A Universal Density Matrix Functional from Molecular Orbital-Based Machine Learning: Transferability across Organic Molecules

Lixue Cheng,1 Matthew Welborn,1 Anders S. Christensen,2 and Thomas F. Miller III1*  
1) Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA  
2) Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials, Department of Chemistry, University of Basel, Basel, Switzerland

(Dated: 8 April 2019)

We address the degree to which machine learning can be used to accurately and transferably predict post-Hartree-Fock correlation energies. Refined strategies for feature design and selection are presented, and the molecular-orbital-based machine learning (MOB-ML) method is applied to several test systems. Strikingly, for the MP2, CCSD, and CCSD(T) levels of theory, it is shown that the thermally accessible (350 K) potential energy surface for a single water molecule can be described to within 1 millihartree using a model that is trained from only a single reference calculation at a randomized geometry. To explore the breadth of chemical diversity that can be described, MOB-ML is also applied to a new dataset of thermalized (350 K) geometries of 7211 organic models with up to seven heavy atoms. In comparison with the previously reported Δ-ML method, MOB-ML is shown to reach chemical accuracy with three-fold fewer training geometries. Finally, a transferability test in which models trained for seven-heavy-atom systems are used to predict energies for thirteen-heavy-atom systems reveals that MOB-ML reaches chemical accuracy with 36-fold fewer training calculations.

Machine learning (ML) has recently seen wide application in chemistry, including in the fields of drug discovery,13 materials design,14 and reaction prediction.15 In the context of quantum chemistry, much work has focused on predicting electronic energies or densities based on atom- or geometry-specific features,14,15 although other strategies have also been employed.16 Recently, we reported an accurate and transferable molecular-orbital-based machine learning (MOB-ML) approach to the prediction of correlated wavefunction energies based on input features from a self-consistent field calculation, such as the Hartree-Fock (HF) method.17 In this communication, we present refinements to the MOB-ML method with comparisons to test cases from our previous work. We then demonstrate the performance of MOB-ML across a broad swath of chemical space, as represented by the QM7b18 and GDB-1319 test sets of organic molecules.

I. THEORY

The current work aims to predict post-Hartree-Fock correlated wavefunction energies using features of the Hartree-Fock molecular orbitals (MOs). The starting point for the MOB-ML method is that the correlation energy can be decomposed into pairwise occupied MO contributions

$$E_c = \sum_{ij} \varepsilon_{ij}, \tag{1}$$

where the pair correlation energy $\varepsilon_{ij}$ can be written as a functional of the full set of MOs, $\{\phi_i\}$, appropriately indexed by $i$ and $j$

$$\varepsilon_{ij} = \mathcal{E}[\{\phi_i\}] \tag{2}$$

The functional $\mathcal{E}$ is universal across all chemical systems; for a given level of correlated wavefunction theory, there is a corresponding $\mathcal{E}$ that maps the HF MOs to the pair correlation energy, regardless of the molecular composition or geometry. Furthermore, $\mathcal{E}$ simultaneously describes the pair correlation energy for all pairs of occupied MOs (i.e., the functional form of $\mathcal{E}$ does not depend on $i$ and $j$). For example, in second-order Møller-Plesset perturbation theory (MP2),20 the pair correlation energies are

$$\varepsilon_{ij}^{\text{MP2}} = \frac{1}{4 \sum_{ab}} |\langle ij | ab \rangle|^2 (e_a + e_b - e_i - e_j) \tag{3}$$

where $a$ and $b$ index virtual MOs, $e_p$ is the Hartree-Fock orbital energy corresponding to MO $\phi_p$, and $|ij|ab$ are antisymmetrized electron repulsion integrals20 A corresponding expression for the pair correlation energy exists for any post-Hartree-Fock method, but it is typically costly to evaluate in closed form.

In MOB-ML, a machine learning model is constructed for the pair energy functional

$$\varepsilon_{ij} \approx \varepsilon_{ij}^{\text{ML}}[\mathbf{f}_{ij}] \tag{4}$$

where $\mathbf{f}_{ij}$ denotes a vector of features associated with MOs $i$ and $j$ Eq. 4 thus presents the opportunity for the machine learning of a universal density matrix functional for correlated wavefunction energies, which can be evaluated at the cost of the MO calculation.

Following our previous work,25 the features $\mathbf{f}_{ij}$ correspond to unique elements of the Fock ($\mathbf{F}$), Coulomb ($\mathbf{J}$), and exchange ($\mathbf{K}$) matrices between $\phi_i$, $\phi_j$, and the set of virtual orbitals. In the current work, we additionally include features associated with matrix elements between pairs of occupied orbitals for which one member of the pair differs from $\phi_i$ or $\phi_j$ (i.e., non-$i,j$ occupied MO pairs). The feature vector takes the
form

\[ f_{ij} = (F_{ii}, F_{ij}, F_{jj}, F_{i}^{\nu}, F_{j}^{\nu}, F_{jj}^{\nu}), \]
\[ J_{ii}, J_{jj}, J_{ij}^{\sigma}, J_{ij}^{\pi}, J_{jj}^{\pi}, J_{ij}^{\pi}, \]
\[ K_{ij}, K_{ii}^{\sigma}, K_{ii}^{\pi}, K_{ij}^{\sigma}, K_{ij}^{\pi}, K_{jj}^{\pi}). \]

where for a given matrix (F, J, or K) the superscript o denotes a row of its occupied–occupied block, the superscript v denotes a row of its occupied–virtual block, and the superscript vv denotes its virtual–virtual block. Redundant elements are removed, such that the virtual–virtual block is represented by its upper triangle and the diagonal elements of K (which are identical to those of J) are omitted. To increase transferability and accuracy, we choose \( \phi_i \) and \( \phi_j \) to be localized molecular orbitals (LMOs) rather than canonical MOs and employ valence virtual LMOs in place of the set of all virtual MOs (as detailed in Ref. 33). We separate Eq. 3 to independently machine learn the cases of \( i = j \) and \( i \neq j \),

\[ \varepsilon_{ij} \approx \begin{cases} \varepsilon_{ij}^{\text{ML}}[f_i] & \text{if } i = j \\ \varepsilon_{ij}^{\text{ML}}[f_{ij}] & \text{if } i \neq j \end{cases} \]

where \( f_i \) denotes \( f_{ii} \) (Eq. [5]) with redundant elements removed; by separating the pair energies in this way, we avoid the situation where a single ML model is required to distinguish between the cases of \( i = j \) and \( \phi_i \) being nearly degenerate to \( \phi_j \), a distinction which can represent a sharp variation in the function to be learned.

In the current work, several technical refinements are introduced to improve training efficiency (i.e., the accuracy and transferability of the model as a function of the number of training examples). These are now described.

**Occupied LMO symmetrization.** The feature vector is pre-processed to specify a canonical ordering of the occupied and virtual LMO pairs. This reduces permutation of elements in the feature vector, resulting in greater ML training efficiency. Matrix elements \( M_{ij} (M = F, J, K) \) associated with \( \phi_i \) and \( \phi_j \) are rotated into gerade and ungerade combinations

\[ M_{ii} \leftarrow \frac{1}{2} M_{ii} + \frac{1}{2} M_{jj} + M_{ij} \]
\[ M_{jj} \leftarrow \frac{1}{2} M_{ii} + \frac{1}{2} M_{jj} - M_{ij} \]
\[ M_{ij} \leftarrow \frac{1}{2} M_{ii} - \frac{1}{2} M_{jj} \]
\[ M_{ii} \leftarrow \frac{1}{2} M_{ii} + \frac{1}{2} M_{jj} + M_{ii} \]
\[ M_{ij} \leftarrow \frac{1}{2} M_{ii} - \frac{1}{2} M_{jj} \]
\[ M_{ij} \leftarrow \frac{1}{\sqrt{2}} M_{ii} + \frac{1}{\sqrt{2}} M_{jj} \]

with the sign convention that \( F_{ij} \) is negative. Here, \( p \) indexes any LMO other than \( i \) or \( j \) (i.e., an occupied LMO \( k \), such that \( i \neq k \neq j \), or a valence virtual LMO).

**LMO sorting.** The virtual LMO pairs are sorted by increasing distance from occupied orbitals \( \phi_i \) and \( \phi_j \). Sorting in this way ensures that features corresponding to valence virtual LMOs are listed in decreasing order of heuristic importance, and that the mapping between valence virtual LMOs and their associated features is roughly preserved. We recognize this issue could also potentially be addressed through the use of symmetry functions [31], but these are not employed in the current work.

For purposes of sorting, distance is defined as

\[ R_{ij} = \left\| \langle \phi_i | \hat{R} | \phi_i \rangle - \langle \phi_i | \hat{R} | \phi_i \rangle \right\| + \left\| \langle \phi_j | \hat{R} | \phi_j \rangle - \langle \phi_i | \hat{R} | \phi_i \rangle \right\|, \]

where \( \phi_i \) is a virtual LMO, \( \hat{R} \) is the Cartesian position operator, and \( \| \cdot \| \) denotes the 2-norm. \( \| \langle \phi_i | \hat{R} | \phi_i \rangle - \langle \phi_i | \hat{R} | \phi_i \rangle \| \) represents the Euclidean distance between the centroids of orbital \( i \) and orbital \( j \). Previously [32], distances were defined based on Coulomb repulsion, which was found to sometimes lead to inconsistent sorting in systems with strongly polarized bonds. The non-\( i,j \) occupied LMO pairs are sorted in the same manner as the virtual LMO pairs.

**Orbital localization.** We employ Boys localization [24] to obtain the occupied LMOs, rather than IBO localization [22], which was employed in our previous work [33]. Particularly for molecules that include triple bonds or multiple lone pairs, it is found that Boys localization provides more consistent localization as a function of small geometry changes than IBO localization; and the chemically unintuitive mixing of \( \sigma \) and \( \pi \) bonds in Boys localization ("banana bonds") [35] does not present a problem for the MOB-ML method.

**Feature selection.** Prior to training, automatic feature selection is performed using random forest regression [33] with the mean decrease of accuracy criterion (sometimes called permutation importance) [35]. This technique was found to be more effective than our previous use [33] of the Gini importance score [35] which led to worse accuracy and failed to select any features for the case of methane.

The reason for using feature selection in this way is twofold. First, GPR performance is known to degrade for high-dimensional datasets (in practice 50–100 features) and second, the use of the full feature set with small molecules can lead to overfitting as features can become correlated.

A reference PYTHON implementation for generating MOB-ML features is provided online [34].

II. COMPUTATIONAL DETAILS

Results are presented for a single water molecule; a series of alkane molecules; a thermalized version of the QM7b set of 7211 molecules with up to seven C, O, N, S, and Cl heavy atoms; and a thermalized version of the GDB-13 set of molecules with thirteen C, O, N, S, and Cl heavy atoms. All datasets employed in this work are provided in the Supporting Information.

Training and test geometries are sampled at 50 fs intervals from ab initio molecular dynamics trajectories performed with the Q-CHEM 5.0 software package [33] using the B3LYP [49] 6-31g* [53] level of theory and a Langevin thermostat at 350 K.

The features and training pair energies associated with these features are computed using the MOLPRO 2018.0 software package [53] in a cc-pVTZ basis set unless otherwise
noted. Valence virtual orbitals used in feature construction are determined pair correlation energies are computed with second-order Møller-Plesset perturbation theory (MP2) and coupled cluster with singles and doubles (CCSD) as well as with perturbative triples (CCSD(T)). Density fitting for both Coulomb and exchange integrals is employed for all results below except those corresponding to the water molecule. The frozen core approximation is used in all cases.

Gaussian process regression (GPR) is employed to machine learn $\epsilon_0^{\text{ML}}$ and $\epsilon_d^{\text{ML}}$ (Eq. 6) using the GPy 1.9.6 software package. The GPR kernel is Matérn 5/2 with white noise regularization. Kernel hyperparameters are optimized with respect to the log marginal likelihood objective for the water and alkane series results, as well as for $\epsilon_d^{\text{ML}}$ of the QM7b results. We use the Matérn 3/2 kernel instead of the Matérn 5/2 kernel for the case of $\epsilon_0^{\text{ML}}$ for QM7b results, as it was empirically found to yield slightly better accuracy.

Feature selection is performed using the random forest regression implementation in the scikit-learn v0.20.0 package.

### III. RESULTS

The ML model of Eq. 6 is a universal functional for any molecular Hamiltonian. In principle, with an adequate feature list and unlimited training data (and time), it should accurately and simultaneously describe all molecular systems. In practice, we must train the ML model using a truncated feature list and finite data. These choices determine the accuracy of the model.

Below, we examine the performance of the MOB-ML method in three increasingly broad regions of chemical space: (i) training on randomized water molecule geometries and predicting the energies of other water molecule geometries; (ii) training on geometries of short alkanes and predicting the energies of longer alkanes, and (iii) training on a small set of organic molecules and predicting the energies of a broader set of organic molecules. The first two test cases were introduced in our previous work, and we explore how the refined methodology reported herein leads to improvements in accuracy and transferability. The last case represents a demanding new test of transferability across chemical space. In all cases, we report the ML prediction accuracy as a function of the number of training examples.

As a first example, we consider the performance of MOB-ML for a single water molecule. A separate model is trained to predict the correlation energy at the MP2, CCSD, and CCSD(T) levels of theory, using reference calculations on a subset of 1000 randomized water geometries to predict the correlation energy for the remainder. Feature selection with an importance threshold of $1 \times 10^{-3}$ results in 12, 11 and 10 features for $\epsilon_0^{\text{ML}}$ for MP2, CCSD and CCSD(T), respectively; ten features are selected for $\epsilon_d^{\text{ML}}$ for all three post-Hartree-Fock methods.

Figure 1 presents the test set prediction accuracy of each MOB-ML model as a function of the number of training geometries for MP2, CCSD, and CCSD(T), and the model shows the same level of accuracy for all three methods. Remarkably, all three models achieve a prediction mean absolute error (MAE) of 1 mH when trained on only a single water geometry, indicating that only a single reference calculation is needed to provide chemical accuracy for the remaining 999 geometries at each level of theory. Since it contains 10 distinct LMO pairs, this single geometry provides enough information to yield a chemically accurate MOB-ML model for the global thermally accessible potential energy surface.

For all three methods (Fig. 1), the learning curve exhibits the expected power-law behavior as a function of training data, and the total error reaches microhartree accuracy with tens of water training geometries. As compared to our previous results, where training on 200 geometries resulted in a prediction MAE of 0.027 mH for the case of CCSD the current implementation of the MOB-ML model is substantially improved; the improvement for this case stems primarily from the use of Boys localization which specifies unique and consistent LMOs corresponding to the oxygen lone pairs.
The predictions of MOB-ML in the current work (Fig. 2) are markedly improved. First, the overall prediction accuracy is improved for all four summary statistics (inset in Fig. 2) despite substantial reduction in the number of training examples used. (The current work uses only 50 geometries of ethane, 20 geometries of propane, and no methane data.) Second, n-butane and isobutane are predicted with nearly identical accuracy. Finally, the prediction errors are no longer skewed as a function of true correlation energy. The primary methodological sources of these improvements are found to be symmetrization of occupied orbitals (Eq. 7) and the improved feature selection methodology. The MOB-ML features in the current work are selected with an importance threshold of $1 \times 10^{-4}$, resulting in 27 features for $\varepsilon_{\text{ML}}^c$ and 12 features for $\varepsilon_{\text{ML}}^{c}$. results presented in Fig. 2 for CCSD(T) are qualitatively identical to those obtained for CCSD (not shown).

We now examine the transferability of the MOB-ML method across a broad swath of chemical space. Specifically, we consider the QM7b dataset which is comprised of 7,211 plausible organic molecules with 7 or fewer heavy atoms. The chemical elements in QM7b are limited to those likely to be found in drug-like compounds: C, H, O, N, S, and Cl. We refer to the dataset used herein as QM7b-T to reflect the fact that it contains geometries sampled at a temperature of 350 K (as described in Sec. 11), as opposed to DFT-optimized geometries. The MOB-ML model is trained on a randomly chosen subset of QM7b-T molecules and used to predict the correlation energy of the remainder. Active learning was also tested as a training data selection strategy, but was not found to improve the predictions in the regime of chemical accuracy, and in fact led to slightly worse transferability.

For comparison, a $\Delta$-ML model was trained on the same molecules using kernel-ridge regression using the FCHL representation with a Gaussian kernel function (FCHL/\$\Delta$-ML), as implemented in the QML package. All hyperparameters of the model were set to those obtained in Ref. 70 which have previously been demonstrated to work well for datasets containing structures similar to those in QM7b-T. A possible source of concern for MOB-ML is that the number of selected features would grow with the chemical complexity of the training data. For example, 27 features for $\varepsilon_{\text{ML}}^c$ and 12 features for $\varepsilon_{\text{ML}}^{c}$ were selected in the alkane test case using ethane + propane training data (Fig. 2b), whereas only 10 features for $\varepsilon_{\text{ML}}^c$ and 10 features for $\varepsilon_{\text{ML}}^c$ were selected for the water test case at the CCSD(T) level of theory (Fig. 1). To examine this, we perform feature selection on increasing numbers of randomly selected molecules from the QM7b-T dataset. Table 2 presents two statistics on the feature importance as a function of the number of training molecules: (i) the number of “important features” (i.e., those whose permutation importance exceeds a set threshold of $2 \times 10^{-4}$ and $5 \times 10^{-5}$ for $\varepsilon_{\text{ML}}^c$ and $\varepsilon_{\text{ML}}^{c}$, respectively) and (ii) the inverse participation ratio of the feature importance scores. The latter is a threshold-less measure of the number of important features; it takes a value of 1 when only 1 feature has nonzero importance and $N$ when all $N$ features have equal importance. Although the QM7b-T dataset contains many different chemical elements and bonding motifs, Table 2 reveals that the selected features remain compact and do not grow with the number of training molecules. Indeed, for a large number of training molecules, the number of selected features slightly decreases, reaching 42 and 24 selected features for $\varepsilon_{\text{ML}}^c$ and $\varepsilon_{\text{ML}}^{c}$, respectively, for the largest training sizes considered.

The learning curves for MOB-ML models trained at the MP2/cc-pVTZ and CCSD(T)/cc-pVDZ levels of theory are shown in Fig. 3, as well as the FCHL/\$\Delta$-ML learning curve for MP2/cc-pVTZ. At the MP2 level of theory, the MOB-ML model achieves an accuracy of 2 mH with 110 training calculations (representing 1.5% of the molecules in the QM7b-T dataset), whereas the FCHL/\$\Delta$-ML requires over 300 training geometries to reach the same accuracy threshold. Fig. 3 also illustrates the relative insensitivity of MOB-ML to the level of electronic structure theory, with the learning curve for CCSD(T)/cc-pVDZ reaching 2 mH accuracy with 140 training calculations. An analysis of the sensitivity of the MOB-
ML predictions to the number of selected features is presented in supporting information Fig. S1, which indicates that the reported results are robust with respect to the number of selected features.

As a final test of transferability of the MOB-ML and FCHL/Δ-ML methods across chemical space, Figs. 3b and 3c show results in which the ML methods are trained on QM7b-T molecules and then used to predict results for a dataset of 13-heavy-atom organic molecules at thermalized geometries, GDB-13-T, which includes six thermally sampled geometries each of 1,000 13-heavy-atom organic molecules chosen randomly from the GDB-13 database. Like QM7b, the members of GDB-13 contain C, H, N, O, S, and Cl. The size of these molecules precludes the use of coupled cluster theory to generate reference data; we therefore make comparison at the MP2/cc-pVTZ level of theory, noting that MOB-ML has consistently been shown to be insensitive to the employed post-Hartree-Fock method (as in Fig. 3a). Transfer learning results as a function of the number of training molecules are presented in Figs. 3b (on a linear-linear scale) and 3c (on a log-log scale).

Using the MOB-ML model that is trained on 110 seven-heavy-atom molecules (corresponding to a prediction MAE of 1.89 mH for QM7b-T), we observe a prediction MAE of 3.88 mH for GDB-13-T. Expressed in terms of size-intensive quantities, the prediction MAE per heavy atom is 0.277 mH and 0.298 mH for QM7b-T and GDB-13-T, respectively, indicating that the accuracy of the MOB-ML results are only slightly worse when the model is transferred to the dataset of larger molecules. On a per-heavy-atom basis, MOB-ML reaches chemical accuracy with the same number of QM7b-T training calculations (approximately 100), regardless of whether it is tested on QM7b-T or GDB-13-T.

In contrast with MOB-ML, the FCHL/Δ-ML method is found to be significantly less transferable from QM7b-T to GDB-13-T. For models trained using 100 seven-heavy-atom molecules, the MAE per heavy atom of FCHL/Δ-ML is over twice that of MOB-ML (Fig. 3b). Moreover, whereas MOB-ML reaches the per-heavy-atom chemical accuracy threshold with 140 training calculations, the FCHL/Δ-ML method only reaches that threshold with 5000 training calculations.

### IV. CONCLUSIONS

Molecular-orbital-based machine learning (MOB-ML) has been shown to be a simple and strikingly accurate strategy for predicting correlated wavefunction energies at the cost of a Hartree-Fock calculation, benefiting from the intrinsic transferability of the localized molecular orbital representa-
tion. The starting point for the MOB-ML method is a rigorous mapping from the Hartree-Fock molecular orbitals to the total correlation energy, which ensures that the use of sufficient training data and molecular orbital features will produce a model that matches the corresponding correlated wavefunction method across the entirety of chemical space. The current work explores this possibility within the subspace of organic molecules. It is shown that MOB-ML predicts energies of the QM7b-T dataset to within a 2 millihartree accuracy using only 110 training calculations at the MP2/cc-pVTZ level of theory and using 140 training calculations at the CCSD(T)/cc-pVDZ level of theory. Direct comparison with FCHL/Δ-ML reveals that MOB-ML is threefold more efficient in reaching chemical accuracy for describing QM7b-T. Further, a transferability test of a MOB-ML model trained on QM7b-T to GDB-13-T reveals that MOB-ML exhibits negligible degradation in accuracy; as a result, chemical accuracy is achieved with 36-times fewer training calculations using MOB-ML versus FCHL/Δ-ML. These results suggest that MOB-ML provides a promising approach toward the development of density matrix functionals that are applicable across broad swaths of chemical space.

ACKNOWLEDGMENTS

We thank Daniel Smith (Molecular Sciences Software Institute) and Alberto Gobbi (Genentech) for a helpful discussion about available training datasets. T.F.M. acknowledges support from AFOSR award no. FA9550-17-1-0102. A.S.C. acknowledges support from the National Centre of Competence in Research (NCCR) Materials Revolution: Computational Design and Discovery of Novel Materials (MARVEL) of the Swiss National Science Foundation (SNSF). We additionally acknowledge support from the Resnick Sustainability Institute postdoctoral fellowship (M.W.) and the Camille Dreyfus Teacher-Scholar Award (T.F.M.). Computational resources were provided by the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the DOE Office of Science under contract DE-AC02-05CH11231.

SUPPORTING INFORMATION

The datasets used in this work are available for download[1] they include MOB-ML features, HF energies, pair correlation energies, and geometries. MOB-ML and FCHL/Δ-ML predictions corresponding to Fig. 3 and an analysis of the sensitivity of the results of Fig. 3 to the number of selected features are available in the attached supporting information. A reference PYTHON implementation for generating MOB-ML features is available at https://github.com/thomasfmiller/MOB-ML.

1. A. Lavecchia, “Machine-learning approaches in drug discovery: methods and applications,” Drug Discov. Today 20, 318–331 (2015)
2. E. Gawehn, J. A. Hiss, and G. Schneider, “Deep learning in drug discovery,” Mol. Inform. 35, 3–14 (2016)
3. M. Popova, O. Isayev, and A. Tropsha, “Deep reinforcement learning for de novo drug design,” Sci. Adv. 4, eaap7885 (2018)
4. E. Kim, K. Huang, S. Jegelka, and E. Olivetti, “Virtual screening of inorganic materials synthesis parameters with deep learning,” npj Comput. Mater. 3, 53 (2017)
5. F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers, and A. Mehta, “Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments,” Sci. Adv. 4, eaau1566 (2018)
6. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, “Machine learning for molecular and materials science,” Nature 559, 547–555 (2018)
7. B. Sanchez-Lengeling and A. Aspuru-Guzik, “Inverse molecular design using machine learning: Generative models for matter engineering,” Science 361, 360–365 (2018)
8. J. N. Wei, D. Duvenaud, and A. Aspuru-Guzik, “Neural networks for the prediction of organic chemistry reactions,” ACS Cent. Sci. 2, 725–732 (2016)
9. P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, “Machine-learning-assisted materials discovery using failed experiments,” Nature 553, 73–76 (2016)
10. Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, “To address surface reaction network complexity using scaling relations machine learning and DFT calculations,” Nat. Commun. 8, 14621 (2017)
11. M. H. S. Segler and M. P. Waller, “Neural-symbolic machine learning for retrosynthesis and reaction prediction,” Chem. - A Eur. J. 23, 5966–5971 (2017)
12. M. H. S. Segler, M. Preuss, and M. P. Waller, “Planning chemical syntheses with deep neural networks and symbolic AI,” Nature 555, 604–610 (2018)
13. J. S. Smith, O. Isayev, and A. E. Roitberg, “ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost,” Chem. Sci. 3, 3192–3203 (2017)
14. J. S. Smith, B. T. Nebgen, Z. Rubatysk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. Roitberg, “Outsmarting quantum chemistry through transfer learning,” (2018)
15. N. Lubbers, J. S. Smith, and K. Barros, “Hierarchical modeling of molecular energies using a deep neural network,” J. Chem. Phys. 148, 241715 (2018)
16. A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, “Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons,” Phys. Rev. Lett. 104, 136403 (2010)
17. M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, “Fast and accurate modeling of molecular atomization energies with machine learning,” Phys. Rev. Lett. 108, 58302 (2012)
18. G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, “Machine learning of molecular electronic properties in chemical compound space,” New J. Phys. 15, 9503 (2013)
19. K. Hansen, G. Montavon, F. Biegler, S. Fuzli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, “Assessment and validation of machine learning methods for predicting molecular atomization energies,” J. Chem. Theory Comput. 9, 3404 (2013)
20. P. Gasparotto and M. Ceriotti, “Recognizing molecular patterns by machine learning: an agnostic structural definition of the hydrogen bond,” J. Chem. Phys. 141, 174110 (2014)
21. R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, “Big data meets quantum chemistry approximations: the Δ-machining approach,” J. Chem. Theory Comput. 11, 2087 (2015)
22. J. Behler, “Perspective: Machine learning potentials for atomistic simulations,” J. Chem. Phys. 145, 170901 (2016)
23. S. Kearnes, K. Mccloskey, M. Berndl, V. Pande, and P. Riley, “Molecular graph convolutions: moving beyond fingerprints,” J. Comput. Aided Mol. Des. 30, 595 (2016)
24. E. Paesani, “Getting the right answers for the right reasons: toward predictive molecular simulations of water with many-body potential energy functions,” Acc. Chem. Res. 49, 1844 (2016)
25. K. T. Schütt, F. Arbabzadah, S. Chmiela, K.-R. Müller, and A. Tkatchenko, “Quantum-chemical insights from deep tensor neural networks,” Nat. Commun. 8, 13890 (2017)
In principle, the smoothness of the Matérn kernel could be taken as a kernel hyperparameter; however, this possibility was not explored in this work.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: machine learning in python,” Journal of Machine Learning Research 12, 2825 (2011).

C. Cortes, L. D. Jackel, S. A. Solla, V. Vapnik, and J. S. Denker, “Learning curves: Asymptotic values and rate of convergence,” in Advances in Neural Information Processing Systems 6 edited by J. D. Cowan, G. Tesauro, and J. Alspector (Morgan-Kaufmann, 1994) pp. 327–334.

K. Pearson, “Mathematical contributions to the theory of evolution. III. regression, heredity, and panmixia,” Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 187, 253–318 (1896).

G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, “Machine learning of molecular electronic properties in chemical compound space,” New J. Phys. 15, 095003 (2013).

F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, “Alchemical and structural distribution based representation for universal quantum machine learning,” J. Chem. Phys. 148, 241717 (2018).

A. S. Christensen, F. A. Faber, and O. A. von Lilienfeld, “Operators in quantum machine learning: Response properties in chemical space,” J. Chem. Phys. 150, 064105 (2019).

B. Kramer and A. MacKinnon, “Localization: theory and experiment,” Rep. Prog. Phys. 56, 1469 (1993).

L. Cheng, M. Welborn, A. S. Christensen, and T. F. Miller III, “Thermalized (350K) QM7b, GDB-13, water, and short alkane quantum chemistry dataset including MOB-ML features,” (2019), 10.22002/D1.1177.