THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

Cite This: J. Phys. Chem. Lett. 2019, 10, 4129−4135

pubs.acs.org/JPCL

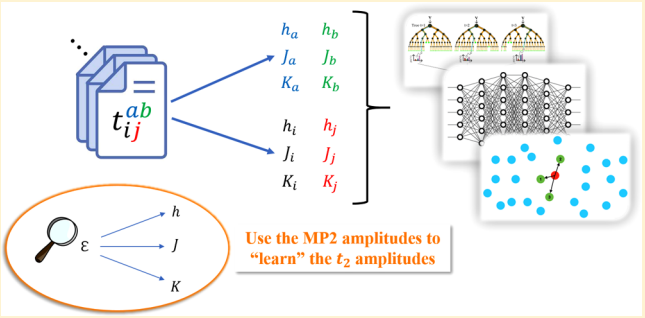
Data-Driven Acceleration of the Coupled-Cluster Singles and Doubles Iterative Solver

Jacob Townsend and Konstantinos D. Vogiatzis*

Department of Chemistry, University of Tennessee, Knoxville, Tennessee 37996, United States

Supporting Information

ABSTRACT: Solving the coupled-cluster (CC) equations is a cost-prohibitive process that exhibits poor scaling with system size. These equations are solved by determining the set of amplitudes (t) that minimize the system energy with respect to the coupled-cluster equations at the selected level of truncation. Here, a novel approach to predict the converged coupled-cluster singles and doubles (CCSD) amplitudes, thus the coupled-cluster wave function, is explored by using machine learning and electronic structure properties inherent to the MP2 level. Features are collected from quantum chemical data, such as orbital energies, one-electron Hamiltonian, Coulomb, and exchange terms. The data-driven CCSD (DDCCSD) is not an alchemical method because the actual iterative coupled-cluster equations are solved. However, accurate energetics can also be obtained by bypassing solving the CC equations entirely. Our preliminary data show that it is possible to achieve remarkable speedups in solving the CCSD equations, especially when the correct physics are encoded and used for training of machine learning models.



Coupled-cluster singles and doubles with perturbative triples (CCSD(T)) has become the "gold standard" in computational chemistry because it can provide accurate electronic energies and properties for molecules with a single-reference ground state.[1,2] The CCSD(T) scheme has been applied in a plethora of computational studies for the calculation of accurate atomization energies,[3,4] reaction barriers,[5−7] noncovalent interactions,[8,9] and more. However, conventional CCSD(T) is a computationally intensive method in terms of CPU time and disk space, which limits its applicability to small- and medium-size molecular systems; a large basis close to the complete basis set (CBS) limit is needed to achieve the expected accuracy. Solutions to the slow convergence with respect to the basis set include extrapolation schemes[3,10−13] and addition of two-electron functions that explicitly depend on the interelectronic r_{12} distance (explicitly correlated methods or R12/F12 methods).[14−17] Alternative methodologies that speed-up CC schemes are exploration of the locality of electron excitations,[18] fragmentation schemes,[19,20] pair-natural orbital (PNO) expansions,[21−25] linear scaling methods,[26] and high-performance computing.[27] Herein, we present a novel approach toward implementing low-cost, highly accurate CC calculations through machine learning (ML). The data-driven coupled-cluster (DDCC) scheme uses ML for accurate prediction of the double excitation amplitudes, t_{ij}^{ab}. A new wave function is generated that aims to replicate the converged CCSD wave function. The machine-learned wave function can be used either directly for accurate CCSD-level observables, such as energy, at MP2-level cost, or as an improved guess for the t_{ij}^{ab} amplitudes, which reduces the number of iterations to converge the CCSD equations.

In recent years, ML has had an increased role in chemistry as a promising technique because it is fast and scalable for a broad range of applications. ML has been widely applied in determining potential energy surfaces[28−40] and has been recently applied as an efficient method for screening the chemical space for quantum-level accuracy prediction and materials discovery.[41−47] Achieving quantum-level accuracy at low computational cost has received much attention due to developments in machine-learnable molecular representations such as Coulomb matrices,[48] Bag-of-Bonds,[49] and many other fingerprinting methods.[50−57] However, these methods ultimately rely on mapping molecular geometries to properties, which can require extensive amounts of training data and may have limited transferability between systems. Additionally, this may require optimization of chemical structures in order to properly relate molecular structure to function.

In contrast to mapping properties from molecular structure, ML has been also used to improve the electronic structure description directly. For density functional theory (DFT)-based approaches, ML has been applied to directly learn the electron density and for the development of exchange−correlation functionals.[58−61] Recently, there have also been developments in applying ML to wave function-based

Received: May 20, 2019
Accepted: June 25, 2019
Published: June 25, 2019

DOI: 10.1021/acs.jpclett.9b01442
J. Phys. Chem. Lett. 2019, 10, 4129−4135

methods. For instance, Coe developed a technique for choosing important electronic configurations in a selected configuration interaction (CI) expansion.[62] Additionally, there have been several implementations to predict coupled-cluster energies using electronic structure techniques. McGibbon et al. developed spin-network-scaled MP2 (SNS-MP2), which uses Hartree−Fock, second-order Møller−Plesset perturbation theory (MP2), and symmetry-adapted perturbation theory (SAPT0) calculations coupled with an artificial neural network (ANN) to predict noncovalent interaction energies at the CCSD(T) level.[63] Margraf and Reuter developed a method to predict the CC correlation energy on systems by implementing a transformation on a subset of the largest MP2 amplitudes modified using a transformed tensor representation.[64] Their technique uses features independent of the number of atoms and includes only electronic structure features but does not demonstrate transferability across different molecular systems. Lastly, Welborn et al. presented a ML method for predicting CCSD energies based on Hartree−Fock localized molecular orbital (MO) features, yielding an accurate method with promising chemical transferability.[65] These previous ML-CC methods have used MP2 or HF wave functions to predict CC energies. Building on these ideas, we present a data-driven approach that uses the MP2 wave function and its implicit properties to predict the CCSD wave function. This new data-driven CCSD (DDCCSD) wave function can be used as a starting guess for the CCSD solver for faster convergence of the CC equations or implemented as a low-cost improvement to the MP2 wave function for more accurate properties and energetics.

The CC ansatz uses an exponential form, where the cluster operator typically acts on a canonical Hartree−Fock (HF) reference

$$|\psi_{CC}\rangle = \exp(\hat{T})|\psi_0\rangle \tag{1}$$

$\hat{T}$ is the cluster operator, and ψ_0 represents the HF reference wave function. The cluster operator is truncated for practical purposes, and the most common implementation uses only explicit singles and doubles excitations (CCSD). The correlation energy is given by

$$E_{corr}^{CCSD} = \sum_{\substack{a<b \\ i<j}} \langle ij||ab\rangle t_{ij}^{ab} + \sum_{\substack{a<b \\ i<j}} \langle ij||ab\rangle t_i^a t_j^b \tag{2}$$

where $\{ij\}$ and $\{ab\}$ refer to occupied and virtual orbitals, respectively, and t_i^a and t_{ij}^{ab} are the one- and two-excitation amplitudes that are determined through the iterative solution of the CC equations. Conventional CCSD uses the MP2 amplitudes as initial guess

$$t_{ij(MP2)}^{ab} = \frac{\langle ab||ij\rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{3}$$

where ε_k corresponds to the energy of the respective orbital, k. The amplitudes and properties of these orbitals will ultimately comprise the feature set used to design a ML-based wave function.

For developing a transferable model to molecular systems of varying size, orientation, and atom composition, we chose to restrict the features to electronic structure properties. Therefore, the features do not explicitly incorporate the identity or position of any of the atoms. Additionally, all of the features consist of values already obtained in the HF or MP2 calculation, adding no additional cost for the generation of the feature set. An obstacle in applying ML to quantum chemical applications is dimensionality: i.e., how we frame an input feature matrix of amplitudes when the number of amplitudes changes for all systems. Molecular representations of molecules such as the Coulomb matrix have used padding, where all empty cells are filled with zeros.[48] Here, the size of the input and output matrices will vary, and for the case of amplitudes, the size of the matrix of amplitudes is $(M_{occ})^2(M_{vir})^2$, where M_{occ} and M_{vir} corresponds to the number of occupied and virtual MOs, respectively. However, this is overcome by predicting each amplitude independently. Therefore, only the properties of one instance of a double excitation will be considered by the ML model at a time.

For a given double excitation, elements from the Fock, Coulomb, and exchange matrices are collected, which can be considered as an extension of the approach taken by Welborn and co-workers.[65] For each double excitation, the following are collected:

- orbital energies, as well as the broken-down contributions to these energies from the one-electron Hamiltonian matrix (h), Coulombic matrix (J), and exchange matrix (K), which come from diagonalization of their respective matrices
- Coulombic and exchange integrals $(J_i^a, J_j^b, K_i^a, K_j^b)$
- whether two promoted electrons go into the same virtual orbital or not (value of 1 or 0)
- two-electron integral $\langle ij||ab\rangle$
- MP2 denominator $(\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b)$
- initial MP2 amplitude (see eq 3)

These features and some mathematical transformations discussed in the Supporting Information comprise the 30 features for each doubles excitation. Such selection of features provides a representation that contains properties analogous to the HF method used to generate the features; the predictions are invariant to rotation and translation. However, the predictions may not be invariant toward different localization schemes, and therefore, it necessitates the use of the same scheme for the generation of the MOs. It may be beneficial to use localized MOs as many of the amplitudes would tend toward zero values at increasing distances between atoms, which is encoded in the feature set by proxy through the Coulombic integrals.[65] The standard MOs are delocalized, and this may lead to more nonvanishing amplitudes. However, the performance change for such an implementation will be examined in a future study.

The k-Nearest Neighbors (k-NN) algorithm was used for the ML prediction of the double excitation amplitudes. k-NNs have several distinct advantages for a proof-of-concept. First, they are conceptually simple: they measure the distance of a test sample in the feature space from the training samples, take the k closest neighbors, and predict with the average value of those neighbors. In this way, it does not make assumptions about the data and is strictly instance-based learning. Most ML methods fit with a mean-squared error loss function, and this would lead to fitting the largest amplitudes more accurately and ultimately neglecting the smaller amplitudes. This was shown to be problematic when attempting to use the ML-predicted wave function as a starting point for iterations to calculate the converged CCSD wave function. However, k-NNs avoid the need for scaling amplitudes as they are not trained via a loss function. Due to their simplicity, k-NN

DOI: 10.1021/acs.jpclett.9b01442
J. Phys. Chem. Lett. 2019, 10, 4129−4135

models provide an excellent proof-of-concept which can be considered a first step towards a large-scale implementation of DDCC.

By default, *k*-NN algorithms weigh the distance in feature space for all features equivalently. This involves calculation of the Euclidean distance

$$d(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \ldots + (q_n - p_n)^2} \quad (4)$$

$d(\boldsymbol{p},\boldsymbol{q})$ represents the distance measured by the *k*-NN, $\boldsymbol{p}$ and $\boldsymbol{q}$ represent a given test and training sample, respectively, and the subscripts denote the feature number up to n features. In practice, this means that all feature distances contribute equally in their importance toward the calculated total distance used to decide the nearest neighbor, which is undesirable for our implementation, as all of the features may not contribute equally or perhaps a small error in a certain feature is more significant than another. This can be visualized by the modified Euclidean distance

$$d(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{[x_1(q_1 - p_1)]^2 + [x_2(q_2 - p_2)]^2 + \ldots + [x_n(q_n - p_n)]^2} \quad (5)$$

where the feature weights, x_n, were determined via a grid search on the scaled data to optimize their weights. The most significant features include the corresponding two-electron integral, the MP2 amplitude, the orbital energy differences, and whether or not the two promoted electrons are being excited into the same virtual orbital. Optimized weights are listed in the Supporting Information. The features were scaled using a MinMaxScaler in the Sci-Kit Learn[27] module to accentuate and normalize the distances between points in the feature space to allow for a truer recognition of neighbors. All *k*-NN models were implemented with one nearest neighbor (k = 1), which was shown to have the best performance and is listed in the Supporting Information.

All HF calculations were performed in the Psi4[66] program. The MP2 and coupled-cluster calculations were performed using the Psi4Numpy[67] spin-factored CCSD implementation.[68] The implementation of the DDCC code described herein uses the CCSD code modified to generate the ML features and predict the CC wave function using *k*-NNs. All ML results included use the ML predicted double excitation amplitudes (t_{ij}^{ab}) as an initial guess, and both t_i^a and t_{ij}^{ab} are further converged as in the conventional CCSD scheme. All structures were obtained from simulations at 200 K using the OPLS-AA force field,[69] as generated by LigParGen[70] and implemented in the LAMMPS[71] software.

Single-Molecule Prediction: H_2O as a Test Case. There are two metrics that will be used to evaluate the effectiveness of DDCC: (1) energy differences between the converged CCSD energy and the starting energy before performing iterations and (2) the number of iterations until convergence. Errors in energy will be reported as the mean absolute error in mHartree (mE_h). Another metric to determine the accuracy of the method will be the percent improvement over MP2

$$\text{Percent Improvement} = \left(1 - \frac{(\Delta E_{ML})}{(\Delta E_{MP2})}\right) \times 100 \quad (6)$$

ΔE corresponds to the energy difference between the converged CCSD energy and the initial energy calculated before iterations, and the subscripts ML and MP2 correspond to the respective method's energies. The CCSD iterations include also the relaxation of the single excitation amplitudes, but because their number is significantly smaller than the double excitation amplitudes, we followed their standard initialization (set to zero and solved separately). The projected equations for the t_i^a amplitudes were solved for all schemes discussed in the next paragraphs.

The method was investigated on a set of water molecules at different geometries generated with OPLS-AA.[69] The performance of the method was evaluated on a sample of 1, 5, 10, 20, 40, and 100 water molecules in the training set, and the trained models were tested on a different set of 50 water molecules. The results for both energy differences and iterations for the STO-3G,[72] cc-pVDZ,[73] aug-cc-pVDZ,[74] and aug-cc-pVTZ[74] basis sets are presented in Figures 1 and 2. Even training on

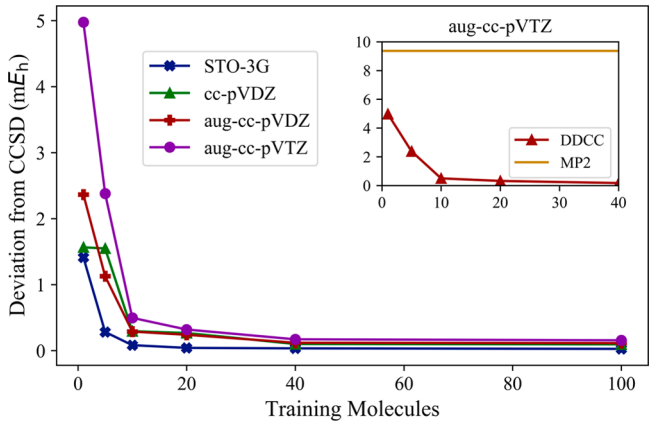


Figure 1. Average absolute deviation between the initial DDCC energy and the converged CCSD energies for 50 water molecules in the test set by varying the size (number of waters) in the training set. For comparison, the MP2 deviation from the CCSD results is included for the aug-cc-pVTZ basis set (inset).

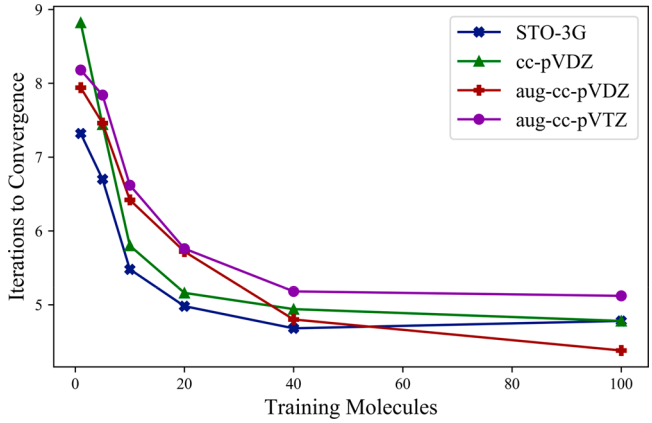


Figure 2. Number of steps to converge the CC equations based on the number of molecules in the training set.

one molecule, the error in energy is significantly lower than the MP2 error, which is depicted for the aug-cc-pVTZ basis set (Figure 1, inset). By systematically increasing the number of molecules in the training set, the average deviation from exact CCSD significantly decreases. The larger basis sets require more training examples to reduce error due to the increased number of amplitudes; however, they all show systematic improvement in the expansion of training molecules, and with

DOI: 10.1021/acs.jpclett.9b01442
J. Phys. Chem. Lett. 2019, 10, 4129−4135

10 training structures, a sub-mE_h average deviation was obtained. With 40 training examples, the average deviation for all basis sets is below 0.2 mE_h, or approximately 98% of the difference between MP2 and CCSD. By using 40 molecules for training, the energy difference from the true CCSD correlation energy has nearly converged (Figure 1), and thus, 40 training examples will be used in the following applications, unless otherwise noted.

Another important consideration is the number of steps (iterations) taken to solve the CCSD equations. By building the wave function with the DDCC amplitudes, the number of steps to convergence is reduced significantly (Figure 2). Conventional CCSD converges in eight iterations for all basis sets, except cc-pVDZ, which converges in seven steps. The 100 water *k*-NN models result in up to a 40% reduction in iterations in the test set.

Transferability: Data from Seven Small Molecules. DDCC is able to predict better starting amplitudes for the coupled-cluster wave function than MP2. To demonstrate transferability and model performance, we trained ML models where electronic structure data from other molecules are used, and tested on a set of seven small molecules. Eight different models were created using electronic structure data from seven different molecules obtained with the 6-311G basis set:[75] water, methane, ethylene, ethane, methanol, ammonia, and formaldehyde. Each model was trained on one molecule type and tested on all seven (ML-*molecule name* of Figure 3). The

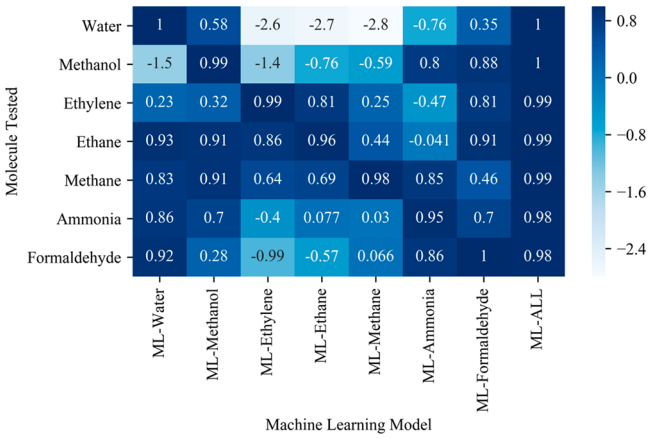


Figure 3. Heat map showing the performance of different *k*-NNs models, where the value being plotted is the percent improvement (eq 6).

eighth model was trained using electronic structure data from all of the molecules combined (ML-All of Figure 3). Figure 3 shows a summary of the results from these calculations on sets of 50 test molecules in the form of a heat map that plots the percent improvement.

A remarkable percent improvement is achieved when each molecule uses a model trained from data of the same type of molecule (diagonal of Figure 3). Off-diagonal cases show varying degrees of accuracy. For instance, formaldehyde test cases see a large percent improvement from the model trained on water, whereas it does not perform better with the model trained on ethylene. Such performance enhancements are not intuitive, and it is unclear when a performance enhancement will occur, but it is probably related to lack of data that capture the physics (*vide infra*) of a specific atom type. For example, the ML-water model predicts well (80% improvement or greater) for five of the seven molecules, but its energy is not well predicted by the models generated from other molecules. However, the ML model trained on the data with all seven molecules shows impressive performance for all molecules with at least 98% improvement, which corresponds to less than a 0.5 mE_h error for all molecules. Additionally, a 15−60% reduction in iterations to convergence was observed (see Table S4 of the Supporting Information).

Molecular Clusters: Extrapolating from Few Water Molecules to Larger Clusters. While it is useful to be able to predict the energies of isolated molecules, as in the seven molecule heat map (Figure 3), it is also useful to generate accurate energies for systems containing several interacting molecules. A successful model must be able to capture the physics of a larger interacting system based on those provided by smaller subsystems. To demonstrate this, models have been trained using conformations containing one, two, and three water molecules to predict the amplitudes of four different clusters containing six (book, prism, cage, and ring conformations), eight (D_{2d} and S_4), and 16 water molecules (Figure 4)

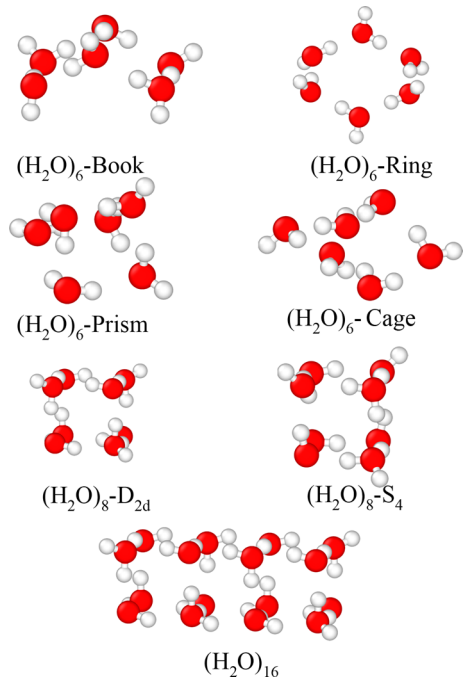


Figure 4. Molecular structures of water hexamers in the book, ring, prism, and cage conformation, water octamers in D_{2d} and S_4 conformations, and of the ordered water decahexamer conformation.

optimized by Miliordos et al.[76] An optimized monomer, 11 dimers generated through a relaxed dissociation, and 24 trimer systems were used for training (see the Supporting Information). The goal was to demonstrate that these smaller systems are capable of capturing the physics of the larger clusters. Due to the larger system sizes, the STO-3G basis was used for testing purposes.

For demonstration purposes, models that use data from each supersystem structure will be added on top of the previous model's data; for example, the trimer model will contain data from the monomer, dimer, and trimer systems. Therefore, we can examine how interactions of the larger supersystem are captured given the increasing system complexity through the monomer, dimer, and trimer systems. The deviations and

DOI: 10.1021/acs.jpclett.9b01442
J. Phys. Chem. Lett. 2019, 10, 4129−4135

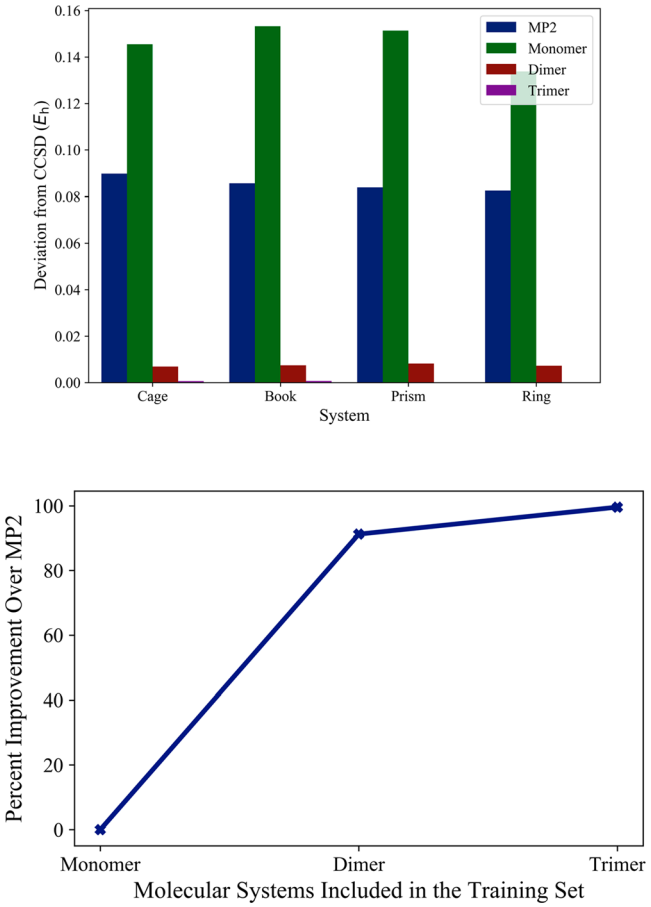


Figure 5. DDCC absolute deviation (E_h) from the converged CCSD plotted for the four systems containing data with the monomer, dimer, and trimer systems, as well as from CCSD using from MP2 amplitudes (top). The average percent improvement for the four systems at each level of physics is included (below).

improvements for each of the models on the six-molecule supersystems are shown in Figure 5. The model based on only isolated water failed to predict hexamer test cases, and none of the ML systems showed an improvement on the MP2 energy. This finding is in line with intuition that the physics of the intermolecular interactions between the water molecules are not captured from an isolated water. However, by adding data from dimer conformations in the training set, the errors drop drastically (less than 9.0 mE_h for all systems), which results in a 90% improvement over MP2. By including trimer systems in the training set, the DDCC deviations from the CCSD energy drop to below 1.0 mE_h for all systems, which results in greater than 99% percent improvement for all of the six-water systems.

To further demonstrate the extrapolation, the 3 water *k*-NN model was used to predict the CCSD energy on two different 8 water clusters (D_{2d} and S_4) and a 16 water cluster. The absolute errors were 9.65, 15.92, and 8.33 mE_h, respectively. These errors correspond to 90.5, 81.3, and 95.7% improvement over the MP2 energies, which demonstrates that the model of three waters can be extrapolated to much larger systems and the DDCC scheme demonstrates a significant improvement to the MP2 wave function.

This study demonstrates the prediction of the CC amplitudes with a great deal of efficiency. In the first test case (convergence of CCSD for one water molecule), the ML-predicted amplitudes resulted in energies that are two orders of magnitude closer to the true CCSD correlation energy and in 40% reduction of the number of iterations on average for all basis sets tested. To demonstrate the transferability of the proposed methodology, seven ML models were trained on data from different molecules, plus a model that incorporated data from all seven molecules. Across the board, all molecules showed to have improvement over the MP2 predicted amplitudes using the model trained on all seven molecules, meaning that more data, in general, should lead to more accurate models using *k*-NNs. This shows that the model can be extrapolated to new molecules with good accuracy. Lastly, subsystems containing few waters (1−3) were used to predict larger clusters of water (6,8,16). It was found that the models including one, two, and three waters could predict six water systems with a 99% improvement over MP2 because the chemical interactions between water molecules were encoded successfully. The 8 and 16 water systems also showed significant improvements. In conclusion, we have shown that machine-learned coupled-cluster amplitudes can provide a better starting approximation to the CC wave function than MP2 amplitudes and are capable of extrapolating to larger systems. The suggested methodology is not alchemical because the exact CCSD energies can be computed by solving the exact CC equations and can be used in a plethora of chemical applications (such as the calculation of full potential surfaces from only a few structures, and the efficient computation of energy and properties of molecular clusters from smaller subclusters). Our methodology is transferable to other methods that use iterative solvers, such as the complete active space second-order perturbation theory (CASPT2) methods, which are currently being examined in our group. Finally, a data-driven scheme for the accurate computation of the (T) term is also underway.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.9b01442.

Additional details about feature selection, feature weighting, convergence behavior, and coordinates to water systems (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kvogiatz@utk.edu.

ORCID

Konstantinos D. Vogiatzis: 0000-0002-7439-3850

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the National Science Foundation (CHE-1800237) for financial support of this work and the Advanced Computer Facility (ACF) of the University of Tennessee for computational resources. The authors would like to thank Justin Kirkland for fruitful discussions on the machine learning aspects of this project.

■ REFERENCES

(1) Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent Advances in Wave Function-Based Methods of

DOI: 10.1021/acs.jpclett.9b01442
J. Phys. Chem. Lett. 2019, 10, 4129−4135

Molecular-Property Calculations. *Chem. Rev.* **2012**, *112* (1), 543−631.
(2) Townsend, J.; Kirkland, J. K.; Vogiatzis, K. D. Post-Hartree-Fock Methods: Configuration Interaction, Many-Body Perturbation Theory, Coupled-Cluster Theory. In *Mathematical Physics in Theoretical Chemistry*; 2019; pp 63−117.
(3) Feller, D.; Peterson, K. A.; Grant Hill, J. On the Effectiveness of CCSD(T) Complete Basis Set Extrapolations for Atomization Energies. *J. Chem. Phys.* **2011**, *135*, 044102.
(4) Klopper, W.; Ruscic, B.; Tew, D. P.; Bischoff, F. A.; Wolfsegger, S. Atomization Energies from Coupled-Cluster Calculations Augmented with Explicitly-Correlated Perturbation Theory. *Chem. Phys.* **2009**, *356* (1−3), 14−24.
(5) Pu, J.; Truhlar, D. G. Benchmark Calculations of Reaction Energies, Barrier Heights, and Transition-State Geometries for Hydrogen Abstraction from Methanol by a Hydrogen Atom. *J. Phys. Chem. A* **2005**, *109* (5), 773−778.
(6) Zhang, J.; Valeev, E. F. Prediction of Reaction Barriers and Thermochemical Properties with Explicitly Correlated Coupled-Cluster Methods: A Basis Set Assessment. *J. Chem. Theory Comput.* **2012**, *8* (9), 3175−3186.
(7) Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2019**, *119* (4), 2453−2523.
(8) Sirianni, D. A.; Alenaizan, A.; Cheney, D. L.; Sherrill, C. D. Assessment of Density Functional Methods for Geometry Optimization of Bimolecular van Der Waals Complexes. *J. Chem. Theory Comput.* **2018**, *14* (6), 3004−3013.
(9) Lee, H. M.; Youn, I. S.; Saleh, M.; Lee, J. W.; Kim, K. S.; et al. Interactions of CO_2 with Various Functional Molecules. *Phys. Chem. Chem. Phys.* **2015**, *17* (16), 10925−10933.
(10) Martin, J. M. L. Ab Initio Total Atomization Energies of Small Molecules - Towards the Basis Set Limit. *Chem. Phys. Lett.* **1996**, *259* (5−6), 669−678.
(11) Pansini, F. N. N.; Varandas, A. J. C. Toward a Unified Single-Parameter Extrapolation Scheme for the Correlation Energy: Systems Formed by Atoms of Hydrogen through Neon. *Chem. Phys. Lett.* **2015**, *631*−*632*, 70−77.
(12) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997**, *106* (23), 9639−9646.
(13) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-Set Convergence in Correlated Calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett.* **1998**, *286* (3−4), 243−252.
(14) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112* (1), 4−74.
(15) Kong, L.; Bischoff, F. A.; Valeev, E. F. Explicitly Correlated R12/F12 Methods for Electronic Structure. *Chem. Rev.* **2012**, *112* (1), 75−107.
(16) Tew, D. P.; Klopper, W.; Helgaker, T. Electron Correlation: The Many-Body Problem at the Heart of Chemistry. *J. Comput. Chem.* **2007**, *28* (8), 1307−1320.
(17) Köhn, A.; Tew, D. P. Explicitly Correlated Coupled-Cluster Theory Using Cusp Conditions. I. Perturbation Analysis of Coupled-Cluster Singles and Doubles (CCSD-F12). *J. Chem. Phys.* **2010**, *133* (17), 174117.
(18) Saebo, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.
(19) Stoll, H. The Correlation Energy of Crystalline Silicon. *Chem. Phys. Lett.* **1992**, *191* (6), 548−552.
(20) Vogiatzis, K. D.; Klopper, W.; Friedrich, J. Non-Covalent Interactions of CO_2 with Functional Groups of Metal-Organic Frameworks from a CCSD(T) Scheme Applicable to Large Systems. *J. Chem. Theory Comput.* **2015**, *11* (4), 1574−1584.
(21) Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2015**, *11* (4), 1525−1539.
(22) Pavošević, F.; Peng, C.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps - A Systematic Infrastructure for Reduced Scaling Electronic Structure Methods. V. Linear Scaling Explicitly Correlated Coupled-Cluster Method with Pair Natural Orbitals. *J. Chem. Phys.* **2017**, *146* (17), 174108.
(23) Brabec, J.; Lang, J.; Saitow, M.; Pittner, J.; Neese, F.; Demel, O. Domain-Based Local Pair Natural Orbital Version of Mukherjee's State-Specific Coupled Cluster Method. *J. Chem. Theory Comput.* **2018**, *14* (3), 1370−1382.
(24) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013**, *139* (13), 134101.
(25) Vogiatzis, K. D.; Barnes, E. C.; Klopper, W. Interference-Corrected Explicitly-Correlated Second-Order Perturbation Theory. *Chem. Phys. Lett.* **2011**, *503* (1−3), 157−161.
(26) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. Linear-Scaling Methods in Quantum Chemistry. In *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 2007; pp 1−82.
(27) De Jong, W. A.; Bylaska, E.; Govind, N.; Janssen, C. L.; Kowalski, K.; Müller, T.; Nielsen, I. M. B.; Van Dam, H. J. J.; Veryazov, V.; Lindh, R. Utilizing High Performance Computing for Chemistry: Parallel Computational Chemistry. *Phys. Chem. Chem. Phys.* **2010**, *12* (26), 6896−6920.
(28) 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.
(29) Handley, C. M.; Popelier, P. L. A. Potential Energy Surfaces Fitted by Artificial Neural Networks. *J. Phys. Chem. A* **2010**, *114* (10), 3371−3383.
(30) Manzhos, S.; Carrington, T. Using Neural Networks to Represent Potential Surfaces as Sums of Products. *J. Chem. Phys.* **2006**, *125* (19), 194105.
(31) Dawes, R.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. Interpolating Moving Least-Squares Methods for Fitting Potential Energy Surfaces: A Strategy for Efficient Automatic Data Point Placement in High Dimensions. *J. Chem. Phys.* **2008**, *128* (8), 084107.
(32) Lorenz, S.; Scheffler, M.; Gross, A. Descriptions of Surface Chemical Reactions Using a Neural Network Representation of the Potential-Energy Surface. *Phys. Rev. B* **2006**, *73* (11), 1−13.
(33) Dawes, R.; Guo, Y.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. Interpolating Moving Least-Squares Methods for Fitting Potential Energy Surfaces: Improving Efficiency via Local Approximants. *J. Chem. Phys.* **2007**, *126* (18), 184108.
(34) Behler, J.; Lorenz, S.; Reuter, K. Representing Molecule-Surface Interactions with Symmetry-Adapted Neural Networks. *J. Chem. Phys.* **2007**, *127* (1), 014705.
(35) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145* (17), 170901.
(36) Pukrittayakamee, A.; Malshe, M.; Hagan, M.; Raff, L. M.; Narulkar, R.; Bukkapatnum, S.; Komanduri, R. Simultaneous Fitting of a Potential-Energy Surface and Its Corresponding Force Fields Using Feedforward Neural Networks. *J. Chem. Phys.* **2009**, *130* (13), 134101.
(37) Witkoskie, J. B.; Doren, D. J. Neural Network Models of Potential Energy Surfaces: Prototypical Examples. *J. Chem. Theory Comput.* **2005**, *1* (1), 14−23.
(38) Morawietz, T.; Behler, J. A Density-Functional Theory-Based Neural Network Potential for Water Clusters Including van Der Waals Corrections. *J. Phys. Chem. A* **2013**, *117* (32), 7356−7366.
(39) Jose, K. V. J.; Artrith, N.; Behler, J. Construction of High-Dimensional Neural Network Potentials Using Environment-Dependent Atom Pairs. *J. Chem. Phys.* **2012**, *136* (19), 194111.
(40) Botu, V.; Ramprasad, R. Learning Scheme to Predict Atomic Forces and Accelerate Materials Simulations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92* (9), 094306.
(41) Meyer, B.; Sawatlon, B.; Heinen, S.; von Lilienfeld, O. A.; Corminboeuf, C. Machine Learning Meets Volcano Plots: Computa-

DOI: 10.1021/acs.jpclett.9b01442
J. Phys. Chem. Lett. 2019, 10, 4129−4135

tional Discovery of Cross-Coupling Catalysts. *Chem. Sci.* **2018**, *9* (35), 7069−7077.
(42) Janet, J. P.; Kulik, H. J. Predicting Electronic Structure Properties of Transition Metal Complexes with Neural Networks. *Chem. Sci.* **2017**, *8* (7), 5137−5152.
(43) Janet, J. P.; Kulik, H. J. Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure-Property Relationships. *J. Phys. Chem. A* **2017**, *121* (46), 8939−8954.
(44) Janet, J. P.; Chan, L.; Kulik, H. J. Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. *J. Phys. Chem. Lett.* **2018**, *9* (5), 1064−1071.
(45) Janet, J. P.; Gani, T. Z. H.; Steeves, A. H.; Ioannidis, E. I.; Kulik, H. J. Leveraging Cheminformatics Strategies for Inorganic Discovery: Application to Redox Potential Design. *Ind. Eng. Chem. Res.* **2017**, *56* (17), 4898−4910.
(46) Fooshee, D.; Mood, A.; Gutman, E.; Tavakoli, M.; Urban, G.; Liu, F.; Huynh, N.; Van Vranken, D.; Baldi, P. Deep Learning for Chemical Reaction Prediction. *Mol. Syst. Des. Eng.* **2018**, *3* (3), 442−452.
(47) Goldsmith, B. R.; Esterhuizen, J.; Liu, J. X.; Bartel, C. J.; Sutton, C. Machine Learning for Heterogeneous Catalyst Design and Discovery. *AIChE J.* **2018**, *64* (7), 2311−2323.
(48) 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), 058301.
(49) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Muller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6* (12), 2326−2331.
(50) Faber, F. A.; Christensen, A. S.; Huang, B.; von Lilienfeld, O. A. Alchemical and Structural Distribution Based Representation for Universal Quantum Machine Learning. *J. Chem. Phys.* **2018**, *148* (24), 241717.
(51) Seko, A.; Hayashi, H.; Nakayama, K.; Takahashi, A.; Tanaka, I. Representation of Compounds for Machine-Learning Prediction of Physical Properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95* (14), 144110.
(52) von Lilienfeld, O. A.; Ramakrishnan, R.; Rupp, M.; Knoll, A. Fourier Series of Atomic Radial Distribution Functions: A Molecular Fingerprint for Machine Learning Models of Quantum Chemical Properties. *Int. J. Quantum Chem.* **2015**, *115* (16), 1084−1093.
(53) Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. Machine Learning of Molecular Properties: Locality and Active Learning. *J. Chem. Phys.* **2018**, *148* (24), 241727.
(54) Choi, S.; Kim, Y.; Kim, J. W.; Kim, Z.; Kim, W. Y. Feasibility of Activation Energy Prediction of Gas-Phase Reactions by Machine Learning. *Chem. - Eur. J.* **2018**, *24* (47), 12354−12358.
(55) Pronobis, W.; Schütt, K. T.; Tkatchenko, A.; Müller, K. R. Capturing Intensive and Extensive DFT/TDDFT Molecular Properties with Machine Learning. *Eur. Phys. J. B* **2018**, *91* (8), 178.
(56) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K. R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9* (8), 3404−3419.
(57) Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K. R.; von Lilienfeld, O. A. Machine Learning of Molecular Electronic Properties in Chemical Compound Space. *New J. Phys.* **2013**, *15*, 095003.
(58) 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.
(59) 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.
(60) 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.
(61) Seino, J.; Kageyama, R.; Fujinami, M.; Ikabata, Y.; Nakai, H. Semi-Local Machine-Learned Kinetic Energy Density Functional with Third-Order Gradients of Electron Density. *J. Chem. Phys.* **2018**, *148* (24), 241705.
(62) Coe, J. P. Machine Learning Configuration Interaction. *J. Chem. Theory Comput.* **2018**, *14* (11), 5739−5749.
(63) 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.
(64) Margraf, J. T.; Reuter, K. Making the Coupled Cluster Correlation Energy Machine-Learnable. *J. Phys. Chem. A* **2018**, *122*, 6343−6348.
(65) Welborn, M.; Cheng, L.; Miller, T. F. Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis. *J. Chem. Theory Comput.* **2018**, *14* (9), 4772−4779.
(66) 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.
(67) Smith, D. G. A.; Burns, L. A.; Sirianni, D. A.; Nascimento, D. R.; Kumar, A.; James, A. M.; Schriber, J. B.; Zhang, T.; Zhang, B.; Abbott, A. S.; et al. Psi4NumPy: An Interactive Quantum Chemistry Programming Environment for Reference Implementations and Rapid Development. *J. Chem. Theory Comput.* **2018**, *14* (7), 3504−3511.
(68) Stanton, J. F.; Gauss, J.; Watts, J. D.; Bartlett, R. J. A Direct Product Decomposition Approach for Symmetry Exploitation in Many-Body Methods. I. Energy Calculations. *J. Chem. Phys.* **1991**, *94* (6), 4334−4345.
(69) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(70) Dodda, L. S.; Cabeza de Vaca, I. C.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands. *Nucleic Acids Res.* **2017**, *45* (W1), W331−W336.
(71) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1−19.
(72) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules. *J. Chem. Phys.* **1970**, *52* (5), 2769.
(73) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007−1023.
(74) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96* (9), 6796−6806.
(75) McLean, A. D.; Chandler, G. S. Contracted Gaussian Basis Sets for Molecular Calculations. I. Second Row Atoms, Z = 11−18. *J. Chem. Phys.* **1980**, *72* (10), 5639−5648.
(76) Miliordos, E.; Xantheas, S. S. An Accurate and Efficient Computational Protocol for Obtaining the Complete Basis Set Limits of the Binding Energies of Water Clusters at the MP2 and CCSD(T) Levels of Theory: Application to $(H_2O)_m$, m = 2−6, 8, 11, 16, and 17. *J. Chem. Phys.* **2015**, *142* (23), 234303.

DOI: 10.1021/acs.jpclett.9b01442
J. Phys. Chem. Lett. 2019, 10, 4129−4135