Extrapolation in an implicit many-body expansion: Protonated water neural network potential applied to the extended Zundel cation

Christoph Schran,1,2,3,a) Fabien Brieuc,1 and Dominik Marx1, b)  
1) Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany  
2) Department of Physics and Astronomy, University College London, London, WC1E 6BT, UK  
3) Present address: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK  

(Dated: October 30, 2020)

A previously published neural network potential for the description of protonated water clusters up to the protonated water tetramer, H+(H2O)4, at essentially converged coupled cluster accuracy (J. Chem. Theory Comput. 16, 88 (2020)) is applied to the protonated water hexamer, H+(H2O)6, in its extended Zundel conformation – a system that the neural network has never seen before. Although being in the extrapolation regime, it is shown that the potential not only allows for stable quantum simulations from ultra-low temperatures ~ 1 K up to 100 K but that it is able to describe the new system very accurately compared to explicit coupled cluster calculations. Compared to the interpolation regime the quality of the model is reduced by roughly one order of magnitude, but most of the difference to the coupled cluster reference comes from global shifts of the potential energy surface, while local energy fluctuations are well recovered. These results suggest that the application of neural network potentials in extrapolation regimes can provide useful results and might be more general than usually thought.

I. INTRODUCTION

In recent years, machine learning has become a compelling tool for the representation of potential energy surfaces.1–6 The first family of such machine learning potentials based on artificial neural networks which scale to arbitrary system sizes were high-dimensional neural network potentials7,8 (NNPs). Since then, many distinctly different approaches either also based on artificial neural networks9–15 or on kernel methods16–21 have been introduced over the years, while recent development following the principles of deep-learning has allowed one to incorporate parts of the description of the chemical environments in the architecture of the model.22,23 However, there is usually overarching consensus that all these models can only be used in order to interpolate between a meaningful set of training points, while application outside the predefined configurational or chemical space of these structures is doomed to fail. This is commonly rationalized by the missing physical insight used in the construction of these models, since highly flexible general functions are used to represent the structure–energy relation.

In this communication we show that carefully constructed machine learning models can still be applied outside their originally considered scope of application. For that purpose we use our previously published NNP (exactly as reported in the Supporting Information in Ref. 24) for the description of protonated water clusters, which has been developed in an automated and adaptive workflow. It has been trained on essentially converged coupled cluster reference data for protonated water clusters up to H+(H2O)4 — including also the water monomer itself — providing what can be called an implicit many-body expansion. This notion implies that the very same model is able to describe the potential energy surface of all differently sized clusters H+(H2O)n (here from n = 1 up to 4) on equal footing by virtue of including all of them explicitly and simultaneously in the training. In contrast, explicit many-body expansions, which have also been successfully applied to protonated water clusters,25 expand the potential energy as a sum of many-body corrections to the previous terms.

Here, we apply our model to quantum simulations of the protonated water hexamer, H+(H2O)6, in its extended Zundel conformation (i.e. the hydrated protonated water dimer), a system that was not included in training the NNP. The structures generated by these stable quantum simulations are afterwards validated with respect to single-point coupled cluster calculations of the same quality as those used to generate the NNP up to n = 4 only. We show that due to the implicit nature of the many-body expansion used for the construction of this particular NNP, the model is able to provide meaningful and accurate predictions in this extrapolation regime.

II. RESULTS AND DISCUSSION

In order to test the application of the previously published NNP for the description of protonated water clusters beyond what was considered in the original develop-
ment, we ran path integral simulations at various temperatures starting from the optimized minimum energy structure of the extended Zundel cation. The computational details of these simulation are described in the appendix. All tested quantum simulations from close to the ground state at 1.67 K up to 100 K were stable — although all being evidently in the extrapolation regime of the NNP. The model was indeed exclusively extrapolating since for each and every configuration encountered during the simulations the associated values of the atom-centered symmetry functions, used here as descriptors of the atomic environments, were outside the range of values present during training. This clearly indicates that exclusively unknown configurations were encountered in all steps of these simulations. All simulations up to 100 K exclusively sampled the extended Zundel isomer and no rearrangements to other isomers were observed at these ground-state dominated temperatures. This stable application of the NNP in a deep extrapolation regime is a first promising indication that the potential could be applied beyond the originally considered cluster sizes.

In Fig. 1 we show the probability distribution functions of three main structural properties of the hydrogen bond from the simulation of (H$_5$O$_2^+$)(H$_2$O)$_4$ at 1.67 K and compare them to the respective distributions of the much smaller bare Zundel cation, H$_2$O$_2^+$. As seen therein, the model provides the expected bimodal distributions of the donor-acceptor distance r$_{OO}$ and the proton-sharing coordinate δ for the extended Zundel cation, caused by the two distinctly different types of hydrogen bonds in the system: The central ultra-strong hydrogen bond (shown in blue) and the four additional hydrogen bonds to the dangling water molecules (shown in gray). In comparison, the bare Zundel complex has qualitatively similar distributions in case of the central hydrogen bond, but features a distinct shift towards shorter donor-acceptor distances as well as a slightly more localized proton-sharing coordinate. Finally, the distribution of the HOO angle for the central, ultra-strong hydrogen bond in the extended Zundel cation is again close to the one in the bare Zundel complex, while the four weaker hydrogen bonds to the dangling water molecules feature the expected broader distribution that is shifted to larger angles. These results are in substantial agreement with previous studies on the extended Zundel cation regarding the symmetric nature of the central ultra-strong hydrogen bond and its close match with that in the smaller bare Zundel cation. This highlights the physically meaningful nature of the quantum structures generated with our model operating here in extrapolation mode, even close to the quantum ground state at 1.67 K.

In a next step, the quality of the NNP prediction during the simulations is validated by explicitly evaluating the coupled cluster reference method (namely CCSD(T)-F12a/aug-cc-pVTZ, see the appendix for details) for 300 randomly selected configurations at both, 1.67 and 100 K. We note that due to the system size such coupled cluster calculations are increasingly demanding in view of their steep scaling not only in terms of computation time, but also when it comes to memory resources. This growth is such that it would have been challenging to explicitly include the protonated water hexamer when training the NNP in the first place. The resulting correlation of the NNP prediction and the actual coupled cluster reference for (H$_5$O$_2^+$)(H$_2$O)$_4$ is shown in the left panels of Fig. 2. In the same figure we include in the right panels the equivalent test for the largest cluster considered during the development of that NNP, namely the protonated water tetramer, H$_5$O$_5^+$, in its Eigen structure (corresponding to the hydrated hydronium cation / protonated water monomer H$_3$O$^+$). Note that the same energy scales are used for all axes to allow for one-to-one comparisons. Overall, a high correlation between refer-
Extrapolation: $\text{H}^+(\text{H}_2\text{O})_6$

Interpolation: $\text{H}^+(\text{H}_2\text{O})_4$

Figure 2. Correlation of the energy per atom from explicit CCSD(T)-F12a/aug-cc-pVTZ calculations (CC) and the NNP predictions for 300 randomly selected configurations of the protonated water hexamer in its extended Zundel conformation (left panels) and of the protonated water tetramer in its Eigen conformation (right panels) at 1.67 (red) and 100 K (black), respectively. The mean absolute difference (MAD) for both temperatures are reported in black and red, respectively. In the upper left panel we also include the shifted MAD values, see text, for which the systematic bias (as quantified in the upper left inset) is removed from the NNP prediction by shifting the energies by the MAD at the respective temperature (but without showing the underlying shifted data themselves). The lower panels show the energy differences between CC reference and NNP prediction over the whole range of reference energies, while the inset in the upper panels show the histograms of the energy differences including the corresponding standard deviations $\sigma$ in the respective color.

To further validate the application of the NNP for the extended Zundel cation when used in realistic simulations, we additionally re-evaluated a short segment of the trajectory associated with one replica, or bead, of the quantum path integral molecular dynamics simulations at 1.67 K with the coupled cluster reference method. The resulting potential energy profile along these 25 fs of the simulation is shown in the left panel of Fig. 3. We also carried out the same analysis for the largest cluster explicitly considered in the construction of the NNP, the protonated water tetramer, in the right panel to allow for one-to-one comparison to the interpolation regime. As before, the NNP yields surprisingly good results for the extended Zundel cation, although working exclusively in its extrapolation mode. The NNP is found to be able to recover the overall energy profile along the short segment of the trajectory and correctly reproduces the energy fluctuations (black line). As already seen for the randomly selected structures, the NNP is affected by an overall bias that shifts the prediction to slightly larger energies. However, if we use the mean absolute difference of the 300 random structures from the simulation at 1.67 K to shift the energy prediction according to the NNP, the coupled cluster reference energy profile is recovered perfectly by the NNP shifted on the physically relevant scale as set by the potential energy fluctuations (blue line with squares). As known from the originally published benchmarking of the NNP, the Eigen cation does not suffer from such a bias since it has been used explicitly to train that NNP as shown in the right panel of Fig. 3. Overall,
Figure 3. Potential energy along one replica of quantum PIMD trajectories at 1.67 K of the extended Zundel cation (left) and the protonated water tetramer in the Eigen conformation (right) using the original and shifted (see text) neural network potentials (NNP, NNP_shifted only in the left panel). The CCSD(T*)-F12a/aug-cc-pVTZ reference (CC) was obtained by recomputing the energies for each configuration along the NNP trajectories and are shown as red dotted lines (with only a few circles added since the CC energies mostly superimpose the NNP data). These energies are reported relative to the respective equilibrium structure. The bottom panels highlight the respective energy differences of the NNP predictions to the CC reference method.

This analysis reveals that the fully extrapolating NNP is able to almost perfectly recover the correct energy fluctuations of the extended Zundel cation even in the deep quantum regime at 1.67 K close to the ground state after correcting for the global energy shift. Even without any such shift of the energies, the NNP performs unexpectedly well in such an extrapolation regime, which suggests that there can be great potential in exploiting extrapolation capabilities for building more complex machine learning potentials.

III. CONCLUSION AND OUTLOOK

In summary, our previously published neural network potential, trained for simulations up to the protonated water tetramer, performs unexpectedly well for the extended Zundel cation, a conformer of the protonated water hexamer. This is notable since these simulations are carried out entirely in the extrapolation regime, which means that only unknown configurations are encountered in each and every simulation step. This NNP does not only allow one to run stable path integral quantum simulations, but also recovers correctly the quantum-thermal energy fluctuations in direct comparison to the coupled cluster reference data of the extended Zundel complex. Yet, the NNP suffers from a slight global shift of the potential energy, which however does not influence the relative fluctuations. Moreover, we show how this shift can be systematically corrected \textit{a posteriori} based on rather few additional reference calculations of the extended complex.

These promising results are explained by the implicit many-body expansion intrinsic to the development of this NNP model, which accurately treats a class of molecular complexes consisting of three up to 13 constituting atoms. Since the training set of the neural network potential is composed of clusters from the hydronium cation up to the protonated water tetramer and does include the water monomer as well, sufficient local environments are considered in its construction to yield these reassuring results.

Still, we would like to stress that these promising results should not be taken for granted and using machine learning models beyond their considered scope of application during their construction and parameterization requires great care and sufficient validation. Evidently, in most cases that are too far away from the chemical space spanned by the training set, machine learning models will provide unphysical results. We expect this in the present case at higher temperatures, where significant topological rearrangements are known to occur that generate distinctly different hydrogen bonded network topologies of protonated water clusters compared to those covered by the (extended) Zundel and Eigen motifs. At the same time, our results reveal that there can be cases were carefully constructed machine learning models are applicable much beyond interpolation regimes. This could be a promising route for the development of more complex machine learning models which implement the “implicit many-body expansion” idea by construction. Thus, we hope that our short report will stimulate methodological work to explore systematically and fundamentally the power of machine learning models to safely conquer unknown territory.
ACKNOWLEDGMENTS

We are thankful to Harald Forbert for insightful discussions. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2033 – 390677874 – RESOLV. C.S. acknowledges partial financial support from the Alexander von Humboldt-Stiftung. The computational resources were provided by HPC@ZEMOS, HPC-RESOLV, and BoViLab@RUB.

DATA AVAILABILITY

The data that supports the findings of this study are available within the article.

APPENDIX: COMPUTATIONAL DETAILS

Path integral molecular dynamics simulations of the extended Zundel cation down to 1.67 K have been performed with the CP2k program package. The potential energy surface was described using a recently developed and published NNP fitted to coupled cluster reference calculations. This NNP describes all protonated water clusters, from the protonated water monomer (hydrion cation) up to the protonated water tetramer considered in the development, on equal footing and, in particular, also explicitly includes the water monomer. It has been shown to not only match the reference coupled cluster theory with very high precision, but is also able to accurately describe proton transfer in the considered clusters. We apply this NNP in the present study to the larger protonated water hexamer in its extended Zundel structures, which is thus entirely in the extrapolation regime of the model.

The extended Zundel cation was simulated at temperatures of 100, 20, 10, and 1.67 K including the quantum nature of the nuclei using the PIQTB thermostat which is used here exclusively in its extrapolation regime of the model.

Explicit validations of the predictive power of the NNP which is used here exclusively in its extrapolation regime was achieved by reevaluating the energies of many configurations of the extended Zundel complex with the same coupled cluster method as used for the development of that NNP. These calculations of the CCSD(T) reference energies were performed with the Molpro program package by employing the explicitly correlated F12a method to correct for the basis set incompleteness error. As suggested we additionally employed the size-consistent scaling of the perturbative triples, together with the aug-cc-pVTZ basis set. This so-called CCSD(T**)-F12a/aug-cc-pVTZ electronic structure setup has been shown to provide energies very close to the complete basis set (CBS) limit.

REFERENCES

1. J. Behler, J. Chem. Phys. 145, 170901 (2016).
2. A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, C. Csányi, and M. Ceriotti, Sci. Adv. 3, e1701816 (2017).
3. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Nature 559, 547 (2018).
4. V. L. Deringer, M. A. Caro, and G. Csányi, Adv. Mater. 31, 1902765 (2019).
5. T. Mueller, A. Hernandez, and C. Wang, J. Chem. Phys. 152, 050902 (2020).
6. S. Manzhos and T. Carrington, Chem. Rev. 120, 101021/acs.chemrev.b006655 (2020).
7. J. Behler and M. Parrinello, Phys. Rev. Lett. 98, 144601 (2007).
8. J. Behler, Angew. Chemie - Int. Ed. 56, 12828 (2017).
9. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker, Phys. Rev. B 92, 045131 (2015).
10. A. Khoshidi and A. A. Peterson, Comput. Phys. Commun. 207, 310 (2016).
11. N. Artrith, A. Urban, and G. Ceder, Phys. Rev. B 96, 014112 (2017).
12. J. S. Smith, O. Isayev, and A. E. Roitberg, Chem. Sci. 8, 3192 (2017).
13. M. Gastegger, J. Behler, and P. Marquetand, Chem. Sci. 8, 6924 (2017).
14. O. T. Unke and M. Meuwly, J. Chem. Theory Comput. 15, 3678 (2019).
15. Y. Shao, M. Hellström, P. D. Mitev, L. Knijff, and C. Zhang, J. Chem. Inf. Model. 60, 1184 (2020).
16. A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. 104, 136403 (2010).
17. M. Rupp, A. Tkatchenko, K. R. Müller, and O. A. Von Lilienfeld, Phys. Rev. Lett. 108, 058301 (2012).
18. A. Thompson, L. Swiler, C. Trott, S. Foiles, and G. Tucker, J. Comput. Phys. 285, 316 (2015).
19. A. V. Shapeev, Multiscale Model. Simul. 14, 1153 (2015).
20. Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. 114, 096405 (2015).
21. S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K. R. Müller, Sci. Adv. 3, e1603015 (2017).
22. K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Nat. Commun. 8, 13890 (2017).
23. L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, Phys. Rev. Lett. 120, 143001 (2018).
24. C. Schran, J. Behler, and D. Marx, J. Chem. Theory Comput. 16, 88 (2020).
25. P. Pinski and G. Csányi, J. Chem. Theory Comput. 10, 68 (2014).
26. A. Luzar and D. Chandler, Phys. Rev. Lett. 76, 928 (1996).
27. J. Behler, J. Chem. Phys. 134, 074106 (2011).
28. C. Schran and D. Marx, Phys. Chem. Chem. Phys. 21, 24967 (2019).
29. J. C. Jiang, Y. S. Wang, H. C. Chang, S. H. Lin, Y. T. Lee, G. Niedner-Schatteburg, and H. C. Chang, J. Am. Chem. Soc. 122, 1398 (2000).
30R. A. Christie and K. D. Jordan, J. Phys. Chem. B 106, 8376 (2002).
31Q. C. Nguyen, Yew-Soon Ong, and J. L. Kuo, J. Chem. Theory Comput. 5, 2629 (2009).
32N. Heine, M. R. Fagiani, M. Rossi, T. Wende, G. Berden, V. Blum, and K. R. Asmis, J. Am. Chem. Soc. 135, 8266 (2013).
33F. Mouhat, Fully quantum dynamics of protonated water clusters, Ph.D. thesis, Sorbonne Université (2018).
34J. P. Heindel, Q. Yu, J. M. Bowman, and S. S. Xantheas, J. Chem. Theory Comput. 14, 4553 (2018).
35“CP2K, freely available at the URL https://www.cp2k.org, released under GPL license,” (2020).
36J. Hutter, M. Iannuzzi, F. Schiffmann, and J. Vandevondele, Wiley Interdiscip. Rev. Comput. Mol. Sci. 4, 15 (2014).
37F. Brieuc, H. Dammak, and M. Hayoun, J. Chem. Theory Comput. 12, 1351 (2016).
38C. Schran, F. Brieuc, and D. Marx, J. Chem. Theory Comput. 14, 5068 (2018).
39H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O’Neill, P. Palmieri, D. Peng, K. Pfäger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, “MOLPRO, version 2012.1, a package of ab initio programs,” (2015).
40T. B. Adler, G. Knizia, and H. J. Werner, J. Chem. Phys. 127, 221106 (2007).
41G. Knizia, T. B. Adler, and H. J. Werner, J. Chem. Phys. 130, 054104 (2009).
42R. A. Kendall, T. H. D. Jr, and R. J. Harrison, J. Chem. Phys. 96, 6796 (1992).
43D. E. Woon and T. H. Dunning Jr., J. Chem. Phys. 100, 2975 (1994).