PAPER • OPEN ACCESS

Interpretable delta-learning of GW quasiparticle energies from GGA-DFT

To cite this article: Artem Fediai *et al* 2023 *Mach. Learn.: Sci. Technol.* **4** 035045

View the article online for updates and enhancements.

MACHINE LEARNING
Science and Technology

PAPER

OPEN ACCESS

Interpretable delta-learning of GW quasiparticle energies from GGA-DFT

Artem Fediai[1,3,4,*], Patrick Reiser[1,4] , Jorge Enrique Olivares Peña[1], Wolfgang Wenzel[1] and Pascal Friederich[1,2,*]

[1] Institute of Nanotechnology, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
[2] Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Engler-Bunte-Ring 8, 76131 Karlsruhe, Germany
[3] Current address: Nanomatch GmbH, Griesbachstraße 5, 76185 Karlsruhe, Germany.
[4] Authors contributed equally.
[*] Authors to whom any correspondence should be addressed.

E-mail: artem.fediai@nanomatch.com and pascal.friederich@kit.edu

Abstract

Accurate prediction of the ionization potential and electron affinity energies of small molecules are important for many applications. Density functional theory (DFT) is computationally inexpensive, but can be very inaccurate for frontier orbital energies or ionization energies. The GW method is sufficiently accurate for many relevant applications, but much more expensive than DFT. Here we study how we can learn to predict orbital energies with GW accuracy using machine learning (ML) on molecular graphs and fingerprints using an interpretable delta-learning approach. ML models presented here can be used to predict quasiparticle energies of small organic molecules even beyond the size of the molecules used for training. We furthermore analyze the learned DFT-to-GW corrections by mapping them to specific localized fragments of the molecules, in order to develop an intuitive interpretation of the learned corrections, and thus to better understand DFT errors.

1. Introduction

Accurate prediction of the frontier orbital energies for molecular systems is a persistent challenge for the computational screening [1] of new materials in the field of organic electronics [2–5], functional materials [6], thermo-electrics [7], molecular catalysis [8] or charge donor/acceptor materials [9, 10]. Consequently, there is a high interest in efficient and accurate computational methods to predict electronic structure properties of molecules. However, the demands of large scale screening are often not satisfied by current quantum chemistry methods. Density functional theory (DFT) [11], a relatively computationally 'cheap' quantum chemistry method, suffers from poor prediction of ionization potential and electron affinities (errors of an order of 1 eV are normally observed) [12]. On the other hand, post-Hartree–Fock methods such as coupled cluster [13] or configuration interaction (CI) [14] reach chemical accuracy (1 kcal/mole = 0.0434 eV)[15], but are only applicable to the smallest molecules and scale poorly with molecule size, ranging from $O(n^6)$ to $O(n^{10})$. The $O(n^{3.4})$-scaled GW [16–21] method fills the gap between DFT and more accurate and computationally infeasible post-Hartree–Fock methods, since numerical GW implementations can handle molecules with hundreds of atoms and yield an accuracy of 0.1 eV… 0.2 eV for HOMO/LUMO quasiparticle energy levels. HOMO and LUMO stand for highest occupied and lowest unoccupied molecular orbitals, respectively.

However, high-throughput simulations using the GW method are still problematic for routine usage due to its at least $O(n^{3.4})$ scaling, as tens, hundreds or even thousands of processors are needed to simulate molecules with 10–100 atoms. Here, we explore the possibility of substituting actual GW simulations by various recently published machine learning (ML) methods [22] based on neural networks (NNs) [23], kernel-ridge regression [24] and graph NNs (GNNs) [25–29]. The prospect of improving the absolute

accuracy of cheaper computational methods with ML corrections has been proposed by Ramakrishnan *et al* and is referred to as Δ-learning [30]. It is estimated that by starting from e.g. DFT energies, ML predictions are less prone to generalization errors [31]. GW quasiparticle energies are well suited for Δ-learning as suggested by Çaylak and Baumeier [32] since GW can be applied on e.g. DFT calculations and produce training data for simulations or molecular screenings. A standard data set to test the performance of molecular ML methods is the QM9 dataset that consists of about 133 885 molecules computed at B3LYP level [33, 34].

Here, we use recently reported [35] HOMO and LUMO quasiparticle energies computed at GW level for the entire QM9 dataset to train ML models on the difference between DFT and GW. Using GNNs, we reach a mean absolute error (MAE) of 22 and 31 meV, respectively. Furthermore, we check if the method generalizes towards larger molecules, by testing the ML predictions on molecules with up to 17 atoms other than hydrogen from the GDB-17 dataset [33]. It was found that in particular DimeNet++ [36] model generalizes very well at least up to 17 atoms, preserving the MAE of ∼0.1 eV. An uncertainty of 0.1 eV would be still sufficient considering the accuracy of the GW method being optimistically estimated to be also approximately 0.1 eV.

Since ML models are often considered 'black boxes' [37] and in the light of recent efforts to develop explainable artificial intelligence [38], we furthermore seek interpretable correlations in the learned atom-wise embeddings of trained GNNs and analyze their distribution across model and hyperparameter space, in order to provide simple human-understandable rules to estimate the effect of GW calculations, like for example the assigned atomic delta value of oxygen atoms, which was found to lead to a systematic shift of the delta values towards lower HOMO and higher LUMO energies.

Moreover, by comparing the learning curves, we find better accuracy for delta learning on QM9 with little data compared to directly predicting GW energies. A MAE of 70 meV can be achieved on QM9 using a feed-forward deep NN and RDKit fingerprints[39], which is technically much simpler than GNNs. From a practical point of view, delta-learning is an attractive approach because DFT simulations are computationally cheap, and can be used for high-throughput screening, while aiming for GW accuracy.

2. Results and discussion

2.1. Dataset
Dataset generation is described by Fediai *et al* [35]. Properties of interest learned here are HOMO and LUMO energies, computed at GW level. Eigenvalue self-consistent GW simulations were performed using PBE DFT as a starting point (ev-GW@PBE) as implemented in cp2k [40]. HOMO and LUMO were computed using the augmented correlation-consistent (aug-cc) basis set family extrapolated to the basis set limit [41].

2.2. ML
The learning curve in figure 1 evaluates and compares the performance of typical literature-known ML models for the direct prediction of DFT and GW orbital energies. We trained a set of Kernel ridge regression models on different molecular representations, namely the Coulomb matrix [42], the Spectrum of London and Axillrod–Teller–Muto potential [43] and the Bag of Bonds representation [44]. We used a Laplacian kernel with a small regularization and a hyper-parameter optimized gamma-value as implemented by the *scikit-learn* [45] python package. We computed the representations with implementations in the *QML* [46] and *DScribe* [47] python package using default values and the corresponding size and atom species present in the QM9 dataset. For graph models we chose the well-known SchNet [48] and DimNet++ [36, 49] graph convolutional model as implemented in *kgcnn* [50]. Hyperparameters are chosen according to literature-published values and are given in detail in the SI. All models reach the respective benchmark literature values for the original QM9 dataset in figure 1(a).

Whereas kernel-based methods are in principle fast in learning, they do require a large amount of memory to invert the full kernel matrix. Possible alternatives to reduce memory consumption are sparse kernel or support vector methods, but which were not used in this work. In contrast, NNs or GNNs are slower in learning, even when trained on a GPU, but require less memory and are typically much faster in prediction.

The learning curves in figure 1 suggest that the ML-model performance behaves similarly on all datasets for the multi-task objective of simultaneously predicting orbital energies of HOMO, LUMO and GAP (the model is predicting the gap as one of its output but the GAP labels are simply the difference between HOMO and LUMO energies). The learning curves in figure 1(b) are slightly less steep than the learning curves for the original DFT data. However, this may be also attributed to a non-optimal set of hyper-parameters or the distribution of target values. The GNNs SchNet and DimNet++ performed best for both DFT and GW energies for large amounts of training data. With little data, kernel methods trained on specialized representations could outperform graph networks under certain conditions.
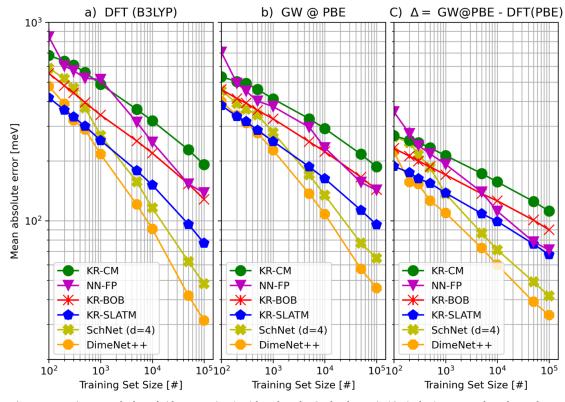
Figure 1. Learning curve for kernel ridge regression (KR) based on the Coulomb matrix (CM), the Spectrum of London and Axillrod–Teller–Muto potential (SLATM) and the Bag of Bonds (BOB) representation as well as for SchNet and DimeNet++ graph neural networks (GNN). Fully connected neural networks (NN) are also trained on molecular fingerprint vectors (FP) that do not include geometric information. The test mean absolute error (MAE) of combined HOMO, LUMO and gap is plotted versus the training set size. Panel (a) shows the MAE values for a model that predicts directly the original DFT labels computed with a B3LYP functional as a baseline. In panel (b), the corresponding MAE of the direct prediction of the GW @ PBE energies are shown. In contrast to that, the model used in panel (c) only predicts the energy difference between GW and DFT with PBE functional, and the MAE values for GW energies are shown.

For delta-learning the ML-model predicts the difference of the quasi-particle energies of GW in the basis set limit and the corresponding DFT orbitals:

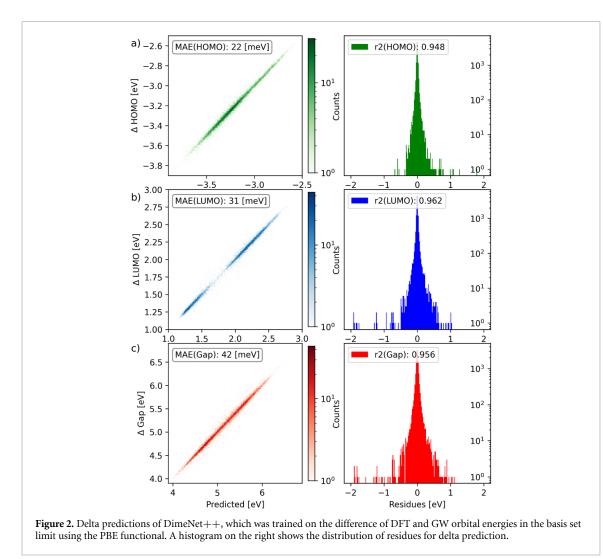$$\Delta = E_{\mathrm{GW}} - E_{\mathrm{DFT}}.$$

Note that in the GW reference calculations of Fediai *et al* [35] both GW and DFT are computed with PBE functional in contrast to the original QM9 DFT labels of Ramakrishnan *et al* [34] which used the B3LYP functional.

The MAE of delta prediction, which is also the MAE of GW from delta predictions given DFT labels, in figure 1(c) is lower compared to the error of the direct GW predictions in figure 1(b) for all models and training set sizes.

We want to point out that fewer training samples in figure 1(c) are required (even 100–1000) to achieve GW accuracy compared to figure 1(b), which reflects the advantage of the delta learning approach as suggested by Ramakrishnan *et al* [30]. This opens the prospect to generate a dataset of orbital energies for molecules of practical relevance, and to perform screening tasks using the combination of DFT and ML models with GW accuracy. In contrast to more expensive post-Hartree–Fock methods, GW can still be applied to large molecules relevant for example as organic semiconductors [51].

For delta learning we trained GNNs such as SchNet and DimeNet++ which will be analyzed more closely in the following. The multi-task model predictions for DimeNet++ are shown in figure 2. The model of DimeNet++ in figure 2 is set up with 4 convolutional layers, expansion of the input into 7 spherical and 6 radial Bessel basis functions, an embedding size of 128 and a range cutoff of 5 Å. Training was performed with the moving-average ADAM optimizer with a learning rate of 1×10^{-3} and an exponential decrease over 872 epochs and a warm-up phase of 3000 steps.

The results in figure 2 shows that the MAE is better for each task than the direct prediction of GW orbital energies in figure 1. Therefore, delta predictions for GW based on comparably cheap DFT calculations can serve as a post-processing step to obtain more realistic energies for screening tasks [52].

Figure 2. Delta predictions of DimeNet++, which was trained on the difference of DFT and GW orbital energies in the basis set limit using the PBE functional. A histogram on the right shows the distribution of residues for delta prediction.
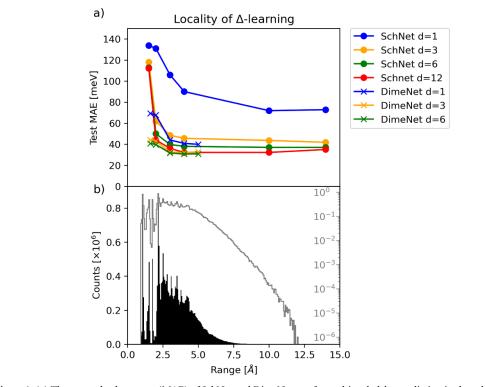
ML approaches in materials science are often criticized for the lack of interpretability. Their complexity makes them a black box from a human perspective. By analyzing the trained model, we seek to find interpretable results that could give us information to understand the difference or the contribution of GW with respect to DFT. For example, the non-locality of orbital energies and more specifically of the GW-DFT delta values can be indirectly probed by checking the ML model performance as a function of the effective field of view defined by the convolutional units. With effective field of view, we denote the radius within which information can be exchanged between atoms in the message passing phase of deep GNNs [28], before in the global pooling step all atomic contributions are summed up.
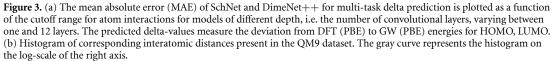
The degree of locality or the necessary range of information being passed to correctly reproduce the delta values are visualized in figure 3. We trained SchNet on the difference of DFT to GW energies with an embedding dimension of 128 and a varying depth from 1 to 12 layers. The interatomic distance of the SchNet interaction block was expanded in a Gaussian basis of 100 bins over 20 Å and a sigma of 0.4 Å for all models. The cutoff, however, restricting the interaction or message passing range was varied between 1.5 and 15 Å. It is to note that the effective field of view for deep convolutional models is a combination of window size and depth, since multiple message passing steps lead to a longer path information can travel on.

We find that having a bond range of about 5 Å is required to achieve good model performance for SchNet. For the 1 layer SchNet model, the MAE continues to decrease until the maximum distance present in the dataset is reached. For deeper models this seems not necessary and the MAE slightly decreases with depth even for fully connected molecules.

Although figure 3 shows that the learned GW-delta prediction of SchNet requires non-local information exchange, it can be instructive to analyze the (local) atomic and structural contributions to the difference of DFT and GW. This can help to find some interpretation on which groups or atoms cause larger shifts to GW orbitals as fitted by the ML model and therefore helps to understand what local environments lead to which DFT errors compared to GW. Most GNNs, including SchNet and DimeNet++, obtain the final graph embeddings from node-level embeddings which are aggregated by a sum- or mean-pooling operation. In the

Figure 3. (a) The mean absolute error (MAE) of SchNet and DimeNet++ for multi-task delta prediction is plotted as a function of the cutoff range for atom interactions for models of different depth, i.e. the number of convolutional layers, varying between one and 12 layers. The predicted delta-values measure the deviation from DFT (PBE) to GW (PBE) energies for HOMO, LUMO. (b) Histogram of corresponding interatomic distances present in the QM9 dataset. The gray curve represents the histogram on the log-scale of the right axis.

case of SchNet and DimeNet++, the final node or atom-embedding is a single value which is averaged for each molecule, or in other words, the output delta prediction is a mean of atom-wise delta-predictions. In figure 4, we plotted the distribution of atomic delta values generated for the QM9 dataset.

We observe different non-standard distributions for each atom species, increasing or decreasing the average energy delta. While some overall aspects of the atom-wise corrections are consistent across models and hyperparameters and allow for interpretable analysis, there are also variations in the specific distributions dependent on the model and its parameters since we do not expect the ML models to converge to a unique global minimum but rather to a good, probably close-lying local minimum.
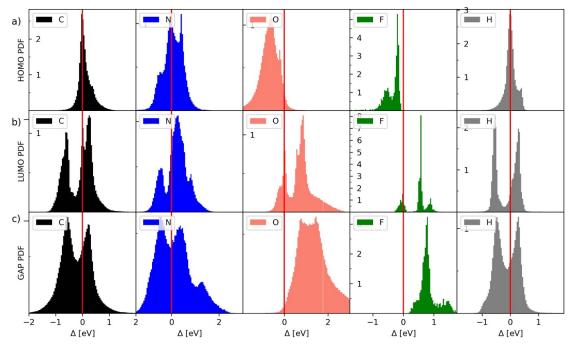
This means that the distribution of delta contributions depends on the model's decision to assign atomic energies based on the messages exchanged with its neighborhood. In the following we give a short discussion on the experimental findings.
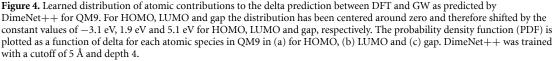
Both SchNet and DimeNet++ show a broader distribution of atomic delta values for shallow models of e.g. figures S1.1 and S2.3 compared to figure 4. This is either due to the reduced long-range interaction or the reduced number of parameters of the model.

The distribution of HOMO delta energies of carbon and hydrogen shows roughly a single component centered at zero, whereas the distribution of LUMO features two components which map the distribution of target delta values. Interestingly, we find a systematic shift of delta values assigned to oxygen towards lower HOMO and higher LUMO values across all models and hyperparameters (compare figures S1 and S2). In contrast, the distribution of fluorine changes depending on the model type but remains comparable across hyperparameters within one model type. The difference of oxygen and fluorine has its origin in the dataset itself. Since fluorine atoms are only bound to carbon atoms, its influence on the delta value can be either attributed to the respective carbon atom or the fluorine atom itself, so the model can assign the delta energy contribution either to a carbon atom that is connected to fluorine, or to fluorine that is connected to carbon. This ambiguity can only be resolved if more perfluorinated structures are added to the training data.

Therefore, it is instructive to generate a more refined view which can be achieved when further separating the atomic contributions of figure 4 into their chemical environment. For example, the chemical environment of a carbon atom is defined as the set of neighboring atoms and their respective bond types, i.e. single, double or aromatic bond. All possible combinations occurring in the QM9 dataset are too extensive to show here, but can be found in the SI. The distribution of all chemical environments for each atom in QM9

Figure 4. Learned distribution of atomic contributions to the delta prediction between DFT and GW as predicted by DimeNet++ for QM9. For HOMO, LUMO and gap the distribution has been centered around zero and therefore shifted by the constant values of -3.1 eV, 1.9 eV and 5.1 eV for HOMO, LUMO and gap, respectively. The probability density function (PDF) is plotted as a function of delta for each atomic species in QM9 in (a) for HOMO, (b) LUMO and (c) gap. DimeNet++ was trained with a cutoff of 5 Å and depth 4.
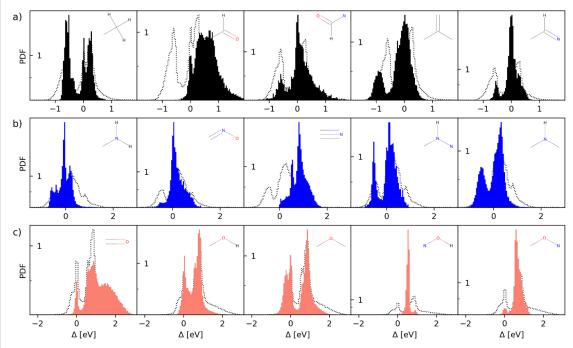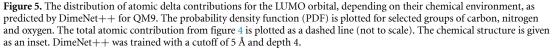
is plotted in the SI (figures S5–S8). In figure 5, we plotted selected groups for HOMO, LUMO and gap. The distribution for a specific chemical group, i.e. the set of bond neighbors, clearly differs from the simple atom-wise delta distribution. As expected, with increasing resolution of the substructure, the distribution of delta values gets more diverse for each central atom. We can identify more trends in the distribution of chemical groups for the central atom across models and hyperparameters like the shift towards positive deltas for aldehydes (figures S3 and S4). Other consistent observations include that nitrile groups (across all models, see figure S3) have substantially underestimated LUMO energies requiring positive corrections of up to 1–2 eV, while amine groups have slightly overestimated LUMO energies in DFT and thus need slightly negative corrections. This presumes that there is a preferred assignment in a local fit of the true GW delta corrections. However, since there is no unique global solution due to the ambiguity to distribute the atomic delta contributions, there are also considerable variations in some chemical environments (like e.g. isobutylene), but since trained GNNs need to recover the correct molecular delta value by being restricted to atomic delta contributions, and no distribution is forced upon on the loss function, the learned distribution should be related to the task, if the model generalizes well. For example, in bad-performing models, the deviations are typically more pronounced with regard to a group of accurate models (e.g. comparing the accurate models in figures S3.2, S3.3, S4.1 of high similarity with inaccurate models of figures S3.1, S3.5 of larger deviations). Nonetheless, the presented data offers insight into the model's fitted solution of attributing atomic delta corrections for the QM9 dataset.

2.3. Generalization to larger molecules

It is known that very often ML models trained on datasets such as QM9 do not generalize very well to larger molecules. Here we want to test the delta ML model trained on QM9 to predict the HOMO, LUMO and GAP energies as a three dimensional target for a set of organic semiconductors. To this end, we have chosen molecules from the *chemical universe* database-project [53] named GDB-17 [33] that consist of 166 billion organic small molecules with up to 17 atoms of C, N, O, S, and halogens. This dataset is the successor of the previously published GDB-11 [53] and GDB-13 [54] dataset with molecules up to 11 and 13 atoms, respectively. More specifically, we took a certain subset of 50 million molecules from the GDB-17, the GDB-17-Set, as generated in Ruddigkeit *et al* [33], available online, and further selected 100 molecules of every size (i.e. different counts of non-hydrogen atoms), starting from size 10–17. To achieve a uniform sampling of the molecular space for every molecular size, molecules were represented as RDkit fingerprints [39], 100 centers were found using *k*-means clustering, and for every center the closest molecule (in the sense

Figure 5. The distribution of atomic delta contributions for the LUMO orbital, depending on their chemical environment, as predicted by DimeNet++ for QM9. The probability density function (PDF) is plotted for selected groups of carbon, nitrogen and oxygen. The total atomic contribution from figure 4 is plotted as a dashed line (not to scale). The chemical structure is given as an inset. DimeNet++ was trained with a cutoff of 5 Å and depth 4.

of Euclidean distance) is finally selected and used for as a test set. Molecules with halogens were excluded from consideration.

As the molecules stored in GDB-17 are molecular graphs, represented in SMILES format, we made a three-step procedure to obtain geometry of those molecules. First an initial conformer is guessed using RDkit and then a conformer search has been carried out using CREST. Finally, the lowest lying conformer for each molecule was optimized at DFT level with the 6–311G** basis set and the functional B3-LYP (Gaussian [55]) in TURBOMOLE [56, 57], mimicking the level of theory for geometry optimizations in the QM9 data set (6–31G(s,p,d)/B3-LYP). Note that final single point calculations were performed in this work for the entire QM9 dataset as well as the GDB-17 subset on the same level of theory. Geometries of the resulting molecules can be found in the Supporting Information. Finally, we conducted GW and DFT calculations for the GDB-17-Subset with the PBE functional according to Fediai *et al* [35].

Figure 6 demonstrates how DimeNet++ model trained on the QM9 dataset generalizes to molecules with up to 17 atoms (excluding hydrogen). While having a chemical accuracy < 0.05 eV in predicting HOMO and LUMO for 10-atom molecules, the model accuracy degrades to approximately 0.1 eV for HOMO and LUMO and 0.15 for gaps in case of larger molecule sizes (figure 6(c)). Considering an optimistic estimation of GW accuracy to be 0.1 eV, the generalization to 'QM17' molecules is sufficiently good. Note that a general decrease of accuracy for out-of-distribution predictions is to be expected and that improving the generalization capability of GNNs is a main focus of current research. In practice, active learning approaches are used to systematically extend the training data distribution and to improve the GNN performance on the specific task at hand [58, 59].

Furthermore, figure S9 demonstrates that the Schnet model generalizes slightly worse towards larger molecules than the DimeNet++ model. There is a clear trend for increasing MAE for larger molecules, and MAE of ∼0.17 eV for HOMO and LUMO and ∼0.32 eV for gaps in case of molecules with 17 non-H atoms. Thus, being almost equally accurate for the QM9 dataset, DimeNet++ outperforms Schnet for larger molecules. We do not have a conclusive explanation for this. There is a spread in performance for different hyperparameters for SchNet, and we hypothesize that with a larger message passing range and sum operations for aggregation, a change in the average number of interacting atoms can lead to shifted hidden representation, since we find that a SchNet model with smaller cutoff radius and medium depth extrapolates better than the best SchNet model of figure 3 having largest depth and highest cutoff radius (compare table S1 and figure S13).

As a side remark, the SchNet model is cheaper and faster to train than DimeNet++, which requires angle computations. Furthermore, a fully-connected NN on fingerprints can hardly generalize to larger molecules
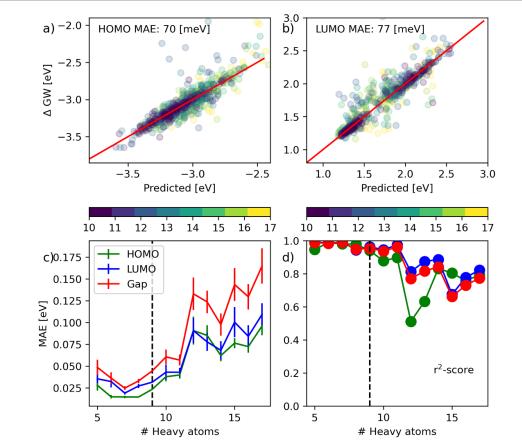
Figure 6. Predictions of the DimeNet++ model on the delta energies between GW and DFT for molecules with 10–17 heavy atoms, where the model was trained on QM9 only, i.e. molecules with up to 9 heavy atoms. Scatter plot for HOMO and LUMO orbital delta energies in (a) and (b). The color map for (a) and (b) encodes the number of heavy atoms. The mean absolute error and r^2-score of the predictions is plotted as a function of the size of the molecule in (c) and (d), respectively. The legend in (c) matches the curves in (d).

at all, as shown in figure S10. Errors of up to 0.25–0.4 eV are observed compared to the errors of 0.1 eV for the DimeNet++ model.

3. Conclusion

We applied delta ML to GW orbital energies of small molecules and tested its generalization to larger molecules, which is a highly demanding task for screening applications in materials science. A set of ML models was trained on the QM9 benchmark set reproducing literature performance. The prediction of delta energies with respect to DFT orbitals was found to be more precise than directly predicting GW orbital energies, becoming especially distinct for little training data, thus showing the advantage of the delta-learning approach. We investigated the fitted distribution of atomic delta energies with respect to their chemical environment, offering some interpretability and insight into ML models such as GNNs. Finally, we tested the generalization capabilities of the ML models on larger molecules up to 17 atoms, excluding hydrogens. We find that DimeNet++ can sufficiently generalize to larger molecules with a mean error of <0.1 eV, on par with the approximate accuracy of GW. In summary, we believe that delta-learning can be used in connection with GW for real-world screening applications requiring accurate orbital energies.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

ORCID iDs

Patrick Reiser ⦿ https://orcid.org/0000-0002-7052-696X
Pascal Friederich ⦿ https://orcid.org/0000-0003-4465-1465

References

[1] Curtarolo S, Hart G L W, Nardelli M B, Mingo N, Sanvito S and Levy O 2013 The high-throughput highway to computational materials design *Nat. Mater.* **12** 191–201

[2] Jacobs I E and Moulé A J 2017 Controlling molecular doping in organic semiconductors *Adv. Mater.* **29** 1703063

[3] Reiser P, Konrad M, Fediai A, Léon S, Wenzel W and Friederich P 2021 Analyzing dynamical disorder for charge transport in organic semiconductors via machine learning *J. Chem. Theory Comput.* **17** 3750–9

[4] Armleder J, Strunk T, Symalla F, Friederich P, Enrique Olivares Peña J, Neumann T, Wenzel W and Fediai A 2021 Computing charging and polarization energies of small organic molecules embedded into amorphous materials with quantum accuracy *J. Chem. Theory Comput.* **17** 3727–38

[5] Symalla F, Fediai A, Armleder J, Kaiser S, Strunk T, Neumann T and Wenzel W 2020 43–3: ab-initio simulation of doped injection layers *SID Symp. Digest of Technical Papers* vol 51 pp 630–3

[6] Qu X, Jain A, Rajput N N, Cheng L, Zhang Y, Ong S P, Brafman M, Maginn E, Curtiss L A and Persson K A 2015 The electrolyte genome project: a big data approach in battery materials discovery *Comput. Mater. Sci.* **103** 56–67

[7] Liang Z *et al* 2018 Influence of dopant size and electron affinity on the electrical conductivity and thermoelectric properties of a series of conjugated polymers *J. Mater. Chem.* A **6** 16495–505

[8] Gaggioli C A, Stoneburner S J, Cramer C J and Gagliardi L 2019 Beyond density functional theory: the multiconfigurational approach to model heterogeneous catalysis *ACS Catal.* **9** 8481–502

[9] Biswal D, Jha A and Sen A 2021 Screening donor and acceptor groups for organic azo-based dyes for dye sensitized solar cells *J. Mol. Struct.* **1228** 129776

[10] Zhao K, Omar Ö H, Nematiaram T, Padula D and Troisi A 2021 Novel thermally activated delayed fluorescence materials by high-throughput virtual screening: going beyond donor–acceptor design *J. Mater. Chem.* C **9** 3324–33

[11] Burke K 2012 Perspective on density functional theory *J. Chem. Phys.* **136** 150901

[12] Zhang G and Musgrave C B 2007 Comparison of DFT methods for molecular orbital eigenvalue calculations *J. Phys. Chem.* A **111** 1554–61

[13] Bartlett R J and Musiał M 2007 Coupled-cluster theory in quantum chemistry *Rev. Mod. Phys.* **79** 291–352

[14] Szalay P G, Müller T, Gidofalvi G, Lischka H and Shepard R 2012 Multiconfiguration self-consistent field and multireference configuration interaction methods and applications *Chem. Rev.* **112** 108–81

[15] van Leeuwen R and Baerends E J 1994 Exchange-correlation potential with correct asymptotic behavior *Phys. Rev.* A **49** 2421–31

[16] Aryasetiawan F and Gunnarsson O 1998 The GW method *Rep. Prog. Phys.* **61** 237–312

[17] Kaplan F 2015 Quasiparticle self-consistent GW-approximation for molecules. Calculation of single-particle excitation energies for molecules *Dissertation* Karlsruher Institut für Technologie (KIT) (https://doi.org/10.5445/IR/1000052904)

[18] van Setten M J, Weigend F and Evers F 2013 The GW-method for quantum chemistry applications: theory and implementation *J. Chem. Theory Comput.* **9** 232–46

[19] Wilhelm J, Del Ben M and Hutter J 2016 GW in the Gaussian and plane waves scheme with application to linear acenes *J. Chem. Theory Comput.* **12** 3623–35

[20] Hüser F, Olsen T and Thygesen K S 2013 Quasiparticle GW calculations for solids, molecules, and two-dimensional materials *Phys. Rev.* B **87** 235132

[21] Hedin L 1965 New method for calculating the one-particle Green's function with application to the electron-gas problem *Phys. Rev.* **139** A796–823

[22] Ramakrishnan R and von Lilienfeld O A 2017 Machine learning, quantum chemistry, and chemical space *Reviews in Computational Chemistry* (Wiley) pp 225–56

[23] Behler J 2015 Constructing high-dimensional neural network potentials: a tutorial review *Int. J. Quantum Chem.* **115** 1032–50

[24] Rupp M 2015 Machine learning for quantum mechanics in a nutshell *Int. J. Quantum Chem.* **115** 1058–73

[25] Reiser P *et al* 2022 Graph neural networks for materials science and chemistry *Commun. Mater.* **3** 93

[26] Scarselli F, Gori M, Tsoi A C, Hagenbuchner M and Monfardini G 2009 The graph neural network model *IEEE Trans. Neural Netw.* **20** 61–80

[27] Xu K, Hu W, Leskovec J and Jegelka S 2019 How powerful are graph neural networks? *7th Int. Conf. on Learning Representations ICLR 2019* (*New Orleans, USA*, *6–9 May*) (available at: https://openreview.net/forum?id=ryGs6iA5Km)

[28] Gilmer J, Schoenholz S S, Riley P F, Vinyals O and Dahl G E 2017 Neural message passing for quantum chemistry *Proc. 34th Int. Conf. on Machine Learning* (*Sydney, Australia*, *6–11 August*) vol 70 (available at: https://proceedings.mlr.press/v70/gilmer17a.html)

[29] Kipf T N and Welling M 2017 Semi-supervised classification with graph convolutional networks *5th Int. Conf. on Learning Representations* (*Toulon, France*, *24–26 April*) (available at: https://openreview.net/forum?id=SJU4ayYgl)

[30] Ramakrishnan R, Dral P O, Rupp M and von Lilienfeld O A 2015 Big data meets quantum chemistry approximations: the Δ-machine learning approach *J. Chem. Theory Comput.* **11** 2087–96

[31] Dandu N, Ward L, Assary R S, Redfern P C, Narayanan B, Foster I T and Curtiss L A 2020 Quantum-chemically informed machine learning: prediction of energies of organic molecules with 10–14 non-hydrogen atoms *J. Phys. Chem.* A **124** 5804–11

[32] Çaylak O and Baumeier B 2021 Machine learning of quasiparticle energies in molecules and clusters *J. Chem. Theory Comput.* **17** 4891–900

[33] Ruddigkeit L, van Deursen R, Blum L C and Reymond J-L 2012 Enumeration of 166 billion organic small molecules in the Chemical Universe Database GDB-17 *J. Chem. Inf. Model.* **52** 2864–75

[34] Ramakrishnan R, Dral P O, Rupp M and von Lilienfeld O A 2014 Quantum chemistry structures and properties of 134 kilo molecules *Sci. Data* **1** 140022

[35] Fediai A, Reiser P, Peña J E O, Friederich P and Wenzel W 2023 Accurate GW frontier orbital energies of 134 kilo molecules (arXiv:2303.08708)

[36] Klicpera J, Giri S, Margraf J T and Günnemann S 2020 Fast and uncertainty-aware directional message passing for non-equilibrium molecules (arXiv:2011.14115 [cs.LG])

[37] Adadi A and Berrada M 2018 Peeking inside the black-box: a survey on explainable artificial intelligence (XAI) *IEEE Access* **6** 52138–60

[38] Barredo Arrieta A *et al* 2020 Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI *Inf. Fusion* **58** 82–115

[39] Landrum G 2016 RDKit: open-source cheminformatics software (available at: https://github.com/rdkit/rdkit/releases/tag/Release_2016_09_4)

[40] Kühne T D *et al* 2020 CP2K: an electronic structure and molecular dynamics software package—quickstep: efficient and accurate electronic structure calculations *J. Chem. Phys.* **152** 194103

[41] Jensen F 2005 Estimating the Hartree—Fock limit from finite basis set calculations *Theor. Chem. Acc.* **113** 267–73

[42] Rupp M, Tkatchenko A, Müller K-R and von Lilienfeld O A 2012 Fast and accurate modeling of molecular atomization energies with machine learning *Phys. Rev. Lett.* **108** 058301

[43] Huang B and von Lilienfeld O A 2020 Quantum machine learning using atom-in-molecule-based fragments selected on the fly *Nat. Chem.* **12** 945–51

[44] Hansen K, Biegler F, Ramakrishnan R, Pronobis W, von Lilienfeld O A, Müller K-R and Tkatchenko A 2015 Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space *J. Phys. Chem. Lett.* **6** 2326–31

[45] Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30

[46] Christensen A S, Faber F A, Huang B, Bratholm L A, Tkatchenko A, Müller K-R and Lilienfeld O A 2017 QML: a Python toolkit for quantum machine learning (https://doi.org/10.5281/zenodo.817332)

[47] Stuke A, Todorović M, Rupp M, Kunkel C, Ghosh K, Himanen L and Rinke P 2019 Chemical diversity in molecular orbital energy predictions with kernel ridge regression *J. Chem. Phys.* **150** 204121

[48] Schütt K T, Sauceda H E, Kindermans P-J, Tkatchenko A and Müller K-R 2018 SchNet—a deep learning architecture for molecules and materials *J. Chem. Phys.* **148** 241722

[49] Klicpera J, Groß J and Günnemann S 2020 Directional message passing for molecular graphs *8th Int. Conf. on Learning Representations (ICLR)* (available at: https://arxiv.org/abs/2003.03123)

[50] Reiser P, Eberhard A and Friederich P 2021 Graph neural networks in tensorflow-keras with raggedtensor representation (Kgcnn) *Softw. Impacts* **9** 100095

[51] Marom N 2017 Accurate description of the electronic structure of organic semiconductors by GW methods *J. Phys.: Condens. Matter* **29** 103003

[52] Yang C and Zhang Y 2022 Delta machine learning to improve scoring-ranking-screening performances of protein–ligand scoring functions *J. Chem. Inf. Model.* **62** 2696–712

[53] Fink T and Reymond J-L 2007 Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery *J. Chem. Inf. Model.* **47** 342–53

[54] Blum L C and Reymond J-L 2009 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13 *J. Am. Chem. Soc.* **131** 8732–3

[55] Frisch M J *et al* 2009 Gaussian 09, revision D.01 (Gaussian, Inc.)

[56] Balasubramani S G *et al* 2020 TURBOMOLE: modular program suite for ab initio quantum-chemical and condensed-matter simulations *J. Chem. Phys.* **152** 184107

[57] Ahlrichs R, Bär M, Häser M, Horn H and Kölmel C 1989 Electronic structure calculations on workstation computers: the program system turbomole *Chem. Phys. Lett.* **162** 165–9

[58] Behler J 2021 Four generations of high-dimensional neural network potentials *Chem. Rev.* **121** 10037–72

[59] Smith J S, Nebgen B, Lubbers N, Isayev O and Roitberg A E 2018 less is more: sampling chemical space with active learning *J. Chem. Phys.* **148** 241733