

ARE NEURAL SCALING LAWS LEADING QUANTUM CHEMISTRY ASTRAY?

Anonymous authors

Paper under double-blind review

ABSTRACT

Neural scaling laws are driving the machine learning community toward training ever-larger foundation models across domains, assuring high accuracy and transferable representations for extrapolative tasks. We test this promise in quantum chemistry by scaling model capacity and training data from quantum chemical calculations. As a generalization task, we evaluate the resulting models’ predictions of the bond dissociation energy of neutral H_2 , the simplest possible molecule. We find that, regardless of dataset size or model capacity, models trained only on stable structures fail dramatically to even qualitatively reproduce the H_2 energy curve. Only when compressed and stretched geometries are explicitly included in training do the predictions roughly resemble the correct shape. Nonetheless, the largest foundation models trained on the largest and most diverse datasets containing dissociating diatomics exhibit serious failures on simple diatomic molecules. Most strikingly, they cannot reproduce the trivial repulsive energy curve of two bare protons, revealing their failure to learn the basic Coulomb’s law involved in electronic structure theory. These results suggest that scaling alone is insufficient for building reliable quantum chemical models.

1 INTRODUCTION

What makes a physics model “good”? At its essence, any model is, by construction, a simplification of complex reality. Yet, a useful one would ideally reflect our current understanding of physics by incorporating as many known first-principle laws as possible while applying Occam’s razor to reproduce the correct empirical observations *for the correct reasons*. For example, in quantum chemistry, Kohn-Sham density functional theory (Hohenberg & Kohn, 1964; Kohn & Sham, 1965) is an exact theory that states the existence of an exact universal functional that yields the exact energy of a given electron density. In practice, this exact functional’s closed-form analytical expression is unknown and must be approximated. Encouragingly, it has been observed that non-empirical functionals *without any fitted parameters* tend to not only become more accurate as they satisfy more known exact constraints on the hierarchy of “Jacob’s ladder” of density functional approximations but are also more accurate and generalizable than parameterized counterparts (Perdew & Schmidt, 2001; Medvedev et al., 2017; Goerigk & Grimme, 2011; Kaplan et al., 2023; Khan, 2025).

In contrast, machine learning (ML) models attempt to represent physical interactions from data, often by learning a direct statistical mapping from input to labeled outputs. Therefore, the central premise of using ML models to predict chemical properties is that their inference can be accurate while being orders-of-magnitude faster than *ab initio* methods that solve the Schrödinger equation (Rupp et al., 2012). The ML community, particularly the deep learning subfield, has also broadly observed across several domains encompassing *e.g.* natural language processing, computer vision, protein structure prediction, and chemistry that models trained with larger amounts of data, compute, and model parameters systematically improve performance in a power-law fashion (Cortes et al., 1993; Bahri et al., 2024; Hestness et al., 2017; Kaplan et al., 2020; Alabdulmohsin et al., 2022; Jumper et al., 2021; Dubey et al., 2024; Hoffmann et al., 2022; Cheng et al., 2024; Frey et al., 2023; Jiang et al., 2025; Hattori et al., 2025). Such “neural scaling laws” are encouraging a paradigm toward the development of large “foundation models” trained on vast amounts of diverse data (Bommasani, 2021). Accordingly, these models can demand enormous amounts of monetary, natural, and compute resources in their training (Cottier et al., 2024; Hao, 2025) but are claimed to have produced meaningful representations with strong generalization capabilities, as implicated

by universal approximation theorems (Augustine, 2024). However, it remains unclear whether large-scale quantum chemical foundation models can truly learn essential physics required to reliably generalize to novel molecules and materials beyond their training set.

In this work, we investigate whether large neural networks trained on quantum chemical data can indeed capture the governing physics needed for reliable generalization, by scaling models on two large datasets of equilibrium-geometry molecules. Our scaling experiments agree with the literature that increasing the number of training samples improves performance — at least, on the hold-out test split. We find that irrespective of training data quantity and model capacity, there is no discernible improvement in the models’ reproductions of the bond dissociation energy (BDE) curve of H_2 , the smallest and simplest possible molecule. The models improve modestly only when non-ground-state structures are included in the training set. However, even the largest foundation models trained on over 101M structures that include dissociating bonds can fail to qualitatively describe the BDE curves of H_2 and other diatomic molecules. More disconcertingly, they cannot reproduce the trivial energy curve for a system of two bare protons produced from the fundamental and classical Coulomb’s law. Contrary to the current pervasive inclination in the ML community to favor scaling, our results indicate that training ever-larger models on ever-larger datasets does not necessarily lead to “good” quantum chemical models that give correct answers for the correct reasons.

2 BACKGROUND

We provide a brief overview of the following topics that pertain to our experiments described in section 3: theoretical quantum chemistry, ML applications in quantum chemistry, neural scaling laws, and foundation models.

2.1 QUANTUM CHEMISTRY

Quantum chemistry uses quantum mechanical principles to accurately calculate properties of an arbitrary chemical system composed of electrons, protons, and neutrons. This requires solving the Schrödinger equation (Schrödinger, 1926), which for N electrons is a linear partial differential equation, given in its time-independent form as

$$\hat{H}\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (1)$$

where $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ is the antisymmetric and normalized wavefunction describing the state, E is the total energy, and each $\mathbf{r}_i$ denotes the spatial-spin coordinate of the i th electron. In atomic units, the non-relativistic Hamiltonian operator, $\hat{H}$, corresponds to E of Ψ and is defined under the Born-Oppenheimer approximation (Born & Oppenheimer, 1927) of stationary nuclei as

$$\hat{H} = -\sum_i \frac{1}{2} \nabla_{\mathbf{r}_i}^2 - \sum_i \sum_j \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} + \sum_i \sum_{k>i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_k|} + \sum_j \sum_{l>j} \frac{Z_j Z_l}{|\mathbf{R}_j - \mathbf{R}_l|} \quad (2)$$

where $\mathbf{R}_j$ is the position of the j th nucleus and Z_j is its nuclear charge. From left to right, the terms respectively represent the electronic kinetic energy, the potential energy from electron-nucleus interactions, the potential energy from electron-electron repulsions, and the potential energy from nucleus-nucleus repulsions. Unfortunately, a closed-form solution to the Schrödinger equation in its differential form does not exist for multi-electron systems due to the electronic repulsion term (Tew et al., 2007). Moreover, the exponential increase with N in the dimension of the Hilbert space spanned by the different electronic configurations (Kohn, 1999) necessitates alternative solution methods that are computationally tractable.

2.1.1 KOHN-SHAM DENSITY FUNCTIONAL THEORY

One such widely-employed method (Jones, 2015) is density functional theory (DFT). It is an exact theory rooted in the universal Hohenberg-Kohn theorems (Hohenberg & Kohn, 1964) that concerns any system of electrons in any external potential:

- **Theorem 1:** To within an additive constant, the external potential, and thereby the total system energy is a unique functional of the ground-state electron density, n .

- **Theorem 2:** From the set of v -representable trial densities (Levy, 1979) associated with antisymmetric ground-state wavefunctions that are non-negative everywhere and normalize to N , the ground-state energy of the system is minimized at its ground-state density, n_0 .

Consequently, DFT dramatically reduces the concerned configuration space to simply three spatial dimensions that define n . Then, by introducing single-electron orbitals through an auxiliary non-interacting system that exactly reproduces n_0 (Kohn & Sham, 1965) to solve the non-interacting electronic kinetic energy, one has Kohn-Sham DFT (KSDFT). Attractively, KSDFT requires just the exchange-correlation functional to be approximated, which comprises just a small percentage of E . It is also relatively inexpensive compared to wavefunction-based methods as its computational time complexity is formally at least $\mathcal{O}(N^3)$ due to orthonormalization of the orbitals and diagonalization of the Kohn-Sham Hamiltonian matrix (Lee & Dieng, 2025). However, its accuracy can vary widely depending on the utilized functional (Goerigk & Grimme, 2011).

2.1.2 G4(MP2) THEORY

Gaussian-4(MP2) theory (Pople et al., 1989; Curtiss et al., 2011; 2007b;a), G4(MP2), is a composite wavefunction method (Helgaker et al., 2008). It can robustly predict energies to within chemical accuracy of 1 kcal/mol relative to experimental data for molecules containing first-, second-, and third-row main-group elements. It achieves this by performing several calculations at different levels of theory, including: Hartree-Fock (Slater, 1930) energies extrapolated to the complete basis set limit; coupled cluster singles and doubles perturbative triples (CCSD(T)) (Csirik & Laestadius, 2023) energies calculated with the 6-31G* basis set (Rassolov et al., 1998); and frozen-core energy corrections calculated with second-order Møller-Plesset perturbation theory (MP2) (Møller & Plesset, 1934) using the G3MP2LargeXP (Curtiss et al., 2007b) and 6-31G(d) basis sets (Rassolov et al., 1998).

2.2 MACHINE LEARNING IN QUANTUM CHEMISTRY

Although KSDFT and G4(MP2) are highly-accurate, their computational time complexity of $\mathcal{O}(N^3)$ or worse (e.g. CCSD(T) scales as $\mathcal{O}(N^7)$ (Ratcliff et al., 2017)) means *ab initio* quantum chemical calculations are prohibitively expensive for even small molecules. Therefore, ML has been applied extensively to quantum chemistry to bypass the computational barrier of repeatedly having to explicitly solve the Schrödinger equation for each new molecule and material. In other words, ML models can learn a statistical mapping from the external potential defined by the sets of nuclear charges, $\{\mathcal{Z}_j\}$, and nuclear coordinates, $\{\mathbf{R}_j\}$, to any given property (Rupp et al., 2012). This is akin to the universal functional stated in Hohenberg and Kohn’s first theorem. Of course, the important distinction is that these ML models can be orders-of-magnitude faster than traditional quantum chemical methods without much loss in accuracy. We refer the reader to many excellent references that comprehensively review the numerous ML applications in chemistry (Schütt et al., 2020; Westermayr et al., 2021; Keith et al., 2021; Sajjan et al., 2022; Huang et al., 2023; Huang & Von Lilienfeld, 2021; von Lilienfeld & Burke, 2020; Von Lilienfeld, 2018; Hansen et al., 2013; von Lilienfeld et al., 2020; Dral, 2020; Malica et al., 2025).

2.2.1 NEURAL SCALING LAWS

The accuracy of a ML model’s predictions has been empirically observed to depend on the quality and quantity of the training data, and the employed learning algorithm. As mentioned in section 1, for neural network architectures, neural scaling laws refer to observations that more training data and larger models reduce prediction errors as a power-law. For example, Frey et al. (2023) showed that state-of-the-art graph neural network architectures like SchNet (Schütt et al., 2017), PaiNN (Schütt et al., 2021), Allegro (Musaelian et al., 2023), and SpookyNet (Unke et al., 2021a) tend to improve their predictions of atomic forces with increasing model capacity, c , defined as

$$c = d \times w \quad (3)$$

where d is the number of convolutional layers and w is the embedding dimension. Moreover, they show monotonic improvements to the loss with increasing dataset size.

2.2.2 FOUNDATION MODELS

Neural scaling laws are motivating the development of foundation models (Bommasani, 2021), which can be broadly defined as large models trained on large and diverse datasets that can subsequently be fine-tuned for a wide range of downstream tasks. Examples include BERT (Devlin et al., 2019), GPT-3 (Brown et al., 2020), and CLIP (Radford et al., 2021) for text- and image-related tasks. The assumption is that they have learned general representations that can offer potential solutions to data-scarce and extrapolative regimes. Unsurprisingly, this is appealing for chemical applications and many chemical foundation models have accordingly been developed (Yuan et al., 2025; Choi et al., 2025; Pyzer-Knapp et al., 2025; Alampara et al., 2025; Ahmad et al., 2022; Batatia et al., 2023; Chen & Ong, 2022; Unke et al., 2021b; Wood et al., 2025; Soares et al., 2024; Beaini et al., 2023; Merchant et al., 2023; Yang et al., 2024).

3 EXPERIMENTS

Here, we are interested in whether neural scaling laws and foundation models indeed produce “good” quantum chemical models by having learned general and physically-meaningful representations that would allow for these large models to be reliably applied to novel molecules and materials.

3.1 H₂ AS A FUNDAMENTAL TEST

To investigate this, we consider whether scaling deep neural networks trained on molecules in their equilibrium geometries can lead to accurate predictions of the bond dissociation energy (BDE) curve at different bond lengths, R , of neutral H₂ — the smallest and simplest possible molecule consisting of only two protons, two electrons, and one bond. The BDE curve for a diatomic plots the total energy versus bond length, R , normalized by subtracting the summed energies of the two isolated atoms from the total system energy. This offers a simple evaluation of the generalization capabilities of the models since there are no complications arising from functional groups, conformational degrees of freedom, relativistic effects (Pyykkö, 2012), unexpected spin multiplicities, core electrons, dipole moments, or the need for large basis sets (Boese et al., 2003).

However, the BDE curve of H₂ is still challenging to accurately describe because of the strong-correlation limit approached in the exchange-correlation energy as the bond dissociates, at which each electron should be localized to their individual nuclei. This regime is accurately described as a multi-reference system where many Slater determinants contribute to the exact wave function. Consequently, single-determinant methods like KSDFT, which assumes the electron density can be described by one Slater determinant, would often be inadequate to provide quantitatively accurate estimates of the H₂ BDE. Therefore, it may perhaps be unrealistic for the trained models to exactly reproduce the exact BDE curve. Nonetheless, we expect that they are capable of reproducing the BDE curve to the accuracy of the level of theory used to generate the training data. Further, we maintain that a model that has properly learned the basic physics involved, as may be anticipated from neural scaling laws, should capture some essential features: no kinks and divergent behavior in the BDE curve; an asymptotic plateau in the energy as the inter-nuclear distance increases; and the obvious limit of $\lim_{R \rightarrow 0} E \rightarrow +\infty$ as the nuclear-nuclear repulsion energy, E_{nn} , dominates and diverges due to the inverse- R relation given by

$$E_{nn} = \frac{Z_j Z_l}{R} \quad (4)$$

In our work, we obtain reference curves from spin-restricted calculations performed using PySCF (Sun et al., 2018).

3.2 SCALING

Here, we describe our scaling experiments in which SchNet models are trained on large, high-quality quantum chemical datasets. Our objective is to systematically explore how training models of larger capacity on larger datasets affects predictive accuracy of predicting molecular energies.

3.2.1 SCHNET

We perform the scaling experiments by training SchNet models, implemented in PyTorch Geometric (Fey & Lenssen, 2019; pyt, a;b), of various c by taking combinations of $d \in \{3, 6\}$ and $w \in \{100, 250, 500\}$. For all models, the number of hidden channels is set to w , 50 Gaussians are used to expand the pairwise distances, the cutoff distance for the interatomic interactions is 5.0 Å, and each node collects at most 32 neighbors within this cutoff. Each model is trained on several training set sizes, $N_{\text{train}} \in \{10^x, 5 \cdot 10^x \mid x \geq 2, x \in \mathbb{Z}\}$, where x is capped by the dataset size. For $x \leq 3$, the batch size is set to 10 samples and for $x \geq 4$, the batch size is set to 50 samples. “Warm-up” is performed where the learning rate for a given epoch, ϵ , within the first ten epochs is set to $\epsilon/10$. After this warm-up period, the learning rate is decreased to and kept fixed at $5 \cdot 10^{-4}$. However, if the tracked value of the lowest validation loss fails to decrease after twenty epochs, the learning rate is scaled by 0.5. If the validation loss does not decrease after 50 epochs, training is terminated. All models are trained with gradient descent using PyTorch (Paszke et al., 2019) for up to 1000 epochs using the AdamW optimizer (Loshchilov & Hutter, 2017) with mean squared errors as training and validation losses. We also clip gradients by their Euclidean norms with a threshold value of 1000. Each model is trained on a single node on compute clusters, utilizing one NVIDIA RTX A6000 GPU, 16 CPU cores, and 64 GB of RAM. The model parameters at the epoch where the validation loss is the smallest during training are saved as the best-performing model. 1000 samples are removed from the dataset and fixed as the test set across all N_{train} . For each model, 1000 samples are randomly chosen from the remaining dataset and used as the validation set.

3.2.2 DATASETS

We train these models in a supervised fashion where nuclear charges and coordinates are used to predict total atomization energies, E_{TAE} , reported in the GDB-9-G4(MP2) (Narayanan et al., 2019) and VQM24 datasets (Khan et al., 2024). E_{TAE} is the difference between a molecule’s E and the sum of E of its isolated constituent atoms:

$$E_{\text{TAE}} = \sum_j E_{\text{atom},j} - E_{\text{molecule}} \quad (5)$$

GDB-9-G4(MP2) contains 133,296 stable closed-shell and neutral organic molecules in their equilibrium geometries containing no more than nine heavy atoms of C, N, O, and F in the GDB-9 database (Ramakrishnan et al., 2014; Ruddigkeit et al., 2012). These molecules’ properties were computed using the highly-accurate G4(MP2) theory. VQM24 reports closed-shell and neutral organic and inorganic molecules with no more than five heavy atoms of C, N, O, F, Si, P, S, Cl, and Br. 835,947 converged molecules were computed at the ω B97X-D3/cc-pVDZ (Chai & Head-Gordon, 2008; Dunning Jr, 1989) level of KSDF. ω B97X-D3 is a range-separated hybrid functional with D3 dispersion correction (Grimme et al., 2010) that has excellent performance in main-group thermochemistry, kinetics, and noncovalent interaction benchmarks (Goerigk & Grimme, 2011). Note that 51,072 of the molecules in the dataset converged to saddle points, as identified by vibrational frequency calculations. Such structures correspond to transition states, which are of highest potential energies along reaction coordinates and contain bonds that are partially broken or formed (Braddock et al., 2024).

For both datasets, we convert E_{TAE} to units of kcal/mol and scale by 1/25, which we find to be beneficial in training.

3.3 FOUNDATION MODELS FOR QUANTUM CHEMISTRY

We also examine five “foundation” machine-learned interatomic potentials that can estimate the energy of an atomic structure: UMA-S-1.1 (Wood et al., 2025), UMA-M-1.1 (Wood et al., 2025), OMol25 eSEN-sm-cons. (Levine et al., 2025), Orb v3 conservative OMol25 (Rhodes et al., 2025; git), and AIMNet2 (Anstine et al., 2025). As of September 2025, Rowan Benchmarks (Rowan) reported that these five models were the most accurate ML models in predicting molecular energies from $\{\mathcal{Z}_j\}$ and $\{\mathbf{R}_j\}$ on tasks involving thermochemistry, kinetics, and non-covalent interactions while being more accurate than popular density functionals, composite density functional methods, and semi-empirical methods. UMA-S-1.1 (150M parameters), UMA-M-1.1 (1.4B parameters), OMol25 eSEN-sm-cons. (6.3M parameters), and Orb v3 conservative OMol25 (25M parameters) were all trained on the OMol25 dataset (Levine et al., 2025), which contains over 101M

ω B97M-V/def2-TZVPD (Mardirossian & Head-Gordon, 2016; Rappoport & Furche, 2010; Hellweg & Rappoport, 2015) KSDFT calculations that required at least 6 billion CPU hours to compute. OMol25 includes small molecules, biomolecules, metal complexes, and electrolytes that span 83 elements, a wide range of intra- and inter-molecular interactions, conformers, reactive structures, spin multiplicities, and charges. AIMNet2 supports H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, I, and can handle charged molecules. Its model size is not publicly reported but it was trained on $2 \cdot 10^7$ ω B97M-D3/def2-TZVPP KSDFT calculations of bioactive molecules and small organic molecules sourced from ChEMBL (Mendez et al., 2019), PubChem (Kim et al., 2023), ANI-2x (Devereux et al., 2020), and OrbNet (Qiao et al., 2020) datasets.

3.4 SCALING DOES NOT NECESSARILY HELP H_2

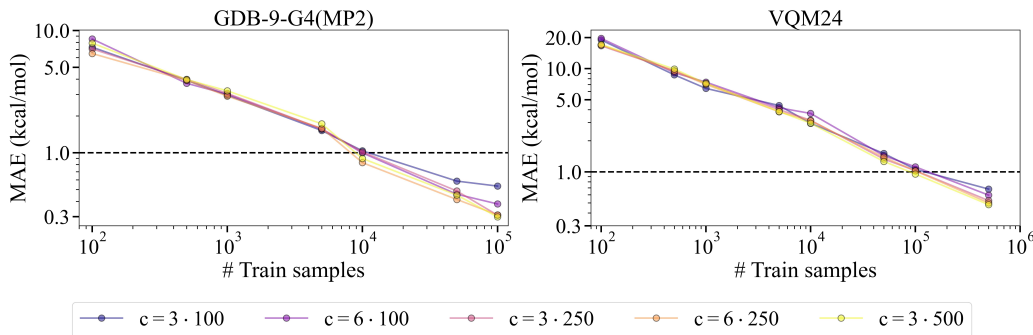


Figure 1: Prediction errors on test set splits of SchNet models of varying model capacity, c , trained on different numbers of training samples from (left) GDB-9-G4(MP2) and (right) VQM24 datasets. The mean absolute error is plotted against the number of training samples on a log-log plot. Horizontal dashed line denotes baseline accuracy of 1 kcal/mol.

As seen in Figure 1, we find that training SchNet models on both the GDB-9-G4(MP2) and VQM24 datasets on increasingly larger numbers of training samples systematically reduces all models’ test set prediction errors. The mean absolute error decreases linearly on the log-log plot plotting error against number of training samples, as expected from prior works in the literature regarding scaling. In addition, models of larger values of c tend to show lower errors, which is also not unsurprising.

What is more interesting is that the SchNet models trained on GDB-9-G4(MP2) show no systematic improvement in their abilities to even qualitatively reproduce the H_2 BDE curve. As shown in Figure 2 (top, left), all models trained on 100k samples predict essentially the same BDE curve, regardless of c . Near the bond length of 0.77 Å, the ω B97M-V/cc-pVQZ KSDFT (Dunning Jr, 1989) and “exact” all-electron CCSD(T) (cc-pVQZ basis) calculations identify a minimum in BDE. However, the SchNet models rather predict curves that are practically completely horizontal and incorrectly peak to their maximum energies near 0.77 Å, despite having trained on molecules calculated using the highly-accurate G4(MP2) method. This immediately suggests an inability of scaling to allow the models trained on stable molecules to produce meaningful representations that allow for generalization to simple bond stretching. This is further apparent as the models do not obviously become more accurate with more data or greater model capacity in predicting either the equilibrium bond length (Figure 2, top, middle) or bond dissociation energy (Figure 2, top, right).

In contrast, the models trained using VQM24 exhibit improved BDE curves (Figure 2, bottom, left). Although they show many imperfections such as incorrect asymptotic behavior in energy as $R \rightarrow 0$, many kinks, and too large of a curvature near the equilibrium bond length, they are at least able to exhibit some resemblance to the reference curve of ω B97M-V. There is still no clear and systematic improvement with more training data in the models’ abilities to predict the equilibrium bond lengths (Figure 2, bottom, middle) and dissociation energies (Figure 2, bottom, right), but they are able to produce results reasonably close to the KSDFT baseline after training on 500k samples. This improvement relative to the models trained on GDB-9-G4(MP2) appears to arise from the saddle-point structures contained in VQM24. For comparison, GDB-9-G4(MP2) only contains stable, non-saddle-point structures without any diatomic molecules; its smallest molecules are H_2O and HCN . Moreover, as seen in Figure 3, the smallest inter-atomic distance found in each

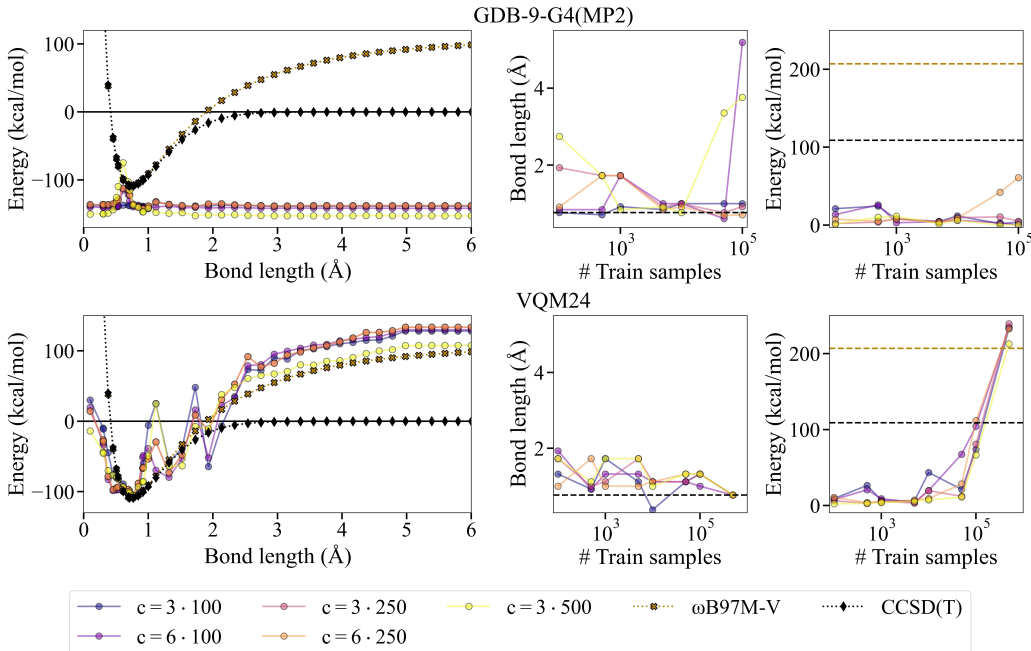


Figure 2: Neutral H_2 bond dissociation energy curves (left column), equilibrium bond lengths (middle column; black dashed line indicates reference CCSD(T) and KSDFT values of 0.77 Å), and dissociation energies (right column; dashed dark and light lines indicate reference CCSD(T) and KSDFT values of 109 and 207 kcal/mol, respectively) obtained from CCSD(T) (cc-pVQZ basis), ω B97M-V/cc-pVQZ KSDFT, and SchNet models of various c trained on (**top**) 100k GDB-9-G4(MP2) samples and on (**bottom**) 500k VQM24 samples.

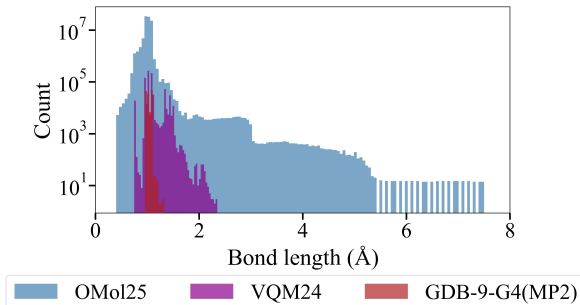


Figure 3: Distributions of smallest inter-atomic distance found in each sample in OMol25, VQM24, and GDB-9-G4(MP2) datasets.

molecule ranges between 0.96 Å and 1.33 Å with a mean of 1.03 Å. In contrast, VQM24 has more diverse bonding motifs as its smallest inter-atomic distance ranges between 0.75 Å and 2.35 Å with a mean of 1.08 Å (Figure 3). It also contains 15 diatomic molecules of HBr, HCl, HF, N_2 , SiO, F_2 , O_2 , Si_2 , BrF, ClF, Cl_2 , Br_2 , BrCl, PN, and P_2 .

3.5 ON THE LIMITS OF FOUNDATION MODELS FOR QUANTUM CHEMISTRY

From the previous comparisons of SchNet models trained on GDB-9-G4(MP2) and VQM24 in subsection 3.4, we may expect improved performances from models trained on more diverse datasets like OMol25. Therefore, we also assess the five state-of-the-art foundation models listed in subsection 3.3 in reproducing the BDE curves of H_2 , along with those of Li_2 , Be_2 , N_2 , O_2 , and F_2 . Indeed, those models produce BDE curves that closely match the ω B97M-V H_2 curve (Figure 4).

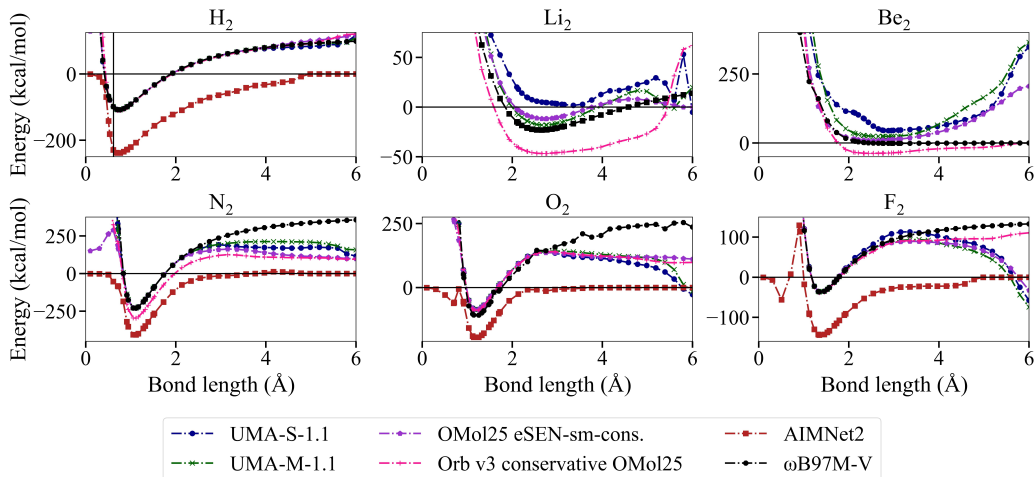


Figure 4: Bond dissociation energy curves of neutral H_2 , Li_2 , Be_2 , N_2 , O_2 , F_2 , calculated using $\omega B97M-V/def2-TZVPD$ KSDFT and estimated using UMA-S-1.1, UMA-M-1.1, OMol25 eSEN-sm-cons., Orb v3 conservative OMol25, and AIMNet2. AIMNet2 was only applied to H_2 , N_2 , O_2 , and F_2 because it does not support elements of the other diatomics.

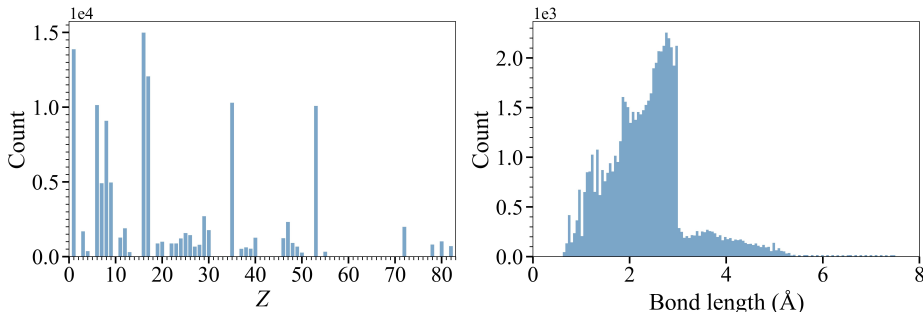


Figure 5: Distribution of (left) elements (denoted by nuclear charge, Z) involved in the 61,498 diatomic systems found in OMol25 dataset and (right) corresponding distribution of bond lengths.

However, there is lack of transferability to other diatomic molecules as the different models show many peculiar and unphysical features, such as kinks and divergent behaviors. These limitations occur although OMol25 contains a vast assortment of molecules with compressed, equilibrium, and dissociating bonds. This is seen in Figure 3 where its distribution of smallest inter-atomic distances range from 0.40 Å and 7.50 Å, with a mean of 1.01 Å. Also, as seen in Figure 5, it encompasses 61,498 diatomic systems involving 45 different elements spanning the periodic table from H to Pb, with H, C, O, S, Cl, Br, and I being the most represented; its bond lengths range from 0.62 Å to 7.50 Å, with a mean of 2.38 Å. In fact, it contains diatomic systems involving all the elements that we have examined in Figure 4 but the models still perform poorly in accurately predicting their molecular dissociation curves. The model failures are interesting since the reference $\omega B97M-V$ functional used to create these models’ training data has, by construction, the correct long-range asymptotic behavior in its exchange potential that guarantees non-divergence in energies. Further, this poor performance is unexpected considering the reported energy error of ~ 0.66 kcal/mol achieved by OMol25 eSEN-sm-cons. for bonds shorter than 6 Å (Levine et al., 2025), which is the range for the bond lengths considered in our BDE curves.

In particular, for H_2 , N_2 , O_2 , and F_2 , AIMNet2 correctly produces energies that asymptote to 0 kcal/mol for large bond lengths. This behavior may perhaps be accredited to AIMNet2’s explicit calculation of dispersion energy (Grimme et al., 2010; 2016) included in a system’s total energy as

$$E = E_{\text{local}} + E_{\text{dispersion}} + E_{\text{Coulomb}} \quad (6)$$

where E_{local} is the local configurational interaction energy, $E_{\text{dispersion}}$ is the dispersion interaction energy, and E_{Coulomb} is the Coulomb interaction energy from partial atomic charges (Anstine et al., 2025). However, it otherwise performs very poorly as it severely underestimates relative energies at equilibrium bond lengths and wrongly plateaus in energy as $R \rightarrow 0$, despite its E_{Coulomb} term.

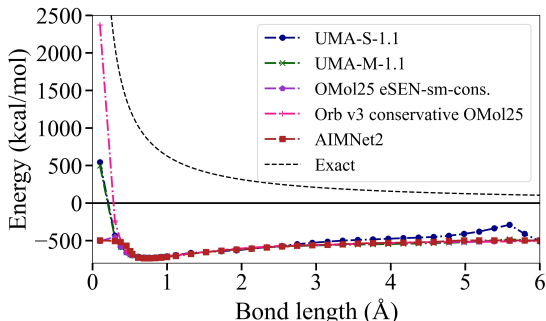


Figure 6: Total energies for a system of two bare protons as a function of inter-proton distance, estimated using UMA-S-1.1, UMA-M-1.1, OMol25 eSEN-sm-cons., Orb v3 conservative OMol25, and AIMNet2. Black dashed line indicates the exact energy profile.

We also isolate the foundation models’ abilities to have learned the fundamental Coulomb’s law that lies at the core of electronic structure theory by analyzing their predicted total energy curves for a system of two bare protons without any electrons. As seen in Figure 6, all of the models predict that the two-proton system is extremely stable by ~ 500 kcal/mol even at long inter-proton distances. In fact, the predicted energy has the wrong, negative sign for practically the entire curve where the inter-proton distance is greater than ~ 0.2 Å, whereas the exact energy should always be positive.

These results imply that scaling model capacity and training data is insufficient to attain accurate and generalizable quantum chemical property predictions. This is further concerning since even the explicit inclusion of inductive biases such as E_{Coulomb} in AIMNet2 produces very poor results on the diatomic systems. These deficiencies cannot be attributed to weak model expressibility since the good results for H_2 in Figure 4 clearly reveal otherwise. Thus, it is not entirely obvious what fast, alternative ML methods can achieve this objective. Yet, it may be worthwhile to consider techniques like Δ -ML (Ramakrishnan et al., 2015) to correct relatively-cheap semi-empirical quantum chemistry methods, and “similarity-based learning” to select minimal amounts of data in training models on-the-fly in data-scarce scenarios (Lemm et al., 2023; Lee et al., 2025).

4 CONCLUSION

Our experimental findings suggest that current deep neural network approaches rooted in scaling face significant challenges in producing accurate and generalizable quantum chemical models, even when trained on large, chemically diverse datasets of high-quality calculations. This limitation is evident in the bond dissociation energy curves for H_2 , which are predicted to be essentially flat across all bond lengths by SchNet models trained on the GDB-9-G4(MP2) dataset. Training on the VQM24 dataset, which comprises off-equilibrium bonding, modestly improves reproduction of the H_2 curves. However, this improvement appears to stem from exposure to dissociated structures rather than from learning underlying physical principles. This is further corroborated by the observation of persistent issues in the predictions of state-of-the-art machine-learned interatomic potentials. Despite being trained on the most diverse collection of elements, chemistry, and structures to date — 101M KSDFT calculations, including 61k covering diatomic systems — their predicted energy curves for simple diatomic molecules degrade significantly outside equilibrium bonding regions and exhibit unphysical features. Moreover, they all incorrectly predict that two protons form a strongly-bound system, underscoring their failure to capture the basic Coulomb’s law even when it is incorporated as an inductive bias. These shortcomings highlight the difficulty of achieving true physical generalization through scaling alone and suggest that current large-scale models risk acting solely as data-driven interpolators. Overall, the intersection of quantum chemistry and ML is in need of new strategies to make fast and accurate property predictions of novel molecules and materials.

CODE AND DATA AVAILABILITY

Refer to this anonymized Zenodo repository for

- nuclear charges, nuclear coordinates, and atomization energies used as training data in our scaling experiments
- model checkpoints from our scaling experiments
- Python code to implement our scaling experiments
- Python code for utilizing the foundation machine-learned interatomic potentials

REFERENCES

- GitHub - orbital-materials/orb-models: ORB forcefield models from Orbital Materials — github.com. <https://github.com/orbital-materials/orb-models>. [Accessed September 2025].
- PyG Documentation & 2014; pytorch-geometric documentation — pytorch-geometric.readthedocs.io. <https://pytorch-geometric.readthedocs.io/en/latest/>, a. [Accessed September 2025].
- torch-geometric.nn.models.SchNet & 2014; pytorch-geometric documentation — pytorch-geometric.readthedocs.io. https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.models.SchNet.html, b. [Accessed September 2025].
- Walid Ahmad, Elana Simon, Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta-2: Towards chemical foundation models. *arXiv preprint arXiv:2209.01712*, 2022.
- Ibrahim M Alabdulmohsin, Behnam Neyshabur, and Xiaohua Zhai. Revisiting neural scaling laws in language and vision. *Advances in Neural Information Processing Systems*, 35:22300–22312, 2022.
- Nawaf Alampara, Anagha Aneesh, Martiño Ríos-García, Adrian Mirza, Mara Schilling-Wilhelmi, Ali Asghar Aghajani, Meiling Sun, Gordan Prastalo, and Kevin Maik Jablonka. General purpose models for the chemical sciences. *arXiv preprint arXiv:2507.07456*, 2025.
- Dylan M Anstine, Roman Zubatyuk, and Olexandr Isayev. Aimnet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs. *Chemical Science*, 16(23):10228–10244, 2025.
- Midhun T Augustine. A survey on universal approximation theorems. *arXiv preprint arXiv:2407.12895*, 2024.
- Yasaman Bahri, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma. Explaining neural scaling laws. *Proceedings of the National Academy of Sciences*, 121(27):e2311878121, 2024.
- Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *arXiv preprint arXiv:2401.00096*, 2023.
- Dominique Beaini, Shenyang Huang, Joao Alex Cunha, Zhiyi Li, Gabriela Moisescu-Pareja, Oleksandr Dymov, Samuel Maddrell-Mander, Callum McLean, Frederik Wenkel, Luis Müller, et al. Towards foundational models for molecular learning on large-scale multi-task datasets. *arXiv preprint arXiv:2310.04292*, 2023.
- A Daniel Boese, Jan ML Martin, and Nicholas C Handy. The role of the basis set: Assessing density functional theory. *The Journal of chemical physics*, 119(6):3005–3014, 2003.
- Rishi Bommasani. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

- M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927. doi: <https://doi.org/10.1002/andp.19273892002>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19273892002>.
- D Christopher Braddock, Siwoo Lee, and Henry S Rzepa. Modelling kinetic isotope effects for swern oxidation using dft-based transition state theory. *Digital Discovery*, 3(8):1496–1508, 2024.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics*, 10(44):6615–6620, 2008.
- Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022.
- Xingyi Cheng, Bo Chen, Pan Li, Jing Gong, Jie Tang, and Le Song. Training compute-optimal protein language models. *Advances in Neural Information Processing Systems*, 37:69386–69418, 2024.
- Junyoung Choi, Gunwook Nam, Jaesik Choi, and Yousung Jung. A perspective on foundation models in chemistry. *JACS Au*, 5(4):1499–1518, 2025.
- Corinna Cortes, Lawrence D Jackel, Sara Solla, Vladimir Vapnik, and John Denker. Learning curves: Asymptotic values and rate of convergence. *Advances in neural information processing systems*, 6, 1993.
- Ben Cottier, Robi Rahman, Loredana Fattorini, Nestor Maslej, Tamay Besiroglu, and David Owen. The rising costs of training frontier ai models. *arXiv preprint arXiv:2405.21015*, 2024.
- Mihaly Andras Csirik and Andre Laestadius. Coupled-cluster theory revisited. 2023.
- Larry A Curtiss, Paul C Redfern, and Krishnan Raghavachari. Gaussian-4 theory using reduced order perturbation theory. *The Journal of chemical physics*, 127(12), 2007a.
- Larry A Curtiss, Paul C Redfern, and Krishnan Raghavachari. Gaussian-4 theory. *The Journal of chemical physics*, 126(8), 2007b.
- Larry A Curtiss, Paul C Redfern, and Krishnan Raghavachari. Gn theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):810–825, 2011.
- Christian Devereux, Justin S Smith, Kate K Huddleston, Kipton Barros, Roman Zubatyuk, Olexandr Isayev, and Adrian E Roitberg. Extending the applicability of the ani deep learning molecular potential to sulfur and halogens. *Journal of chemical theory and computation*, 16(7):4192–4202, 2020.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186, 2019.
- Pavlo O Dral. Quantum chemistry in the age of machine learning. *The journal of physical chemistry letters*, 11(6):2336–2347, 2020.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv e-prints*, pp. arXiv–2407, 2024.
- Thom H Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of chemical physics*, 90(2):1007–1023, 1989.
- Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

- Nathan C Frey, Ryan Soklaski, Simon Axelrod, Siddharth Samsi, Rafael Gomez-Bombarelli, Connor W Coley, and Vijay Gadepally. Neural scaling of deep chemical models. *Nature Machine Intelligence*, 5(11):1297–1305, 2023.
- Lars Goerigk and Stefan Grimme. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Physical Chemistry Chemical Physics*, 13(14):6670–6688, 2011.
- Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of chemical physics*, 132(15), 2010.
- Stefan Grimme, Andreas Hansen, Jan Gerit Brandenburg, and Christoph Bannwarth. Dispersion-corrected mean-field electronic structure methods. *Chemical reviews*, 116(9): 5105–5154, 2016.
- Katja Hansen, Grégoire Montavon, Franziska Biegler, Siamac Fazli, Matthias Rupp, Matthias Scheffler, O Anatole Von Lilienfeld, Alexandre Tkatchenko, and Klaus-Robert Muller. Assessment and validation of machine learning methods for predicting molecular atomization energies. *Journal of chemical theory and computation*, 9(8):3404–3419, 2013.
- Karen Hao. *Empire of AI: Dreams and Nightmares in Sam Altman’s OpenAI*. Penguin Group, 2025.
- Shinnosuke Hattori, Kohei Shimamura, Aiichiro Nakano, Rajiv K Kalia, Priya Vashishta, and Ken-ichi Nomura. Beyond scaling: Chemical intuition as emergent ability of universal machine learning interatomic potentials. *arXiv preprint arXiv:2506.07579*, 2025.
- Trygve Helgaker, Wim Klopper, and David P Tew. Quantitative quantum chemistry. *Molecular Physics*, 106(16-18):2107–2143, 2008.
- Arnim Hellweg and Dmitrij Rappoport. Development of new auxiliary basis functions of the karlsruhe segmented contracted basis sets including diffuse basis functions (def2-svpd, def2-tzvpd, and def2-qvpd) for ri-mp2 and ri-cc calculations. *Physical Chemistry Chemical Physics*, 17(2):1010–1017, 2015.
- Joel Hestness, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou. Deep learning scaling is predictable, empirically. *arXiv preprint arXiv:1712.00409*, 2017.
- Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022.
- Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.
- Bing Huang and O Anatole Von Lilienfeld. Ab initio machine learning in chemical compound space. *Chemical reviews*, 121(16):10001–10036, 2021.
- Bing Huang, Guido Falk von Rudorff, and O Anatole von Lilienfeld. The central role of density functional theory in the ai age. *Science*, 381(6654):170–175, 2023.
- Du Jiang, Xuelan Wen, Yixiao Chen, Ruichen Li, Weizhong Fu, Hung Q Pham, Ji Chen, Di He, William A Goddard III, Liwei Wang, et al. Neural scaling laws surpass chemical accuracy for the many-electron schrödinger equation. *arXiv preprint arXiv:2508.02570*, 2025.
- Robert O Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of modern physics*, 87(3):897–923, 2015.
- John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

- Aaron D Kaplan, Mel Levy, and John P Perdew. The predictive power of exact constraints and appropriate norms in density functional theory. *Annual Review of Physical Chemistry*, 74(1): 193–218, 2023.
- Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.
- John A Keith, Valentin Vassilev-Galindo, Bingqing Cheng, Stefan Chmiela, Michael Gastegger, Klaus-Robert Muller, and Alexandre Tkatchenko. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chemical reviews*, 121 (16):9816–9872, 2021.
- Danish Khan. Non-linear and non-empirical double hybrid density functional. *arXiv preprint arXiv:2503.22661*, 2025.
- Danish Khan, Anouar Benali, Scott YH Kim, Guido Falk von Rudorff, and O Anatole von Lilienfeld. Quantum mechanical dataset of 836k neutral closed shell molecules with upto 5 heavy atoms from cnohipsclbr. *arXiv preprint arXiv:2405.05961*, 2024.
- Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. Pubchem 2023 update. *Nucleic acids research*, 51(D1):D1373–D1380, 2023.
- Walter Kohn. Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Reviews of modern physics*, 71(5):1253, 1999.
- Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.
- Siwoo Lee and Adji Bouso Dieng. Muapbek: An improved analytical kinetic energy density functional for quantum chemistry. *arXiv preprint arXiv:2505.04559*, 2025.
- Siwoo Lee, Jason Hattrick-Simpers, Young-June Kim, and Anatole von Lilienfeld. High- T_c superconductor candidates proposed by machine learning. *Machine Learning: Science and Technology*, 2025.
- Dominik Lemm, Guido Falk von Rudorff, and O Anatole Von Lilienfeld. Improved decision making with similarity based machine learning: applications in chemistry. *Machine Learning: Science and Technology*, 4(4):045043, 2023.
- Daniel S Levine, Muhammed Shuaibi, Evan Walter Clark Spotte-Smith, Michael G Taylor, Muhammad R Hasyim, Kyle Michel, Ilyes Batatia, Gábor Csányi, Misko Dzamba, Peter Eastman, et al. The open molecules 2025 (omol25) dataset, evaluations, and models. *arXiv preprint arXiv:2505.08762*, 2025.
- Mel Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v -representability problem. *Proceedings of the National Academy of Sciences*, 76(12):6062–6065, 1979.
- Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- Cristiano Malica, Kostya S Novoselov, Amanda S Barnard, Sergei V Kalinin, Steven R Spurgeon, Karsten Reuter, Maite Alducin, Volker L Deringer, Gábor Csányi, Nicola Marzari, et al. Artificial intelligence for advanced functional materials: exploring current and future directions. *Journal of Physics: Materials*, 8(2):021001, 2025.
- Narbe Mardirossian and Martin Head-Gordon. ω b97m-v: A combinatorially optimized, range-separated hybrid, meta-gga density functional with vv10 nonlocal correlation. *The Journal of chemical physics*, 144(21), 2016.

- Michael G Medvedev, Ivan S Bushmarinov, Jianwei Sun, John P Perdew, and Konstantin A Lyssenko. Density functional theory is straying from the path toward the exact functional. *Science*, 355(6320):49–52, 2017.
- David Mendez, Anna Gaulton, A Patrícia Bento, Jon Chambers, Marleen De Veij, Eloy Félix, María Paula Magariños, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic acids research*, 47(D1):D930–D940, 2019.
- Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gwooon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023.
- Chr Møller and Milton S Plesset. Note on an approximation treatment for many-electron systems. *Physical review*, 46(7):618, 1934.
- Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023.
- Badri Narayanan, Paul C Redfern, Rajeev S Assary, and Larry A Curtiss. Accurate quantum chemical energies for 133000 organic molecules. *Chemical science*, 10(31):7449–7455, 2019.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- John P Perdew and Karla Schmidt. Jacob’s ladder of density functional approximations for the exchange-correlation energy. In *AIP conference proceedings*, volume 577, pp. 1–20. American Institute of Physics, 2001.
- John A Pople, Martin Head-Gordon, Douglas J Fox, Krishnan Raghavachari, and Larry A Curtiss. Gaussian-1 theory: A general procedure for prediction of molecular energies. *The Journal of Chemical Physics*, 90(10):5622–5629, 1989.
- Pekka Pyykkö. Relativistic effects in chemistry: more common than you thought. *Annual review of physical chemistry*, 63(1):45–64, 2012.
- Edward O Pyzer-Knapp, Matteo Manica, Peter Staar, Lucas Morin, Patrick Ruch, Teodoro Laino, John R Smith, and Alessandro Curioni. Foundation models for materials discovery—current state and future directions. *Npj Computational Materials*, 11(1):61, 2025.
- Zhuoran Qiao, Matthew Welborn, Animashree Anandkumar, Frederick R Manby, and Thomas F Miller. Orbnet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *The Journal of chemical physics*, 153(12), 2020.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.
- Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.
- Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Big data meets quantum chemistry approximations: the δ -machine learning approach. *Journal of chemical theory and computation*, 11(5):2087–2096, 2015.
- Dmitrij Rappoport and Filipp Furche. Property-optimized gaussian basis sets for molecular response calculations. *The Journal of chemical physics*, 133(13), 2010.
- Vitaly A Rassolov, John A Pople, Mark A Ratner, and Theresa L Windus. 6-31g* basis set for atoms k through zn. *The Journal of chemical physics*, 109(4):1223–1229, 1998.

- Laura E Ratcliff, Stephan Mohr, Georg Huhs, Thierry Deutsch, Michel Masella, and Luigi Genovese. Challenges in large scale quantum mechanical calculations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7(1):e1290, 2017.
- Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: atomistic simulation at scale. *arXiv preprint arXiv:2504.06231*, 2025.
- Rowan. Rowan Benchmarks — benchmarks.rowansci.com. <https://benchmarks.rowansci.com/>. [Accessed September 2025].
- Lars Ruddigkeit, Ruud Van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of chemical information and modeling*, 52(11):2864–2875, 2012.
- Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters*, 108(5):058301, 2012.
- Manas Sajjan, Junxu Li, Raja Selvarajan, Shree Hari Sureshbabu, Sumit Suresh Kale, Rishabh Gupta, Vinit Singh, and Sabre Kais. Quantum machine learning for chemistry and physics. *Chemical Society Reviews*, 51(15):6475–6573, 2022.
- Erwin Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical review*, 28(6):1049, 1926.
- Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Saucedo Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.
- Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International conference on machine learning*, pp. 9377–9388. PMLR, 2021.
- Kristof T Schütt, Stefan Chmiela, O Anatole Von Lilienfeld, Alexandre Tkatchenko, Koji Tsuda, and Klaus-Robert Müller. Machine learning meets quantum physics. *Lecture Notes in Physics*, 2020.
- John C Slater. Note on hartree’s method. *Physical Review*, 35(2):210, 1930.
- Eduardo Almeida Soares, Victor Shirasuna, Emilio Ashton Vital Brazil, Renato Fontoura de Gusmao Cerqueira, Dmitry Zubarev, Tiffany Callahan, and Sara Capponi. Molmamba: A large state-space-based foundation model for chemistry. In *American Chemical Society (ACS) Fall Meeting*, 2024.
- Qiming Sun, Timothy C Berkelbach, Nick S Blunt, George H Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D McClain, Elvira R Sayfutyarova, Sandeep Sharma, et al. Pyscf: the python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1340, 2018.
- David P Tew, Wim Klopper, and Trygve Helgaker. Electron correlation: The many-body problem at the heart of chemistry. *Journal of computational chemistry*, 28(8):1307–1320, 2007.
- Oliver T Unke, Stefan Chmiela, Michael Gastegger, Kristof T Schütt, Huziel E Saucedo, and Klaus-Robert Müller. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature communications*, 12(1):7273, 2021a.
- Oliver T Unke, Stefan Chmiela, Huziel E Saucedo, Michael Gastegger, Igor Poltavsky, Kristof T Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021b.
- O Anatole Von Lilienfeld. Quantum machine learning in chemical compound space. *Angewandte Chemie International Edition*, 57(16):4164–4169, 2018.

- O Anatole von Lilienfeld and Kieron Burke. Retrospective on a decade of machine learning for chemical discovery. *Nature communications*, 11(1):4895, 2020.
- O Anatole von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Exploring chemical compound space with quantum-based machine learning. *Nature Reviews Chemistry*, 4(7): 347–358, 2020.
- Julia Westermayr, Michael Gastegger, Kristof T Schütt, and Reinhard J Maurer. Perspective on integrating machine learning into computational chemistry and materials science. *The Journal of Chemical Physics*, 154(23), 2021.
- Brandon M Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Ghahghanyan, John R Kitchin, Daniel S Levine, et al. Uma: A family of universal models for atoms. *arXiv preprint arXiv:2506.23971*, 2025.
- Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, et al. Mattersim: A deep learning atomistic model across elements, temperatures and pressures. *arXiv preprint arXiv:2405.04967*, 2024.
- Eric C-Y Yuan, Yunsheng Liu, Junmin Chen, Peichen Zhong, Sanjeev Raja, Tobias Kreiman, Santiago Vargas, Wenbin Xu, Martin Head-Gordon, Chao Yang, et al. Foundation models for atomistic simulation of chemistry and materials. *arXiv preprint arXiv:2503.10538*, 2025.