

# Machine-Learning Coupled Cluster Properties through a Density Tensor Representation

Benjamin G. Peyton,<sup>†</sup> Connor Briggs,<sup>†</sup> Ruhee D'Cunha,<sup>†</sup> Johannes T. Margraf,<sup>‡</sup>  
and T. Daniel Crawford<sup>\*,†</sup>

<sup>†</sup>*Department of Chemistry, Virginia Tech, Blacksburg, VA 24061, USA*

<sup>‡</sup>*Chair for Theoretical Chemistry, Technische Universität München, Lichtenbergstrae 4,  
D-85747 Garching, Germany*

E-mail: [crawdad@vt.edu](mailto:crawdad@vt.edu)

# Abstract

The introduction of machine-learning (ML) algorithms to quantum mechanics enables rapid evaluation of otherwise intractable expressions at the cost of prior training on appropriate benchmarks. Many computational bottlenecks in the evaluation of accurate electronic structure theory could potentially benefit from the application of such models, from reducing the complexity of the underlying wave function parameter space to circumventing the complications of solving the electronic Schrödinger equation entirely. Applications of ML to electronic structure have thus far been focused on learning molecular properties (mainly the energy) from geometric representations. While this line of study has been quite successful, highly accurate models typically require a “big data” approach with thousands of training data points. Herein, we propose a general, systematically improvable scheme for wave function-based ML of arbitrary molecular properties, inspired by the underlying equations that govern the canonical approach to computing the properties. To this end, we combine the established ML machinery of the *t-amplitude tensor representation* with a new reduced density matrix representation. The resulting model provides quantitative accuracy in both the electronic energy and dipoles of small molecules using only a few dozen training points per system.

## 1 Introduction

Machine-learning (ML) has been applied to a number of chemical applications, with excellent (and current) surveys of the field published in 2019.<sup>1,2</sup> While applications in high-throughput chemical screening using computationally “free” descriptors of molecular systems (e.g. the molecular formula or graph) have been fruitful, recently much work has been focused on machine-learning quantum mechanics (MLQM). These methods generally apply machine-learning to either replace or bolster the often expensive algorithms used in quantum mechanics for solving the electronic Schrödinger equation. Schemes for ML-assisted models include

learning force-fields or other parameterizations from *ab initio* data,<sup>3,4</sup> improved guesses for e.g. important contributions to configuration-interaction space<sup>5</sup> or amplitudes in coupled cluster (CC) theory,<sup>6</sup> and direct energy or property prediction using mean-field, correlated, or density-functional-based approaches.<sup>7-11</sup>

The focus of this work will be on the bolstering of wave function-based methods, such as CC theory<sup>12</sup> and many-body perturbation theory (MBPT),<sup>13,14</sup> for predicting energies and properties across potential energy surfaces or molecular dynamics (MD) trajectories. This is an application also covered by recent work facilitating ML for forces in *ab initio* MD,<sup>15</sup> fast numerical gradients for geometry optimization,<sup>16</sup> and general-purpose local PESs.<sup>17</sup> Note, however, that these methodologies may be extended to other applications, such as learning of the wave function itself and predicting properties across large chemical datasets as described in Refs. 18 and 19, respectively.

There has also been some recent interest in alternative, wave function-based representations (as opposed to the more common representations based on the molecular structure), which in principle have several advantages. In particular, we define three desirable properties: first, representations derived from increasingly accurate electronic structure methods should provide more faithful representations of the wave function, resulting in more accurate and efficient ML. This rules out features that are not systematically improvable, such as the Hartree-Fock wave function parameters used by Miller and co.<sup>7,8</sup> Second, the representation should be of reasonable size to be stored for thousands of molecules, with the possibility of reducing their size for extremely large datasets. This rules out a real space representation of the electron density on a grid.<sup>20</sup> Third, a solid theoretical motivation for the representation should be given, with simple mappings from representation to target preferred. This third point is critical, as we will show that the relationship between the electronic wave function (specifically, the corresponding density matrix) and properties expressed as expectation values of one- and two-electron operators suggests a clear recipe for a wave function representation which is appropriate for machine-learning any molecular response property.

One such wave function-based representation is the t-amplitude tensor representation (TATR) introduced by Margraf and Reuter.<sup>21</sup> This representation is based on CC or MP2 amplitudes, paired with the many-body tensor representation (MBTR) formalism.<sup>22</sup> Using the TATR, it was shown that the CC correlation energies of a series of diatomics can be predicted with chemical accuracy ( $\sim 1$  kcal/mol) across the full potential energy surface with only 12 training points. This representation adheres to the three criteria above: amplitudes from increasingly accurate electronic structure methods should provide increasingly accurate TATRs, these amplitudes can be cut-off and stored based solely on their magnitude (as evidenced by retaining only 150 amplitudes for the diatomic representations in Ref. 21), and the relationship between wave function amplitudes and the target function (the correlation energy) is well-defined.

Indeed, the electronic wave function, in principle, gives access to any molecular property. The derivation of arbitrary properties in terms of wave function amplitudes is far from straightforward, however. In order to generalize the mapping from representation to target function (i.e. arbitrary molecular properties), we propose the density tensor representation (DTR), an analogous extension of the MBTR using one- and two-particle reduced density matrices rather than wave function amplitudes. In the following sections we demonstrate that careful choice of the underlying raw wave function features, as laid out in section 2.1, results in a more compact representation with a simple theoretical motivation. A proof-of principle ML model using this representation is detailed in section 2.2. Results from this initial implementation for electronic energies and dipole moments of several small molecules are given in sections 4.1 - 4.3, and details of their possible extensions for molecular transferability and arbitrary property calculations given in section 4.4.

## 2 Theory

### 2.1 Electronic Structure Theory

The time-independent electronic Schrödinger equation (and, therefore, the energy) can be written generally in terms of one- and two-particle reduced density matrices (1-RDM  $D$  and 2-RDM  $\Gamma$ ):<sup>23</sup>

$$\begin{aligned} E &= \langle \Psi | \hat{H} | \Psi \rangle \\ &= D_{pq} h_{pq} + \Gamma_{pqrs} g_{pqrs} \end{aligned} \quad (1)$$

where implicit sums in Einstein summation notation are included over arbitrary orbitals  $p, q, r, s$  for the one- and two-electron integrals of the electronic Hamiltonian,  $h$  and  $g$ .  $D$  and  $\Gamma$  are defined according to the wave function ansatz

$$D_{pq} = \langle \Psi | a_p^\dagger a_q | \Psi \rangle \quad (2a)$$

$$\Gamma_{pqrs} = \langle \Psi | a_p^\dagger a_q^\dagger a_s a_r | \Psi \rangle \quad (2b)$$

where the excitation operators of the second-quantized electronic Hamiltonian are included in the reduced densities. As molecular properties can be expressed in terms of energy derivatives, they can also be expressed in terms of these reduced density matrices. Differentiating Eq. (1) with respect to an arbitrary parameter  $\Omega$ , noting that  $|\Psi\rangle$  carries no dependence on  $\Omega$  (assuming the Hellmann-Feynman theorem<sup>24,25</sup> holds),

$$\begin{aligned} \frac{\partial E}{\partial \Omega} &= \langle \Psi | \frac{\partial \hat{H}}{\partial \Omega} | \Psi \rangle \\ &= D_{pq} \frac{\partial h_{pq}}{\partial \Omega} + \Gamma_{pqrs} \frac{\partial g_{pqrs}}{\partial \Omega} \end{aligned} \quad (3)$$

thus first-order molecular properties can also be described by the RDM's without their differentiation. This can be said even for time-dependent properties, in which a time-dependent Hellmann-Feynman theorem applies.<sup>26</sup> Furthermore, one-electron properties (such as the

electronic dipole moment) can be expressed solely in terms of the 1-RDM and the relevant property integrals,

$$\Omega = D_{pq} \langle p | \hat{\Omega} | q \rangle. \quad (4)$$

The 1-RDM is a matrix with as many elements as the number of basis functions squared, and is diagonally dominant for correlated methods in the molecular orbital (MO) basis. Matrices of this size are amenable to large-scale storage, but still hold a wealth of relevant wave function information which can be improved by simply improving the correlated method used to generate them. Additionally, the additivity of densities dictates an additivity of the property contributions, e.g. for the total CC dipole moment in one direction  $\alpha$ :

$$\mu_\alpha = (D_{pq}^{SCF} + D_{pq}^{CC}) \langle p | \hat{\mu}_\alpha | q \rangle + \mu_\alpha^{nuc} \quad (5)$$

where we have separated the contributions of the uncorrelated (e.g. self-consistent field (SCF)), correlated (e.g. CC theory), and the nuclear ( $\mu_\alpha^{nuc}$ ) components. This separability of uncorrelated, correlated, and nuclear contributions to the property is particularly desirable for machine-learning, as we minimize the contributions learned by the algorithm. In order to probe the efficacy of using RDMs as “raw features” in a machine-learning context, we will outline and employ a method used previously by Margraf and Reuter<sup>21</sup> for using wave function amplitudes as wave function features.

## 2.2 Machine-Learning

### 2.2.1 Kernel Ridge Regression

We will focus on Kernel Ridge Regression (KRR), as a prototypical ML method.<sup>27,28</sup> In KRR, a target property,  $y$ , is approximated as:

$$y(\boldsymbol{\nu}) = \sum_i^M k(\boldsymbol{\nu}'_i, \boldsymbol{\nu}) \alpha_i. \quad (6)$$

Here,  $\boldsymbol{\nu}$  is the representation of the system of interest,  $k(\boldsymbol{\nu}'_i, \boldsymbol{\nu})$  is a kernel function which measures the similarity between two systems (with representations  $\boldsymbol{\nu}$  and  $\boldsymbol{\nu}'$ ), and  $\alpha_i$  is a regression coefficient corresponding to the training point with the representation  $\boldsymbol{\nu}'_i$ . The sum runs over  $M$  training points, so that the target property can be computed as a dot-product between a kernel vector  $\mathbf{k}$  and a coefficient vector  $\boldsymbol{\alpha}$ .

An advantageous feature of KRR is that the coefficients  $\boldsymbol{\alpha}$  can be obtained from a closed-form linear algebra expression:

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}. \tag{7}$$

Here, the regularization hyperparameter  $\lambda$  is used to prevent overfitting.  $\mathbf{K}$  is the Kernel matrix with elements containing the kernel function of all pairs of training points. Herein, we use the radial basis function (or Gaussian) kernel to measure the similarity of representations:

$$k(\boldsymbol{\nu}', \boldsymbol{\nu}) = \exp\left(-\frac{\|\boldsymbol{\nu}' - \boldsymbol{\nu}\|^2}{2\sigma_m}\right) \tag{8}$$

where  $\sigma_m$  is another hyperparameter of the model which adds flexibility to how the kernel measures system similarity. Both hyperparameters are optimized empirically to give the best performance in leave-one-out cross-validation.

Overall, it should be emphasized that the KRR model bears clear similarities to the equations we wish to approximate— Eqs. (1) and (3) — which contain a density-related term multiplied by an additional, property-specific term.

### 2.2.2 Representation

The representations used herein follow the original idea of the many-body tensor representation (MBTR) of Rupp,<sup>22</sup> a geometrical representation proposed as an extension to the Coulomb matrix.<sup>29</sup> Margraf and Reuter recently proposed a modification of the MBTR that uses wave function amplitudes from Møller-Plesset perturbation theory (MP2) or CC theory instead of geometrical features, termed the t-Amplitude Tensor Representation (TATR).<sup>21</sup>

Simply put, these representations can be understood as discretized, broadened histograms. This is achieved via a sum of Gaussian functions (with some fixed width  $\sigma$ ), centered on the raw features  $t_i$  (i.e. wave function amplitudes). This function is evaluated for a discretized range of values  $x \in \chi$ :

$$\nu(x) = \sum_{i < N} g(t_i, \sigma, x) \tag{9}$$

where the sum runs over the  $N$  raw features. The representation is thus an  $l$ -dimensional vector  $\nu$ , where  $\chi$  is discretized into  $l$  points with even step-size  $\Delta x$ . Thus the  $N$  raw features are represented in a vector form which now contains  $l$  new features. Though it could in theory be optimized, we will use  $\sigma = 0.05$  for consistency throughout this work for simplicity. (This value was empirically determined to be optimal for diatomics in Ref. 21).

In Ref. 21 the highest  $N$  amplitudes (by magnitude) were used for each excitation level  $m$  to give a TATR of fixed length  $l$  times the number of excitation levels considered. This number of amplitudes  $N$  could be chosen in a number of ways, but Margraf and Reuter chose the highest 150 for simplicity in their paper. This proved sufficient for diatomics in a reasonable basis set, but is not a universally applicable choice for larger cases (see section 4.1). The discretization range  $\chi$  was also fixed as  $[-1:1]$  with  $l = 150$  and  $\Delta x = \frac{2}{150}$ , which we will also use for comparison.

The main novelty of the current work is to use density matrix elements in place of  $t_i$ . This “density tensor representation” (DTR) has several advantages. For simplicity we will focus on the 1-RDM — while this is sufficient for one-electron properties, the information stored in the 1-RDM is also necessary (and in some cases, as discussed in section 3, perhaps sufficient) for describing the total wave function. The 1-RDM will thus provide much of the wave function character needed to describe the correlation energy. This is the first advantage of using a density matrix based representation: one-electron properties are exactly described by a product of the 1-RDM with property integrals, as seen in Eq. (4). In analogy, Eq. (6) describes the target function as the product of a (now density-dependent) kernel and the regression coefficients. These coefficients thus describe the property integrals and



all approximations in the model (e.g. correlation beyond the level of the density matrix used to build the representation). Thus, a simple mapping from representation to target function is achieved in strict analogy to theory. The 1-RDM is diagonally dominant in the MO basis, meaning the number of significant ( $> 10^{-8}$ ) elements is scarcely larger than the number of basis functions. This is a further advantage of using 1-RDM elements instead of t-amplitudes: even considering the entire matrix, there will never be more elements than the number of basis functions squared, far fewer than the doubles amplitudes of MP2 or CC theory. These few elements are amenable to large-scale storage, and the number of elements retained could even be reduced in the case of large systems. Finally, the density matrix can be easily defined for most electronic structure methods, unlike other descriptors such as the Hartree-Fock-level descriptors (Fock, Coulomb, and exchange matrix elements) employed in Refs. 7 and 8 or wave function amplitudes which are only available for some correlated levels of theory (and depend on the level of “excitation” considered). The DTR is thus systematically improvable, by building it from improved wave functions. Importantly, TATR and DTR based ML models will always use two levels of theory: a lower level method (e.g. MP2) to build the representation and a higher level method (e.g. CCSD) to provide the training data.

### 2.2.3 Algorithm

Rather than spanning chemical compound space, wave function representations span the “wave function space” — that is, the space of possible wave function parameters — of the systems we wish to describe. The space covered by a KRR model is defined by its training set, meaning that the model will be unsuitable for predicting properties of systems that are very different from the training set. Since the high-level *ab initio* calculations for the construction of the training set are by far the most computationally demanding part of the process, we want to ensure that the training set efficiently spans the wave function space of interest in as few points as possible.

To this end, we apply the k-means clustering algorithm as implemented in the Scikit-Learn ML package<sup>30</sup> to an initial “grand set” of relevant structures. This grand set could, e.g., consist of all grid points that will be computed for a final PES of interest or to a representative MD trajectory. k-means is a multi-dimensional partitioning scheme—  $N$  initial points are partitioned into  $M$  clusters by taking the norm difference of the representations and “binning” them based on randomly selected initial cluster centroids. The cluster centers are then updated and the procedure is iterated to self-consistency. Due to the stochastic nature of choosing an initial guess, the algorithm is run repeatedly (30 times for all datasets in this manuscript) and the “best” set of clusters is chosen (as determined by minimizing the standard deviation of the distance between each point and its cluster centroid).

This approach requires that representations (TATRs, DTRs, *etc.*) for each point in the “grand set” of  $N$  structures be calculated first (at the low level of theory). To ensure that the model is truly predictive (i.e. able to reliably predict properties of systems for which no information known during training), a test set of systems is held back from the grand dataset prior to clustering. For the  $M$  representative data points chosen in this manner (i.e., the training set), the target function (energy, dipole, *etc.*) is computed with the high-level method. Once this is complete, the model is trained and subsequent predictions at approximately high-level accuracy and low-level cost are possible.

It should be noted here that the steps outlined in section 2.2 are completely general – no assumption was made about the level of theory besides that a density matrix can be formed. Furthermore, while KRR and k-means are used herein, alternative clustering schemes, hyperparameter optimization routines, and regression techniques can be substituted. The optimization of these choices will be the subject of a future study.

For benchmarking purposes, wave function parameters and high-level target values were computed for all datasets. The performance of a model is generally measured by the mean absolute error (MAE) of the predicted test set; however, the issues of truncation, separability, and learning rate are also considered.

### 3 Computational Details

Second-order MBPT using a restricted closed-shell Hartree Fock reference wave function (i.e. Møller-Plesset perturbation theory, MP2),<sup>31,32</sup> was used to generate both TATRs and DTRs for all systems. The TATR is generated from Eq. (9) using the highest 150 (by magnitude) amplitudes of double “excitations”  $t_{ij}^{ab}$ . For DTRs, the elements of the MP2 1-RDM are used. This choice is advantageous not just for comparison to Ref. 21, but also due to the special form of the MP2 reduced density matrices:<sup>33</sup>

$$D_{pq} = \langle 0 | \hat{T}_2^\dagger a_p^\dagger a_q \hat{T}_2 | 0 \rangle \quad (10a)$$

$$\Gamma_{pqrs} = 2 \langle 0 | a_p^\dagger a_q^\dagger a_s a_r \hat{T}_2 | 0 \rangle. \quad (10b)$$

where the “excitation” operators are defined as  $\hat{T}_n = \left(\frac{1}{n!}\right)^2 t_{ij\dots}^{ab\dots} a_a^\dagger a_b^\dagger \dots a_j a_i$ , and we have again implied Einstein summation notation this time over occupied  $i, j, \dots$  and virtual  $a, b, \dots$  orbital spaces. The form of Eqs. (10a) and (10b) indicates that the reduced density matrices are fully defined by the doubles amplitudes and products thereof. In this light, it is reasonable to suggest that all of the information necessary to reproduce the wave function is coded into either of these and, to reduce the size of the representation, the 1-RDM should suffice for describing the MP2-level correlated wave function. Thus, any correlated one-electron property at the MP2 level can be described by this matrix.

Coupled cluster with single and double excitations (CCSD) was used to generate the high-level reference data, i.e. correlated energies and electronic dipole moments, for all systems. This method represents a marked improvement over MP2 in systems where correlation is strong, such as stretched diatomics, without incurring the additional cost of the gold-standard CCSD with perturbative triples method, CCSD(T). However, the CCSD(T) method could be used to provide higher quality targets (e.g. energies and dipoles) with no additional modifications — as noted in section 2.2.3, the algorithm as presented is completely general.

All electronic structure calculations were performed with the Psi4<sup>34</sup> electronic structure package. Molecular symmetry was used to maximize TATR performance with as few amplitudes as possible (except in section 4.1). All data (energies, dipoles, amplitudes, and densities) were harvested in full machine precision through either the Psi4-JSON interface or the PsiAPI infrastructure, with the exception of wave function amplitudes at the MP2 level, which were printed in the output file at ten decimal places. The def2-TZVP<sup>35</sup> basis was employed for diatomic calculations for comparison with Ref. 21, and the aug-cc-pVDZ (aDZ) basis<sup>36,37</sup> was used for all remaining calculations. Density-fitted integrals were used in the construction of MP2 densities using the default auxiliary basis (the so-called “RI” basis sets for def2-TZVP<sup>38</sup> and aDZ,<sup>39</sup> obtained from the New Basis Set Exchange<sup>40</sup>).

As outlined above, the general procedure from Margraf and Reuter<sup>21</sup> was followed for machine-learning. The hyperparameters  $\sigma_m$  and  $\lambda$  were optimized on a uniform grid from  $10^{-8} : 10^8$  for all models unless otherwise specified, using the “negative-mean-squared” loss function. All regressions were performed using 20 training points, except in section 4.1 where 12 training points were used to emphasize the difference between the two strategies considered. k-means clustering was repeated 30 times to account for stochastic deviations in the algorithm. Test sets consisting of  $N/4$  points are held back before the clustering step. The radial basis function kernel (sometimes called the Gaussian kernel), Eq. (8), was used for all representations. Machine-learning algorithms were performed using Scikit-Learn (skl)<sup>30</sup> and the Machine-Learning Quantum Mechanics (MLQM) python package,<sup>41</sup> which generates a number of molecular representations, provides options for generating Psi4 input files and harvesting their results, and wraps some skl functions. Both of these codes are open-source and freely available on GitHub.

Geometries for diatomics were constructed by taking  $N = 200$  uniform increments from 0.5-2.0 Å. All other geometries were randomly selected from MD simulations ( $N = 150$ ) carried out in the Gromacs<sup>42</sup> software package. Simulations of (*S*)-methyloxirane and (*R*)-methylthiirane surrounded by a 5Å box of water molecules were executed using an all-atom

OPLS/AA forcefield<sup>43</sup> for the solute and the TIP3P model for water.<sup>44</sup> Each 5 ns trajectory was carried out in the NVT ensemble, with the solute and solvent coupled separately to a temperature bath at 300 K using a modified Berendsen thermostat and a coupling time of 0.1 ps. Geometries were selected from a set of 250 evenly-spaced snapshots along the trajectory.

Geometries, Psi4 input and output files, and their accompanying optimized (hyper)parameter values in JSON format can be found at the Virginia Tech Data Repository [data repository with detailed coordinate and other information to be added prior to publication] to permit reproducibility of our results. Scripts for harvesting and manipulating data using MLQM are also included in the form of Jupyter Notebooks.

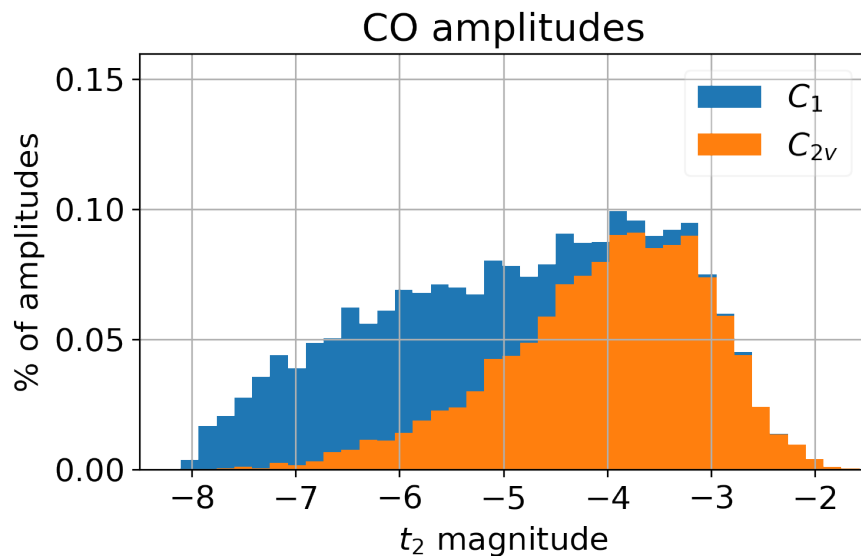
## 4 Results and Discussion

### 4.1 Truncation considerations

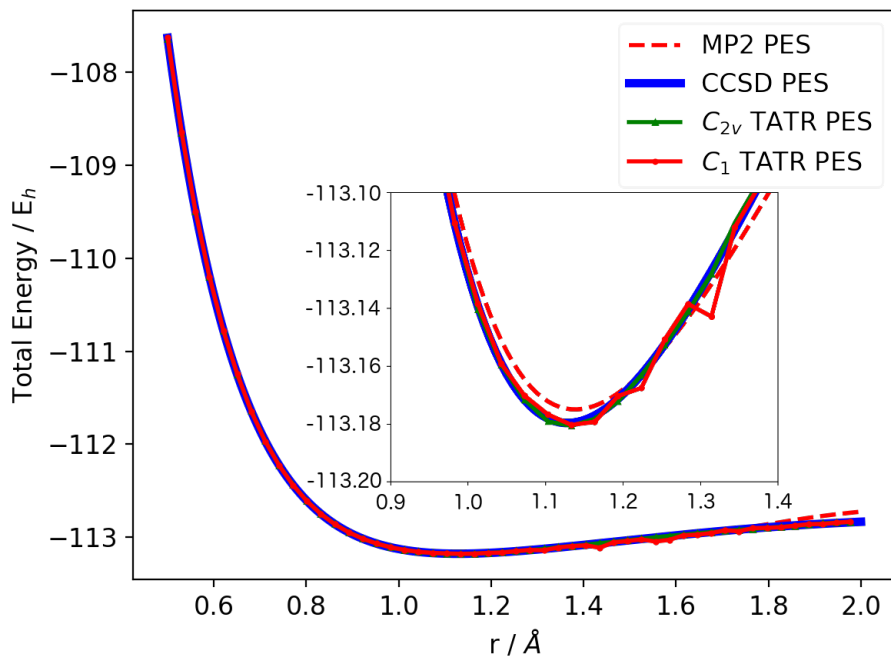
As stated previously, the “raw” wave function features used to build machine-learning representations must be carefully selected. To illustrate this point, the TATR was used to model a carbon monoxide potential energy surface (PES) using the highest 150 amplitudes (by magnitude) and  $M = 12$  training points as in Ref. 21. Figure 1b shows that, when molecular symmetry is considered ( $C_{2v}$ , the largest Abelian subgroup of the full  $C_{\infty v}$  point group), excellent regression results are produced using only 150 amplitudes.

However, to model the effect of decreased sparsity in the amplitudes on the representation, molecular symmetry was then dropped to  $C_1$ . When molecular symmetry is ignored, the number of “significant” ( $> 10^{-8}$ ) amplitudes greatly increases, as shown in Figure 1a. In this case, taking only the highest 150 amplitudes no longer creates an effective representation of the wave function, resulting in poorer performance across the PES, also shown in Figure 1b.

This example shows a considerable disadvantage of the TATR, namely that the T2-amplitude tensor is in general very large (equal to the number of occupied orbitals squared times the number of virtual orbitals squared). It is, of course, in principle possible to build



(a)



(b)

Figure 1: (a) Amplitude distributions and (b) resulting KRR for the carbon monoxide potential energy surface with and without molecular symmetry considered. TATR models used  $M = 12$  training points. All amplitudes with magnitude  $< 10^{-8}$  are set to 0.

the TATR with an arbitrary number of amplitudes, as the size of the representation is independent of the number of amplitudes. However, using all amplitudes is not advisable, as a large number of negligible amplitudes would lead to large (but chemically meaningless) TATR values around zero. Meanwhile, without extensive testing, it is in general unclear how many amplitudes to retain for an effective representation. For this reason, a large number of amplitudes must in general be stored, at least for initial testing and hyperparameter optimization. Here cutoffs based on the amplitude magnitude may be employed, but this can still lead to significant storage requirements for larger molecules and basis sets. Figure 1 also indicates that the TATR will likely not be applicable for larger systems that have no symmetry.

Storing the 1-RDM, on the other hand, only requires storing the number of basis functions squared floating point numbers. Furthermore, it is also known that the majority of significant elements in the MO-basis 1-RDM will be along the diagonal — with some number of off-diagonal contributions. This gives a useful rule-of-thumb for how it could be truncated in extreme cases. For the remainder of this manuscript, the TATR will be computed using the highest 150 amplitudes (by magnitude) computed in the highest symmetry available. The DTR will be computed using all 1-RDM elements.

## 4.2 Energies

In the spirit of comparing to previous results, the same diatomic PESs from Ref. 21 were computed using the DTR. The prediction errors on the test set are shown in the Supporting Information, along with TATR results for comparison. Overall, DTR results do not vary significantly from the  $C_{2v}$ -symmetry TATR results for these systems; however, when molecular symmetry is not considered, a far greater number of amplitudes are required in the TATR to achieve the same accuracy.

To test the applicability of the DTR and TATR models to larger systems which may not benefit from molecular symmetry, geometries for water, methanol, (*S*)-methyloxirane,

and (*R*)-methylthiirane near equilibrium were sampled randomly from a set of molecular dynamics (MD) trajectories and examined using both the TATR and DTR models. Correlation energy data are presented in Figure 2. The prediction errors on the test set are plotted against the true CCSD correlation energy, ordered by increasing energy to evaluate how the model error varies with respect to the magnitude of the correlation energy. Each point represents the prediction error of a single geometry along the MD trajectory. As with the diatomics, most predictions lie well within the bounds of “chemical accuracy” ( $2mE_h$ ). However, for the TATR, linear error centered around zero and which changes sign near the mean correlation energy value suggest that this model is biased towards the mean of the training set. While this still gives a rather agreeable mean-absolute-error across this particular test set, predictions for geometries far from the energetic minimum (i.e., sufficiently different from the training set) may not perform as admirably. This issue is explored further in section 4.4.3.

A summary of energy results is shown for both the TATR and DTR methods in Figure 3. The trained models reproduce the mean average correlation energy of the test set to within two milli-Hartree for every system considered. Furthermore, DTR errors are kept below 0.5 milli-Hartree for all systems except for CO and LiF. Inspection of their individual model performance (see the Supporting Information) reveals some difficulty in modeling the extreme ends of the PES. These examples also cover large ranges of correlation energies (100 and 60 milli-Hartree, respectively), which test both the regression capability and the validity of the underlying MP2 wave function used for the representation.

The energy data suggests that the improvements made by the DTR method result in greater accuracy in the molecular representation, in particular for larger molecules. As shown in Ref. 21, the electronic wave function can be well approximated using this relatively simple functional form when applied to energies. If this functional form is truly representative of the total wave function, rather than simply the parts which are important to the energy, then it should also be possible to compute molecular properties with a similar approach.



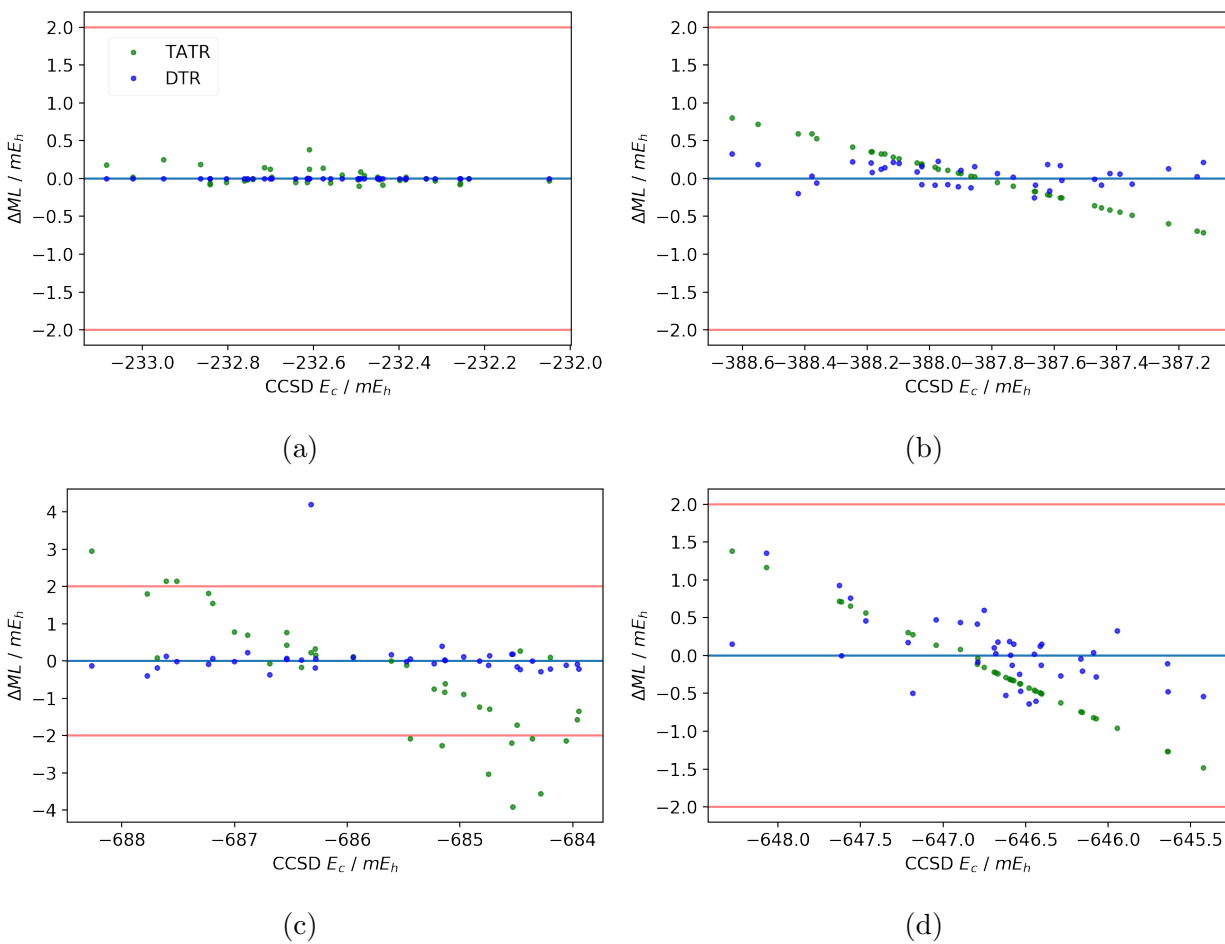


Figure 2: DTR vs TATR errors in  $mE_h$  for small molecule datasets: (a)  $H_2O$ , (b)  $CH_3OH$ , (c) (*S*)-methyloxirane, and (d) (*R*)-methylthiirane. Red lines indicate chemical accuracy ( $2 mE_h$ ).

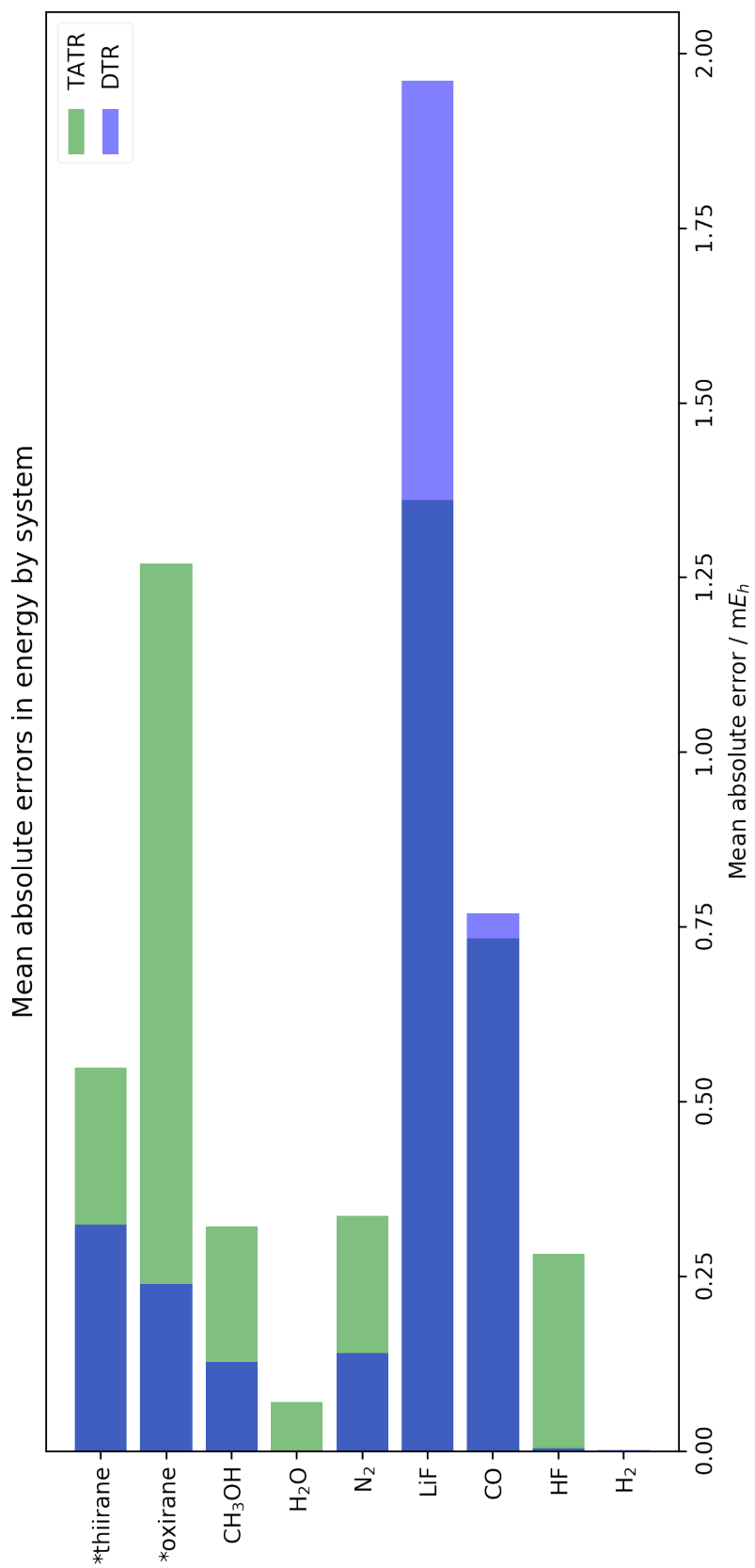


Figure 3: DTR vs TATR errors in  $mE_h$  for all datasets.

### 4.3 Dipole moments

Several schemes for computing dipole moments were employed to emphasize the importance of Eq. (5). Full density matrices ( $D^{SCF} + D^{MP2}$ ) were used as raw features to build DTRs for learning total ( $\mu^T = \mu^e + \mu^{nuc}$ ), electronic ( $\mu^e = \mu^{SCF} + \mu^{CC}$ ), and correlated ( $\mu^{CC}$ ) dipole moments. MAEs (in Debye) for the machine-learned dipole moments, using each of the three possible target values, are given in Table 1. The MAEs for the MP2-level correlated dipole (relative to the CCSD result) for the same predicted sets are given for comparison.

**Table 1: Mean absolute errors (in Debye) relative to CCSD of four machine-learning datasets utilizing full, electronic, and correlated dipole moments as training targets. MP2 dipoles given for comparison.**

Molecule	$\mu^T$	$\mu^e$	$\mu^{CC}$	MP2
H <sub>2</sub> O	$8.2 \times 10^{-4}$	$7.1 \times 10^{-4}$	$1.0 \times 10^{-4}$	$1.8 \times 10^{-2}$
CH <sub>3</sub> OH	$3.7 \times 10^{-2}$	$1.7 \times 10^{-1}$	$2.1 \times 10^{-3}$	$1.8 \times 10^{-2}$
*oxirane	$1.0 \times 10^{-1}$	$4.1 \times 10^{-1}$	$5.7 \times 10^{-3}$	$2.0 \times 10^{-2}$
*thiirane	$5.2 \times 10^{-2}$	$5.2 \times 10^{-1}$	$7.6 \times 10^{-3}$	$3.5 \times 10^{-2}$

As expected, minimizing the extrapolation necessary for the model reduces the error drastically. The error in the electronic component of the dipole moment in e.g. (*S*)-methyloxirane is reduced by an order of magnitude relative to using the full dipole moment. Furthermore, this error is also an order of magnitude lower than the MAE in the MP2 correlated dipole moments on the same predicted set. This is consistent with Ref. 21 where extremely accurate models for the *correlation* energy were built. The same is clearly the case for the (correlated) dipole moment.

A summary of dipole results is shown for both the TATR and DTR methods in Figure 4. The same datasets as for correlation energy learning were considered, except for the homonuclear diatomics. As with the energy, the trained models reproduce the correlated dipole moment of the test sets to reasonable precision. Once again, the challenging cases of CO and LiF produce the maximum errors; however, it is encouraging that even here, errors for states near equilibrium (those modeled by molecular dynamics) are of milliDebye magnitude.

Indeed, it has been shown (e.g. in Ref. 19) that predicting dipole moments with ML is significantly more challenging than predicting energies. Still, our results support the notion that, as long as the reference wave function is sound, the molecular representation will “correctly represent the physics of the problem”<sup>21</sup> as desired with a favorably small number of training points. While these results represent only a marginal improvement over the already sound MP2 approximation, they demonstrate that simple ML models effectively capture the characteristics of the wave function without the need for a complex functional form. They also show that this generalizes beyond the energy to molecular properties as well.

## 4.4 Extensions

The present model can be used to solve a number of potential problems, mostly involving local PESs or dipole moment surfaces of small molecules which are amenable to MP2-level calculations (and a few CCSD training calculations). This is especially useful for generating accurate force field parameters<sup>3,4</sup> and evaluating numerical gradients.<sup>16</sup> Furthermore, the DTR representation has been shown to provide high accuracy with very few training points, resulting in a highly data-efficient method. While this covers a significant space of applications, there are some areas which cannot currently be addressed, but could be well within reach of the proposed algorithm.

### 4.4.1 Transferability

The DTR-based machine-learning model as described above shows that a mapping from the MP2 1-RDM to CCSD electronic properties is feasible using a simple ML model. This allows avoiding the high-cost evaluation of the CC wave function for multiple geometries of the same molecule. However, if the method is to be generalized to different molecular systems (rather than just different geometries as we’ve shown here), additional modifications must be considered. Specifically, two (related) types of transferability are desirable: on one hand,

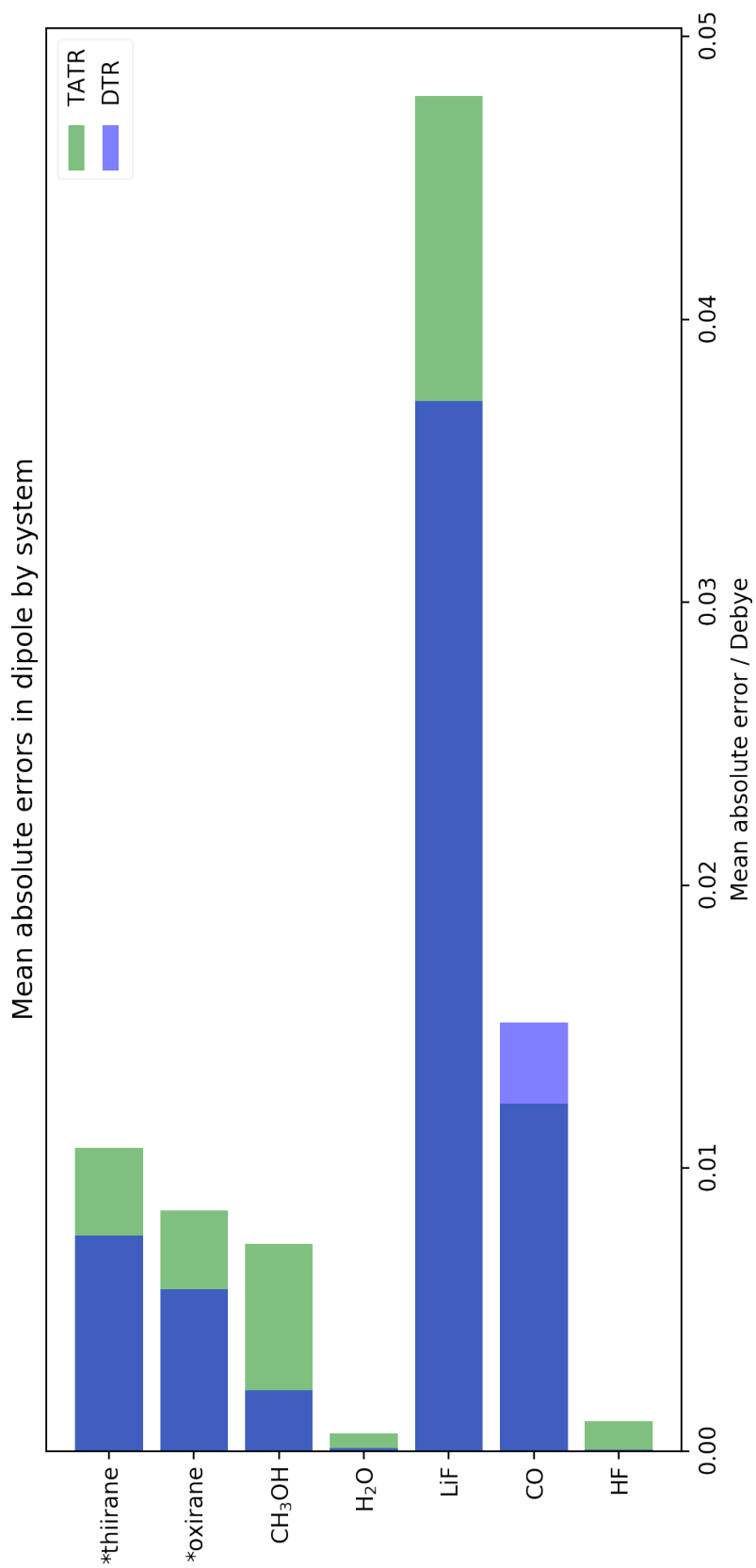


Figure 4: DTR vs TATR errors in *milliDebye* for all datasets.

a model can be transferable between different chemical compositions (i.e. trained on water and methane, applied to methanol).<sup>7</sup> On the other hand, a model can be transferable among different system sizes (i.e. trained on water clusters and applied to bulk water).<sup>45</sup>

The main impediment for both the TATR and the DTR to be transferable is the use of a molecular orbital basis, which is strongly system dependent. An obvious route towards more transferability would be to use the atomic orbital (AO) basis instead. In this representation, the molecular dependence of the representation would stem only from the identity of the basis functions placed on each atom. While this still precludes transferability between different elements, this is not very problematic in practice, as one can always train on the elements of interest. Regarding transferability to different sizes, this essentially boils down to the requirement that the ML model be size extensive. The lack of extensivity is a common weakness of ML models that use global representations such as the DTR or TATR. As recently discussed by Jung et al., global KRR models can be made size extensive, however, if the Kernel function is properly normalized.<sup>45</sup> While that study focused on representations of the molecular geometry, it would certainly be worthwhile exploring how extensive models based on wave function representations would perform.

Finally, a theoretical consideration can also be made with respect to transferability. As mentioned above, one-electron properties can be described by the general form of Eq. (4). Comparing this to the KRR equations (see Eq. (6)), and considering that the kernel only includes information from the 1-RDM, it can be inferred that the coefficients  $\alpha$  are modelling the contributions arising from the property integrals,  $\langle p|\Omega|q\rangle$ . In the present model, however, the coefficients are also doing the job of mapping the low-level (MP2) representation onto the high-level (CCSD) target function. To truly separate the molecule-specific terms from the correlation contributions, the property integrals (in a transferable basis, such as the AO basis) should thus also be included in the representation of the molecule. Modifications such as using product kernels to fold in the contribution from the integrals can be explored to correct this shortcoming.

#### 4.4.2 Additional properties

Properties beyond dipole moments are also accessible, but not without modifications. While first-order one-electron properties are available through Eq. (4), considerations of both the one- and two-electron densities is generally necessary for two-electron properties. Fortunately, we have seen that the 1-RDM suffices to predict at least one property which depends on both the 1- and 2-RDM, namely the energy. This may be related to the special formulation of 1- and 2-RDMs in MP2, as discussed above. It could also be a particular property of the ground-state energy which, according to the Hohenberg and Kohn theorem,<sup>46</sup> formally only depends on the ground state density (and thus the 1-RDM). However, this is not generally the case, for example in CC response theory. Higher-order properties in CC response depend on both the 1- and 2-RDMs. Extensions to include both density matrices could be made, as with a singles-and-doubles TATR, though considerable truncation procedures would need to be implemented in order to keep storage costs reasonable across realistic datasets.

Alternatively, second-order properties which depend only on one-electron operators, such as polarizabilities and chiroptical response, could also be considered using a reformulated CC linear response (CC-LR). It can be shown that, similarly to first-order properties, second-order response properties of this type can be expressed using the 1-RDM only. This is achieved by computing the first-order perturbed left-hand wave function parameters,  $\lambda^{(1)}$ . Using these, second-order properties are computed in CC-LR as products of property integrals with *perturbed* one-electron reduced density matrices.<sup>47</sup> Contributions from one operator are folded into the 1-RDM, and contributions from the other are described by the property integrals. In the context of the DTR, the density representation would now require only this perturbed 1-RDM with no additional changes. Contributions from property integrals may also be included in the representation, perhaps through the same sort of generalizations made to allow for transferability. In this fashion, many linear response properties could be made accessible.

We will acknowledge one final potential oversight in property predictions done in the

preceding manner: a new model must still be trained for each property. To allow for a “one-shot” model prediction of arbitrary properties, the wave function itself (or some usable representation thereof) must be learned. This has recently been explored by Schütt and coworkers,<sup>18</sup> which has proved some efficacy of this strategy. Additional work on the usability of such representations will be considered while algorithms to learn the representation itself are formulated. In practice, however, once the ground state equations (e.g. the amplitudes in CC theory) are solved, it is not necessary to recompute them for additional properties. Instead, the (linear) response function can be evaluated at this time for a multitude of properties. Because of this, we suggest that many properties are computed and stored in a database when proposing and gathering reference or training data for ML algorithms, such as those provided in QCArchive.<sup>48</sup>

#### 4.4.3 ML algorithm optimization

Aside from coding additional physics into the representation and returning more varied properties, the ML workflow itself may also be optimized to be more data-efficient, accurate, and precise across varied molecular representations. To further evaluate model performance and perhaps diagnose potential problems, learning curves were generated for every system considered. These curves plot the error in the test (the “validation curve”) and training (“training curve”) sets with respect to the number of training points used, with values ranging from two to one-hundred training points. The validation curve exposes the limitations of the current model by testing whether the model can be made more accurate using more training points. The training curve, on the other hand, describes how well the model is adhering to the training set, and so reveals problems of under- or over-training. Together, the curves indicate the practical limitations of the model’s accuracy and use of data, suggesting an irreducible error that can only be addressed by changing the model or the representation itself. Learning curves for all systems considered are available in the Supporting Information.

The data discussed within the manuscript so far (excluding section 4.1) has been col-



lected using  $M = 20$  training points, which result in the data points at that value in every validation curve. In a log-log plot, the validation curve should be linear such that an increase in the number of training points results in a commensurate decrease in error. This is well exhibited by, e.g., the validation curve for predicting the correlation energy for (*S*)-methyloxirane, Figure 5. Generally, the DTR validation curve starts lower than the TATR. The DTR model also tends to benefit more from additional training points in the two- to fifty-point range. It is important to note that the most significant outliers to this trend are the dipole moment models for (*S*)-methyloxirane and (*R*)-methylthiirane (see SI), the largest molecules considered and undoubtedly the most complex electronic environments. These systems benefit from larger training sets and are more prone to fortuitously low errors with small training sets, requiring a careful analysis of the validation curve before deciding on an optimal training set size. This is, of course, subject to the constraint that the training points are still taken from near-equilibrium values. Improving the quality of the training set by including more distinct electronic states away from equilibrium may improve performance.

In considering the training curves (see SI), it is revealed that the TATR suffers from additional bias as compared to the DTR, as evidenced by the higher error of the training curves. Bias (sometimes referred to as *inductive bias*)<sup>28</sup> refers to error arising from simplifying assumptions made in the model. As this bias causes high training errors, it suggests that the TATR model is not fitting the training data as well as the DTR. This is likely the cause of the linear errors discussed in section 4.2, as the model is simply learning toward the mean of the training set, rather than properly reproducing each point. These errors can be reduced by improving the model, e.g. by including more amplitudes in the TATR or by the DTR modifications described above. Additional features or a reformulation of the representation may also be explored using automated model-selection routines. These routines would, in theory, result in a more complex model accessing a higher dimensional feature space, so the challenges of extension become ensuring that this model still fits the training data (does not

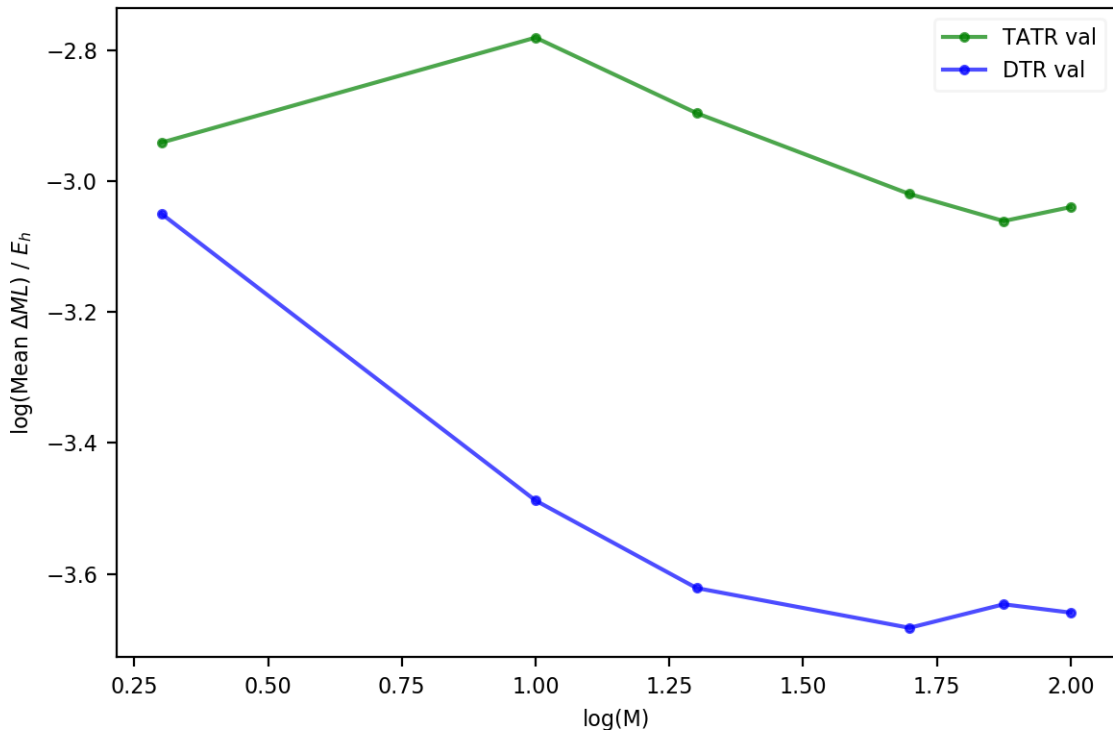


Figure 5: DTR and TATR validation curves for (*S*)-methyloxirane correlation energy.

under-train as a result of bias) and generalizes to unseen data (does not over-train as a result of variance) when compared to a simpler approach.

## 5 Conclusions

Here we introduce the density tensor representation (DTR) for machine-learning quantum mechanics applications. The representation is based on the previous t-amplitude tensor representation (TATR), with improvements made through strictly theoretical considerations of three categories: systematic improvement, storage, and simplified representation-target mapping. Investigating the limits of these categories on small test sets show a number of favorable properties. The DTR is defined for any electronic structure method in which a density can be defined. When compared to the TATR for MP2, it achieves superior accuracy across most test cases when the MP2 wave function is expected to produce reasonable results.

This accuracy is in the sub- $mE_h$  range for correlation energies. Furthermore, applications to molecular properties are both theoretically and operationally justified for representations utilizing electronic densities as raw wave function features. Roughly milliDebye error was achieved for correlated electronic dipole moments of several small molecules near equilibrium. Extensions to include additional properties and molecular transferability are also considered, with the data-efficient DTR model providing a vital stepping stone to these generalizing improvements.

## 6 Supporting Information

A document containing supporting information (`si.pdf`) is available. This contains data not given but referenced in the manuscript (diatomic energy TATR and DTR errors, dipole TATR and DTR errors for all datasets, learning curves for all datasets). All pertinent ML files, including output files and trained model parameters for all datasets, are available through the Virginia Tech Data Repository [data repository with detailed coordinate and other information to be added prior to publication].

## 7 Acknowledgements

This work was supported by the US Department of Energy (DOE) Office of Science, Office of Basic Energy Sciences, Computational Chemical Sciences (CCS) Research Program under work proposal number AL-18-380-057. The authors gratefully acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within the paper.

## References

- (1) Haghghatlari, M.; Hachmann, J. Advances of machine learning in molecular modeling and simulation. *Curr. Opin. Chem. Eng.* **2019**, *23*, 51–57.
- (2) Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep learning for molecular design - A review of the state of the art. *Mol. Syst. Des. Eng.* **2019**, *4*, 828–849.
- (3) Saucedo, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K. R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, 114102.
- (4) Galvelis, R.; Doerr, S.; Damas, J. M.; Harvey, M. J.; De Fabritiis, G. A Scalable Molecular Force Field Parameterization Method Based on Density Functional Theory and Quantum-Level Machine Learning. *J. Chem. Inf. Model.* **2019**, *59*, 3485–3493.
- (5) Coe, J. P. Machine Learning Configuration Interaction. *J. Chem. Theory Comput.* **2018**, *14*, 5739–5749.
- (6) Townsend, J.; Vogiatzis, K. D. Data-Driven Acceleration of the Coupled-Cluster Singles and Doubles Iterative Solver. *J. Phys. Chem. Lett.* **2019**, *10*, 4129–4135.
- (7) 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*, 4772–4779.
- (8) Cheng, L.; Welborn, M.; Christensen, A. S.; Miller, T. F. A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules. *J. Chem. Phys.* **2019**, *150*, 131103.
- (9) Yuan, S.; Jiao, Z.; Quddus, N.; Kwon, J. S. I.; Mashuga, C. V. Developing Quantita-

- tive Structure-Property Relationship Models to Predict the Upper Flammability Limit Using Machine Learning. *Ind. Eng. Chem. Res.* **2019**, *58*, 3531–3537.
- (10) Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A.; Ceriotti, M. Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 3401–3406.
- (11) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 2903.
- (12) Crawford, T.; Schaefer III, H. An introduction to coupled cluster theory for computational chemists. *Rev. Comput. Chem. Vol. 14* **2000**, *14*, 33–136.
- (13) Coester, F.; Kümmel, H. Short-range correlations in nuclear wave functions. *Nucl. Phys.* **1960**, *17*, 477–485.
- (14) Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.
- (15) Chmiela, S.; Sauceda, H. E.; Müller, K. R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.
- (16) Schmitz, G.; Christiansen, O. Gaussian process regression to accelerate geometry optimizations relying on numerical differentiation. *J. Chem. Phys.* **2018**, *148*, 241704.
- (17) Abbott, A. S.; Turney, J. M.; Zhang, B.; Smith, D. G.; Altarawy, D.; Schaefer, H. F. PES-Learn: An Open-Source Software Package for the Automated Generation of Machine Learning Models of Molecular Potential Energy Surfaces. *J. Chem. Theory Comput.* **2019**, *15*, 4386–4398.

- (18) Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K. R.; Maurer, R. J. Unifying machine learning and quantum chemistry – a deep neural network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.
- (19) Christensen, A. S.; Faber, F. A.; Von Lilienfeld, O. A. Operators in quantum machine learning: Response properties in chemical space. *J. Chem. Phys.* **2019**, *150*, 64105.
- (20) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K. R.; Burke, K. Finding density functionals with machine learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.
- (21) Margraf, J. T.; Reuter, K. Making the Coupled Cluster Correlation Energy Machine-Learnable. *J. Phys. Chem. A* **2018**, *122*, 6343–6348.
- (22) Huo, H.; Rupp, M. Unified Representation of Molecules and Crystals for Machine Learning. 2018; arXiv:1704.06439v3.
- (23) Harris, F. E.; Monkhorst, H. J.; Freeman, D. L. *Algebraic and diagrammatic methods in many-fermion theory*; Oxford University Press, 1992.
- (24) Hellmann, H. *Einführung in die Quantenchemie*; Deuticke: Leipzig, 1937.
- (25) Feynman, R. P. Forces in molecules. *Phys. Rev.* **1939**, *56*, 340–343.
- (26) Norman, P. A perspective on nonresonant and resonant electronic response theory for time-dependent molecular properties. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20519.
- (27) Murphy, K. P. *Machine Learning: A Probabilistic Perspective*; The MIT Press, 2012.
- (28) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; The MIT Press, 2006.
- (29) 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*, 058301.

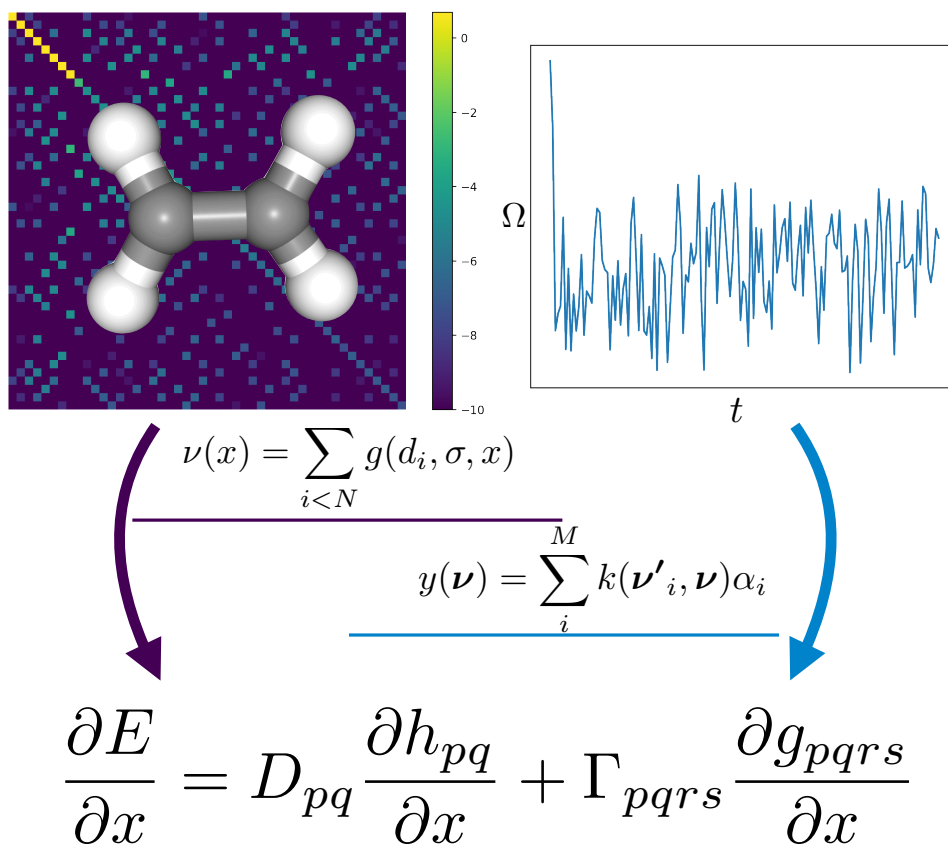
- (30) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (31) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (32) Bartlett, R. J.; Silver, D. M. Pair-correlation energies in sodium hydride with many-body perturbation theory. *Phys. Rev. A* **1974**, *10*, 1927–1931.
- (33) Trucks, G. W.; Salter, E. A.; Sosa, C.; Bartlett, R. J. Theory and implementation of the MBPT density matrix. An application to one-electron properties. *Chem. Phys. Lett.* **1988**, *147*, 359–366.
- (34) Parrish, R. M.; Burns, L. A.; Smith, D. G.; 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*, 3185–3197.
- (35) 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*, 3297–3305.
- (36) 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*, 1007–1023.
- (37) Woon, D. E.; Dunning, T. H. J. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.

- (38) Hättig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple- $\zeta$  basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.
- (39) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- (40) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820.
- (41) Machine-Learning Quantum Mechanics. [github.com/CrawfordGroup/MLQM](https://github.com/CrawfordGroup/MLQM).
- (42) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D. et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.
- (43) 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*, 11225–11236.
- (44) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (45) Jung, H.; Stocker, S.; Kunkel, C.; Oberhofer, H.; Han, B.; Reuter, K.; Margraf, J. T. SizeExtensive Molecular Machine Learning with Global Representations. *ChemSystemsChem* **2020**, *2*, e1900052.
- (46) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.



- (47) Crawford, T. D.; Kumar, A.; Bazanté, A. P.; Di Remigio, R. Reduced-scaling coupled cluster response theory: Challenges and opportunities. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, *9*, 1–25.
- (48) Smith, D. G. A.; Altarawy, D.; Burns, L. A.; Welborn, M.; Naden, L. N.; Ward, L.; Ellis, S.; Crawford, T. D. The MolSSI QCArchive Project: An open-source platform to compute, organize, and share quantum chemistry data. 2020; ChemRxiv:22566e14d96e43f7611a.

for Table of Contents use only



## Machine-Learning Coupled Cluster Properties through a Density Tensor Representation

Benjamin G. Peyton, Connor Briggs, Ruhee D'Cunha, Johannes T. Margraf, and T. Daniel Crawford