

PAPER • OPEN ACCESS

## A recipe for cracking the quantum scaling limit with machine learned electron densities

To cite this article: Joshua A Rackers *et al* 2023 *Mach. Learn.: Sci. Technol.* **4** 015027

View the [article online](#) for updates and enhancements.

You may also like

- [Addressing the dose perturbation of metallic implant in spinal Stereotactic Body Radiotherapy \(SBRT\)](#)

A T Hoang, W L Jong and N M Ung

- [Pulse Measurements on  \$\text{LiNi}\_{0.6}\text{Co}\_{0.2}\text{Mn}\_{0.2}\text{O}\_2\$  Cathodes with Various Electrode Parameters Affecting the Charge-Discharge Rate Performance](#)

Soo Min Hwang, Dong-Won Kim and Young-Jun Kim

- [The embedded atom method ansatz: validation and violation](#)

M I Baskes and S G Srinivasan



## PAPER

## A recipe for cracking the quantum scaling limit with machine learned electron densities

## OPEN ACCESS

## RECEIVED

13 May 2022

## REVISED

26 October 2022

## ACCEPTED FOR PUBLICATION


13 January 2023

## PUBLISHED

27 February 2023

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Joshua A Rackers<sup>1,\*</sup> , Lucas Tecot<sup>2</sup>, Mario Geiger<sup>3</sup> and Tess E Smidt<sup>4,5</sup><sup>1</sup> Center for Computing Research, Sandia National Laboratories, Albuquerque, NM, United States of America<sup>2</sup> Department of Computer Science, University of California, Los Angeles, CA, United States of America<sup>3</sup> Institute of Physics, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland<sup>4</sup> Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America<sup>5</sup> Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, United States of America

\* Author to whom any correspondence should be addressed.

E-mail: [jracker@sandia.gov](mailto:jrackers@sandia.gov)**Keywords:** machine learning, electron density, quantum chemistry, waterSupplementary material for this article is available [online](#)**Abstract**

A long-standing goal of science is to accurately simulate large molecular systems using quantum mechanics. The poor scaling of current quantum chemistry algorithms on classical computers, however, imposes an effective limit of about a few dozen atoms on traditional electronic structure calculations. We present a machine learning (ML) method to break through this scaling limit for electron densities. We show that Euclidean neural networks can be trained to predict molecular electron densities from limited data. By learning the electron density, the model can be trained on small systems and make accurate predictions on large ones. In the context of water clusters, we show that an ML model trained on clusters of just 12 molecules contains all the information needed to make accurate electron density predictions on cluster sizes of 50 or more, beyond the scaling limit of current quantum chemistry methods.

**1. Introduction**

One of the grand challenges of science is to simulate large molecular systems like solvated biological macromolecules from first principles quantum physics. The ability to perform such *ab initio* simulations would enable us to predict protein-drug binding, analyze the behavior of materials, and design new enzymes. Unfortunately, large quantum chemistry calculations like these are currently impossible. This is not because the laws of quantum mechanics are not well understood; it is because the algorithms with which they are implemented on classical computers scale sharply with increasing system size. For instance, the so-called ‘gold standard’ method of quantum chemistry, coupled cluster, scales as  $O(N^6)$  or  $O(N^7)$  with system size. In practical terms this means that while a coupled cluster calculation on a single water molecule might take only a few minutes, an equivalent calculation on a small protein would take several ages of the universe. This behavior defines the scaling limit of about a few dozen atoms for rigorous quantum chemistry.

Despite the scaling of quantum chemistry algorithms, there is no fundamental rule stipulating that large-scale quantum mechanics calculations cannot be done. The origin of the scaling problem is electron correlation. However, we know both empirically and theoretically that this phenomenon of electron correlation cannot have infinite range. The principle of the ‘nearsightedness of electronic matter’ introduced by Walter Kohn in 1996, states that there must be some scale at which electron correlation approaches zero [1, 2]. Prodan and Kohn give a concrete definition of this concept: the nearsightedness of a given molecular system is the smallest distance such that any perturbing charge density outside of this radius produces a change in the local charge density smaller than some tolerance [2]. In formal terms, the nearsightedness range at a given point is:

$$R(r_0, \Delta\rho) = \min r \text{ s.t. } \Delta\rho(r_0, \rho(r')) < \widetilde{\Delta\rho} \quad (1)$$

where  $r_0$  is the point under consideration,  $\Delta\rho$  is the change in charge density at  $r_0$ ,  $r$  is the distance from  $r_0$ ,  $\rho(r')$  is a perturbing charge density lying entirely outside of  $r$  and  $\widehat{\Delta\rho}$  is the tolerance. In Prodan and Kohn's original formulation, this is an absolute definition;  $\rho(r')$  can be any arbitrarily large perturbation. This definition leads to the conclusion that insulators are 'classically farsighted'. In practice, however, we are rarely interested in arbitrarily large perturbations. In calculations of biomolecular systems near room temperature, for instance, we are mainly interested in perturbations due to the rearrangement of a limited set of nearby atoms. This suggests the possibility of an alternative, context-dependent definition of nearsightedness where the set of perturbations is limited to only those relevant to the system we aim to simulate. Finding the range of this context-dependent nearsightedness is essential to the project of large-scale quantum simulation.

The nearsightedness principle is already employed in some techniques in quantum chemistry. Notable examples include linear-scaling density functional theory (DFT) and local coupled cluster methods [3–9]. Although these methods are nominally linear scaling, their considerable computational expense still renders them untenable for large-scale simulation. To overcome this computational expense, a large body of work has focused on developing machine learned potential energy surfaces, or machine learned force fields. These models, however, all suffer from the same problem: although the goal is to simulate large molecular systems, the training data is limited to small-scale systems for which reference calculations are possible. This leads most machine learning (ML) force fields to miss long-range forces, because they are not in the training data. A recent review by Unke and co-workers gives an excellent example [10]. They examined the ability of various ML force fields to predict the torsional energy profile for a sequence of linear cumulene-hydrocarbon chains of increasing length. They found that as the chain length increases past the cutoff radius of the model, the quality of the ML force field rapidly degrades due to the model's inability to capture long-range interactions. This breakdown indicates that the strategy of ML energies and forces may not be fully capitalizing on the nearsightedness of molecular systems.

In this paper we present a ML model that exploits the nearsightedness of electronic matter to perform *ab initio* electron density calculations on systems beyond the scaling limit of quantum chemistry algorithms. The model is trained to predict the electron density, one of the most fundamental quantum properties of a molecular system. We show that a model which learns the electron density can capitalize on the context-dependent nearsightedness of molecular systems in a way that a model which learns energies and forces does not. Groundbreaking work established that it is possible to machine learn an electron density and subsequent work has highlighted the potential for machine learned density models [11–19]. In particular, this study builds off the pioneering work using equivariant Gaussian process regression to predict electron densities of molecules in [12, 13]. Using a new class of ML algorithm, Euclidean neural networks (e3nns), we show that it is possible to train a model on electron densities of small clusters of molecules and make accurate predictions on large clusters.

The results presented here will demonstrate not only that accuracy on large clusters is empirically possible, but also that there exists a maximum training cluster size beyond which the accuracy of the model does not increase. This suggests the existence of a concrete, system-dependent 'radius of electronic nearsightedness' which the ML model can find and exploit to make possible arbitrarily large, systematically improvable electron density predictions. In this way, the model presented here is more than just a surrogate for quantum chemistry. Here, ML gives us a unique tool to examine the nature of electron correlation in condensed phase molecular systems. We conclude by demonstrating high-level, correlated quantum chemistry electron density predictions on systems of more than 1000 atoms, well beyond the quantum chemistry scaling limit.

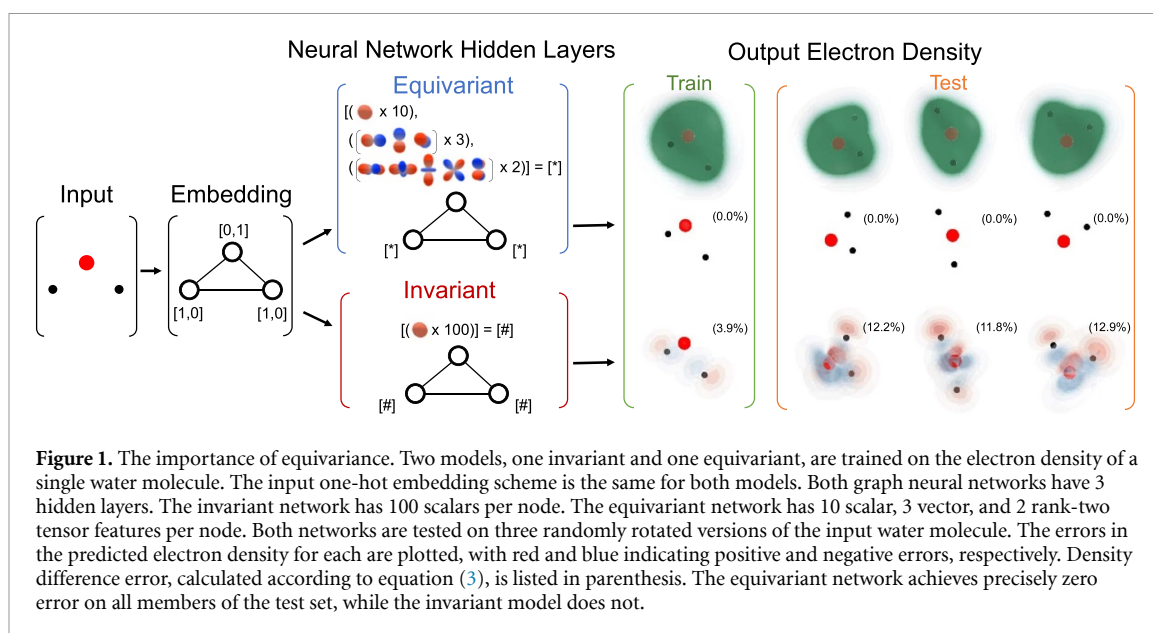
## 2. Results

### 2.1. The importance of equivariance

The basic task of our model is to learn the *ab initio* electron density distribution surrounding a set of molecules. An efficient and well-established way to represent this density is to expand it as a linear combination of atom-centered Gaussian basis functions,

$$\rho(r) = \sum_{i=0}^{N_{\text{atoms}}} \sum_{k=0}^{N_{\text{basis}}} \sum_{l=0}^{l_{\text{max}}} \sum_{m=-l}^l C_{iklm} Y_{l,m} e^{-\alpha_{ikl}(r-r_i)^2} \quad (2)$$

where  $Y_{l,m}$  are the real spherical harmonics,  $\alpha_{ikl}$  are the Gaussian widths of each basis function, and  $C_{iklm}$  are the coefficients for each basis function on each atom [20]. Previous work has shown that standard quantum chemistry 'density fitting' basis sets which specify a list of  $Y_{l,m}$  and  $\alpha_{ikl}$  for each element work well for representing an *ab initio* density [13]. By choosing a density fitting basis set, our ML model is trained simply

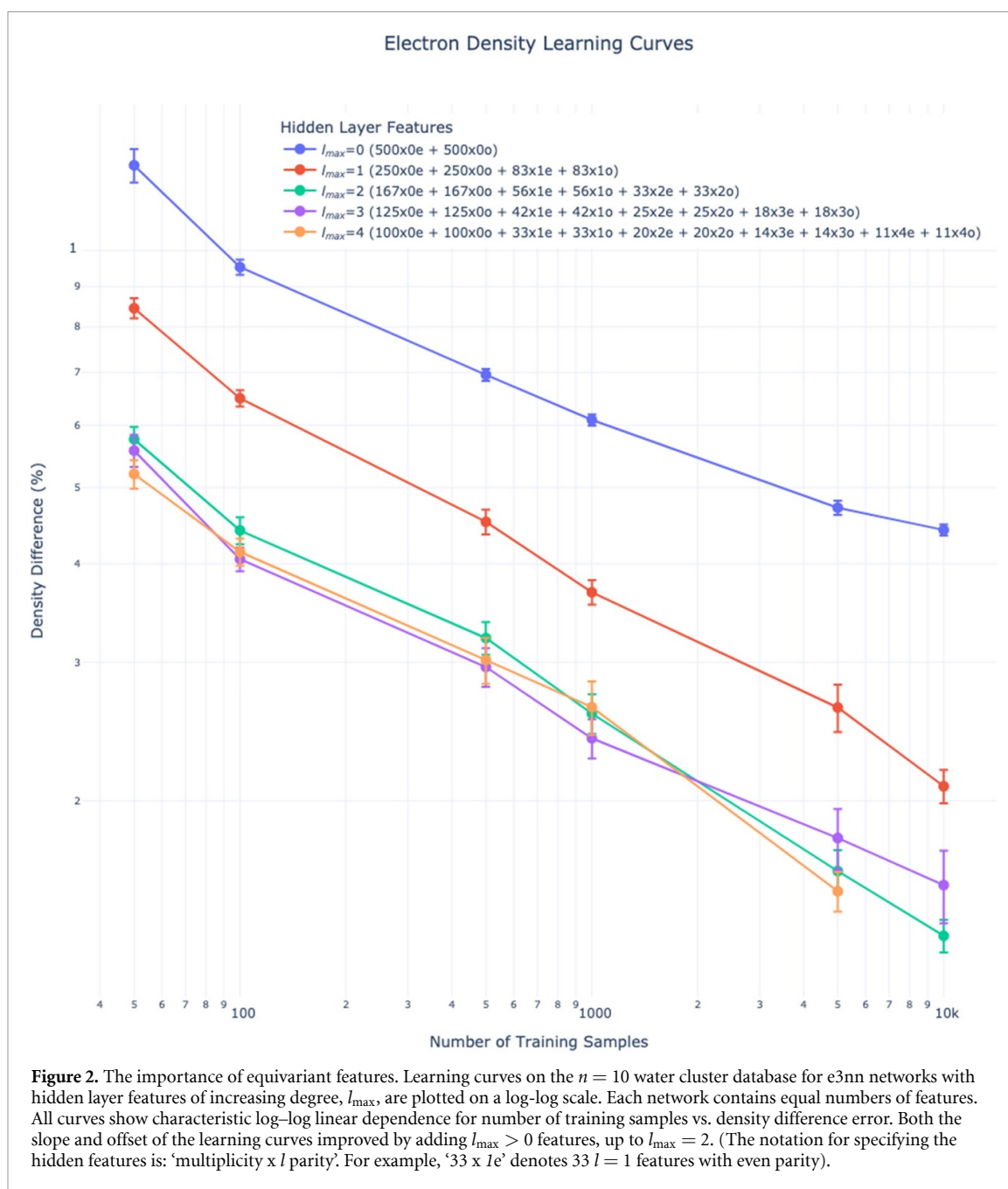


to predict the  $C_{iklm}$  coefficients computed by quantum chemistry. Because each basis function contains a spherical harmonic, each  $C_{iklm}$  coefficient is part of a geometric tensor which represents a 3D contribution to the electron density. For instance, for an  $l = 1$ , or  $p$ -type, basis function there are three  $C_{iklm}$  coefficients that represent the  $x$ ,  $y$ , and  $z$  components of that basis function's contribution to the density.

Learning the coefficients of geometric tensors is a unique task for ML. Objects in 3D space have certain symmetries; namely they are invariant to permutations and translations, and equivariant to rotations. Equivariance is an intuitive concept: if the input coordinates of a molecule are changed by some arbitrary rotation, the various basis functions that compose its density should rotate with it. This symmetry, however, is not a feature of most current ML algorithms. We have recently developed a ML framework, called e3nn, which encodes all the symmetries of 3D Euclidean space, including equivariance [21–23]. Recent work has highlighted the importance of this property for 3D learning problems. The AlphaFold 2, RoseTTAFold, EquiFold, and ARES models for protein and nucleic acid structure prediction use equivariant ML algorithms [24–27]. Our e3nn framework has been shown to reduce the amount of training data needed for 3D data by a factor of 1000, compared with models that do not include symmetry [28]. And previous electron density learning work using Gaussian process regression, another type of ML algorithm, has shown equivariance is a desired property for density predictions [12]. We set out to test the importance of equivariance for this task of neural network-based electron density prediction.

In figure 1 we show that equivariance is required for an electron density model expressed in an atom-centered basis. In this experiment, we overtrained an equivariant e3nn model on the electron density of a single water molecule. Then we tested that model on rotated versions of the same molecule. We also did the same for an invariant model. Invariant ML models are ubiquitous in the ML for chemistry field. They guarantee that the outputs of a model do not change with translation or rotation of a molecule. Figure 1 clearly shows that for this electron density learning task, invariance is not sufficient; equivariance is required. The invariant model produces large, rotation-dependent errors on rotated versions of the water molecule. Because the density representation contains vectors and higher rank tensors, the equivariant model produces identical densities for identical objects, while the invariant model does not.

Beyond this simple example, we set out to find if equivariance confers any practical advantage in data efficiency for a real problem. To test this, we set up a learning problem on electron densities of clusters of ten water molecules. The e3nn framework achieves equivariance by imposing that the features at every hidden layer of the neural network be direct sums of irreducible representations. In 3D space, these can be interpreted as spherical harmonics. For practical construction of networks, the user selects a highest angular frequency or degree,  $l$ , for hidden layer features. Previous work has indicated that the presence of  $l > 0$  features can increase the data efficiency of neural networks [28]. We set out to test the extent to which this is true for electron densities. We trained five different, equal size, neural networks with  $l_{\max}$  of 0, 1, 2, 3, 4, and 5. The results, plotted in figure 2, show a dramatic increase in data efficiency from including  $l_{\max} > 1$  features in the network. A network with  $l_{\max} = 2$ , for instance, achieves a density difference of 0.44% with just 100 training samples, while a  $l_{\max} = 0$  network requires 10 000 training samples, to reach similar accuracy.

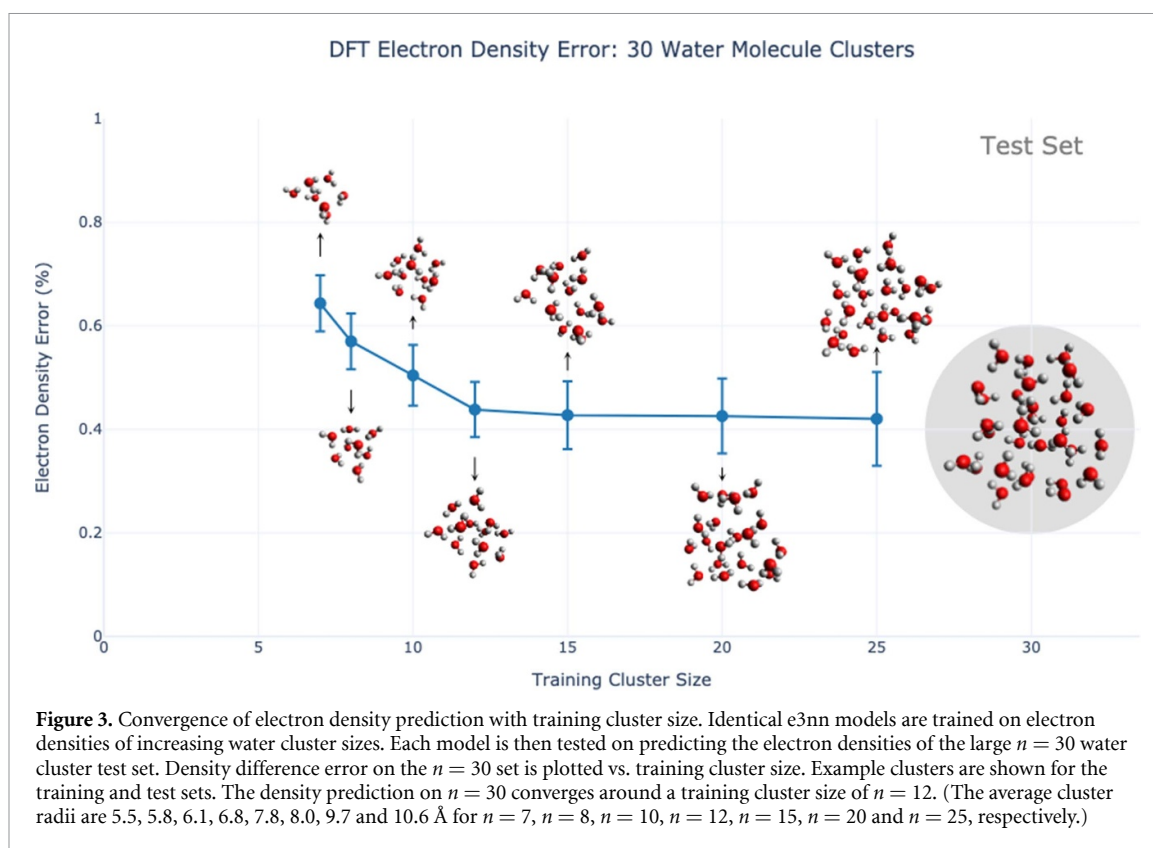


## 2.2. DFT experiments

Our first set of experiments was done using DFT. We chose to use the PBE0 density functional because it produces the most accurate (closest to coupled cluster) electron densities for small test systems [29]. For all experiments presented in this paper, we use the database of water cluster minima, assembled by Rakshit and co-workers [30]. Water clusters make an ideal system of study here because they isolate the problem of learning intermolecular interactions from the problem of learning chemical diversity. The database of water cluster minima contains structures of all known energy minima of water cluster sizes 3–30. All target electron density calculations were performed with the psi4 quantum chemistry package (see supplementary materials) [31]. The scheme of the following experiments is to train an e3nn model on structures of small clusters, and then test the model on the largest clusters.

### 2.2.1. Experiment 1: the effect of cluster size on density prediction

For our first experiment, we set out to investigate the effect of training set cluster size on the predicted electron densities of the largest clusters available, 30 water molecules. We chose to start at a cluster size of  $n = 7$ , since lower cluster sizes all have fewer than 100 structures. We trained separate models on cluster sizes



$n = 7, 8, 10, 12, 15, 20$ , and  $25$ , and then evaluated those models on predictions of the electron density of  $n = 30$ . The quality of the prediction on each structure was calculated using the standard measure of density difference,

$$\epsilon_p(\%) = 100 \frac{\int d\mathbf{r} |\rho_{QM}(\mathbf{r}) - \rho_{ML}(\mathbf{r})|}{\int d\mathbf{r} \rho_{QM}(\mathbf{r})} \quad (3)$$

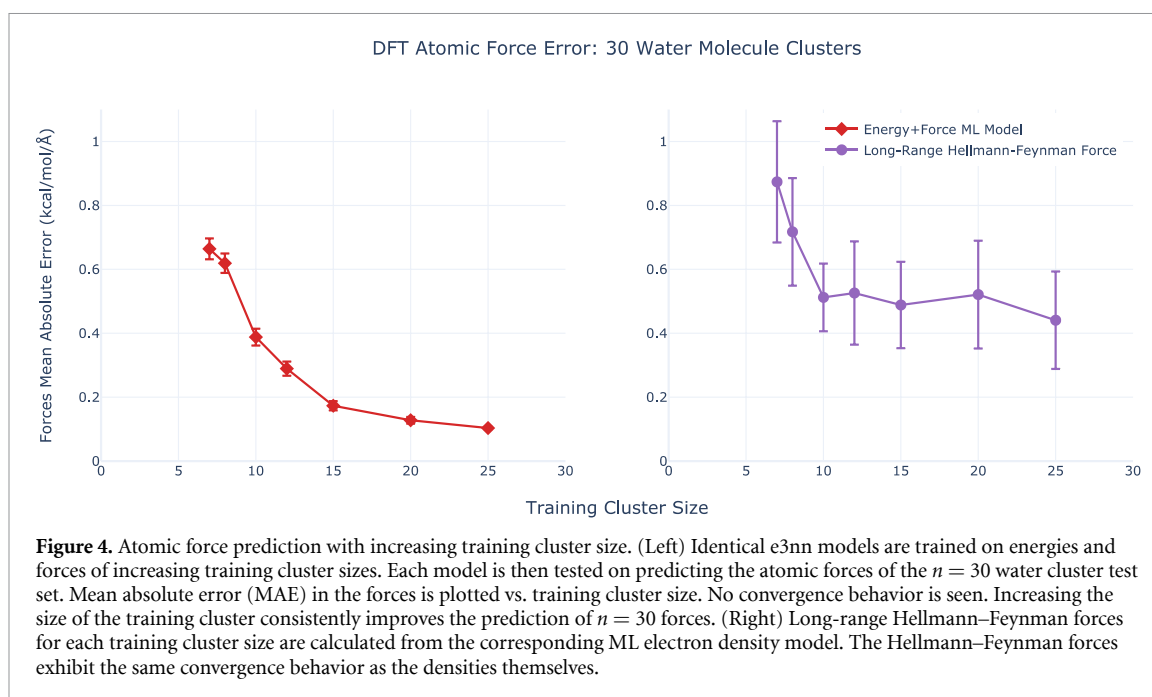
where  $\rho_{QM}$  is the target density,  $\rho_{ML}$  is the predicted density, and the integral is evaluated on a grid of spacing 0.1 Å. A density difference value of 0% indicates identical densities. As a point of reference, superimposing spherical, isolated atom densities typically yields errors of about 20% [12]. For each training set, we made sure to include identical numbers of atoms to normalize the total number of samples in the set.

The results in figure 3 show a striking behavior. For low cluster sizes ( $n = 7-10$ ), the performance on  $n = 30$  predictably increases with increasing cluster size. However, starting at cluster size  $n = 12$ , the performance on  $n = 30$  converges. This means that the e3nn can accurately learn the electron density of any given atom from an environment of just 12 molecules and that this function is virtually unchanged by the addition of more molecules. Even training with  $n = 25$ , a cluster size generally assumed to have more bulk solution properties like  $n = 30$ , does not improve the performance of the model on  $n = 30$ . In this experiment, the model is unable to find any signal of density contributions to a given atom outside of the average radius of a 12 water molecule cluster, about 6.8 Å.

The converged models trained in this experiment achieve average density differences of 0.4%, on par with the current state-of-the-art in electron density prediction. As a point of reference, this error is smaller than the error of the Perdew–Burke–Ernzerhof (PBE) density functional vs. full configuration interaction (FCI) on  $H_2$  in the complete basis set limit [32]. Using the model density trained with cluster size  $n = 12$ , we computed the electrostatic potential at the van der Waals 0.002 e-/bohr<sup>-3</sup> isodensity surface of every  $n = 30$  structure. An example of the accuracy of this electrostatic potential energy surface is shown in figure S5. The average root mean squared error across all structures is 0.006 a.u. This is comparable to state-of-the-art deep neural networks trained directly on the electrostatic potential surface itself [33].

To rule out the possibility that the convergence behavior shown in figure 3 is a result of extrapolation, we performed a set of experiments replicating this experiment for other test set cluster sizes. The results of these tests are analyzed in figures S9 and S10 in the supplementary materials. The tests show identical convergence behavior regardless of test set cluster size, indicating that extrapolation is not a primary cause of convergence.





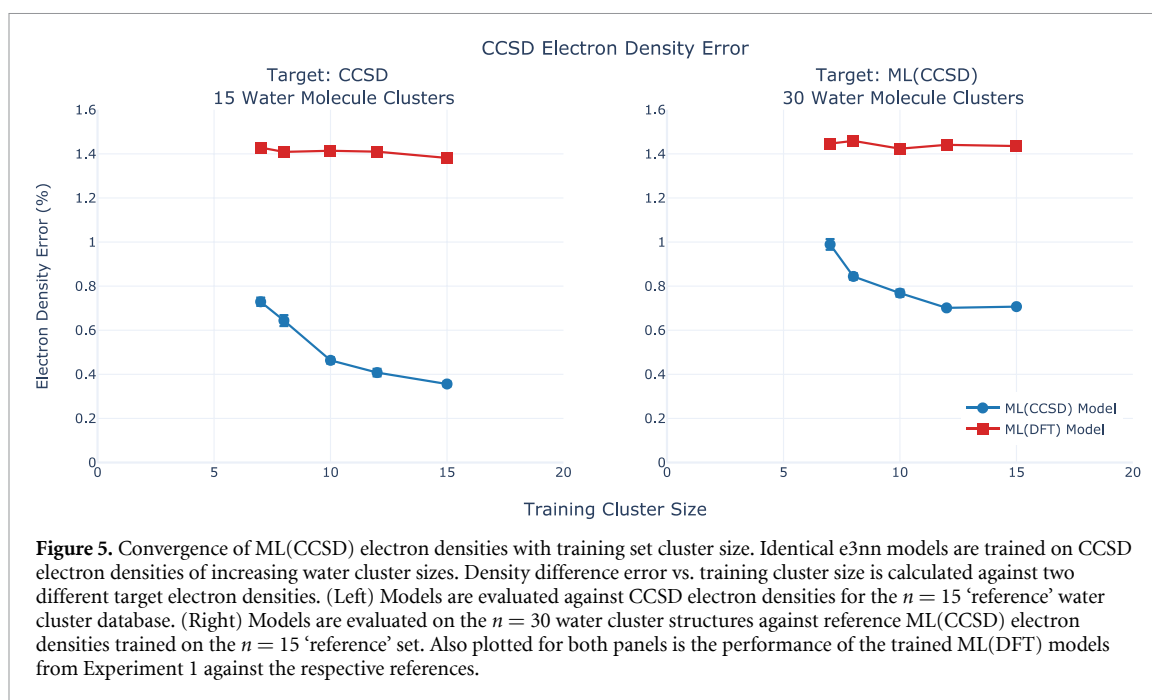
### 2.2.2. Experiment 2: the effect of cluster size on force prediction

The majority of the work in ML for quantum chemistry to date has focused on energies and forces [10]. In our second experiment we set out to see if training an e3nn model on energies and forces would produce the same cluster size convergence behavior as the electron density learning model. The architecture of the model tested is identical to that used in Experiment 1. The only difference is the outputs of this model are energy and forces as opposed to coefficients of basis functions. ML forces are calculated as atomic position derivatives of the energy using backpropagation, ensuring that the learned forces are conservative [34]. We evaluated this model by comparing the predicted force vs. the reference force for the  $n = 30$  water cluster test set.

The results, plotted in figure 4, show that training on energies and forces does not produce the same behavior as training on densities. The accuracy of the resulting models is excellent. The model trained on the  $n = 25$  clusters, achieves a mean absolute error of 0.1 kcal/mol/Å. This error is comparable to state-of-the-art ML force fields, with an error of less than 1% on the 10.2 kcal/mol/Å average total force magnitude of the  $n = 30$  test set [28]. However, there is no convergence as is seen with densities. The mean absolute error in forces on the large system decreases monotonically with system size. This does not mean that convergence will never occur, but it does mean that it has not been reached by  $n = 30$  for this system.

One of the many properties that can be derived from the electron density is forces, via the Hellmann–Feynman theorem. The Hellmann–Feynman theorem states that given an electron density, atomic forces can be calculated exactly from simple electrostatics [35]. In practice, however, the Hellmann–Feynman theorem is not valid for short-range forces with basis sets like the aug-cc-pVTZ basis set used in this work [36–39]. It is perfectly valid, however, to compute long-range forces with the Hellmann–Feynman theorem. To test the behavior of these long-range forces with increasing training set cluster size, we computed the long-range contribution to the Hellmann–Feynman forces (see supplementary materials for details) from the predicted densities of the models trained in Experiment 1. These results are also plotted in figure 4. Like the ML forces, the long-range Hellmann–Feynman forces calculated from the ML electron densities of Experiment 1 are very accurate. For example, the long-range Hellmann–Feynman forces calculated from the ML electron density model trained on the  $n = 12$  dataset give a mean absolute error of 0.5 kcal/mol/Å on the  $n = 30$  test set. This is an error of 3% on the average long-range force magnitude of 16.0 kcal/mol/Å. Note that the average total magnitude of the forces in the long-range case is larger than for the total forces because the structures in the database are close to DFT minima.

Most importantly, figure 4 illustrates that the long-range Hellmann–Feynman forces computed from the electron density model follow the same behavior with respect to cluster size as the densities themselves. The accuracy of the prediction on 30 water molecule clusters converges around a training cluster size of  $n = 12$ . The convergence of these long-range forces corroborates the main observation: equivariant neural networks trained on the electron density converge much more rapidly with cluster size than models trained on energies and forces.



**Figure 5.** Convergence of ML(CCSID) electron densities with training set cluster size. Identical e3nn models are trained on CCSD electron densities of increasing water cluster sizes. Density difference error vs. training cluster size is calculated against two different target electron densities. (Left) Models are evaluated against CCSD electron densities for the  $n = 15$  ‘reference’ water cluster database. (Right) Models are evaluated on the  $n = 30$  water cluster structures against reference ML(CCSID) electron densities trained on the  $n = 15$  ‘reference’ set. Also plotted for both panels is the performance of the trained ML(DFT) models from Experiment 1 against the respective references.

### 2.3. Coupled cluster experiments

Coupled cluster methods are the so-called gold-standard of quantum chemistry. They are accurate and expensive to compute because they systematically include the effects of electron correlation across an entire molecular system. The two most commonly used versions, CCSD (coupled cluster with single and double excitations) and CCSD(T) (coupled cluster with single and double perturbative triple excitations), are used as benchmarks against which other approximate methods are measured [40]. In many cases coupled cluster is accurate enough to be used as a surrogate for experiment when experimental data is unavailable [41]. Because of the  $O(N^6)$  or worse scaling, however, it is virtually impossible to perform these calculations on anything more than a few dozen atoms. We endeavored to see if a ML electron density model can be used to perform correlated coupled cluster calculations that break through this scaling limit.

#### 2.3.1. Experiment 3: coupled cluster density prediction

In Experiment 3, (just as in Experiment 1) we trained ‘experiment’ ML models on CCSD electron densities of training set cluster sizes  $n = 7$ –15. Because CCSD calculations on the  $n = 30$  cluster size are impossible, we were forced to employ two surrogates to evaluate how well these ‘experiment’ models perform on larger structures.

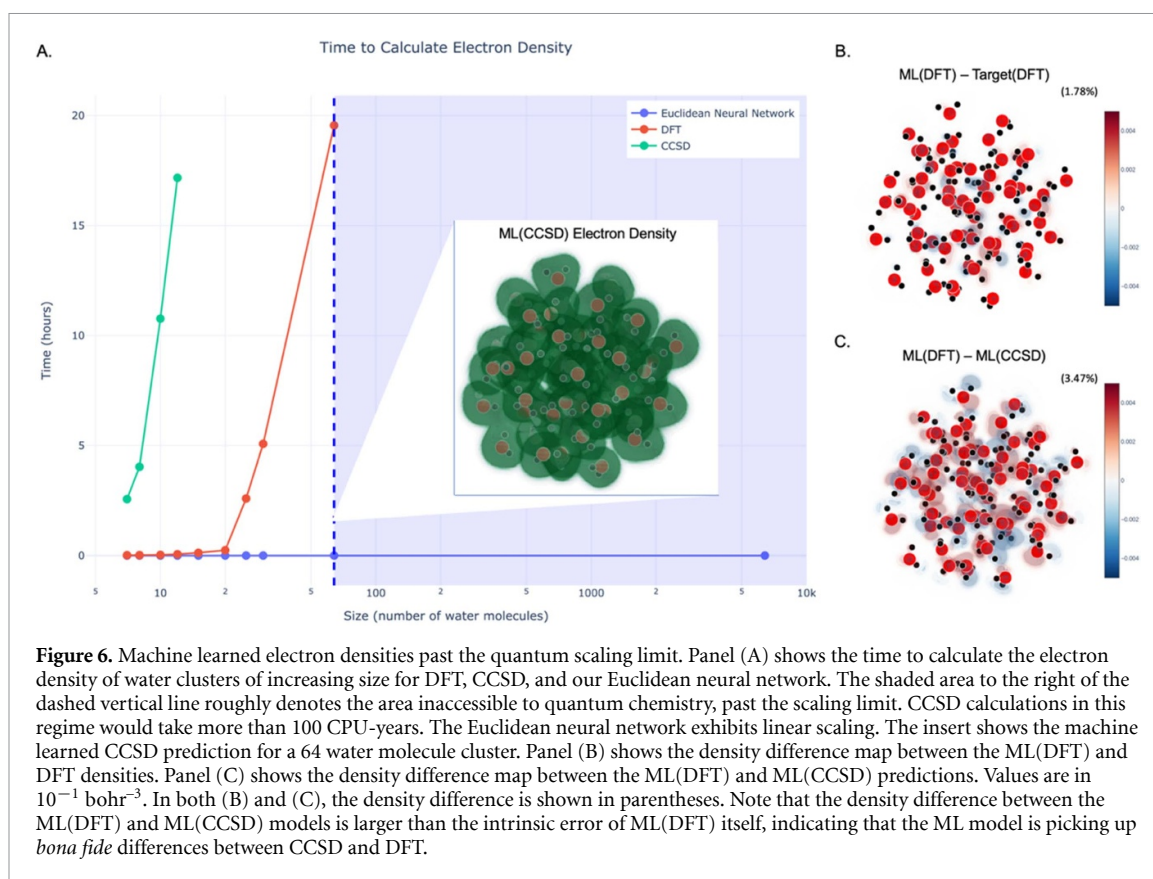
We first created a separate ‘reference’ set of  $n = 15$  CCSD electron density calculations that does not include the data used to train the  $n = 15$  ‘experiment’ model. The left side of figure 5 shows the accuracy of each trained ‘experiment’ model on this  $n = 15$  CCSD ‘reference’ test set. These  $n = 15$  CCSD calculations, however, are lacking as a surrogate for bulk water structure. Therefore, we set out to create a second method to evaluate the quality of the ‘experiment’ models on the  $n = 30$  water cluster structures. To do this, we trained a separate ML model on the ‘reference’  $n = 15$  CCSD electron densities. We then compared the predicted electron densities of the ‘experiment’ models against the electron densities of the ‘reference’ model on the structures of the  $n = 30$  water cluster database. This allows us to make comparisons on  $n = 30$  structures using our best-possible ML(CCSID) estimate. The right side of figure 5 shows this comparison. The curve exhibits the same convergence behavior seen in Experiment 1.

We also tested the ML(DFT) models trained in Experiment 1. We calculated the density difference between the densities predicted by the ML(DFT) and both reference densities. The results show that the ML model can clearly distinguish between CCSD and DFT levels of theory. The error of the ML(DFT) with respect to the reference ML(CCSID) model is consistent with the average density difference error of DFT vs. CCSD from quantum chemistry on the  $n = 7$  set of about 2%.

#### 2.3.2. Experiment 4: large-scale predictions

Lastly, we used the ML(CCSID) electron density network from Experiment 3, trained on clusters of  $n = 12$  water molecules, to predict the electron densities of structures well beyond the scaling limits of quantum





chemistry. In figure 6 we show predictions for clusters of 64 and 6400 molecules. While a calculation on 64 water molecules is outside of the scaling limit for coupled cluster, it is still possible for DFT. Figure 6 shows that the ML model trained on  $n = 12$  water clusters is still accurate for the  $n = 64$  structure. Moreover, the ML models are clearly able to distinguish between DFT and CCSD levels of theory, even for this large structure. On the other hand, 6400 water molecules is beyond the reach of almost any quantum chemistry method. Figure 6 shows that large calculations pose no problem for the neural network. While a full CCSD calculation on 6400 water molecules would take more than the age of the universe, the electron density of this structure can be calculated with near-CCSD accuracy in under one second with a e3nn model on a single graphics processing unit.

### 3. Discussion

The experiments in this work probe the ability of e3nns to learn molecular electron densities. Several interesting observations arise from the results of these experiments.

Equivariance is essential for an accurate ML model of electron densities. The data we are predicting in this task are geometric tensors, and the results in figure 1 show that equivariance is a required attribute for sensible results for this data type. Moreover, figure 2 shows that the presence of higher-order ( $l > 1$ ) tensors in the hidden layers of an equivariant network confers a large increase in data-efficiency. This suggests that having features in the network that have the same data type as the target output is a natural fit for the learning task.

The practical implication from this work is a recipe for how to solve the ‘training data problem’ in the ML for chemistry field. The goal of most ML for chemistry models is to perform calculations on systems that are too large for *ab initio* calculations. The problem, however, is that it is impossible to generate training data for these large systems. This poses an existential conundrum for condensed-phase molecular systems where long-range forces matter. Omitting long-range forces can produce large errors in molecular dynamics simulations [42]. If the training data cannot capture long-range forces, then the model will not be accurate. This work suggests that learning the electron density is a way out of this trap. The data in figures 3 and 5 show that it is possible to train on small clusters of molecules and expect similarly accurate predictions on larger clusters. Importantly, the minimum training cluster radius of 7 Å for water suggests that this strategy is practically viable with current computer hardware.

By restricting the training data used for a given ML model, we were able to articulate a rigorous scientific question: what is the effective radius needed to make accurate predictions of the electronic structure of bulk water? Experiments 1 and 3 directly address this question and show striking convergence behavior. We suggest that this is the result of the neural network picking up an effective ‘radius of electronic nearsightedness’ for water clusters. Figures 3 and 5 show that, up to the sensitivity of the model, electrons outside of a radius of about 7 Å, have no measurable effect on the electron density surrounding an atom. This is not an indirect or implied measurement of nearsightedness. The range of nearsightedness laid out by Prodan and Kohn is couched in terms of the electron density. The machine learned electron density models directly measure the change in electron density due to perturbations in an atom’s surrounding neighborhood.

It is possible that this observation is an artifact of limitations in the model. However, the results of Experiment 2 suggest that this is not the case. In Experiment 2, we used an identical neural network architecture, but predicted atomic forces rather than electron densities. In this case, the network does pick up signal from atoms outside of the 7 Å radius observed for electron densities. This data suggests that learning the electron density may be a better route to the goal of simulating large systems from small-scale training data than learning the forces themselves.

It may seem curious that the density-learning and force-learning experiments produce differing behavior, but perturbation theory offers a straightforward interpretation of this observation. A toy example illustrates this clearly. An atom, A, 10 Å away from another atom, B, can exert a force on atom B without substantively changing atom B’s electronic structure. To compute the force atom A exerts on atom B, the first order contribution is from the electric field generated by A at B. Then, the second and higher order contributions come from the relaxation of the electron density of B in response to the field, field gradient, etc from A. For well-behaved systems at long-range, the first order effect should dominate, with second and higher order contributions producing diminishing effects. This is exactly in line with rigorous symmetry adapted perturbation theory of intermolecular interactions [43]. Moreover, recent work comparing the many-body expansion convergence for energy versus electron density showed that electron densities converge at lower body-order than energies for molecular clusters [44]. Our observation that learned densities converge faster than forces with cluster size is consistent with this result, suggesting that the effects of electron correlation at long range are well-behaved and systematically diminishing for water clusters.

This study demonstrates that ML models can be more than just surrogates. In this case, the ML model does not just reproduce quantum chemistry calculations, it tells us something about quantum chemistry itself. Despite the conceptual appeal of the principle of ‘nearsightedness of electron matter’, applying this principle in practice is challenging. For non-metals, in fact, there is no absolute definition of nearsightedness. It has proved difficult to devise experiments that measure the range of nearsightedness for molecular systems. The ML models here give us an unbiased tool to explore the nature of this range from data. In a general sense, the model probes something unknown (physical electron density correlations) by examining something that is known (statistical correlations in neural networks).

Our results suggest that the same behavior observed for DFT holds for coupled cluster results as well. This is noteworthy because while the density functional used for the ML(DFT) models, PBE0, contains some nonlocal electron correlation, it is not a fully correlated method [45]. CCSD on the other hand, contains full non-local electron correlation (up to double excitations). This makes it a more stringent test of the ‘effective radius of electronic nearsightedness’ argument. The fact that we see the same behavior in figure 5 for CCSD densities as in figure 3 for DFT densities, implies that convergence with training set radius was not an artifact of the DFT method. It suggests that even at higher, inaccessible levels of theory like CCSD(T) or FCI, this trend will hold.

There are limitations to the ML electron density framework presented here. First, the ‘radius of nearsightedness’ suggested here is likely to be highly system dependent. Further work will be required to characterize this limit for more complex systems like biomolecular fragments. Some recent work has started in this area for DNA [46]. Second, while this framework gives a recipe for predicting accurate electron densities of large systems, extending this to forces is not trivial. Using the Hellmann–Feynman theorem to calculate forces with standard basis sets suffers from so-called Pulay errors. These errors are large enough to these forces unusable for applications like molecular dynamics. We note, however, that recent work has shown it is possible to construct basis sets which minimize the Pulay error [47]. Combining these Hellmann–Feynman-optimized basis sets with this ML electron density framework may offer a promising path forward.

## 4. Conclusions

The electron density is one of the most important physical observables of a molecular system. Unfortunately, computing the electron density for large systems, like biomolecules or condensed phase liquids, with

traditional quantum chemistry methods is practically impossible. The work presented here shows that a machine learned electron density model that uses e3nns can break through this quantum scaling limit. In the case of water clusters, our experiments suggest that this is possible because the networks trained on electron densities discover an effective ‘radius of electronic nearsightedness’ which is not apparent from an equivalent model trained on energies and forces. This observation establishes a framework for ML models trained on the electron densities of small systems that can systematically approach previously impossible gold standard quantum chemistry results for arbitrarily large molecular systems. The potential impacts are far-reaching. This framework could be used, for example, to build *ab initio* models of protein dynamics, from a simple training set of protein fragment interactions. These quantum-accurate simulations would revolutionize applications from computational drug design to protein engineering.

### Data availability statement

All quantum chemistry data is publicly available in an online repository at <https://zenodo.org/record/5563139>. All code needed to reproduce the experiments and analysis is publicly available on GitHub at [https://github.com/JoshRackers/equivariant\\_electron\\_density](https://github.com/JoshRackers/equivariant_electron_density).

### Acknowledgments

J A R thanks Shivesh Pathak, Aidan Thompson, Andrew Simmonett, Simon Batzner, and Alex Lee for helpful discussions. L T thanks Cho-Jui Hsieh for mentorship support.

### Funding

J A R and L T were supported by the Harry S. Truman Fellowship and the Laboratory Directed Research and Development Program of Sandia National Laboratories. T E S was supported by the Luis Alvarez Fellowship and the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc. for the US. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the US. Department of Energy or the United States Government.

### Authors contributions

Joshua Rackers: Conceptualization, Methodology, Investigation, Software, Writing—Original Draft.  
Lucas Tecot: Investigation, Formal Analysis, Resources, Writing—Review and Editing.  
Mario Geiger: Software, Writing—Review and Editing.  
Tess Smidt: Methodology, Software, Writing—Review and Editing.

### Conflict of interest

The authors declare no competing interests.

### ORCID iD

Joshua A Rackers  <https://orcid.org/0000-0002-8874-018X>

### References

- [1] Kohn W 1996 Density functional and density matrix method scaling linearly with the number of atoms *Phys. Rev. Lett.* **76** 3168
- [2] Prodan E and Kohn W 2005 Nearsightedness of electronic matter *Proc. Natl Acad. Sci.* **102** 11635–8
- [3] Kudin K N and Scuseria G E 2000 Linear-scaling density-functional theory with Gaussian orbitals and periodic boundary conditions: efficient evaluation of energy and forces via the fast multipole method *Phys. Rev. B* **61** 16440
- [4] Bowler D R, Choudhury R, Gillan M J and Miyazaki T 2006 Recent progress with large-scale *ab initio* calculations: the CONQUEST code *Phys. Status Solidi b* **243** 989–1000
- [5] Skylaris C-K, Haynes P D, Mostofi A A and Payne M C 2005 Introducing ONETEP: linear-scaling density functional simulations on parallel computers *J. Chem. Phys.* **122** 084119
- [6] Shimojo F, Kalia R K, Nakano A and Vashishta P 2001 Linear-scaling density-functional-theory calculations of electronic structure based on real-space grids: design, analysis, and scalability test of parallel algorithms *Comput. Phys. Commun.* **140** 303–14

- [7] Schütz M and Werner H-J 2001 Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD) *J. Chem. Phys.* **114** 661–81
- [8] Riplinger C and Neese F 2013 An efficient and near linear scaling pair natural orbital based local coupled cluster method *J. Chem. Phys.* **138** 034106
- [9] Riplinger C, Sandhoefer B, Hansen A and Neese F 2013 Natural triple excitations in local coupled cluster calculations with pair natural orbitals *J. Chem. Phys.* **139** 134101
- [10] Unke O T, Chmiela S, Sauceda H E, Gastegger M, Poltavsky I, Schütt K T, Tkatchenko A and Müller K-R 2021 Machine learning force fields *Chem. Rev.* **121** 10142–86
- [11] Brockherde F, Vogt L, Li L, Tuckerman M E, Burke K and Müller K-R 2017 Bypassing the Kohn-Sham equations with machine learning *Nat. Commun.* **8** 1–10
- [12] Grisafi A, Fabrizio A, Meyer B, Wilkins D M, Corminboeuf C and Ceriotti M 2018 Transferable machine-learning model of the electron density *ACS Cent. Sci.* **5** 57–64
- [13] Fabrizio A, Grisafi A, Meyer B, Ceriotti M and Corminboeuf C 2019 Electron density learning of non-covalent systems *Chem. Sci.* **10** 9424–32
- [14] Chandrasekaran A, Kamal D, Batra R, Kim C, Chen L and Ramprasad R 2019 Solving the electronic structure problem with machine learning *npj Comput. Mater.* **5** 1–7
- [15] Zepeda-Núñez L, Chen Y, Zhang J, Jia W, Zhang L and Lin L 2021 Deep density: circumventing the Kohn-Sham equations via symmetry preserving neural networks *J. Comput. Phys.* **443** 110523
- [16] Cuevas-Zuñiga B and Pacios L F 2021 Machine learning of analytical electron density in large molecules through message-passing *J. Chem. Inf. Model.* **61** 2658–66
- [17] Ellis J A, Fiedler L, Popoola G A, Modine N A, Stephens J A, Thompson A P, Cangi A and Rajamanickam S 2021 Accelerating finite-temperature Kohn-Sham density functional theory with deep neural networks *Phys. Rev. B* **104** 035120
- [18] Jørgensen P B and Bhowmik A 2020 Deepdft: neural message passing network for accurate charge density prediction (arXiv:2011.03346)
- [19] Jørgensen P B and Bhowmik A 2021 Graph neural networks for fast electron density estimation of molecules, liquids, and solids (arXiv:2112.00652)
- [20] Eichkorn K, Treutler O, Öhm H, Häser M and Ahlrichs R 1995 Auxiliary basis sets to approximate Coulomb potentials *Chem. Phys. Lett.* **240** 283–90
- [21] Thomas N, Smidt T, Kearnes S, Yang L, Li L, Kohlhoff K and Riley P 2018 Tensor field networks: rotation-and translation-equivariant neural networks for 3d point clouds (arXiv:1802.08219)
- [22] Smidt T E 2020 Euclidean symmetry and equivariance in machine learning *Trends Chem.* **3** 82–85
- [23] Geiger M et al 2021 e3nn/e3nn: 2021–12–15 (<https://doi.org/10.5281/ZENODO.3724963>)
- [24] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A and Potapenko A 2021 Highly accurate protein structure prediction with AlphaFold *Nature* **596** 583–9
- [25] Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee G R, Wang J, Cong Q, Kinch L N and Schaeffer R D 2021 Accurate prediction of protein structures and interactions using a three-track neural network *Science* **373** 871–6
- [26] Lee J H, Yadollahpour P, Watkins A, Frey N C, Leaver-Fay A, Ra S, Cho K, Gligorijevic V, Regev A and Bonneau R 2022 EquiFold: protein structure prediction with a novel coarse-grained structure representation *bioRxiv Preprint*
- [27] Townshend R J, Eismann S, Watkins A M, Rangan R, Karelina M, Das R and Dror R O 2021 Geometric deep learning of RNA structure *Science* **373** 1047–51
- [28] Batzner S, Smidt T E, Sun L, Mailoa J P, Kornbluth M, Molinari N and Kozinsky B 2021 SE (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials (arXiv:2101.03164)
- [29] Medvedev M G, Bushmarinov I S, Sun J, Perdew J P and Lyssenko K A 2017 Density functional theory is straying from the path toward the exact functional *Science* **355** 49–52
- [30] Rakshit A, Bandyopadhyay P, Heindel J P and Xantheas S S 2019 Atlas of putative minima and low-lying energy networks of water clusters  $n=3-25$  *J. Chem. Phys.* **151** 214307
- [31] Smith D G, Burns L A, Simmonett A C, Parrish R M, Schieber M C, Galvelis R, Kraus P, Kruse H, Di Remigio R and Alenaizan A 2020 PSI4 1.4: open-source software for high-throughput quantum chemistry *J. Chem. Phys.* **152** 184108
- [32] Bochevarov A D and Friesner R A 2008 The densities produced by the density functional theory: comparison to full configuration interaction *J. Chem. Phys.* **128** 034102
- [33] Rathi P C, Ludlow R F and Verdonk M L 2019 Practical high-quality electrostatic potential surfaces for drug discovery using a graph-convolutional deep neural network *J. Med. Chem.* **63** 8778–90
- [34] Chmiela S, Tkatchenko A, Sauceda H E, Poltavsky I, Schütt K T and Müller K-R 2017 Machine learning of accurate energy-conserving molecular force fields *Sci. Adv.* **3** e1603015
- [35] Feynman R P 1939 Forces in molecules *Phys. Rev.* **56** 340
- [36] Salem L and Wilson E B Jr 1962 Reliability of the Hellmann—Feynman theorem for approximate charge densities *J. Chem. Phys.* **36** 3421–7
- [37] Bakken V, Helgaker T, Klopper W and Ruud K 1999 The calculation of molecular geometrical properties in the Hellmann—Feynman approximation *Mol. Phys.* **96** 653–71
- [38] Fernández Rico J, López R, Ema I and Ramírez G 2004 Accuracy of the electrostatic theorem for high-quality Slater and Gaussian basis sets *Int. J. Quantum Chem.* **100** 221–30
- [39] Fernández Rico J, López R, Ema I and Ramírez G 2002 Density and binding forces in diatomics *J. Chem. Phys.* **116** 1788–99
- [40] Rezáč J, Riley K E and Hobza P 2011 S66: a well-balanced database of benchmark interaction energies relevant to biomolecular structures *J. Chem. Theory Comput.* **7** 2427–38
- [41] Bartlett R J and Musiał M 2007 Coupled-cluster theory in quantum chemistry *Rev. Mod. Phys.* **79** 291
- [42] Loncharich R J and Brooks B R 1989 The effects of truncating long-range forces on protein dynamics *Proteins: Struct. Funct. Bioinform.* **6** 32–45
- [43] Szalewicz K 2012 Symmetry-adapted perturbation theory of intermolecular forces *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2** 254–72
- [44] Schmitt-Monreal D and Jacob C R 2021 Density-based many-body expansion as an efficient and accurate quantum-chemical fragmentation method: application to water clusters *J. Chem. Theory Comput.* **17** 4144–56

- [45] Adamo C and Barone V 1999 Toward reliable density functional methods without adjustable parameters: the PBE0 model *J. Chem. Phys.* **110** 6158–70
- [46] Lee A J, Rackers J A and Bricker W P 2022 Predicting accurate *ab initio* DNA electron densities with equivariant neural networks *Biophys. J.* **121** 3883–95
- [47] Pathak S, Rackers J A, López I E, Fernández R L, Lee A J, Bricker W P and Lehtola S 2022 Accurate Hellmann-Feynman forces with optimized atom-centered Gaussian basis sets (arXiv:2207.03587)