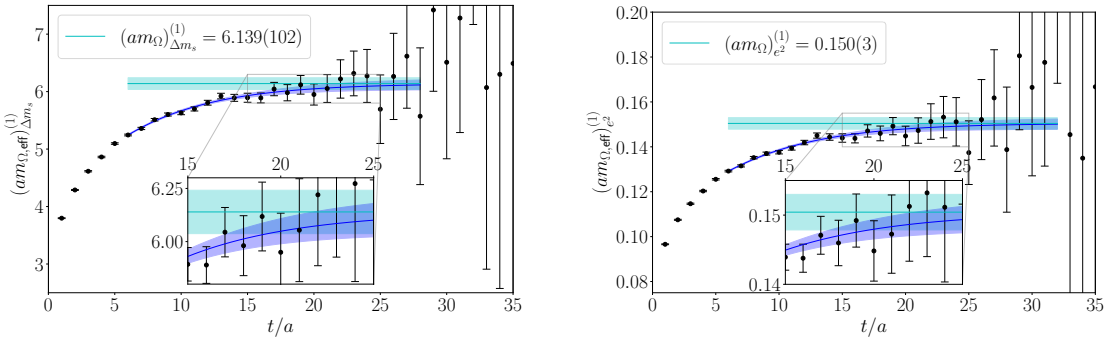

Figure 6: The contributions from the strange quark mass detuning (left panel) and electromagnetic corrections (right panel) to the effective mass of the Ω^- baryon, in lattice units.

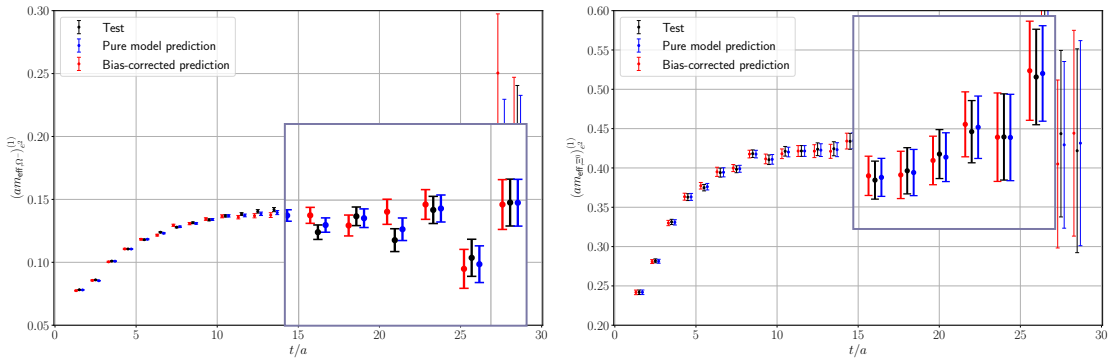

Figure 7: Electromagnetic corrections to the effective mass of the Ω^- (left panel) and Ξ^- (right panel) baryons plotted in lattice units. The insets show the enlarged region for $t/a = 16-21$.

$G(t)$ in the long-distance regime is less straightforward. Although our machine-learned correlator, combined with a bias correction, can reproduce $G(t)$, the uncertainty remains larger than that obtained from an exact evaluation, with the bias correction step being the dominant source of error and computational expense. This outcome suggests that our model has not yet fully exploited the correlations among the various contributions to $G(t)$, highlighting an important area for future improvement. We will continue the search for alternative models designed to increase the correlations between the approximate and exact evaluation of the rest-eigen contribution.

Acknowledgements

Calculations for this project have been performed on HPC platforms at Helmholtz Institute Mainz, Johannes Gutenberg-Universität Mainz, Jülich Supercomputing Centre (JSC) and Höchstleistungsrechenzentrum Stuttgart (HLRS). We gratefully acknowledge the support of the Gauss Centre for Supercomputing (GCS) and the John von Neumann-Institut für Computing (NIC) for projects HMZ21 and HINTSPEC at JSC and project GCS-HQCD at HLRS. This work has been supported by the German Research Foundation through the Cluster of Excellence "Precision Physics, Fundamental Interactions and Structure of Matter" (PRISMA+ EXC 2118/1), funded within the German Excellence strategy (Project ID 39083149). SK is supported by the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101106243. TB is partially supported by the US DOE under grant DE-SC0010339.

References

[1] D. Bernecker and H.B. Meyer, *Vector Correlators in Lattice QCD: Methods and applications*, *Eur. Phys. J. A* **47** (2011) 148 [1107.4388].

[2] G. Parisi, *Strategy for Computing the Hadronic Mass Spectrum*, *Phys. Rept.* **103** (1984) 203.

[3] G.P. Lepage, *The Analysis of Algorithms for Lattice Field Theory*, in *Boulder ASI 1989:97-120*, pp. 97–120, 1989.

[4] L. Giusti, P. Hernandez, M. Laine, P. Weisz and H. Wittig, *Low-energy couplings of QCD from current correlators near the chiral limit*, *JHEP* **04** (2004) 013 [hep-lat/0402002].

[5] T.A. DeGrand and S. Schaefer, *Improving meson two point functions in lattice QCD*, *Comput. Phys. Commun.* **159** (2004) 185 [hep-lat/0401011].

[6] D. Djukanovic, G. von Hippel, S. Kuberski, H.B. Meyer, N. Miller, K. Ottnad et al., *The hadronic vacuum polarization contribution to the muon g − 2 at long distances*, 2411.07969.

[7] T. Blum, T. Izubuchi and E. Shintani, *New class of variance-reduction techniques using lattice symmetries*, *Phys. Rev. D* **88** (2013) 094503 [1208.4349].

[8] B. Yoon, T. Bhattacharya and R. Gupta, *Machine Learning Estimators for Lattice QCD Observables*, *Phys. Rev. D* **100** (2019) 014504 [1807.05971].

[9] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Improving neural networks by preventing co-adaptation of feature detectors*, 1207.0580.

[10] Y.A. LeCun, L. Bottou, G.B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural Networks: Tricks of the Trade: Second Edition*, G. Montavon, G.B. Orr and K.-R. Müller, eds., (Berlin, Heidelberg), pp. 9–48, Springer Berlin Heidelberg (2012), DOI.

[11] M. Della Morte, A. Francis, V. Gülpers, G. Herdoíza, G. von Hippel, H. Horch et al., *The hadronic vacuum polarization contribution to the muon g − 2 from lattice QCD*, *JHEP* **10** (2017) 020 [1705.01775].

[12] A.M. Segner, A. Risch and H. Wittig, *Precision Determination of Baryon Masses including Isospin-breaking*, *PoS* **LATTICE2023** (2023) 044 [2312.09065].

[13] G.M. de Divitiis et al., *Isospin breaking effects due to the up-down mass difference in Lattice QCD*, *JHEP* **04** (2012) 124 [1110.6294].

[14] G.M. de Divitiis, R. Frezzotti, V. Lubicz, G. Martinelli, R. Petronzio, G.C. Rossi et al., *Leading isospin breaking effects on the lattice*, *Phys. Rev. D* **87** (2013) 114505 [1303.4896].