Boosting Quantum Machine Learning Models with a Multilevel Combination Technique: Pople Diagrams Revisited
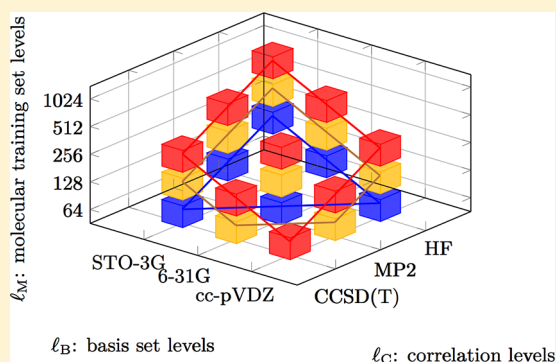
Peter Zaspel,[†] Bing Huang,[‡] Helmut Harbrecht,[*,†] and O. Anatole von Lilienfeld[*,‡]

[†]Department of Mathematics and Computer Science, University of Basel, Spiegelgasse 1, 4051 Basel, Switzerland

[‡]Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland

S *Supporting Information*

ABSTRACT: Inspired by Pople diagrams popular in quantum chemistry, we introduce a hierarchical scheme, based on the multilevel combination (C) technique, to combine various levels of approximations made when molecular energies are calculated. When combined with quantum machine learning (QML) models, the resulting CQML model is a generalized unified recursive kernel ridge regression that exploits correlations implicitly encoded in training data composed of multiple levels in multiple dimensions. Here, we have investigated up to three dimensions: chemical space, basis set, and electron correlation treatment. Numerical results have been obtained for atomization energies of a set of ~7000 organic molecules with up to 7 atoms (not counting hydrogens) containing CHONFClS, as well as for ~6000 constitutional isomers of $C_7H_{10}O_2$. CQML learning curves for atomization energies suggest a dramatic reduction in necessary training samples calculated with the most accurate and costly method. In order to generate millisecond estimates of CCSD(T)/cc-pvdz atomization energies with prediction errors reaching chemical accuracy (~1 kcal/mol), the CQML model requires only ~100 training instances at CCSD(T)/cc-pvdz level, rather than thousands within conventional QML, while more training molecules are required at lower levels. Our results suggest a possibly favorable trade-off between various hierarchical approximations whose computational cost scales differently with electron number.

1. INTRODUCTION

Chemical compound space, the property space spanned by all possible chemical compounds, is unfathomably large due to its combinatorial nature.[1,2] Exploring chemical space from first-principles is desirable in the context of computational materials design[3−5] as well as to fundamentally deepen our understanding of chemistry.[6] Over the last couple of years, overwhelming evidence has been collected indicating that quantum machine learning (QML) models, trained throughout chemical space, hold great promise to dramatically reduce the cost for predicting quantum properties, such as atomization energies of molecules, for arbitrary out-of-sample molecules.[7−20] The core idea of QML is to learn the implicit mapping from geometrical and compositional information encoded in nuclear charges and positions to corresponding electronic properties from a set of training molecules with precomputed properties at a specific level of theory. The knowledge thus obtained from training is then applied to molecules out-of-sample, i.e., molecules not in the training set. Nowadays, QML is a well-established technique and has several supervised learning variants, including mainly neural network[10,11,21] and kernel ridge regression.[7,22,23] Currently, most of the efforts toward QML in literature are devoted to developing more efficient molecular representations[16,20,24] and adapting machine learning models to a growing

number of applications.[11,25,26] Recent overviews on the field were published in refs 27−29 and an entire issue in *J. Chem. Phys.* was recently devoted to the theme of "data-enabled theoretical chemistry".[30]

This progress was made possible due to the advent of modern computers, which enabled routine calculations of electronic properties such as ground state energies for large training sets of medium-sized organic molecules[31−33] using common density functional approximations.[34,35] While QML prediction errors have converged to values smaller than DFT accuracy,[15] the predictive power of any QML model inherently hinges on the accuracy of the employed reference data used for training. However, while the latest machine learning models are now able to make rather accurate and yet efficient predictions, the time required to compute training samples for large data sets with chemical accuracy is still prohibitive. More specifically, in order to routinely match the experimental uncertainty of thermochemistry, the highly coveted "chemical accuracy" of ~1 kcal/mol, typical approximations made within density functional theory do not suffice, and computationally expensive theories, e.g., CCSD(T) in a large basis, have to be used even when dealing
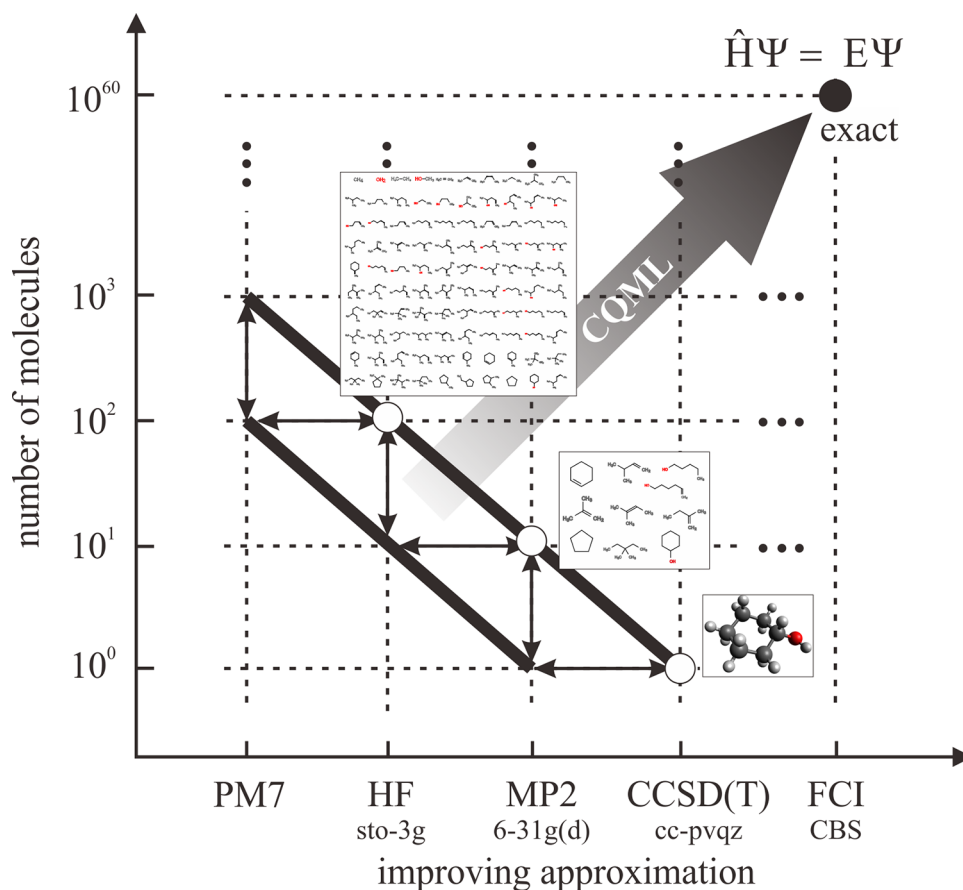
Figure 1. Adaptation of a Pople diagram involving various levels of theory (abscissa) and molecular spaces (ordinate). The wide arrow indicates how to best approximate highly accurate solutions (solid black circle) of Schrödinger's equation by combining ever improving levels of theory with an exponentially decreasing number of molecules used for training of machine learning models. Qualitative estimates of constant cost-benefit ratios (bold diagonals) correspond to Pareto-optimal solutions which can be sampled by using the CQML approach presented herewithin. For example, training data consisting of 1 CCSD(T)/cc-pvqz, 4 MP2/6-31g(d), and 16 HF/sto-3g calculation results can be cheaper *and* more valuable than three CCSD(T)/cc-pvqz results. Two-sided arrows indicate bridges in chemical and method space.

just with closed-shell molecules in relaxed geometries. Unfortunately, due to its substantially larger computational complexity, the routine generation of CCSD(T) numbers in large basis sets for thousands of training molecules remains prohibitive.

The hierarchies encoded in model chemistries, well established in quantum chemistry, can be used to exploit systematic trends in cancellation of errors among different levels of theory, as proposed and demonstrated by Pople and co-workers.[36,37] Composite methods are based on these ideas[38] and include, among many others, Gaussian-*n* theories,[39−41] the Weizmann-*n* methods,[42,43] and complete basis set (CBS) methods.[44−46] They can reach chemical accuracy at the computational cost of combinations of more efficient models. When it comes to chemical space, the Pople diagram is a two-dimensional display of the relationship of the size of *any* molecule and level of theory.[47] Pople diagrams can easily be extended to accommodate additional or other dimensions such as relativistic effects[48] or accuracy.[49] In this study, we apply the idea of a Pople diagram to combine varying levels of theory in the training set of QML models (see Figure 1 for the general idea). More specifically, we apply the sparse-grid combination (C) technique to estimate the optimal balance among (i) electron correlation (HF, MP2, CCSD(T)), (ii) basis set size (sto-3g, 6-31g, cc-pvdz), and (iii) number of organic molecules. We find that

the resulting CQML models require substantially fewer training instances at the computationally most demanding target level of theory.

To showcase our new developments, we will discuss a series of multilevel and multispace machine learning models, as well as results for molecules from the QM7b data set.[50] Using several levels in the space of electron correlation approximations already leads to a very strong improvement in the learning results, with respect to the amount of necessary training data at target accuracy. Further improvement is found by adding different levels of basis sets.

This paper is structured as follows: Section 3 briefly introduces the CQML model, as well as the data sets used for training and testing. In section 4, results of the CQML model are presented and discussed for 2D and 3D CQML models. Finally, section 5 summarizes the main findings, draws general conclusions, and presents an outlook. The Appendix provides detailed methodological information to facilitate reproducibility of our findings.

2. COMPUTATIONAL DETAILS

2.1. Data Sets. Two data sets were used for proof of principle: QM7b[50] and 6k constitutional isomers[51] (dubbed "CI9"); both are subsets of the GDB-17 universe.[52,53] QM7b is composed of molecules with up to seven heavy atoms, including

C, N, O, S, and Cl (H not counted), totaling 7211 molecules. Molecules in CI9 correspond to 6095 constitutional isomers of $C_7H_{10}O_2$.

For QM7b molecules, geometries were first optimized at the level of B3LYP/6-31g(d) with Gaussian 09,[54] then single-point calculations were calculated by using three levels of theory (HF, MP2, CCSD(T)) and three basis sets (sto-3g, 6-31g, and cc-pvdz) with Molpro,[55] resulting in nine single-point energies per molecule.

For the CI9 molecules, three different methods were used: PM7, B3LYP/6-31g(2df,p), and G4MP2. Relaxed geometries and energies were retrieved directly from ref 51 for the last two methods, while PM7 relaxed geometries and energies were obtained by using MOPAC2016.[56]

2.2. QML Details. We used both the sorted Coulomb matrix[57,58] and SLATM[16] for modeling the CI9 data set, while SLATM[16] only was used for QM7b. Though slightly better performing representations have been published previously, such as SOAP,[59,60] aSLATM,[16] or FCHL,[20] comparison between CM and SLATM results indicates that trends are stable and that the conclusions drawn are independent of choice of representation. As kernel functions, we have always chosen the Laplace kernel $e^{-\|\mathbf{R}_q - \mathbf{R}_i\|_1 / \sigma}$ with σ being a hyper-parameter. The hyper-parameter σ was optimized manually and converged to $\sigma = 400$. Furthermore, we use a Lavrentiev regularization of size 10^{-10}. All presented errors are mean absolute error (MAE) comparing the prediction by the CQML method with the *true* solution of the target theory level. The MAE is computed as out-of-sample error over 200 randomly chosen molecules that are not part of the training data set. These results are averaged over 20 training runs. Note that we randomly choose the $N_{l_{M=0}}$ training molecules on the lowest level, while randomly selecting subsets of them on higher levels. This sequence of drawing ensures the nestedness of all the training samples.

3. THEORY

In this section, we start by reviewing systematic error cancellation, composite methods, the CQML approach, and kernel ridge regression based QML and Δ-ML,[61] as well as two- and d-dimensional CQML.

3.1. From Pople Diagrams to CQML. Telescoping series, as a means to systematic convergence of error cancellation, is a well established mathematical tool. In short, if a_n is a sequence of numbers, then

$$\sum_{I=1}^{N} (a_I - a_{I-1}) = a_N - a_0 \tag{1}$$

and if we define $\Delta_{I-1}^{I} = a_I - a_{I-1}$ and $a_0 = 0$, one has

$$a_N = a_0 + \sum_{I=1}^{N} \Delta_{I-1}^{I} \tag{2}$$

Error cancellation is also at the root of many common practices in theoretical chemistry. Most notable are composite methods,[39−46,62−64] recently reviewed in ref 65, which correspond to computational protocols that combine various quantum chemical approximations such that high accuracy (frequently chemical accuracy, i.e., ~1 kcal/mol) is achieved for thermodynamic quantities (e.g., atomization enthalpies). Typically, they combine the results of a high level of theory with a small basis set with methods that employ lower levels of

theory with larger basis sets. Importantly, they impose a computationally much reduced burden in comparison to brute-force convergence in basis set size and electron correlations. For example, an extensively used composite method called Gaussian-2 (G2)[66] approximates the energy as (starting from a geometry optimized at MP2/6-31g(d) level)

$$E_{\text{true}} \approx E^{\text{G2}} := E_{\text{QCISD(T)}/6\text{-311g(d)}} + \Delta_1 + \Delta_2 + \Delta_3 \tag{3}$$

where further correction terms have been neglected. Note that, here and throughout, we denote approximations and reference results by upper and lower indices, respectively. The individual terms read

$$\Delta_1 = E_{\text{MP4}/6\text{-311g(2df,p)}} - E_{\text{MP4}/6\text{-311g(d)}}$$

$$\Delta_2 = E_{\text{MP4}/6\text{-311+g(d,p)}} - E_{\text{MP4}/6\text{-311g(d)}}$$

$$\Delta_3 = E_{\text{MP2}/6\text{-311+g(3df,2p)}} + E_{\text{MP2}/6\text{-311g(d)}}$$
$$- E_{\text{MP2}/6\text{-311g(2df,p)}} - E_{\text{MP2}/6\text{-311g+(d,p)}} \tag{4}$$

with Δ_1 accounting for the effect of adding the polarization functions, Δ_2 correcting for the diffuse functions, and Δ_3 correcting for the larger basis set as well as preventing contributions from being counted twice in Δ_1 and Δ_2, respectively.

Note that the formalism of the composite method corresponds to a sophisticated extension of the telescoping series in eq 2. One could also simply rewrite (2) as

$$E_{\text{CCSD(T)}} = E_{\text{HF}} + \Delta_{\text{HF}}^{\text{MP2}} + \Delta_{\text{MP2}}^{\text{CCSD(T)}} \tag{5}$$

with all terms obtained for some large basis set. The problem then reduces to define efficient yet accurate estimates of the Δ's. Here, we introduce the methodology to solve this problem through generalization of the Δ-ML approach[61] in the form of CQML.

3.2. CQML Approach. To exploit varying levels of theory in order to improve prediction accuracy, and thereby reduce the number of necessary costly training instances, some of us previously introduced the Δ-ML approach.[61] It uses reference data calculated from a computationally efficient but inaccurate method as a baseline and estimates the difference to a more expensive but accurate target level of theory. Numerical results for organic molecules indicated that given an appropriately chosen baseline method, it is possible to achieve orders of magnitude reduction in training set size when compared to results from traditional QML approaches. Many other studies have already shown the usefulness and applicability of the Δ-ML approach.[59,67−74] Alternatively, efforts have been made toward training set size reduction based on training set optimization[16,68,75,76] or improvements in the representations.[17,20,24,77] To the best of our knowledge, no conceptual improvements or generalizations of the Δ-ML approach have been proposed so far.

In this work, we generalize the core ideas of Δ-ML[61] to arrive at a multilevel combination technique QML (CQML) approach. CQML is a unified kernel ridge regression machine learning model incorporating training data from several spaces and levels of information. As proposed by, e.g., John Pople,[36,47] we distinguish between

1. the space of electron correlation (e.g., MP2) and
2. the *space of basis set* (e.g., 6-31g), and we also add
3. the *space of training molecules* (e.g., some training set drawn from QM9[31]) as third degree of freedom that can easily be exploited through machine learning models.

We call a specific choice of training information, e.g., Hartree−Fock calculations on a 6-31g basis set done for 256 molecules, a *subspace*. Within each space, we assume a multilevel hierarchy of growing accuracy and computational complexity. For example, in electron correlation and basis set space, one commonly expects that the degree of approximative nature decays systematically as one goes from HF to MP2 to CCSD(T), or from sto-3g to 6-31g to cc-pvdz, respectively. In chemical space, it is less obvious how to establish a hierarchy of accuracy. For the purpose of our approach, we rely on the well established tenet in statistical learning that the predictive accuracy for out-of-sample molecules increases systematically with training set size,[78] which is applicable to chemical space and quantum chemistry as demonstrated first in 2012.[57] This finding has by now been confirmed and reproduced within multiple studies for various quantum properties and system classes.[27,28] As such, and when drawing training molecules at random, we can consider their number made available to training (e.g., N = 16, 32, 64, ...) as the chemical space equivalent to the space of theory (e.g., HF/MP2/CCSD(T)) or basis set (e.g., sto-3g/6-31g/cc-pvdz). Generally speaking, a CQML model built on low levels of theories/basis sets/small number of training molecules, will result in a model with low accuracy and easily accessible training data. Conversely, including more levels in each dimension, the resulting CQML model will become increasingly more accurate, requiring, however, also access to ever more costly training data. Figure 1 exemplifies these ideas for various levels of electron correlation, basis sets, and molecular training set sizes.

The *sparse grid combination technique* known for high-dimensional approximation[79−89] and quadrature/uncertainty quantification[90,91] in numerical analysis corresponds to a rigorous means to generate QML models constructed on a combination of sets of different subspaces. The general idea is to combine the subspaces such that only very few expensive training samples are needed at target accuracy (e.g., CCSD(T) for cc-pvdz at high sample count), some less expensive subspaces with higher training sample count are needed, and so on. Figure 2 outlines a choice of
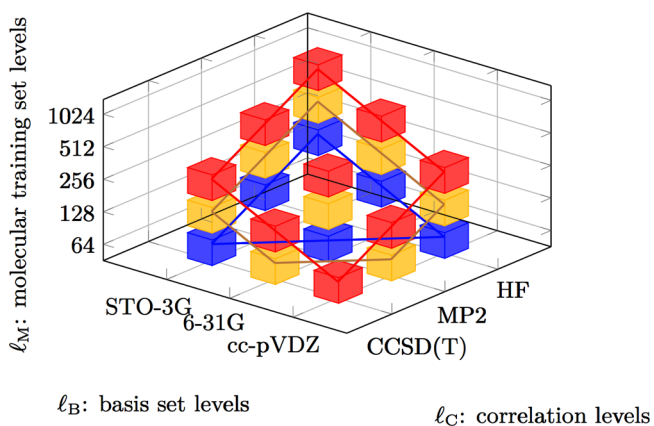

Figure 2. 3D CQML approach combining multiple levels in the spaces of electron correlation, basis sets, and training molecules. Subspaces with $l_C + l_B + l_M = 4$ (see main text for definition of l's) are colored in red, subspaces with $l_C + l_B + l_M = 3$ are colored in yellow, and subspaces with $l_C + l_B + l_M = 2$ are colored in blue. They are given in layers, which are indicated by the colored connecting lines.

subspaces by a modified sparse grid combination technique. Here, each subspace is represented by a colored cube.

In this work, we will first generalize the aforementioned Δ-ML approach to a multilevel approach that incorporates the space

of theories, basis sets, and training molecules. The CQML approach differs from existing multifidelity machine learning models[92] in that it is (a) generalized to multiple dimensions and (b) does not unite the various spaces within one kernel matrix, but rather through a series of independently trained kernels. While the CQML approach accounts for an arbitrary number of information spaces, for the sake of brevity and without any loss of generality, we restrict ourselves only to the three spaces discussed above.

3.3. Kernel Ridge Regression and the Δ-ML Approach. In order to properly discuss CQML, we first need to briefly recall the principal idea of the established kernel ridge regression based QML models. With $\mathbf{R}$ (some) representation of a molecule, we denote by $E_l(\mathbf{R})$ the ML based approximation of the electronic ground state property of that molecule at a certain level of theory l. We train the ML model by using N training molecules $\mathbf{R}_i$ with $i = 1, ..., N$ with corresponding reference energies at the corresponding specified level, $E_l^{\text{ref}}(\mathbf{R})$. The objective is to predict energy E_l^{ref} for an out-of-sample *query* molecule $\mathbf{R}_q$, part of neither training nor validation sets.

The ML model E_l within kernel ridge regression is then given by $E_l^{\text{ref}}(\mathbf{R}_q) \approx E_l(\mathbf{R}_q) := \sum_{i=1}^{N} \alpha_i^{(l)} k(\mathbf{R}_q, \mathbf{R}_i)$, where k is an appropriate unitless kernel function. For this study, we always choose the radial basis kernel function, $\exp[-\|\mathbf{R}_q - \mathbf{R}_i\|_1/\sigma]$ (Laplace) with length scale σ. Optimization of kernel function space could represent yet another potentially interesting dimension for future investigations. As described in detail elsewhere,[27,78] the coefficients α_i are obtained by solving the kernel matrix inversion problem $\boldsymbol{\alpha} = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{e}_l$ for given regularizer λ and reference energy vector $\mathbf{e}_l$. Here, we use matrix notation with capital and small case letters for matrices and vectors, respectively.

The Δ-ML approach[61] models the difference between a baseline and target level of theory, e.g., HF and MP2, respectively. Note that we here have decided to adapt a slightly different notation in contrast to ref 61 in order to facilitate the generalization of the Δ-ML to the CQML approach. Here, $P_{(b)}(\mathbf{R})$ and $P_{(t)}(\mathbf{R})$ represent the properties of interest computed at baseline and target level of theory, respectively. Note that within Δ-ML, for $P_{(b)}$ and $P_{(t)}$, it is not mandatory to estimate the same property; e.g., it could be the ground state energy in the baseline theory and the enthalpy in the target theory. Hence, the Δ-ML model prediction is given by

$$P_{(t)}(\mathbf{R}_q) := P_{(b)}(\mathbf{R}_q) + \Delta_b^t(\mathbf{R}_q) \qquad (6)$$

where $\Delta_b^t(\mathbf{R}_q) = \sum_{i=1}^{N} \alpha_i k(\mathbf{R}_q, \mathbf{R}_i)$. We emphasize that within the Δ-ML model a potentially costly baseline evaluation of the query compound is still necessary when making a prediction. This differs from the CQML approach that recovers the original speed of QML by modeling even the baseline through a machine.

3.4. Two-Dimensional Multilevel Learning. The CQML approach generalizes the Δ-ML model to several spaces *and* levels. This is illustrated in Figure 2 for three dimensions and levels that we have also considered in this study (*vide infra*). To facilitate the discussion, we first discuss the adaptation of the Pople diagram in order to exemplify the general idea of the CQML approach for the simple case of only two dimensions. More specifically, we now consider the space of theory and training molecules. Thereafter, we will also discuss the

generalization to three-dimensional, as well as *d*-dimensional cases, in section 3.5.

We assume L levels of theory with running index $l = 0, 1, ..., L − 1$, for which the calculated energy increases in accuracy a (with respect to an experimentally yet unknown truth) and computational cost with growing theoretical complexity, $a^{l+1} > a^l$, for all $l < L − 1$. Multilevel learning in two dimensions is performed as follows

(1) On level $l = 0$ compute reference energies $E_{l=0}^{\text{ref}}$ for $N_{l=0}$ molecules and train the standard QML kernel ridge regression model $E_{l=0}$.

(2) On level $l = 1$ compute reference energies $E_{l=1}^{\text{ref}}$ for $N_{l=1} < N_{l=0}$ training molecules.

(3) Still on level $l = 1$, train a model of the difference between E_0 and E_1^{ref} for the N_1 molecules.

(4) Repeat recursively until target level $l = L − 1$ is reached.

Note that while N_l and N_{l+1} molecules do not have to be mutual subsets, in this study all N_{l+1} molecules are also part of the N_l molecules out of convenience.

Formally, one can recursively define the intermediate multilevel 2D model E_l for $l = 0, 1, ..., L − 1$ and build on the lowest level baseline ($l = 0$) as

$$E_l(\mathbf{R}_q) := E_{l-1}(\mathbf{R}_q) + \sum_i^{N_l} \alpha_i^{(l-1,l)} k(\mathbf{R}_q, \mathbf{R}_i) \tag{7}$$

where we set $E_{-1} \equiv 0$. For example, the CQML model that combines PM7 ($l = 0$), DFT ($l = 1$), and G4MP2 ($l = 2$) reads

$$E_2(\mathbf{R}_q) = E_1(\mathbf{R}_q) + \sum_i^{N_2} \alpha_i^{(1,2)} k(\mathbf{R}_q, \mathbf{R}_i) \tag{8}$$

where

$$E_1(\mathbf{R}_q) = E_0(\mathbf{R}_q) + \sum_j^{N_1} \alpha_j^{(0,1)} k(\mathbf{R}_q, \mathbf{R}_j) \tag{9}$$

where

$$E_0(\mathbf{R}_q) = \sum_k^{N_0} \alpha_k^{(0)} k(\mathbf{R}_q, \mathbf{R}_k) \tag{10}$$

where the last term corresponds to the conventional direct QML model of the PM7 energy. For numerical results obtained from this model, and their discussion, *vide infra*. To compute the coefficients $\alpha_i^{(l)}$, we solve the previously mentioned kernel ridge regression problem.

Let us briefly compare this approach to the conventional Δ-ML models discussed before in section 3.3. In the single-level case, the resulting model E_0 is the direct conventional QML kernel ridge regression model. In the two-level case, the resulting model E_1 bears similarity with the Δ-ML model, the major difference being that also the baseline is a machine. Thereby, it becomes possible to use different amounts of training information (N_0, N_1) on both levels. Nevertheless, if we chose the training molecules on the first and the second level to be identical and skipped regularization in the regression problem, E_1 and conventional direct QML would be identical. And if we chose the training molecules on the first and the second level to be identical and built only one ML model (namely of the

difference), E_1^{ref} and Δ-ML would be identical. E_2 and higher order approximations have, to the best of our knowledge, not yet been discussed in the literature.

Using above definition, we did not fix yet how to choose the amount of training samples on each level. This choice is based on the sparse grid combination technique.[79−91] Qualitatively, the combination technique implies using many training samples on the lower levels of theory and reducing the number of samples to fewer and fewer samples on higher and higher levels of theory. As we will see, the balance between the amount of training samples per level can be a point of optimization within our method. In section 4, we discuss our choices of level balancing on the basis of the sparse grid combination technique. These choices have been evaluated for different training data, and with respect to two possible optimality measures. Future work will deal with a more systematic assessment of how to tailor and optimize the relative ratios of training molecules at each level and in each dimension.

3.5. Three- and d-Dimensional Multilevel Learning. Extending eq 7 to more than two dimensions results in dimension-dependent levels. Table 1 provides an exemplifying

Table 1. Exemplifying Overview of Levels in Three Dimensional Multilevel Learning for Basis Sets (B), Electron Correlation (C), and Molecular Training Set (M)

level	0	1	2
l_C	HF	MP2	CCSD(T)
l_B	sto-3g	6-31g	cc-pvdz
l_M	N_0	N_1	N_2

overview for three dimensions involving basis set (B), electron correlation (C), and molecular training set (M), with their corresponding levels l_B, l_C, and l_M.

Thus, any given combination of levels can be specified as the ordered triplet l of respective level indices, $l = (l_C, l_B, l_M)$. For example, the combination CCSD(T)/cc-pvdz, N_1 is encoded by the triplet $l = (l_C = 2, l_B = 2, l_M = 1) = (2,2,1)$.

In order to extend this principle to even more dimensions, we now generalize this approach following the lines of the sparse grid combination technique; cf. Appendix A. We introduce for d dimensions with corresponding levels $l_1, ..., l_d$, which we collect together in the d-dimensional *multi-index* $l = (l_1, ..., l_d)$.

Above, $d = 3$ and l_1 correspond to l_C, l_2 corresponds to l_B, and l_3 corresponds to l_M. Following the notation that the last level index refers to molecular training set size, i.e., $l_d = l_M$, we define the energy $E_{(l)}^{\text{ref}}$ given on a subspace l, and the QML model E_l for each subspace,

$$E_l(\mathbf{R}_q) := \sum_{i=1}^{N_{l_d}} \alpha_i^{(l)} k(\mathbf{R}_q, \mathbf{R}_i) \tag{11}$$

Computing the coefficients $\alpha_i^{(l)}$ for a fixed subspace l is done by solving the regression problem

$$E_{(l)}^{\text{ref}}(\mathbf{R}_j) \approx \sum_{i=1}^{N_{l_d}} \alpha_i^{(l)} k(\mathbf{R}_j, \mathbf{R}_i) \tag{12}$$

for all $j = 1, ..., N_{l_d}$.

The generalized CQML machine learning model is then given as

$$E_I(\mathbf{R}_q) := \sum_{l \in \mathcal{I}} \beta_l \sum_{i=1}^{N_{ld}} \alpha_i^{(l)} k(\mathbf{R}_q, \mathbf{R}_i) \tag{13}$$

In fact, it is the combination of the machine learning models from eq 11 for different subspaces l that are collected in the *index set* $\mathcal{I}$. The classical sparse grid combination technique proposes to use the index set

$$\mathcal{I} := \{l \in \mathbb{N}^d \,|\, \|l\|_1 = (L-1) - i, \, i \in \{0, \dots, d-1\}\} \tag{14}$$

with $\|l\|_1 := \sum_{s=1}^{d} l_s$. In the following, the coefficients β_l can always be evaluated as[89]

$$\beta_l := \sum_{z \in \{0,1\}^d} (-1)^{\|z\|_1} \chi_{\mathcal{I}}(l+z) \tag{15}$$

Here, the sum is to be understood in the sense that vector z of size d takes all possible combinations of zeros and ones. Moreover, we define the characteristic function $\chi_{\mathcal{I}}$ of index set $\mathcal{I}$ by

$$\chi_{\mathcal{I}}(l+z) := \begin{cases} 1 & \text{if } (l+z) \in \mathcal{I} \\ 0 & \text{else} \end{cases} \tag{16}$$

It is well-known that the above choice of the index set $\mathcal{I}$ and coefficients β_l in $d = 2$ is equivalent to the multilevel learning approach from section 3.4.

For $d = 3$ the above choice of index set $\mathcal{I}$ would lead to the CQML model $E_{(2,2,2)}$ with

$$\begin{aligned} E_{(2,2,2)}(\mathbf{R}_q) = \; &E_{(0,2,0)}(\mathbf{R}_q) - 2E_{(0,1,0)}(\mathbf{R}_q) + E_{(1,1,0)}(\mathbf{R}_q) \\ &+ E_{(0,1,1)}(\mathbf{R}_q) - 2E_{(1,0,0)}(\mathbf{R}_q) + E_{(0,0,0)}(\mathbf{R}_q) \\ &- 2E_{(0,0,1)}(\mathbf{R}_q) + E_{(2,0,0)}(\mathbf{R}_q) + E_{(1,0,1)}(\mathbf{R}_q) \\ &+ E_{(0,0,2)}(\mathbf{R}_q) \end{aligned} \tag{17}$$

and $l_C, l_B, l_M = 0, \dots, 2$, exemplified with the spaces discussed in section 1. Note, however, that this choice does not use any training data from the target subspace, here CCSD(T) calculations with a cc-pvdz basis set. In practice, it is preferable to include the corresponding subspaces with this accuracy to the training set, at least with a small training set size, in order to include the physics of the corresponding target subspace. To this end, in $d = 3$, we shift the index set $\mathcal{I}$ such that the subspace choice from Figure 2 is achieved. The resulting index set becomes

$$\begin{aligned} \mathcal{I}_{\text{shifted}} := \{(l_C, l_B, l_M) | l_C, \, l_B \in [0,2], \, l_M \in [0,4], \\ l_C + l_B + l_M \in \{2,3,4\}\} \end{aligned} \tag{18}$$

The corresponding CQML model then becomes $E_{(2,2,2)}^{\text{shifted}}$ (cf. Appendix A) and reads

$$\begin{aligned} E_{(2,2,2)}^{\text{shifted}}(\mathbf{R}_q) = \; &-2E_{(1,2,0)}(\mathbf{R}_q) + E_{(0,2,0)}(\mathbf{R}_q) - 2E_{(0,2,1)}(\mathbf{R}_q) \\ &+ E_{(2,2,0)}(\mathbf{R}_q) + E_{(1,2,1)}(\mathbf{R}_q) + E_{(0,2,2)}(\mathbf{R}_q) \\ &- 2E_{(2,1,0)}(\mathbf{R}_q) + E_{(1,1,0)}(\mathbf{R}_q) - 2E_{(1,1,1)}(\mathbf{R}_q) \\ &+ E_{(0,1,1)}(\mathbf{R}_q) - 2E_{(0,1,2)}(\mathbf{R}_q) + E_{(2,1,1)}(\mathbf{R}_q) \\ &+ E_{(1,1,2)}(\mathbf{R}_q) + E_{(0,1,3)}(\mathbf{R}_q) + E_{(2,0,0)}(\mathbf{R}_q) \\ &- 2E_{(2,0,1)}(\mathbf{R}_q) + E_{(1,0,1)}(\mathbf{R}_q) - 2E_{(1,0,2)}(\mathbf{R}_q) \\ &+ E_{(0,0,2)}(\mathbf{R}_q) - 2E_{(0,0,3)}(\mathbf{R}_q) + E_{(2,0,2)}(\mathbf{R}_q) \\ &+ E_{(1,0,3)}(\mathbf{R}_q) + E_{(0,0,4)}(\mathbf{R}_q) \end{aligned} \tag{19}$$

with $l_C, l_B, l_M = 0, \dots, 2$. The reader is referred to Appendix A for more details of the CQML derivation.

4. RESULTS AND DISCUSSION

Before entering the detailed discussion of our results, we now briefly discuss the use of learning curves as a measure of machine learning model quality. Clearly, reporting a single out-of-sample error for any machine learning model is hardly meaningful: It is the very point of machine learning that models should improve with training set size. Vapnik and co-workers[93,95] discussed already in the nineties that prediction errors, i.e., out-of-sample estimates of statistically estimated functions, decay inversely with training set size N. More specifically, for kernel ridge regression models (used throughout this study), the leading prediction error term was shown to be proportional to a/N^b, where a and b are proportionality constant and power law exponent, respectively.[93−95] In order to facilitate comparison among models, it is therefore recommended practice[28] to discuss the performance in terms of learning curves on log−log scales, i.e., for prediction errors decaying linearly with training set size, i.e., $\log(\text{Error}) = \log(a) - b \log(N)$. Saturation of errors indicates failure to learn, and small offsets and steep slopes indicate preferable models.

4.1. Data. For all the $\sim$7000 QM7b molecules,[9] we have calculated total energies for all combinations among the various levels of correlation energies (HF, MP2, CCSD(T)) and basis set sizes (sto-3g, 6-31g, cc-pvdz). Resulting effective atomization energies (see the Supporting Information for the entire data set), are shown within scatter plots in Figure 3. Depending on stoichiometry and size, the molecules spread out over the various levels and dimensions.

More specifically, molecules can be divided into two clusters: the one dominating the distribution is almost sulfur-free and the other cluster of molecules, clearly separated from the majority, containing sulfur atoms (see bottom row in Figure 3). This pattern indicates that sto-3g and 6-31g are too small basis sets and should not be used to describe S containing molecules. By comparing the three figures in each column of the first three rows in Figure 3, one can see that the shape of distribution changes significantly upon introduction of electron correlation (going from HF treatment to the MP2). From MP2 to CCSD(T), however, the change in the distribution is barely noticeable.

Considering the right-hand panel in the third row in Figure 3, the color code corresponds exactly to the correlation energy contribution to the atomization energy, as estimated by CCSD(T)-HF within cc-pvdz basis. As one would expect, the larger the molecule, the more electron correlation energy is being contributed. The 200 molecules with the largest and smallest correlation energy contribution to the atomization energy are on display in Figure 4. We note that molecules with high degrees of saturation exhibit the largest amount of electron correlation in their atomization energy, while atomization energies of molecules with multiple double bonds, triple bonds, and aromatic moieties contain the least electron correlation energy. This trend is to be expected because the electrons in unsaturated bonding patterns can contribute less to binding than in saturated species, thereby also decreasing their electron correlation energy contribution to binding.

The reason for developing the CQML model is based on the hypothesis that it will systematically exploit all these underlying implicit correlations that are on display in these figures.

4.2. 2D Results for QM7b. As a first test, we have investigated our QM7b derived data set for the two-dimensional
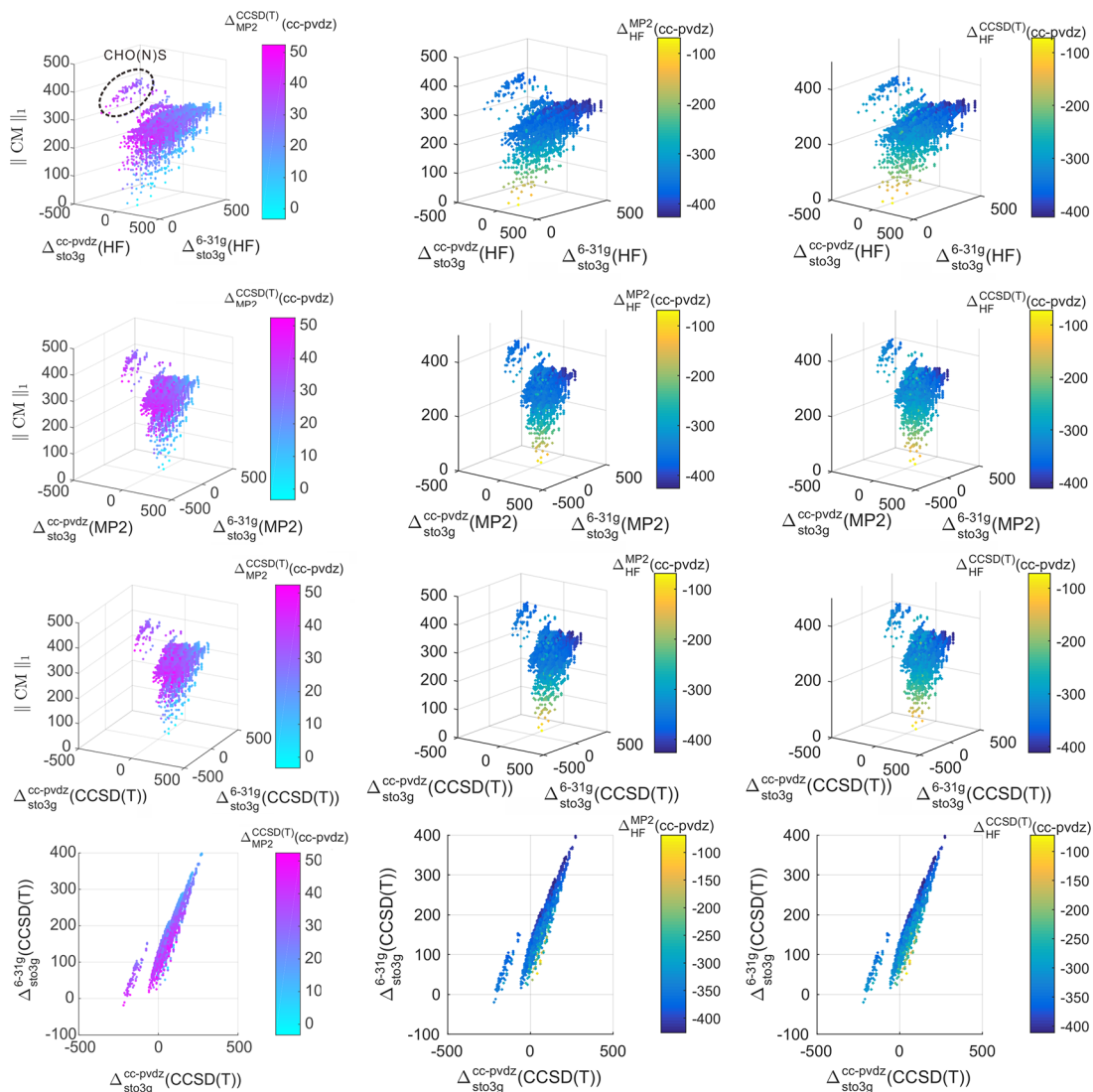
Figure 3. Scatter plots for QM7b. Size in chemical space as measured by 1-norm of Coulomb matrix [au] (i.e., $\|CM\|_1$) vs energy differences [kcal/mol] due to various basis set size differences for HF (first row), MP2 (second row), and CCSD(T) (third row). The color code corresponds to the atomization energy difference Δ [kcal/mol] between electron correlation models at cc-pvdz for MP2 vs CCSD(T) (left), HF vs MP2 (mid), and HF vs CCSD(T). In the upper leftmost panel, the brackets enclosing N indicate that nitrogen atoms may or may not be present. The bottom row corresponds to the 2D projection of the third row.

($d = 2$) case of atomization energies at a fixed basis set (cc-pvdz) for three levels of electron correlation, i.e., HF ($l_C = 0$), MP2 ($l_C = 1$), and CCSD(T) ($l_C = 2$). The second dimension corresponds to three variable molecular training data set sizes ($l_M = 0, 1, 2$). Their relative extent is fixed at ratios independent of absolute training set size. In this study, we considered two such sets of ratios ($s = 1$ and $s = 2$) that reflect different sample size increases for higher levels. These ratios are summarized in Table 2. The number of training molecules N_{l_M} on each level of the CQML with $d = 2$ as a function of training set size at the highest level $N_{l_M=2}$ is thus given by $N_{l_M} = r_{l_M} \times N_{l_M=2}$, where r_{l_M} is the ratio as displayed in Table 2. Recall that all ML model results presented in this section have been obtained by using kernel ridge regression, a Laplacian kernel, and the SLATM[16] representation.

In Figure 5, various learning curves for atomization energies, estimated according to eq 7, are shown. First of all, we note the rapid and systematic lowering for all CQML models as training

set size increases. The models exhibit differing offsets, and similar slopes, in line with previous results for training-set optimization experiments using ensembles of training sets within genetic optimization protocols.[68] The learning curves of conventional QML pass the chemical accuracy threshold (~1 kcal/mol) at ~4000 training molecules calculated at target level, CCSD(T)/cc-pvdz. This learning curve has a slightly larger offset with respect to the original SLATM benchmark results (see the Supporting Information in ref 96) due to the use of (i) the Laplacian instead of a Gaussian kernel function, (ii) B3LYP rather than PBE0 geometries, and (iii) CCSD(T) rather than PBE0 energies.

Addition of MP2 reference energies of further molecules affords a systematic decrease in the learning offset resulting in ~2000 and ~1000 CCSD(T) training molecules necessary to reach chemical accuracy for $s = 1$ and $s = 2$, respectively. The corresponding necessary MP2 training set sizes (not shown in the figure) amount to 4000 molecules for both s-values (see Table 2). Slightly worse results are obtained by replacing MP2 reference energies with HF energies. This result may seem
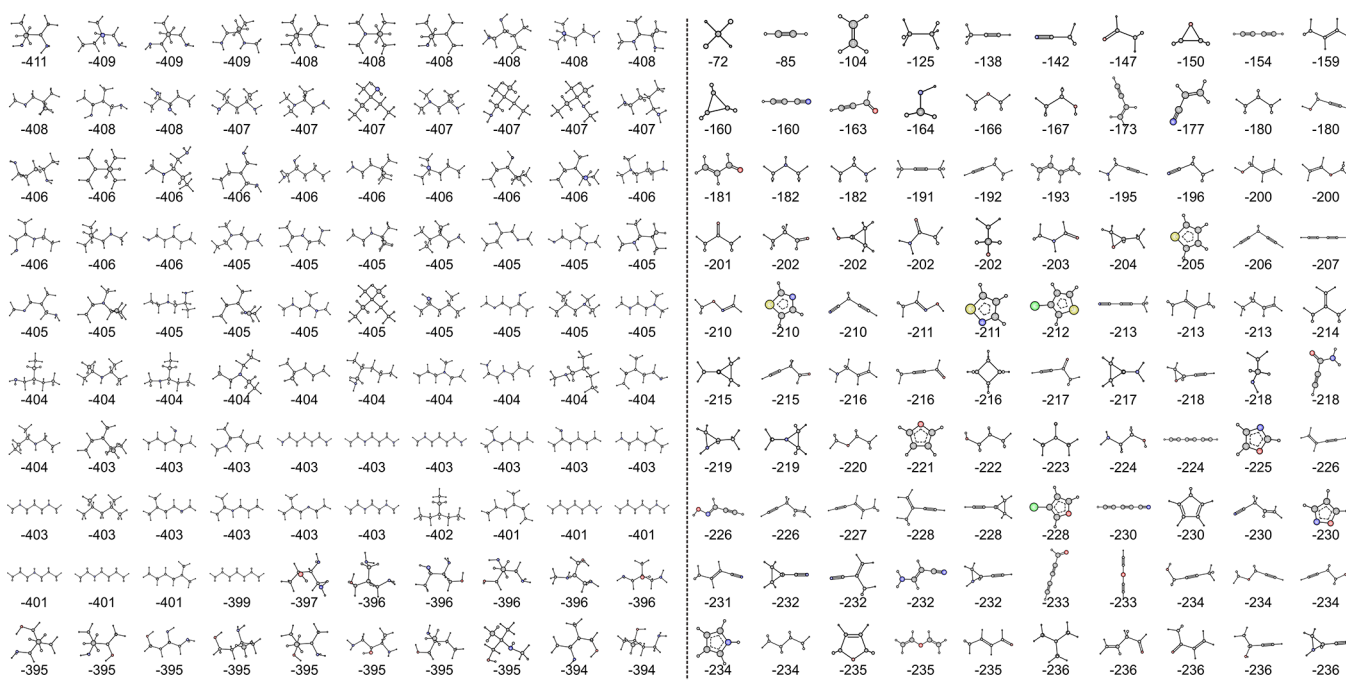
Figure 4. Two hundred QM7b molecules with largest (left) and smallest (right) electron correlation energy contributions to the atomization energy (CCSD(T)-HF within cc-pvdz basis [kcal/mol]), respectively. See the Supporting Information for the complete data set. Color key: white, H; gray, C; yellow, S; red, O; blue, N; green, Cl.

Table 2. Level-Dependent Ratios between Training Set Sizes for the Two Sample Size Increases s Considered[a]

s	$r_{l_M = L-1}$	$r_{l_M = L-2}$	$r_{l_M = L-3}$
1	1	2	4
2	1	4	16

[a] L is the total number of levels.

puzzling but is in full agreement with what we have found in Figure 3; i.e., the values of $\Delta_{MP2}^{CCSD(T)}$ and Δ_{HF}^{MP2} are of the same magnitude. This result also implies the possibility to optimize the levels of theory by minimizing the computational cost, meanwhile retaining the accuracy.

Adding Hartree–Fock treatment for additional training molecules, we observe even further improvement, reaching chemical accuracy already at ~1000 and ~300 CCSD(T) training molecules for $s = 1$ and $s = 2$, respectively. According to the ratios in Table 2, the corresponding necessary MP2 and HF training set sizes (not shown in the figure) amount respectively to 2000 and 4000 for $s = 1$, and to 1200 and 2400 for $s = 2$.

These results are very encouraging; they suggest that reductions by an order of magnitude are possible with respect to high-level reference numbers (from expensive computation or experiment) necessary to reach chemical accuracy. Effectively, the CQML model appears to exploit correlations inherent among the various approximation levels that live within hierarchical spaces of theories.

4.3. 3D Results for QM7b. We have also studied the extension of the 2D-CQML model by a third dimension ($d = 3$), which explicitly introduces the effect of basis set size on atomization energies. More specifically, we have considered sto-3g ($l_B = 0$) as our lowest level, 6-31g ($l_B = 1$) as an intermediate size, and cc-pvdz ($l_B = 2$) as the largest set. Obviously, larger basis set choices as well as additional levels with more subtle differences could have been included as well.

Here, we assume that the general trend and the conclusions drawn are not affected by the relatively modest size of the basis sets employed.

In Figure 5, we show corresponding learning curves of 2D-CQML models that connect the different basis sets according to eq 7 with just one correlation energy model, CCSD(T). In line with the behavior encountered above for the fixed basis set CQML models, a systematic improvement is found for MAE as well as RMSE (root mean square error). The error approaches chemical accuracy already with ~1000 training examples with the largest basis used (cc-pvdz). Again, increasing the ratios between levels by going from $s = 1$ to $s = 2$ (see Table 2) leads to systematic lowering of the learning curve.

Finally, when multiple basis set and electron correlation levels are combined into a single 3D-CQML model, obtained according to eq 19, the most favorable learning curves are obtained for MAE as well as for RMSE (see Figure 5). For $s = 1$ and $s = 2$, extrapolation indicates that chemical accuracy can be reached with just 500 and 100 training instances at the CCSD(T)/cc-pvdz level, respectively. Note that the learning curves end already for relatively small training set sizes because the necessary number of molecules required at lower levels of theory rapidly reaches the maximal number of available molecules in QM7b. For example, for the $s = 2$ case, 100 training molecules at the highest level combination would have required $100 \times 4^4 = 25\,600$ training molecules at the lowest level combination. However, QM7b is composed of only 7211 molecules. As such, this is an artifact of the finite size of QM7b, and we expect these learning curves to further decay linearly when larger data sets are used in the future.

Overall, these results amount to numerical evidence that it is beneficial to include not only multiple levels but also multiple dimensions. The obvious consequence is that an additional substantial reduction in need for high-level reference numbers (from expensive computation or experiment) is possible
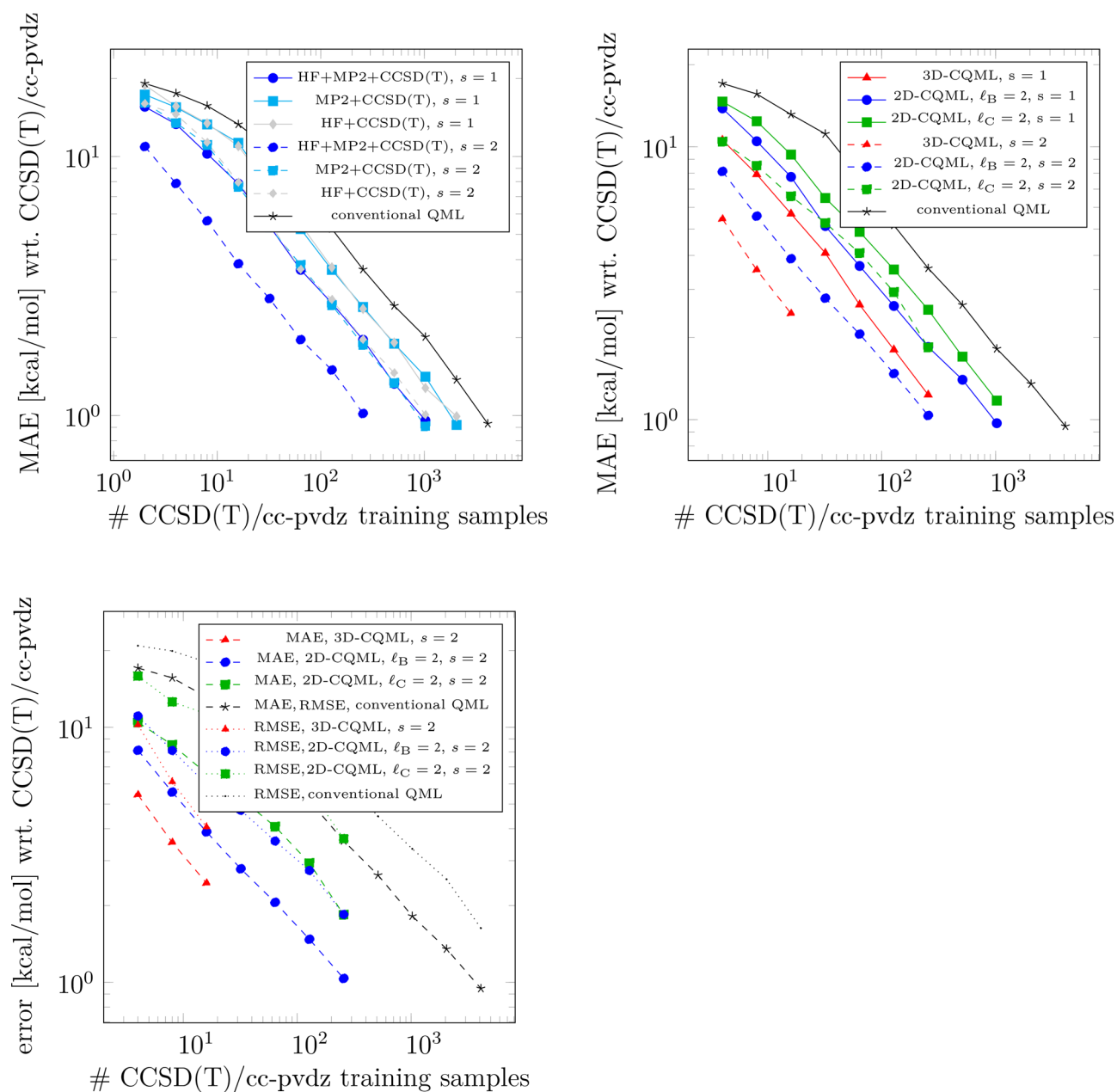
Figure 5. Learning curves for CCSD(T) atomization energies of QM7b molecules for various CQML models. Level ratios considered include $s = 1$ and $s = 2$ (Table 2). Upper left: 2D-CQML at fixed basis set (cc-pvdz) including two (MP2, CCSD(T))/(HF,CCSD(T)) and three levels of electron correlation treatment (HF, MP2, CCSD(T)). Upper right: 2D-CQML (green) at fixed electron correlation treatment (CCSD(T)) for three basis sets (sto-3g, 6-31g, cc-pvdz). 3D-CQML (red) exploiting basis set size (sto-3g, 6-31g, cc-pvdz) *and* electron correlation treatment (HF, MP2, CCSD(T)). Bottom: learning curves for RMSE and MAE for the machines in the upper right panel with $s = 2$. Note that the horizontal axis in all three figures may also be chosen to represent the number of training samples from other levels, which can be obtained by rescaling the current axis with a ratio of s, as shown in Table 2.

through the use of CQML based exploitation of training data obtained for smaller basis sets and more approximate electron correlation models. We believe that this is possible because of inherent error-cancellation between various levels *and* dimensions.

Furthermore, the success of CQML indicates that we can push the efforts further so as to achieve an absolute prediction error on par with experiments by taking advantage of even higher level theory of reference data, such as CCSD(T)-F12/cc-pvqz or quantum Monte Carlo (QMC). More specifically, the highest level of theory adopted in this study is CCSD(T)/cc-pvdz, to which the prediction errors are referenced. As a result, the

predictive power of CQML can match at best CCSD(T)/cc-pvdz. Considering that we need very few training instances at the highest level, we are optimistic that at a very low cost (cf. brute force high-level-of-theory calculation), we can eventually achieve chemical accuracy for energy predictions of arbitrary out-of-sample molecules. This will be explored in future work.

4.4. 2D Results for CI9. For the stoichiometrical isomers $C_7H_{10}O_2$, data set CI9, we have also investigated the 2D-CQML model corresponding to eq 7. The resulting models differ from the previous 2D-CQML models in that they unite energy approximation effects and basis sets into a single dimension (PM7, B3LYP/6-31g(D), G4MP2). Furthermore, and in

analogy to the original Δ-ML model,[61,69,97−100] all small changes in geometry due to use of different levels of theory are also being accounted for through the ML model. As such, only PM7-quality input geometries are required for the 2D-CQML models discussed in this section. Resulting learning curves are shown in Figure 6 for two different representations, the Coulomb
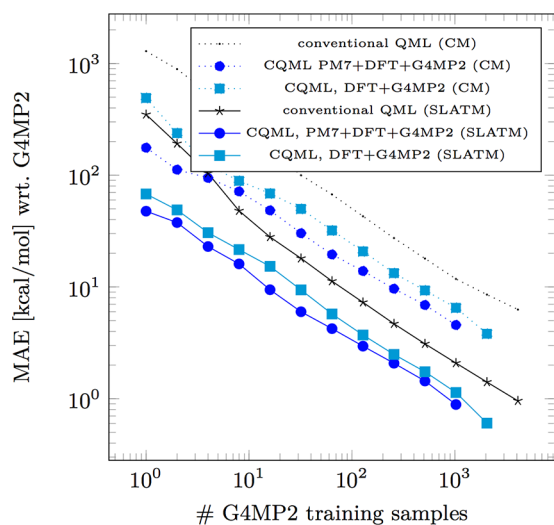

Figure 6. Prediction errors of atomization energies in the CI9 data set (consitutional isomers of $C_7H_{10}O_2$) vs number of training molecules with G4MP2 energies for various 2D-CQML models. Results differ by representation (SLATM vs CM) and number of levels included.

matrix,[57,58] and SLATM,[96] as well as for two different number of levels ($L = 2$ and $L = 3$).

Again, when comparing to conventional QML, we note systematic and improved (through lower offsets) learning as the number of different levels increases from two to three. The relative performance for Coulomb matrix and SLATM meets the expected trend,[20] SLATM systematically leading to a substantially lower offset. These results suggest a certain independence of the CQML methodology from other salient features of QML models, such as training set selection[68,96] or choice of representation.[20,24] In this case, the best 2D-CQML SLATM based model reaches chemical accuracy with respect to G4MP2 based on a training set consisting of ∼1000, 2000, and 4000 at G4MP2, B3LYP/6-31g(d), and PM7 level reference results, respectively.

5. CONCLUSIONS

We have extended the ideas manifested in Pople diagrams within the systematic framework of the multilevel sparse grid combination technique and machine learning. A generalized CQML model has been presented, and we have demonstrated its performance for various 2D variants and for one 3D application using atomization energies of organic molecules as property of interest. Using learning curves to compare models, we have found for all cases investigated that the addition of levels and spaces enables a systematic and substantial reduction in necessary training data at the highest level of theory. As such, we have shown how to construct QML models for which an expensive training molecule can be replaced by multiple cheaper training molecules. Due to the unfavorable polynomial scaling and large prefactors of the more expensive quantum approximations, such trade-offs can deliver significantly more accurate QML models at constant training data compute budget. In conclusion, our numerical findings support the idea that there is an

additional "knob" one can use to improve QML models: In addition to improved representations[20,24] or training set selection,[68,96] one can also exploit the intrinsic correlations among the various hierarchies that exist among different levels of approximations.

For future work, we will consider the inclusion of more intermediate levels, e.g., the various rungs on Jacob's ladder, or MP4, CCSD, CCSDT(Q), etc., or continuous changes in basis set size through plane waves. Other dimensions, such as relativistic effects, spin−orbit coupling, or nuclear quantum effects, can be envisioned. While we have focused on atomization energies only for this study, we will consider CQML models of other quantum properties within subsequent studies. Technical settings can also be investigated, e.g., the relative amount of training data obtained at different levels (currently set globally through parameter s), could still be adapted in a locally optimal manner. Finally, we plan to include this implementation in qmlcode.[101]

■ DERIVATION OF THE COMBINATION TECHNIQUE FOR QUANTUM MACHINE LEARNING

In applied mathematics, the sparse grid combination technique is a means to approximate, e.g., high-dimensional functions. Lets assume that such a function f is in some (function) space $V := V^{(1)} \otimes V^{(2)} \otimes \cdots \otimes V^{(d)}$. That is, it is in the tensor product of d spaces. Then, we introduce for each of the L_m-dimensional function spaces $V^{(m)}$ a series of subspaces of lower dimension

$$V_0^{(m)} \subset V_1^{(m)} \subset V_j^{(m)} \subset ... \subset V_{L_m}^{(m)} \tag{20}$$

(indicated by the lower index). Classic (full tensor product) approximation would now approximate this function f on a level j in the space $V_j := V_j^{(1)} \otimes V_j^{(2)} \otimes \cdots \otimes V_j^{(d)}$. However, this leads to the so-called *curse of dimensionality*, i.e., the exponential growth in computational work with growing dimension d.

In many cases, the sparse grid combination technique allows us to approximate f in a much cheaper way. This is done by recursively introducing the sparse approximation space $\hat{V}_j$ with

$$\hat{V}_j^{(d)} := \sum_{k=0}^{j} (V_{j-k}^{(d)} - V_{j-1-k}^{(d)}) \otimes \hat{V}_k^{(d-1)} \tag{21}$$

where $\hat{V}_k^{(d-1)}$ is the sparse approximation space

$$\hat{V}_j^{(d-1)} := \sum_{k=0}^{j} (V_{j-k}^{(d-1)} - V_{j-1-k}^{(d-1)}) \otimes \hat{V}_k^{(d-2)} \tag{22}$$

That is, it is recursively built from the first $d − 1$ spaces in the same way.

In this work, we transfer this approach to the field of quantum machine learning. To this end, we provide a derivation for the combination technique for quantum machine learning in two and three dimensions/spaces. Let us first briefly introduce a general machine learning model for a given subspace (l_C, l_B, l_M). Note that we assume here that $l_C, l_B, l_M \in \{0, ..., L\}$. The general ML model for a given subspace reads as

$$E_{(l_C, l_B, l_M)}(\mathbf{R}_q) := \sum_{i=1}^{N_{l_M}} \alpha_i^{(l_C l_B l_M)} k(\mathbf{R}_q, \mathbf{R}_i) \tag{23}$$

We identify this model with some subspace $V_{l_C}^{(1)} \otimes V_{l_B}^{(2)} \otimes V_{l_M}^{(3)}$. Following eq 21, the two-dimensional combination technique for QML on level j_2 for the spaces of theory and training set size and a fixed basis set level l_B can be introduced as

$$E_{(j_2, l_B, j_2)}(\mathbf{R}_q) := \sum_{k_2=0}^{j_2} E_{(j_2-k_2, l_B, k_2)}(\mathbf{R}_q) - E_{(j_2-1-k_2, l_B, k_2)}(\mathbf{R}_q) \tag{24}$$

Note that, whenever a level index becomes negative, we assume the machine learning model to be exactly zero, i.e.

$$E_{(-1,\cdot,\cdot)} \equiv E_{(\cdot,-1,\cdot)} \equiv E_{(\cdot,\cdot,-1)} \equiv 0 \tag{25}$$

For the choice of $j_2 = 2$ and $l_B = 2$, we can explicitly derive

$$
\begin{aligned}
E_{(2,2,2)}(\mathbf{R}_q) &= (E_{(2-0,2,0)}(\mathbf{R}_q) - E_{(2-1-0,2,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(2-1,2,1)}(\mathbf{R}_q) - E_{(2-1-1,2,1)}(\mathbf{R}_q)) \\
&\quad + (E_{(2-2,2,2)}(\mathbf{R}_q) - E_{(2-1-2,2,2)}(\mathbf{R}_q)) \\
&= (E_{(2,2,0)}(\mathbf{R}_q) - E_{(1,2,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(1,2,1)}(\mathbf{R}_q) - E_{(0,2,1)}(\mathbf{R}_q)) \\
&\quad + (E_{(0,2,2)}(\mathbf{R}_q) - E_{(-1,2,2)}(\mathbf{R}_q)) \\
&= E_{(2,2,0)}(\mathbf{R}_q) - E_{(1,2,0)}(\mathbf{R}_q) + E_{(1,2,1)}(\mathbf{R}_q) \\
&\quad - E_{(0,2,1)}(\mathbf{R}_q) + E_{(0,2,2)}(\mathbf{R}_q) \tag{26}
\end{aligned}
$$

Note that we have the equalities

$$E_{(2,2,0)}(\mathbf{R}_q) - E_{(1,2,0)}(\mathbf{R}_q) = \sum_i^{N_2} \alpha_i^{(1,2)} k(\mathbf{R}_q, \mathbf{R}_i)$$

$$E_{(1,2,1)}(\mathbf{R}_q) - E_{(0,2,1)}(\mathbf{R}_q) = \sum_i^{N_1} \alpha_i^{(0,1)} k(\mathbf{R}_q, \mathbf{R}_i)$$

$$E_{(0,2,2)}(\mathbf{R}_q) = \sum_i^{N_0} \alpha_i^{(0)} k(\mathbf{R}_q, \mathbf{R}_i) \tag{27}$$

with the notation from section 3.4. That is, model $E_{(2,2,2)}$, as derived here, is exactly the model discussed in section 3.4.

On the basis of the two-dimensional combination technique model, we can now recursively build a three-dimensional combination technique further integrating the space of basis set size and with the global three-dimensional level j_3 as follows

$$E_{(j_3, j_3, j_3)}(\mathbf{R}_q) := \sum_{k_3=0}^{j_3} E_{(k_3, j_3-k_3, k_3)}(\mathbf{R}_q) - E_{(k_3, j_3-1-k_3, k_3)}(\mathbf{R}_q) \tag{28}$$

This construction uses the definition of the two-dimensional combination technique in a recursive fashion.

We finally exemplify the three-dimensional combination technique for $j_3 = 2$. That is, we first expand the recursive model for the three-dimensional combination technique by

$$
\begin{aligned}
E_{(2,2,2)}(\mathbf{R}_q) &= (E_{(0,2-0,0)}(\mathbf{R}_q) - E_{(0,2-1-0,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(1,2-1,1)}(\mathbf{R}_q) - E_{(1,2-1-1,1)}(\mathbf{R}_q)) \\
&\quad + (E_{(2,2-2,2)}(\mathbf{R}_q) - E_{(2,2-1-2,2)}(\mathbf{R}_q)) \\
&= (E_{(0,2,0)}(\mathbf{R}_q) - E_{(0,1,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(1,1,1)}(\mathbf{R}_q) - E_{(1,0,1)}(\mathbf{R}_q)) \\
&\quad + (E_{(2,0,2)}(\mathbf{R}_q) - E_{(2,-1,2)}(\mathbf{R}_q)) \\
&= (E_{(0,2,0)}(\mathbf{R}_q) - E_{(0,1,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(1,1,1)}(\mathbf{R}_q) - E_{(1,0,1)}(\mathbf{R}_q)) + E_{(2,0,2)}(\mathbf{R}_q) \tag{29}
\end{aligned}
$$

Then, we expand each of the term by means of the two-dimensional combination technique. Thus, we compute

$$
\begin{aligned}
E_{(0,2,0)}(\mathbf{R}_q) &= E_{(0-0,2,0)}(\mathbf{R}_q) - E_{(0-1-0,2,0)}(\mathbf{R}_q) \\
&= E_{(0,2,0)}(\mathbf{R}_q) \tag{30}
\end{aligned}
$$

$$
\begin{aligned}
E_{(0,1,0)}(\mathbf{R}_q) &= E_{(0-0,1,0)}(\mathbf{R}_q) - E_{(0-1-0,1,0)}(\mathbf{R}_q) \\
&= E_{(0,1,0)}(\mathbf{R}_q) \tag{31}
\end{aligned}
$$

$$
\begin{aligned}
E_{(1,1,1)}(\mathbf{R}_q) &= E_{(1-0,1,0)}(\mathbf{R}_q) - E_{(1-1-0,1,0)}(\mathbf{R}_q) \\
&\quad + E_{(1-1,1,1)}(\mathbf{R}_q) - E_{(1-1-1,1,1)}(\mathbf{R}_q) \\
&= E_{(1,1,0)}(\mathbf{R}_q) - E_{(0,1,0)}(\mathbf{R}_q) + E_{(0,1,1)}(\mathbf{R}_q) \tag{32}
\end{aligned}
$$

$$
\begin{aligned}
E_{(1,0,1)}(\mathbf{R}_q) &= E_{(1-0,0,0)}(\mathbf{R}_q) - E_{(1-1-0,0,0)}(\mathbf{R}_q) \\
&\quad + E_{(1-1,0,1)}(\mathbf{R}_q) - E_{(1-1-1,0,1)}(\mathbf{R}_q) \\
&= E_{(1,0,0)}(\mathbf{R}_q) - E_{(0,0,0)}(\mathbf{R}_q) + E_{(0,0,1)}(\mathbf{R}_q) \tag{33}
\end{aligned}
$$

$$
\begin{aligned}
E_{(2,0,2)}(\mathbf{R}_q) &= (E_{(2-0,0,0)}(\mathbf{R}_q) - E_{(2-1-0,0,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(2-1,0,1)}(\mathbf{R}_q) - E_{(2-1-1,0,1)}(\mathbf{R}_q)) \\
&\quad + (E_{(2-2,0,2)}(\mathbf{R}_q) - E_{(2-1-2,0,2)}(\mathbf{R}_q)) \\
&= (E_{(2,0,0)}(\mathbf{R}_q) - E_{(1,0,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(1,0,1)}(\mathbf{R}_q) - E_{(0,0,1)}(\mathbf{R}_q)) \\
&\quad + (E_{(0,0,2)}(\mathbf{R}_q) - E_{(-1,0,2)}(\mathbf{R}_q)) \\
&= (E_{(2,0,0)}(\mathbf{R}_q) - E_{(1,0,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(1,0,1)}(\mathbf{R}_q) - E_{(0,0,1)}(\mathbf{R}_q)) + E_{(0,0,2)}(\mathbf{R}_q) \tag{34}
\end{aligned}
$$

Finally, we combine these results with the previous calculations for $E_{(2,2,2)}$ and obtain

$$
\begin{aligned}
E_{(2,2,2)}(\mathbf{R}_q) &= E_{(0,2,0)}(\mathbf{R}_q) - E_{(0,1,0)}(\mathbf{R}_q) + E_{(1,1,1)}(\mathbf{R}_q) \\
&\quad - E_{(1,0,1)}(\mathbf{R}_q) + E_{(2,0,2)}(\mathbf{R}_q) \\
&= E_{(0,2,0)}(\mathbf{R}_q) - E_{(0,1,0)}(\mathbf{R}_q) + E_{(0,1,1)}(\mathbf{R}_q)] \\
&\quad - [E_{(1,0,0)}(\mathbf{R}_q) - E_{(0,0,0)}(\mathbf{R}_q) + E_{(0,0,1)}(\mathbf{R}_q)] \\
&\quad + [(E_{(2,0,0)}(\mathbf{R}_q) - E_{(1,0,0)}(\mathbf{R}_q)) \\
&\quad + (E_{(1,0,1)}(\mathbf{R}_q) - E_{(0,0,1)}(\mathbf{R}_q)) + E_{(0,0,2)}(\mathbf{R}_q)] \\
&= E_{(0,2,0)}(\mathbf{R}_q) - 2E_{(0,1,0)}(\mathbf{R}_q) + E_{(1,1,0)}(\mathbf{R}_q) \\
&\quad + E_{(0,1,1)}(\mathbf{R}_q) - 2E_{(1,0,0)}(\mathbf{R}_q) + E_{(0,0,0)}(\mathbf{R}_q) \\
&\quad - 2E_{(0,0,1)}(\mathbf{R}_q) + E_{(2,0,0)}(\mathbf{R}_q) + E_{(1,0,1)}(\mathbf{R}_q) \\
&\quad + E_{(0,0,2)}(\mathbf{R}_q) \tag{35}
\end{aligned}
$$

This is exactly the spelled out version of eq 17 for $E_{(2,2,2)}$.

As discussed in section 3.5, practical considerations motivate us to modify the machine learning model $E_{(2,2,2)}$ to further include the training data from subspace (2,2,0), i.e., data from the most expensive subspace with CCSD(T) calculations on a

cc-pvdz basis set. Therefore, we introduced the *shifted* machine learning model $E_{(2,2,2)}^{shifted}$ in section 3.5. In principle, it can be computed by using the modified index set $\mathcal{I}_{shifted}$ together with eqs 15 and 16. However, we can also recursively derive the shifted model by first computing the machine learning model $E_{(4,4,4)}$, which, derived similarly to $E_{(2,2,2)}$, becomes

$$
\begin{aligned}
E_{(4,4,4)}(\mathbf{R}_q) = {} & E_{(0,4,0)}(\mathbf{R}_q) + E_{(0,3,1)}(\mathbf{R}_q) + E_{(1,3,0)}(\mathbf{R}_q) \\
& + E_{(0,2,2)}(\mathbf{R}_q) + E_{(1,2,1)}(\mathbf{R}_q) + E_{(2,2,0)}(\mathbf{R}_q) \\
& + E_{(0,1,3)}(\mathbf{R}_q) + E_{(1,1,2)}(\mathbf{R}_q) + E_{(2,1,1)}(\mathbf{R}_q) \\
& + E_{(3,1,0)}(\mathbf{R}_q) + E_{(0,0,4)}(\mathbf{R}_q) + E_{(1,0,3)}(\mathbf{R}_q) \\
& + E_{(2,0,2)}(\mathbf{R}_q) + E_{(3,0,1)}(\mathbf{R}_q) + E_{(4,0,0)}(\mathbf{R}_q) \\
& - 2E_{(0,3,0)}(\mathbf{R}_q) - 2E_{(0,2,1)}(\mathbf{R}_q) - 2E_{(1,2,0)}(\mathbf{R}_q) \\
& - 2E_{(0,1,2)}(\mathbf{R}_q) - 2E_{(1,1,1)}(\mathbf{R}_q) - 2E_{(2,1,0)}(\mathbf{R}_q) \\
& - 2E_{(0,0,3)}(\mathbf{R}_q) - 2E_{(1,0,2)}(\mathbf{R}_q) - 2E_{(2,0,1)}(\mathbf{R}_q) \\
& - 2E_{(3,0,0)}(\mathbf{R}_q) + E_{(0,2,0)}(\mathbf{R}_q) + E_{(0,1,1)}(\mathbf{R}_q) \\
& + E_{(1,1,0)}(\mathbf{R}_q) + E_{(0,0,2)}(\mathbf{R}_q) + E_{(1,0,1)}(\mathbf{R}_q) \\
& + E_{(2,0,0)}(\mathbf{R}_q)
\end{aligned}
\tag{36}
$$

In a second step, we exclude those subspaces that are not contained in $\mathcal{I}_{shifted}$, i.e., the subspaces for the multi-indices

$$
(0,4,0), \ (0,3,1), \ (1,3,0), \ (3,1,0)
$$

$$
(3,0,1), \ (4,0,0), \ (0,3,0), \ (3,0,0)
\tag{37}
$$

For symmetry reasons and to keep a valid combination technique rule for arbitrary index sets, we also have to exclude one contribution of the subspaces with multi-indices

$$
(2,0,0), \ (2,1,0), \ (2,0,1)
\tag{38}
$$

The resulting shifted machine learning model $E_{(2,2,2)}^{shifted}$ is given in eq 19.

ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00832.

> Geometries as xyz files; two types of energy data for each of the three basis sets (sto-3g, 6-31g, and cc-pvdz) [i.e., total energy (E) and effective averaged atomization energies (E^*), the latter defined as $E - \sum_I n_I e_I$, where n_I is the number of atom I in the molecule and e_I is the effective atomic energy of atom I obtained through a linear least-square fit of $E = \sum_I n_I e_I$ for all molecules in the data set]; free atom energies for all basis sets and electron methods; energy data for any basis set used (txt format), consisting of three columns representing HF, MP2, and CCSD(T) energies, respectively (ZIP)

AUTHOR INFORMATION

Corresponding Authors
*H. Harbrecht. E-mail: helmut.harbrecht@unibas.ch.
*O. A. von Lilienfeld. E-mail: anatole.vonlilienfeld@unibas.ch.
ORCID 🄳
O. Anatole von Lilienfeld: 0000-0001-7419-0466

Notes
The authors declare no competing financial interest.

REFERENCES
(1) Kirkpatrick, P.; Ellis, C. Chemical space. *Nature* **2004**, *432*, 823.

(2) Mullard, A. The drug-maker's guide to the galaxy. *Nature* **2017**, *549*, 445.

(3) Ceder, G. Predicting properties from scratch. *Science* **1998**, *280*, 1099−1100.

(4) Hafner, J.; Wolverton, C.; Ceder, G.; Editors, G. Toward computational materials design: The impact of density functional theory on materials research. *MRS Bull.* **2006**, *31*, 659.

(5) von Lilienfeld, O. A. In *Many-Electron Approaches in Physics, Chemistry and Mathematics*; Bach, V., Site, L. D., Eds.; Mathematical Physics Studies; Springer, 2014; Vol. *IX*, p 417.

(6) von Lilienfeld, O. A. First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties. *Int. J. Quantum Chem.* **2013**, *113*, 1676−1689.

(7) 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.

(8) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **2013**, *3*, 2810.

(9) 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.

(10) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.

(11) Pyzer-Knapp, E. O.; Li, K.; Aspuru-Guzik, A. Learning from the Harvard Clean Energy Project: The Use of Neural Networks to Accelerate Materials Discovery. *Adv. Funct. Mater.* **2015**, *25*, 6495−6502.

(12) Faber, F. A.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R. Machine Learning Energies of 2 Million Elpasolite (ABC_2D_6) Crystals. *Phys. Rev. Lett.* **2016**, *117*, 135502.

(13) De, S.; Bartok, A. P.; Csanyi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754−13769.

(14) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, No. e1603015.

(15) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Prediction errors of molecular machine learning models lower than hybrid DFT error. *J. Chem. Theory Comput.* **2017**, *13*, 5255−5264.

(16) Huang, B.; von Lilienfeld, O. The DNA of chemistry: Scalable quantum machine learning with amons. *arXiv:1707.04146* **2017**.

(17) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Science Advances* **2017**, *3*, e1701816.

(18) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *Proceedings of the*

34th International Conference on Machine Learning, ICML 2017; ICML, 2017.

(19) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet—A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(20) 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*, 241717.

(21) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106.

(22) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 184115.

(23) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(24) Huang, B.; von Lilienfeld, O. A. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *J. Chem. Phys.* **2016**, *145*, 161102.

(25) Janet, J. P.; Kulik, H. J. Predicting electronic structure properties of transition metal complexes with neural networks. *Chemical Science* **2017**, *8*, 5137−5152.

(26) Gastegger, M.; Behler, J.; Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chemical science* **2017**, *8*, 6924−6935.

(27) Ramakrishnan, R.; von Lilienfeld, O. A. *Reviews in Computational Chemistry*; John Wiley & Sons, Inc., 2017; Vol. *30*, pp 225−256.

(28) von Lilienfeld, O. A. Quantum Machine Learning in Chemical Compound Space. *Angew. Chem., Int. Ed.* **2018**, *57*, 4164.

(29) Huang, B.; Symonds, N. O.; Lilienfeld, O. A. v. Quantum Machine Learning in Chemistry and Materials. *Handbook of Materials Modeling: Methods: Theory and Modeling*; Springer, 2018; pp 1−27.

(30) Rupp, M.; von Lilienfeld, O. A.; Burke, K. Guest Editorial: Special Topic on Data-Enabled Theoretical Chemistry. *J. Chem. Phys.* **2018**, *148*, 241401.

(31) Ramakrishnan, R.; Dral, P.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.

(32) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Sci. Data* **2017**, *4*, 170193.

(33) Ghahremanpour, M. M.; Van Maaren, P. J.; Van Der Spoel, D. The Alexandria library, a quantum-chemical database of molecular properties for force field development. *Sci. Data* **2018**, *5*, 180062.

(34) Koch, W.; Holthausen, M. C. *A chemist's guide to density functional theory*; John Wiley & Sons, 2015.

(35) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289−320.

(36) Pople, J. A. Nobel lecture: Quantum chemical models. *Rev. Mod. Phys.* **1999**, *71*, 1267.

(37) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons, LTD., 2000.

(38) Karton, A. A computational chemist's guide to accurate thermochemistry for organic molecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2016**, *6*, 292−310.

(39) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first-and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221−7230.

(40) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. Gaussian-3 theory using reduced Møller-Plesset order. *J. Chem. Phys.* **1999**, *110*, 4703−4709.

(41) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Gaussian-3 theory using scaled energies. *J. Chem. Phys.* **2000**, *112*, 1125−1132.

(42) Martin, J. M. L.; Oliveira, G. d. *J. Chem. Phys.* **1999**, *111*, 1843.

(43) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129.

(44) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A., Jr. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *J. Chem. Phys.* **1996**, *104*, 2598−2619.

(45) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* **1999**, *110*, 2822−2827.

(46) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VII. Use of the minimum population localization method. *J. Chem. Phys.* **2000**, *112*, 6532−6542.

(47) Pople, J. Two-Dimensional Chart of Quantum Chemistry. *J. Chem. Phys.* **1965**, *43*, S229−S230.

(48) Kefalidis, C.; Castro, L.; Yahia, A.; Perrin, L.; Maron, L. In *Computational Methods in Lanthanide and Actinide Chemistry*; Dolg, M., Ed.; Wiley, 2015.

(49) Karplus, M. Three-dimensional "Pople diagram". *J. Phys. Chem.* **1990**, *94*, 5435−5436.

(50) 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.

(51) Ramakrishnan, R.; Dral, P.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.

(52) Ruddigkeit, L.; van Deursen, R.; Blum, L.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864.

(53) Fink, T.; Bruggesser, H.; Reymond, J.-L. *Angew. Chem., Int. Ed.* **2005**, *44*, 1504−1508.

(54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperst, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 09*; Gaussian Inc.: Wallingford, CT, 2009.

(55) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Györffy, W.; Kats, D.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. *MOLPRO, version 2015.1, a package of ab initio programs.* 2015; http://www.molpro.net.

(56) Stewart, James J. P. *MOPAC2009*; Stewart Computational Chemistry: Colorado Springs, CO, USA, 2008.

(57) 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.

(58) 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*, 3404−3419.

(59) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Science advances* **2017**, *3*, No. e1701816.

(60) Willatt, M. J.; Musil, F.; Ceriotti, M. A Data-Driven Construction of the Periodic Table of the Elements. *arXiv preprint arXiv:1807.00236* **2018**.

(61) Ramakrishnan, R.; Dral, P.; Rupp, M.; von Lilienfeld, O. A. Big Data meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087.

(62) Griebel, M.; Hamaekers, J.; Heber, F. A bond order dissection ANOVA approach for efficient electronic structure calculations. *Oberwolfach Report* **2011**, *32*, 1804−1808.

(63) Griebel, M.; Hamaekers, J.; Heber, F. A bond order dissection anova approach for efficient electronic structure calculations. In *Extraction of Quantifiable Information from Complex Systems*; Springer, 2014; pp 211−235.

(64) Chinnamsetty, S. R.; Griebel, M.; Hamaekers, J. An Adaptive Multiscale Approach for Electronic Structure Methods. *Multiscale Model. Simul.* **2018**, *16*, 752−776.

(65) Karton, A. A computational chemist's guide to accurate thermochemistry for organic molecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2016**, *6*, 292−310.

(66) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first-and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221−7230.

(67) Faber, F.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R. Crystal Structure Representations for Machine Learning Models of Formation Energies. *Int. J. Quantum Chem.* **2015**, *115*, 1094.

(68) Browning, N. J.; Ramakrishnan, R.; von Lilienfeld, O. A.; Roethlisberger, U. Genetic Optimization of Training Sets for Improved Machine Learning Models of Molecular Properties. *J. Phys. Chem. Lett.* **2017**, *8*, 1351.

(69) Ramakrishnan, R.; Hartmann, M.; Tapavicza, E.; von Lilienfeld, O. A. Electronic spectra from TDDFT and machine learning in chemical space. *J. Chem. Phys.* **2015**, *143*, 084111.

(70) Rupp, M.; Ramakrishnan, R.; von Lilienfeld, O. A. Machine Learning for Quantum Mechanical Properties of Atoms in Molecules. *J. Phys. Chem. Lett.* **2015**, *6*, 3309.

(71) Shen, L.; Yang, W. Molecular Dynamics Simulations with Quantum Mechanics/Molecular Mechanics and Adaptive Neural Networks. *J. Chem. Theory Comput.* **2018**, *14*, 1442−1455.

(72) Dral, P. O.; Owens, A.; Yurchenko, S. N.; Thiel, W. Structure-based sampling and self-correcting machine learning for accurate calculations of potential energy surfaces and vibrational levels. *J. Chem. Phys.* **2017**, *146*, 244108.

(73) Simm, G. N.; Reiher, M. Error-Controlled Exploration of Chemical Reaction Networks with Gaussian Processes. *J. Chem. Theory Comput.* **2018**, *14*, 5238.

(74) Schmitz, G.; Christiansen, O. Gaussian process regression to accelerate geometry optimizations relying on numerical differentiation. *J. Chem. Phys.* **2018**, *148*, 241704.

(75) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733.

(76) Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. Machine learning of molecular properties: Locality and active learning. *J. Chem. Phys.* **2018**, *148*, 241727.

(77) Huo, H.; Rupp, M. Unified Representation for Machine Learning of Molecules and Crystals. *arXiv preprint arXiv:1704.06439* **2017**.

(78) Rasmussen, C. E.; Williams, C. K. I. In *Gaussian Processes for Machine Learning*; Dietterich, T., Ed.; MIT Press: Cambridge, 2006; www.GaussianProcess.org.

(79) Bungartz, H.-J.; Griebel, M. Sparse grids. *Acta Numerica* **2004**, *13*, 147−269.

(80) Griebel, M.; Schneider, M.; Zenger, C. A combination technique for the solution of sparse grid problems. In *Iterative methods in linear algebra (Brussels, 1991)*; North-Holland: Amsterdam, 1992; pp 263−281.

(81) Griebel, M.; Harbrecht, H. On the construction of sparse tensor product spaces. *Mathematics of Computation* **2013**, *82*, 975−994.

(82) Griebel, M.; Harbrecht, H. A note on the construction of *L*-fold sparse tensor product spaces. *Constructive Approximation* **2013**, *38*, 235−251.

(83) Griebel, M.; Harbrecht, H. On the convergence of the combination technique, In *Sparse grids and applications—Munich 2012*; Lect. Notes Comput. Sci. Eng.; Springer: Cham, 2014; Vol. *97*, pp 55−74.

(84) Griebel, M.; Knapek, S. Optimized tensor-product approximation spaces. *Constructive Approximation. An International Journal for Approximations and Expansions* **2000**, *16*, 525−540.

(85) Griebel, M.; Knapek, S. Optimized general sparse grid approximation spaces for operator equations. *Mathematics of Computation* **2009**, *78*, 2223−2257.

(86) Hegland, M.; Garcke, J.; Challis, V. The combination technique and some generalisations. *Linear Algebra and its Applications* **2007**, *420*, 249−275.

(87) Pflaum, C. Convergence of the combination technique for second-order elliptic differential equations. *SIAM Journal on Numerical Analysis* **1997**, *34*, 2431−2455.

(88) Reisinger, C. Analysis of linear difference schemes in the sparse grid combination technique. *IMA Journal of Numerical Analysis* **2013**, *33*, 544−581.

(89) Rüttgers, A.; Griebel, M. Multiscale simulation of polymeric fluids using the sparse grid combination technique. *Applied Mathematics and Computation* **2018**, *319*, 425−443.

(90) Harbrecht, H.; Peters, M.; Siebenmorgen, M. On multilevel quadrature for elliptic stochastic partial differential equations. In *Sparse grids and applications*; Lect. Notes Comput. Sci. Eng.; Springer: Heidelberg, 2013; Vol. *88*, pp 161−179.

(91) Harbrecht, H.; Peters, M.; Siebenmorgen, M. Combination technique based *k*-th moment analysis of elliptic problems with random diffusion. *J. Comput. Phys.* **2013**, *252*, 128−141.

(92) Pilania, G.; Gubernatis, J. E.; Lookman, T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. *Comput. Mater. Sci.* **2017**, *129*, 156−163.

(93) Cortes, C.; Jackel, L. D.; Solla, S. A.; Vapnik, V.; Denker, J. S. Learning curves: Asymptotic values and rate of convergence. *Advances in Neural Information Processing Systems*; Morgan-Kaufmann, 1994; pp 327−334.

(94) Müller, K. R.; Finke, M.; Murata, N.; Schulten, K.; Amari, S. A numerical study on learning curves in stochastic multilayer feedforward networks. *Neural Comp* **1996**, *8*, 1085.

(95) Vapnik, V. *The nature of statistical learning theory*; Springer Science & Business Media, 2013.

(96) Huang, B.; von Lilienfeld, O. A. The DNA of chemistry: Scalable quantum machine learning with amons. *arXiv preprint arXiv:1707.04146* 2017, submitted to Nature.

(97) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. Outsmarting Quantum Chemistry Through Transfer Learning. *ChemRxiv. Preprint* **2018** DOI: 10.26434/chemrxiv.6744440.v1.

(98) Rupp, M.; Ramakrishnan, R.; von Lilienfeld, O. A. Machine learning for quantum mechanical properties of atoms in molecules. *J. Phys. Chem. Lett.* **2015**, *6*, 3309−3313.

(99) Bereau, T.; DiStasio, R. A., Jr.; Tkatchenko, A.; Von Lilienfeld, O. A. Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning. *J. Chem. Phys.* **2018**, *148*, 241706.

(100) Kennedy, M. C.; O'Hagan, A. Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **2000**, *87*, 1−13.

(101) Christensen, A. S.; Faber, F. A.; Huang, B.; Bratholm, L. A.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. *QML: A Python Toolkit for Quantum Machine Learning* 2017; https://github.com/qmlcode/qml.