THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

Cite This: J. Phys. Chem. A 2018, 122, 6343−6348

pubs.acs.org/JPCA

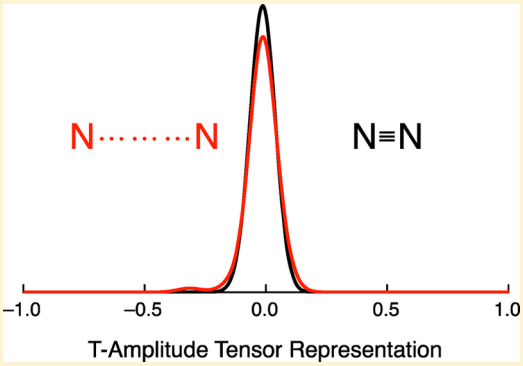
Making the Coupled Cluster Correlation Energy Machine-Learnable

Johannes T. Margraf* and Karsten Reuter

Chair of Theoretical Chemistry, Technische Universität München, Lichtenbergstrasse 4, D-85747 Garching, Germany

Supporting Information

ABSTRACT: Calculating the electronic structure of molecules and solids has become an important pillar of modern research in diverse fields of research from biology and materials science to chemistry and physics. Unfortunately, increasingly accurate and thus reliable approximate solution schemes to the underlying Schrödinger equation scale steeply in computational cost, rendering most accurate approaches like "gold standard" coupled cluster theory, CC, quickly intractable for larger systems of interest. Here we show that this scaling can be significantly reduced by applying machine-learning to the CC correlation energy. We introduce a vector-based representation of CC wave functions and use potential energy surfaces of a small molecule test set to learn the correlation energy from this representation. Our results show that the CC correlation energy can be efficiently learned, even when the representation is constructed from approximate amplitudes provided by computationally less demanding Møller−Plesset (MP2) perturbation theory. Exploiting existing linear scaling MP2 implementations, this potentially opens the door to CC-quality molecular dynamics simulations.



INTRODUCTION

Machine-learning (ML) has recently received a great deal of attention in physics, chemistry, and materials science.[1,2] While ML and interpolation techniques have extensively and successfully been applied to fitting potential energy surfaces (PES) for many years,[3−8] a hallmark of the recent revival of interest in ML is the development of powerful methods for encoding molecular structure in a "learnable" representation. Examples include the Coulomb matrix (CM),[9] atom-centered symmetry functions,[10,11] the smooth overlap of atomic positions (SOAP),[12] and the many-body tensor representation (MBTR).[13] Tkatchenko, Müller, et al. have even used deep tensor neural networks to learn the representation itself.[14]

The principle objective of most of these efforts is to learn the mapping from structural information to energies and forces. While the reference data for such studies usually come from density functional theory (DFT) calculations, the ML methodology is in principle agnostic to the underlying electronic structure method.[15] In contrast to this, Müller, Burke, and co-workers have applied ML to different aspects of DFT on a more fundamental level, e.g., to learn the kinetic energy functional or the Hohenberg−Kohn mapping of external potential to density.[16−18] Their work indicates that it can be more robust to learn such fundamental relationships, rather than "directly" learning the total energy. Similar observations have been made in the context of semiempirical MO theory and Møller−Plesset perturbation theory.[19,20]

In a similar spirit, we aim to apply ML to coupled cluster (CC) theory. CC is the most accurate generally applicable electronic structure method.[21] Unfortunately, it is too computationally expensive for most applications (e.g., to drive molecular dynamics simulations), although developments in localized CC methods have strongly accelerated single-point calculations.[22,23] As a first step toward a ML-CC approach, we show that an economical vector representation of the CC wave function can be used to learn CC correlation energies with high accuracy. Importantly, approximate wave functions from perturbation theory can also be used to construct this representation, with virtually no loss in accuracy.

THEORY

CC theory starts from an exponential wave function ansatz:

$$\Psi_{CC} = e^{T}\phi_0 \tag{1}$$

with a reference determinant ϕ_0. The cluster operator T is defined as

$$T = T_1 + T_2 + \ldots \tag{2}$$

$$T_m\phi_0 = \sum_p t_p\phi_p \tag{3}$$

where the sum on the right-hand side of eq 3 goes over all m-fold excitations and p is a composite index enumerating them. The coefficients t_p are known as T-amplitudes, and they need to be determined from a complex set of nonlinear equations. In practice, eq 2 is truncated. So, to obtain for instance the CC singles and doubles method (CCSD), we use $T = T_1 + T_2$.

Received: May 10, 2018
Revised: July 5, 2018
Published: July 9, 2018

DOI: 10.1021/acs.jpca.8b04455
J. Phys. Chem. A 2018, 122, 6343−6348

From this, one can derive the CC correlation energy expression (for a canonical HF reference):[24]

$$E_C = \frac{1}{4}\sum_{aijb} \langle ij||ab\rangle t_{ij}^{ab} + \frac{1}{2}\sum_{aijb} \langle ij||ab\rangle t_i^a t_j^b \quad (4)$$

Here, the composite index p has been exchanged by the explicit indices of occupied (i, j) and virtual spin−orbitals (a, b), which define single and double excitations with respect to a HF reference determinant. The correlation energy therefore depends on the amplitudes of the T_1 and T_2 operators (t_i^a and t_{ij}^{ab}), and the antisymmetrized two-electron integrals *ijab*. In practice, we are of course interested in the total energy, which is simply a sum of E_C and the total energy of a single determinant reference wave function E_{SCF} (e.g., from a Hartree−Fock calculation). Determining E_C (more specifically, converging the T-amplitudes) is, however, by far the computationally limiting step, and therefore the focus of this work. Below we will use total energies when showing potential energy surfaces, but all error plots and discussions refer to correlation energies.

In a nutshell, the goal of this work is to find a machine-learnable form of eq 4. We understand ML as an interpolation technique, which, given a set of known inputs and outputs ($\boldsymbol{x}$, $\boldsymbol{y}$) (the training data), defines a model that predicts the unknown output $\boldsymbol{y}^*$ for any input $\boldsymbol{x}^*$. The output in our case is the CC correlation energy, but what is the input (or, to keep with ML jargon, the *representation*)?

The representation encodes the relevant information about the system and should respect the same invariances (e.g., rotational and translational) and variances as the original problem. As our objective is to stay as close as possible to the physics of CC theory (*vide supra*), we choose not to infer the energy from structural descriptors, such as the Coulomb matrix (CM). Following eq 4, we instead propose a representation that uses the magnitude of the T-amplitudes. In principle, the most straightforward such representation would be a list of the amplitudes. However, constant-size vector representations are more practical, in particular for the vector-based kernel regression method used in this work.

In analogy to Rupp's many-body tensor representation (MBTR),[13] we therefore define a T-amplitude tensor representation (TATR), which meets these requirements. We start with a smooth function:

$$f_m(x) = \sum_{i\le N_{amp}} g(t_i,\sigma_T,x) \quad (5)$$

where the sum goes over the N_{amp} largest amplitudes t_i of the excitation level m (e.g., singles or doubles) and $g(t_i,\sigma_T,x)$ is a Gaussian function centered at t_i with variance σ_T^2. To obtain a constant-size vector-representation $\boldsymbol{v}$ (the TATR), the functions $f_m(x)$ are discretized in the range $[-1{:}1]$, with a step-size Δx. The CC wave function of any arbitrarily large system can then be encoded into a vector $\boldsymbol{v}$ with $2/\Delta x$ elements per excitation level. A convenient side effect of this is that the TATR allows visualizing coupled cluster wave functions (see Figure 1). In the figure we use $\sigma_T = 0.05$, which was found to be a good value for diatomics (see the Supporting Information). In general, it can be worthwhile to optimize σ_T for specific cases. For example, the potential energy surface of water (*vide infra*) is more accurately interpolated with $\sigma_T = 0.005$. As a rule of thumb, a larger value of σ_T is appropriate if the model is supposed to interpolate between very different electronic situations (e.g., equilibrium and strongly stretched bonds), whereas a smaller value should be used for relatively similar wave functions (e.g., a potential energy surface around equilibrium), for a given size of the training set. This is because large σ_T values smear out the individual features of each training point, thus favoring a smoother interpolation between relatively dissimilar points, whereas small values allow for a more faithful distinction between similar points. If large range of electronic structures is to be covered with very high accuracy around the equilibrium, a small value of σ_T needs to be combined with a dense enough training set (see below for a numerical demonstration).

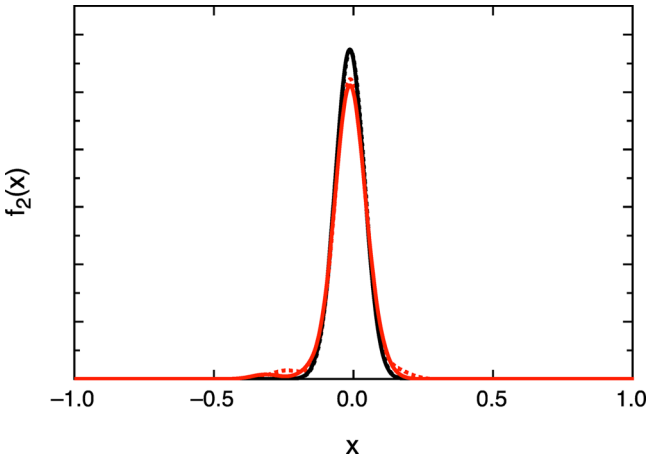


Figure 1. CCSD T2-amplitude representation for the nitrogen molecule at r = 1.1 Å (black) and r = 1.9 Å (red). Static correlation effects upon bond dissociation lead to larger T-amplitudes. The corresponding MP2 amplitudes are also shown in dotted lines.

Given this vector representation, we now approximate the CC correlation energy as a regression problem:

$$E_C^{ML}(\boldsymbol{v}) \approx \sum_i^M \alpha_i k(\boldsymbol{v_i},\boldsymbol{v}) \quad (6)$$

where the coefficients α_i need to be determined and the kernel function $k(\boldsymbol{v_i},\boldsymbol{v})$ quantifies the similarity between the TATRs of the wave function of interest and the wave functions known from the training set (of size M). We use the popular Gaussian kernel,

$$k(\boldsymbol{v}_1,\boldsymbol{v}_2) = \exp\left(-\frac{\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|^2}{2\sigma^2}\right) \quad (7)$$

The coefficients α_i are determined via kernel ridge regression (KRR),[25] as described in the Supporting Information.

RESULTS AND DISCUSSION

It is instructive to first consider the simplest two-electron molecule, dihydrogen. CCSD wave functions and total energies were calculated at 200 evenly spaced interatomic distances between 0.5 and 2.0 Å (see Figure 2 and the Supporting Information for a corresponding plot of correlation energies). The resulting potential energy surface shows the well-known features of this prototypical system. As expressed by the significant difference between the CCSD curve and a simple Hartree−Fock (HF) calculation, there is a fair amount of dynamical correlation around the equilibrium bond-length and a larger amount of static correlation in the dissociation limit.

DOI: 10.1021/acs.jpca.8b04455
J. Phys. Chem. A 2018, 122, 6343−6348

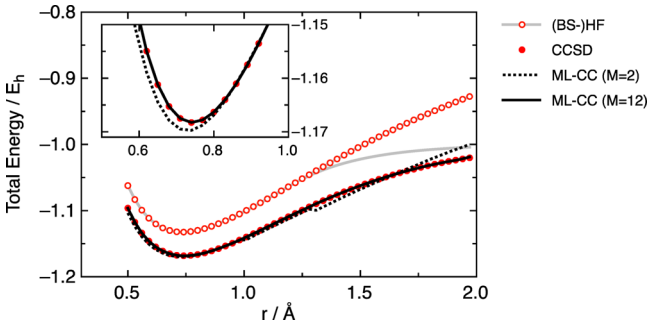


Figure 2. Dissociation curves of molecular hydrogen calculated using Hartree–Fock (HF), CCSD, broken-symmetry (BS)-HF and ML models (with different training sets M = 2 and 12, σ_T = 0.05). Note the deviation of the BS-HF curve (gray line) from the restricted HF curve (open circles) beyond the Coulson–Fischer point. A detailed view of the equilibrium region of the curve is shown in the inset.

Additionally, the restricted HF solution becomes unstable at around 1.3 Å, in favor of a broken-symmetry, unrestricted solution (the so-called Coulson–Fischer point).[26]

In the same figure, the dissociation curves predicted by ML-CC models trained on 2 and 12 data points are shown (for more information on training, see the Supporting Information). The M = 12 predictions virtually overlap the CCSD curve, with a mean absolute error of 6×10^{-5} E_h. Even the undertrained (M = 2) model performs surprisingly well around the equilibrium bond distance. Note also that the errors of this model display a discontinuity at the Coulson–Fischer point, as the character of the wave function changes and static correlation begins to kick in. This shows that the TATR correctly represents the physics of the problem.

These results are encouraging, as the accuracy of the predicted energies is high, even with fairly sparse training sets. Furthermore, the TATR is obviously a faithful representation of the CC wave function, as the accuracy of the correlation energy can be reproduced with sub-μHa accuracy, if the training set is large enough (see the Supporting Information). This is not a trivial point, as it is an approximation to assume that the energy can be inferred from the magnitude of the largest T-amplitudes alone.

Turning to more complex systems, ML-CC models for the dinitrogen, carbon monoxide, lithium fluoride and hydrogen fluoride dissociation curves were generated, again using 200 reference data points between 0.5 and 2.0 Å. In all cases, the TATR can be applied to accurately reproduce the correlation energy across the full range of interatomic distances (see Figure 3). This shows that CC wave functions of many-electron systems can also be faithfully encoded in this representation. Beyond the error statistics, it is also important to note that the resulting curves are in almost all cases smooth fits (see the Supporting Information). Discontinuities in the predicted curves indicate regions of insufficient training (as for the M = 2 H_2 model in Figure 2). When fitting potential energy surfaces, this information could potentially be fed back into the program in a more sophisticated "self-correcting" ML scheme.

It should be stressed that this is achieved solely using the T-amplitudes without any structural descriptors. It is obviously quite easy to interpolate diatomic potential energy surfaces with any standard regression technique using the bond-length as a descriptor. However, this quickly becomes intractable for higher dimensional systems, whereas the dimension of the TATR representation is *independent* of the number of atoms in the system. Additionally, while we use potential energy surfaces as a target in this paper, encoding the wave function in principle allows describing all electronic properties of a system, even for cases where structural descriptors cannot be applied.

While the ML-CC models can clearly be used to predict CCSD correlation energies with high accuracy, this is arguably not a very useful exercise. The computational bottleneck of CC is the determination of the amplitudes and not the evaluation of the energy. To overcome this, we can define a slightly different, perturbation theory based model (abbreviated as ML-PT in the following). Here, we still use CCSD *energies* for training, but the TATRs are constructed from MP2 *amplitudes*, which are typically used as the initial guess for CC calculations. As shown in Figure 1, the corresponding TATRs are in general different, but qualitatively similar to the ML-CC ones.

In other words, we infer the CCSD energy from MP2 amplitudes, effectively bypassing the CC amplitude equations. ML-PT models are similarly accurate as the models using CC amplitudes (shown in Figure 3, right). As can be seen from the example of hydrogen dissociation (see the Supporting Information), ML-PT is accurate even at large interatomic distances, where the MP2 *energy* is qualitatively wrong.

Going beyond diatomics, we consider the water molecule as a simple triatomic system. Here, 400 equidistant reference data

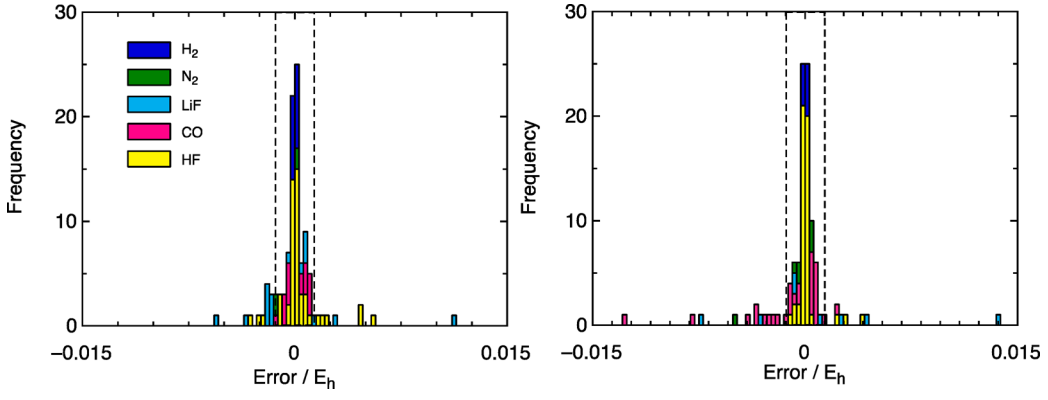


Figure 3. Histograms of error distributions for ML-CC (left) and ML-PT (right) dissociation curves of diatomic molecules. Shown are the errors for 50 predicted data points for each molecule. The ML models were trained on 12 data points (σ_T = 0.05). The dashed lines delimit ±0.0015 E_h (~1 kcal mol^{-1}).

DOI: 10.1021/acs.jpca.8b04455
J. Phys. Chem. A 2018, 122, 6343–6348

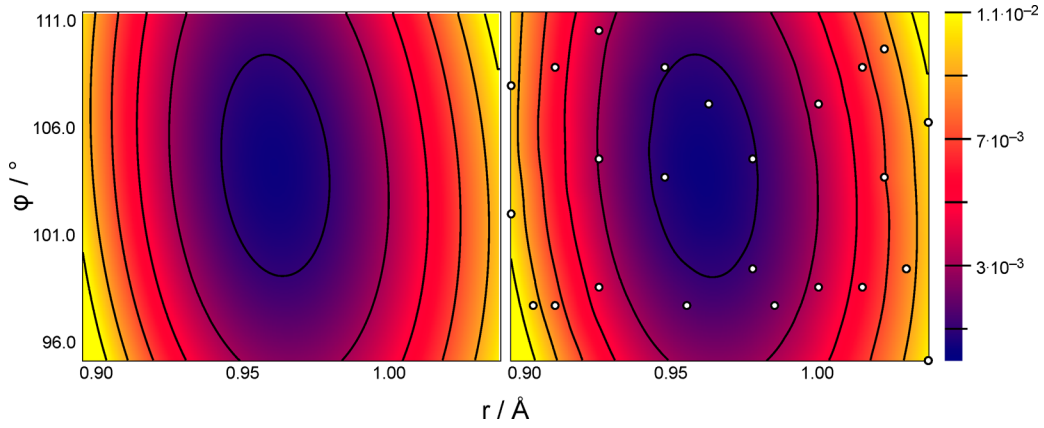


Figure 4. Potential energy surface of the water molecule with respect to angle bending (φ) and symmetric bond stretching (r). Shown is the CCSD reference (left) and the surface obtained for 100 predicted data-points from a ML-PT model (right, M = 24, σ_T = 0.005). The geometries of the training points are marked with white circles.

points were generated by scanning along the symmetric bond-stretch (0.97 ± 0.075 Å) and HOH angle (104.3 ± 8.59°). A ML-PT model was trained using the MP2 amplitudes and CCSD energies of 24 data points. The resulting potential energy surface constructed from the 100 unseen data points in the prediction set is shown in Figure 4, along with the full CCSD reference data (see the Supporting Information for corresponding plots of the correlation energy). The performance of the ML model is again very promising for this higher dimensional system. Here, it is important to emphasize that the ML model is completely unaware of any structural information, as the TATR is constructed from the wave function alone. Nevertheless, the potential energy surface is smooth, concave, the minimum is at the correct location, and the mean absolute energetic error to the CCSD reference is ca. 3×10^{-5} E_h (~0.02 kcal mol^{-1}; see Table 1).

Table 1. Error Statistics for the Prediction of the Water PES Scanning along the Symmetric Bond Stretch (0.97 ± 0.075 Å) and HOH Angle (104.3 ± 8.59°)

	E_h	(kcal mol^{-1})
root mean squared error	4.4×10^{-5}	0.027
mean absolute error	3.0×10^{-5}	0.019
mean signed error	5.3×10^{-6}	0.003
maximum deviation	1.6×10^{-4}	0.10

As a second triatomic, the hydrogen cyanide molecule (HCN) was also investigated (see the Supporting Information). Here, the potential energy surface was scanned along the CN and CH bond stretches. While the results are overall comparable to H_2O, this is an interesting case because it demonstrates that the TATR can be used to simultaneously describe the rather soft CH and the much more rigid CN mode. This indicates that the TATR is flexible enough to describe more complex chemical situations as well. We also include training curves for this PES using σ_T = 0.05 and 0.005. This provides a numerical illustration of the discussion above, regarding the choice of this parameter. In the limit of a very sparse training set, a large value of σ_T performs significantly better, whereas a small value leads to ultimately more accurate models for a dense training set.

The most intriguing prospect of this work is that the methodology could be applied to run CC-quality molecular dynamics simulations. However, this is only feasible if forces on the atoms can be calculated. While the forces can in principle be obtained from numerical differentiation, this might be rather error-prone as the ML models discussed so far are not trained to capture the very small energy differences that occur in this context.

An appealing alternative is to instead learn the forces directly.[27] To this end, the TATR representation needs to be modified, to reflect how the wave function changes upon a small perturbation of the molecular geometry. As a proof of principle to this end, we use a simple numerical representation $\delta\boldsymbol{\nu}$, for the perturbed wave function:

$$\delta\boldsymbol{\nu} = \boldsymbol{\nu}_+ - \boldsymbol{\nu}_- \tag{8}$$

where $\boldsymbol{\nu}_+$ and $\boldsymbol{\nu}_-$ designate the TATRs constructed at slightly displaced geometries (±1.0 × 10^{-7} Å along the respective degree of freedom).

It is then straightforward to apply the same ML-PT methodology used above to learn residual forces (i.e., the difference between CCSD and HF forces). In Figure 5 we show the residual forces for the dissociation of the hydrogen molecule.

Just like energies, CCSD forces can be learned from MP2 amplitudes with high accuracy, even for stretched geometries

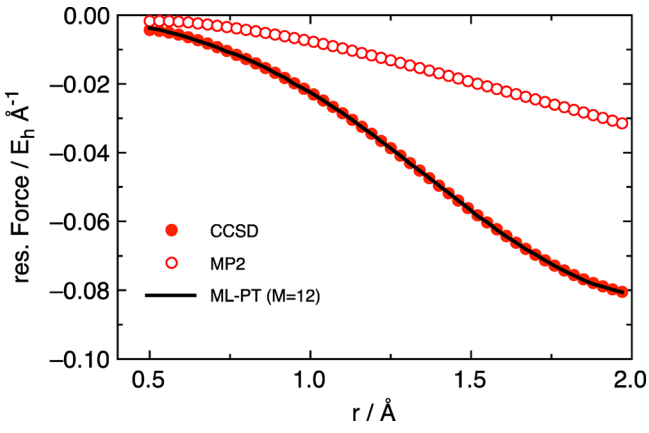


Figure 5. Residual forces (from the difference between correlated and HF gradients) from CCSD, MP2, and ML-PT (trained on MP2 amplitude differences, M = 12, σ_T = 0.05), for the dissociation of the hydrogen molecule.

DOI: 10.1021/acs.jpca.8b04455
J. Phys. Chem. A 2018, 122, 6343−6348

where the MP2 forces are qualitatively wrong. The numerical representation used here is obviously not ideal, as it is quite expensive to calculate (i.e. as expensive as numerical MP2 gradients). However, a similar representation based on analytical MP2 gradients could also be defined. This will be the focus of future work.

CONCLUSIONS

In conclusion, we have introduced a machine-learnable vector representation of CC wave functions. The representation was used to train ML models for the CCSD correlation energy, using both CCSD and MP2 amplitudes. For a series of small molecules, the CCSD potential energy surfaces could be accurately and efficiently predicted in this way. We also showed that the ML approach can be applied directly to forces. This opens the door toward affordable CC-quality molecular dynamics simulations of molecular systems.

As the ML prediction has negligible computational cost compared to the determination of the amplitudes, the computational scaling of such a simulation is already formally improved by one order of magnitude ($O(n^5)$ vs $O(n^6)$) within the present framework. Additionally, highly efficient (up to $O(n)$) MP2 implementations exist, which would further reduce this scaling.[28,29]

Given the proof-of-principle nature of our study, it is appropriate to consider the limitations of the proposed approach at the current stage. First, the CCSD/def2-TZVP reference data used herein does not correspond to "gold-standard" CC, which would also include perturbative triples and extrapolation to the complete basis set limit (i.e., the CCSD(T)/CBS model). However, the data are well suited for the purpose of this study, namely, to illustrate that the coupled cluster correlation energy is machine learnable on the basis of a representation of the T-amplitudes.

A general shortcoming of global representations like the TATR or MBTR is the transferability of the corresponding ML-models between similar systems with different sizes. In this context, a potential further benefit of using linear-scaling CC and PT frameworks to generate the references (see above) lies in the fact that these approaches typically rely on some form of orbital localization. This would potentially allow splitting the TATR into atomic or pair contributions, with the advantage that the scheme then becomes transferable from small to large systems. It should be stressed, however, that a TATR based on localized CC or MP2 cannot be compared with a canonical one, and additional work is necessary to develop such a localized scheme.

A related issue is that the TATR is not invariant to changes in the phase of molecular orbitals. In other words, even for canonical orbitals, the representation may differ when generated from different initial guess orbitals, or with different codes. For the current study, this is not problematic since training and prediction steps were performed jointly. However, in general, there is no guarantee that the phase convention is consistent throughout training and prediction, particularly if different codes or SCF solvers are used. There are two possible solutions to this issue, namely, enforcing some phase convention (which may be difficult in practice) or modifying the TATR to ensure invariance with respect to the phase of molecular orbitals.

The remaining challenge for treating larger systems is that, in general, the training data cannot be generated by simply scanning along all degrees of freedom, as we have done here. However, the work of Brockherde et al. has shown that simple molecular mechanics trajectories provide sufficiently representative geometries for this purpose (in the context of DFT-ML models).[16] Beyond this, we will also explore if the TATR itself is machine-learnable, which would allow us to completely bypass the explicit calculation of the T-amplitudes. This is in fact quite likely, as the TATR changes smoothly with changes in geometry.

COMPUTATIONAL DETAILS

The KRR hyperparameters σ (in eq 7) and λ (see the Supporting Information) are determined via cross-validation. Reference data are generated by evenly sampling the relevant degrees of freedom (as described in the main text). In each case, 50 data points are randomly chosen to be the prediction set. These are exclusively used for validation. To obtain a representative training set with M data points, the remaining reference data (the grand training set) are clustered using the k-means algorithm (see the Supporting Information for details).[30] The reference data point closest to the center of each cluster is then included in the training set. TATRs are calculated using the 150 largest T_1 and T_2 amplitudes. For convenience, all energies are internally shifted by the average correlation energy of the training set (i.e., the mean of the training data is set to zero). All CC calculations are performed with the Psi4 code, using the def2-TZVP basis set, as are corresponding calculations at the level of Møller−Plesset (MP2) perturbation theory to second order.[31,32] Further methodological details are provided in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.8b04455.

Additional methodological details about the KRR, choice of σ_T, and training sets; ML dissociation curves HCN PES and its learning curves; correlation energy plots for all studied diatomics (PDF)

AUTHOR INFORMATION

Corresponding Author

*J. Margraf. E-mail: johannes.margraf@ch.tum.de.

ORCID

Johannes T. Margraf: 0000-0002-0862-5289

Karsten Reuter: 0000-0001-8473-8659

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Funding by the Humboldt Foundation and the TUM University Foundation is gratefully acknowledged.

REFERENCES

(1) Rupp, M. Machine Learning for Quantum Mechanics in a Nutshell. *Int. J. Quantum Chem.* **2015**, *115* (16), 1058−1073.

(2) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114* (10), 105503.

(3) Lorenz, S.; Groß, A.; Scheffler, M. Representing High-Dimensional Potential-Energy Surfaces for Reactions at Surfaces by Neural Networks. *Chem. Phys. Lett.* **2004**, *395* (4−6), 210−215.

DOI: 10.1021/acs.jpca.8b04455
J. Phys. Chem. A 2018, 122, 6343−6348

(4) Prudente, F. V.; Acioli, P. H.; Neto, J. J. S. The Fitting of Potential Energy Surfaces Using Neural Networks: Application to the Study of Vibrational Levels of H3+. *J. Chem. Phys.* **1998**, *109* (20), 8801−8808.
(5) Ischtwan, J.; Collins, M. A. Molecular Potential Energy Surfaces by Interpolation. *J. Chem. Phys.* **1994**, *100* (11), 8080−8088.
(6) Lee, K.; Bowman, J. M. Rotational Distributions and Collision Lifetimes in H+CO Scattering. *J. Chem. Phys.* **1986**, *85* (10), 6225−6226.
(7) Behler, J.; Lorenz, S.; Reuter, K. Representing Molecule-Surface Interactions with Symmetry-Adapted Neural Networks. *J. Chem. Phys.* **2007**, *127* (1), 014705.
(8) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104* (13), 136403.
(9) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108* (5), 58301.
(10) Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2011**, *134* (7), 074106.
(11) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192−3203.
(12) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87* (18), 184115.
(13) Huo, H.; Rupp, M. Unified Representation of Molecules and Crystals for Machine Learning. **2017**, arXiv:1704.06439.
(14) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890.
(15) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13* (11), 5255−5264.
(16) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham Equations with Machine Learning. *Nat. Commun.* **2017**, *8* (1), 872.
(17) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108* (25), 253002.
(18) Li, L.; Baker, T. E.; White, S. R.; Burke, K. Pure Density Functional for Strong Correlation and the Thermodynamic Limit from Machine Learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94* (24), 245129.
(19) McGibbon, R. T.; Taube, A. G.; Donchev, A. G.; Siva, K.; Hernández, F.; Hargus, C.; Law, K.-H.; Klepeis, J. L.; Shaw, D. E. Improving the Accuracy of Møller-Plesset Perturbation Theory with Neural Networks. *J. Chem. Phys.* **2017**, *147* (16), 161725.
(20) Dral, P. O.; von Lilienfeld, O. A.; Thiel, W. Machine Learning of Parameters for Accurate Semiempirical Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 2120−2125.
(21) Bartlett, R. J.; Musiał, M. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79* (1), 291−352.
(22) Schwilk, M.; Ma, Q.; Köppl, C.; Werner, H. J. Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD). *J. Chem. Theory Comput.* **2017**, *13* (8), 3650−3675.
(23) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse Maps - A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory. *J. Chem. Phys.* **2016**, *144* (2), 024109.
(24) Purvis, G. D.; Bartlett, R. J. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76* (1982), 1910−1918.
(25) Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; MIT Press, 2006.
(26) Coulson, C. A.; Fischer, I. XXXIV. Notes on the Molecular Orbital Treatment of the Hydrogen Molecule. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **1949**, *40* (303), 386−393.
(27) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3* (5), e1603015.
(28) Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. Sparse maps—A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. I. An Efficient and Simple Linear Scaling Local MP2Method That Uses an Intermediate Basis of Pair Natural Orbitals. *J. Chem. Phys.* **2015**, *143* (3), 034108.
(29) Maurer, S. A.; Clin, L.; Ochsenfeld, C. Cholesky-Decomposed Density MP2 with Density Fitting: Accurate MP2 and Double-Hybrid DFT Energies for Large Systems. *J. Chem. Phys.* **2014**, *140* (22), 224112.
(30) Lloyd, S. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28* (2), 129−137.
(31) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297−3305.
(32) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13* (7), 3185−3197.

DOI: 10.1021/acs.jpca.8b04455
J. Phys. Chem. A 2018, 122, 6343−6348