**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

A Machine Learning Approach for MP2 Correlation Energies and Its Application to Organic Compounds

Han, Ruocheng ; Rodríguez-Mayorga, Mauricio ; Luber, Sandra

Abstract: A proper treatment of electron correlation effects is indispensable for accurate simulation of compounds. Various post-Hartree–Fock methods have been adopted to calculate correlation energies of chemical systems, but time complexity usually prevents their usage in a large scale. Here, we propose a density functional approximation, based on machine learning using neural networks, which can be readily employed to produce results comparable to second-order Møller–Plesset perturbation (MP2) ones for organic compounds with reduced computational cost. Various systems have been tested and the transferability across basis sets, structures, and nuclear configurations has been evaluated. Only a small number of molecules at the equilibrium structure has been needed for the training, and generally less than 5% relative error has been achieved for structures outside the training domain and systems containing about 140 atoms. In addition, this approach has been applied to make predictions on correlation energies of nuclear configurations extracted from density functional theory-based molecular dynamics trajectories with only one or two structures as training data.

DOI: <https://doi.org/10.1021/acs.jctc.0c00898>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-207266>

Journal Article

Accepted Version

Originally published at:

Han, Ruocheng; Rodríguez-Mayorga, Mauricio; Luber, Sandra (2021). A Machine Learning Approach for MP2 Correlation Energies and Its Application to Organic Compounds. *Journal of Chemical Theory and Computation*, 17(2):777-790.

DOI: <https://doi.org/10.1021/acs.jctc.0c00898>

A machine learning approach for MP2 correlation energies and its application to organic compounds

R. Han, M. Rodríguez-Mayorga, and S. Lubert*

Department of Chemistry A, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

E-mail: sandra.luber@chem.uzh.ch

Abstract

A proper treatment of electron correlation effects is indispensable for accurate simulation of compounds. Various post-Hartree–Fock methods have been adopted to calculate correlation energies of chemical systems, but time complexity usually prevents their usage in large scale. Here, we propose a density functional approximation, based on machine learning using neural networks, which can be readily employed to produce results comparable to 2nd order Møller–Plesset perturbation (MP2) ones for organic compounds with reduced computational cost. Various systems have been tested and the transferability across basis sets, structures, and nuclear configurations has been evaluated. Only a small number of molecules at the equilibrium structure has been needed for the training, and generally less than 5 % relative error has been achieved for structures outside the training domain and systems containing about 140 atoms. Besides, this approach has been applied to make predictions on correlation energies of nuclear configurations extracted from density functional theory-based molecular dynamics trajectories with only one or two structures as training data.

Introduction

The electron correlation energy $E_C = E^{exact} - E^{HF}$, where E^{exact} is the exact electronic energy and E^{HF} the Hartree–Fock electronic energy, has been a central topic in the field of quantum chemistry.^{1–16}

Wavefunction-based methods like Møller–Plesset perturbation¹⁷ up to n th order (MP n), coupled cluster¹⁸ (CC), and configuration interaction¹⁹ (CI) were developed to calculate the correlation energies on top of Hartree–Fock (HF) calculations in an *ab initio* way. However, the usage of these methods is usually limited by the high computational cost, that for example for CCSD(T) may scale up to $\mathcal{O}(O^3V^4)$ (O is the number of occupied orbitals and V is the number of virtual orbitals). Density functional theory-based (DFT-based) methods,^{20,21} on the other hand, account for the correlation energy as a functional of electron density that scales computationally usually only as $\mathcal{O}(N^3)$ (N is the number of orbitals). Multiple types of exchange–correlation functionals have been developed with different levels of strategies: local density approximation^{22,23} (LDA) which only depends on the electron density, generalized gradient approximation^{24–26} (GGA) which depends on both electron density and its gradient, meta-GGA^{27,28} (mGGA) which depends on electron density, its gradient and kinetic energy density, and hybrid functionals^{29,30} in which a fixed fraction of HF exchange is included. Recently, machine learning approaches have been applied to various aspects for the development of computer hardware and speed-up of algorithms. A number of atom-based machine learning methods^{31–82} have been reported to reproduce energies/forces at the DFT or CCSD(T) level using atom type, atom positions, and/or their transformations as descriptors. These methods, despite their fast prediction, have limitations in the transferability among different chemical environments. Studies of machine-learned exchange–correlation energies/potentials^{83–92} are emerging as a powerful approach to provide chemical properties at the post-HF level universally for a large range of systems while keeping the computational cost at the DFT level. Molecular orbital/basis function based models^{83–88} have been focused on designing descriptors from the density matrix,⁸³ Fourier basis,⁸⁴ atomic-centered basis,⁸⁵

or selecting feature vectors from Fock matrix, Coulomb matrix, or exchange matrix.⁸⁶⁻⁸⁸ Grid-based models⁸⁹⁻⁹² have been targeted on the construction of descriptors directly from the density grid, including density kernels,⁸⁹ density derivatives,^{90,91} and projected density grid.⁹² In the spirit of the work by T. Nuddejima *et al.*,⁹¹ this study has been focused on proposing functionals for the electron correlation at the post-HF level.

While it is not possible to absolutely separate dynamic and nondynamic electron correlation, both parts have been investigated and analyzed over the past years.⁹⁻¹⁵ Dynamic electron correlation corresponds to the correction of the HF method arising from instantaneous interaction between electrons. It plays an important role in the electron correlation of (near-)equilibrium structures for many compounds (except for e.g. transition metal systems in which electronic configurations have similar energy and static correlation thus dominates) because the HF frontier orbitals of these structures are often far from being electronically degenerate and a multiconfigurational treatment is not necessary. Based on the fact that dynamic electron correlation effects introduce only small changes on the electronic density *w.r.t* to the HF one, its effects on the energy can be estimated locally (local-energy) in the real space as a density functional. Besides, each grid point required for the integration of the electronic energy in DFT can be used as a data point in the machine learning (ML) scheme so we do not have the problem of data scarcity and even training on a single structure can give reasonable ML models. MP2 as a widely used electron correlation method usually recovers most of the dynamic electron correlation and we have chosen this as the target and the reference of our ML models in this work. The MP2 local-energy used was first proposed by Y. Imamura and H. Nakai⁹³ and the idea behind is a grid-based one-body energy decomposition scheme introduced earlier.^{94,95} Previous work^{91,96} has also applied this scheme for fitting of density functionals.

We introduce here an orbital energy-weighted electron density as the descriptor in addition to those often taken from mGGA functionals. Multiple neural networks (NNs) have been constructed to reproduce the correlation energy calculated at MP2 level with different basis

sets. A variety of molecules and some DFT-based molecular dynamics (MD) trajectories have been tested for the transferability across nuclear structures and configurations. Moreover, our NN functionals can also be used in the DFT context to reduce the computational cost of double-hybrid functionals (like B2PLYP).

Theory

The goal of this study is to obtain the correlation energy at the MP2 level (E_C^{MP2}) defined in Eq. 1 ($\langle ij|ab \rangle$ is the two-electron integral expressed in Eq. 2 and e is molecular orbital (MO) energy), which can be expressed as the integration of the correlation energy density (ε_C^{MP2}) in the real space (see Eq. 3). In order to estimate the value of ε_C^{MP2} on each grid, the functional dependence of the electron density (ρ), the Euclidean norm of the density gradient ($|\nabla\rho|$), the kinetic energy density (τ), and the orbital energy-weighted electron density (ρ') are calculated (see Eq. 3). The definitions of ρ , $|\nabla\rho|$, and τ are known from DFT approximations (see Eqs. 4 – 6). Note that all formula used in this work are in spin-restricted form and the approach proposed has been only used in the prediction of correlation energy of closed-shell MP2 calculations. In order to resemble the density functional approximation, we considered electron density contribution from all occupied orbitals, i.e. no frozen core approximation is applied.

$$E_C^{MP2} = E^{MP2} - E^{HF} = \sum_{ij}^{occ} \sum_{ab}^{vir} \frac{2 \langle ij|ab \rangle \langle ab|ij \rangle - \langle ij|ba \rangle \langle ab|ij \rangle}{e_i + e_j - e_a - e_b} \quad (1)$$

$$\langle ij|ab \rangle = \int \varphi_i^*(\mathbf{r}_1) \varphi_j^*(\mathbf{r}_2) \frac{1}{r_{12}} \varphi_a(\mathbf{r}_1) \varphi_b(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (2)$$

$$E_C^{MP2} = \int \varepsilon_C^{MP2}(\mathbf{r}) d\mathbf{r} \simeq \int f(\rho(\mathbf{r}), \rho'(\mathbf{r}), |\nabla\rho(\mathbf{r})|, \tau(\mathbf{r})) d\mathbf{r} \quad (3)$$

$$\rho(\mathbf{r}) = 2 \sum_i^{occ} \varphi_i^*(\mathbf{r}) \varphi_i(\mathbf{r}) \quad (4)$$

$$|\nabla \rho(\mathbf{r})| = 2 \left| \sum_i^{occ} \nabla \varphi_i^*(\mathbf{r}) \varphi_i(\mathbf{r}) + \varphi_i^*(\mathbf{r}) \nabla \varphi_i(\mathbf{r}) \right| \quad (5)$$

$$\tau(\mathbf{r}) = \sum_i^{occ} \nabla \varphi_i^*(\mathbf{r}) \cdot \nabla \varphi_i(\mathbf{r}) \quad (6)$$

To define ρ' , we introduce an orbital energy-weighted coefficient k_i (see Eq. 7) which depends on the energy of occupied MO i (e_i) and energies of all virtual MOs a (e_a). $\frac{1}{N_k}$ is the normalization factor of k_i which can be calculated as in Eq. 8 (N_e is the total number of electrons). Then the orbital energy-weighted electron density ρ' can be defined as in Eq. 9.

$$k_i = \frac{1}{N_k} \sum_a^{vir} \frac{1}{e_a - e_i} \quad (7)$$

$$N_k = \frac{2}{N_e} \sum_i^{occ} \sum_a^{vir} \frac{1}{e_a - e_i} \quad (8)$$

$$\rho'(\mathbf{r}) = 2 \sum_i^{occ} k_i \varphi_i^*(\mathbf{r}) \varphi_i(\mathbf{r}) \quad (9)$$

ρ' is designed under the following considerations:

Implication of transition probability: From the denominator in the equation of the MP2 energy correction (see right hand side of Eq. 1), we notice that the correlation energy depends on the MO energy differences, and we would like to transfer this effect to the designed property ρ' . Occupied MOs should be weighted in a way so that lower orbital energy leads to less transition probability to virtual MOs and thus contributes less to ρ' , while higher MO energy has a higher transition probability to virtual MOs and thus contributes more to ρ' . This is done by including a coefficient k_i for each occupied MO in Eq. 9.

Scale invariance: Orbital energy-weighted electron density should be invariant under different energy units. This guarantees the uniqueness of ρ' . In order to fulfill this criteria, the coefficient k_i has to be unique under different energy units. We note, however, this needs to be carefully constructed and not all dimensionless quantities can satisfy the request.

Be normalized: The integrated orbital energy-weighted electron density should be the number of electrons. ρ' only changes the distribution of electron density in the real space but not the total number of electrons, so we have to include a normalization factor $\frac{1}{N_k}$.

For a more intuitive view of ρ' , we provide isosurface maps which imply the difference between ρ and ρ' of C_2H_4 and NH_2OH in Fig. 1 (red wireframe represents the region where $\rho > \rho'$ and blue wireframe represents the region where $\rho' > \rho$). As one may see, ρ' can offer a more diffused feature compared to ρ .

Figure 1: Isosurface which imply the difference between ρ and ρ' of (a) C_2H_4 and (b) NH_2OH . Red wireframe represents the region where $\rho > \rho'$ and blue wireframe represents the region where $\rho' > \rho$. Isovalue is set to 0.005 a.u. in both cases.

The reference MP2 local-energy ε_C^{MP2} is calculated by Eqs. 10 and 11 given two-electron integral $\langle ij|ab \rangle$, MO energy e , MO to primitive coefficient C , and primitive χ . The quantities mentioned above can be extracted from the FCIDUMP⁹⁷ file (produced by e.g. *OpenMolcas*,⁹⁸ *Psi4*⁹⁹) and wavefunction file (e.g. WFN or MOLDEN file). $\mathbf{r}_g$ represents the grid point on which the correlation energy density is calculated (see the Appendix for more details about the integration of Eq. 11).

$$\varepsilon_C^{MP2}(\mathbf{r}_g) = \sum_{ij}^{occ} \sum_{ab}^{vir} \frac{2\langle ij|ab\rangle - \langle ij|ba\rangle}{e_i + e_j - e_a - e_b} \varphi_i^*(\mathbf{r}_g) \varphi_a(\mathbf{r}_g) A_{jb}(\mathbf{r}_g) \quad (10)$$

$$A_{jb}(\mathbf{r}_g) = \sum_{\mu\nu} C_{\mu j}^* C_{\nu b} \int \frac{\chi_\mu^*(\mathbf{r}) \chi_\nu(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_g|} d\mathbf{r} \quad (11)$$

For the integration of $\varepsilon_C(\mathbf{r}_g)$ in the whole space, Legendre quadrature¹⁰⁰ and Lebedev quadrature¹⁰¹ are applied. 70000 grid points (200 radial and 350 angular) per atom are used for integration in Eq. 3. The scheme in Ref.¹⁰² is implemented for the integration of polyatomic systems in which the grids are sampled around multiple atomic centers.

Methods

The general idea is to use the electron density and density-related variables (as described in Eqs. 4–6 and 9) at each grid point to predict correlation energy density at the same grid point (namely, design of a local functional of functions of electron density and density-related variables). Our training data is based on the descriptors $\{\rho, \rho', |\nabla\rho|, \tau\}$ and they are evaluated on every grid point $\mathbf{r}_g$ set in the real space. They are calculated by Eqs. 4 to 6 and 9 given the wavefunction file from a HF calculation. Our training label is $\{\varepsilon_C^{MP2}\}$, which is also evaluated on the same grid points calculated by Eq. 10 given both wavefunction and FCIDUMP files from a MP2 calculation. All $\mathbf{r}_g$ points are selected by the quadratures. We note that the preparation of the descriptors determines the scaling of this ML method, with the time complexity of $\mathcal{O}(N_{grid}N)$ (N_{grid} is the number of density grid points) and the space complexity (e.g. memory usage/disk storage) of $\mathcal{O}(N^2)$.

For preprocessing, we take the logarithm of both training data and training label so the input of the ML model is $\{\log_{10}(\rho), \log_{10}(\rho'), \log_{10}(|\nabla\rho|), \log_{10}(\tau)\}$ and output of the ML model is $\{\log_{10}(-\varepsilon_C^{MP2})\}$.

Four NN models (NN1 $\sim$ NN4) have been designed for the machine learning (see Fig.

2). All NN models have a main structure of two dense layers which include 48 neurons and 4 neurons (48/4), respectively. NN1 and NN2 are densenet type NNs with only sequential structures and the difference is the use of ρ (NN1) or ρ' (NN2). NN3 and NN4 aim for the consideration of both ρ and ρ' by adding a branch structure. NN3 uses a simple dense layer for mixing ρ and ρ' while NN4 introduces a parameter $\lambda \in (0, 1)$ for linear combination of ρ and ρ' as ρ_{cmb} (see Eq. 12). We would like to have the parameter λ decided by the NN without further knowledge, so it is directly assumed as a function of ρ_{cmb} estimated by a hidden layer with 2 neurons (f in Eq. 13). To optimize λ , Eqs. 12 and 13 are calculated recursively four times before $\rho_{cmb}^{(4)}$ is fed into the next NN layer. In practice, we train each NN with 10 initial guesses of weights by setting different random seeds and the predicted result is averaged over values given by the NN with these 10 different parameter sets. Hyperparameters in NN models, namely, the number of neurons in the dense layers, are chosen based on the idea to keep the NN as simple as possible in order to avoid overfitting. We have tested several combinations of neuron numbers and there is no or not much improvement above 48/4. The comparison of NNs with 200/200 is presented in the last section.

$$\rho_{cmb}^{(l)} = (1 - \lambda^{(l)})\rho^{(l)} + \lambda^{(l)}\rho'^{(l)} \quad (12)$$

$$\lambda^{(l)} = f(\rho_{cmb}^{(l-1)}) \quad (13)$$

Initially, 13 molecules shown in Table 1 (*abbr.* **13mol**) are chosen as the training data, considering they cover common bond types between C , H , N , and O atoms. Totally 2590000 $\mathbf{r_g}$ points are used for training: 37 atom types with different chemical environment and 70000 $\mathbf{r_g}$ points per atom. Calculations with three different basis sets, cc-pVDZ, cc-pVTZ, and cc-pVQZ,¹⁰³ are used for the training, respectively. Note that it is also possible to train only on a selected subset of the 13 molecules (*abbr.* **sel**) based on the bond type(s) present in the molecular system of interest, e.g. training on C_2H_6 and C_2H_4 while testing on benzene.

Figure 2: Four NN models are used for the machine learning of correlation energy density ε_C^{MP2} (ε shown in the figure). Note that the input and output shown in the figures are after logarithmic transformation mentioned in Data preparation section. (a) NN1: electron density (ρ), Euclidean norm of density gradient ($|\nabla\rho|$), and kinetic energy density (τ) are three neurons in the input layer, followed by 48 neurons as the first hidden layer and 4 neurons as the second hidden layer. (b) NN2: similar to NN1 but use orbital energy-weighted electron density (ρ') rather than ρ as neuron in the input layer. (c) NN3: ρ and ρ' are first "mixed" by a hidden layer with 4 neurons, then the resulting neuron is combined with $|\nabla\rho|$ and τ as the input layer of the same NN structure as NN1 and NN2. (d) NN4: similar to NN3 but ρ and ρ' are "mixed" by a parameter λ , which is calculated recursively for 4 times as the function of the resulting neuron by a hidden layer with 2 neurons. All NN layers are followed by a *ReLu* activation function except for the one before λ which is a *Sigmoid* activation function.

As a by-product from the preparation of the training data, we may define an "atomic contribution" (AC) of E_C^{MP2} by integrating ε_C^{MP2} for each atom (see Eq. 14, where Ω_{atom} is the atomic domain determined by the scheme in Ref.¹⁰²). We can also use AC to approximate E_C^{MP2} for the test systems by simply summing up $E_C^{AC}(atom)$ (obtained from training systems) of all atoms in the molecule. In this work, only four AC values (for the atom types C, H, N, and O, respectively, see Table S1) are used and the prediction based on AC only depends on these values and the stoichiometry of a given molecule. Each atom type is averaged over all of its chemical environments in **13mol**.

$$E_C^{AC}(atom) = \int_{\Omega_{atom}} \varepsilon_C^{MP2}(\mathbf{r}) d\mathbf{r} \quad (14)$$

Molecular structures in this work were optimized with HF and cc-pVDZ basis set¹⁰³ (STO-3G¹⁰⁴ for chignolin and β -cyclodextrin) using the *Gaussian09*¹⁰⁵ package. For the **13mol** training data, FCIDUMP and wavefunction files were generated using the *Psi4*⁹⁹ package. For HF, MP2, and DFT calculations of all testing cases, wavefunction files were generated using the *Gaussian09*¹⁰⁵ package. DFT-MD frames were extracted from NVT Born–Oppenheimer DFT-based MD simulations obtained at 300 K with Goedecker–Teter–Hutter norm-conserving pseudopotentials,^{106–108} the BLYP^{24,25} exchange–correlation functional, and the corresponding DZVP-GTH basis sets using the Nosé–Hoover thermostat^{109,110} and the *CP2K/QUICKSTEP*^{111–113} package. In total, a 10 ps simulation was performed with 0.5 fs timestep and 100 frames were extracted uniformly from the last 10,000 frames for the testing. The NNs used in this work were constructed and trained using the *Keras*¹¹⁴ and *TensorFlow*¹¹⁵ packages. The most time consuming part in the whole ML procedure is the preparation of ε_C^{MP2} , which is also implemented for GPU in our code. A general speed-up of ~ 2.5 times compared to a 16 cores CPU running at 2.0 GHz has been achieved using a RTX 2070 GPU, and parallelization can be performed across multiple GPUs and CPU-cores with almost ideal scaling. CUDA based packages scikit-cuda¹¹⁶ and PyCUDA¹¹⁷ are adopted in the GPU code.

Table 1: Bonding types of 13 molecules that are chosen for the training data.

Molecule	Bonding types
C_2H_6	C—C, C—H
C_2H_4	C=C, C—H
C_2H_2	C≡C, C—H
CH_3NH_2	C—N, C—H, N—H
CH_2NH	C=N, C—H, N—H
HCN	C≡N, C—H
CH_3OH	C—O, C—H, O—H
$HCHO$	C=O, C—H
N_2H_4	N—N, N—H
N_2H_2	N=N, N—H
NH_2OH	N—O, N—H, O—H
HNO	N=O, N—H
H_2O_2	O—O, O—H

Results and Discussion

In this section, we test the proposed approach with respect to different aspects, including the stability of the ML functional under various basis sets, and the transferability across different molecular systems and nuclear configurations.

Test on small molecules

As our first test, we employed the 32 small molecules selected in Ref.⁹¹ The relative error of the predicted E_C with respect to E_C^{MP2} is calculated for each molecule using different NNs or AC (see Table S1) as shown in Fig. 3. Comparing different NNs, the relative error of NN2, NN3, and NN4 is smaller than that of NN1, suggesting the important role of ρ' as a descriptor. From cc-pVDZ to cc-pVQZ, the relative error decreases, which may be related to the fact that MP2 calculations are rather sensitive to the size of basis set. We also notice that molecules containing one or more oxygen atoms (like $CH_2(OH)_2$) have a larger relative error than the other molecules in the test set. The error in kcal/mol can be found in Fig. S1.

We have tested the predictions from small basis set to large basis set (see Table 2). A $\rightarrow$

(a)

(b)

(c)

Figure 3: Relative error of the predicted E_C with respect to E_C^{MP2} using NNx (x=1-4) and AC. (a) cc-pVDZ, (b) cc-pVTZ, or (c) cc-pVQZ basis set are adopted in both training and prediction. The shading region indicates the relative error up to 5 %.

B represents the test case that the descriptors are extracted from the HF calculation with basis set A while the NN functional for the prediction is the one optimized with basis set B (see Eq. 16 for the definition, $f^{NNx/B}$ represents NNx functional optimized with basis set B, $\rho^{HF/A}$ represents electron density extracted from the HF calculation with basis set A, etc.). The prediction is also compared to E_C^{MP2} with basis set B ($E_C^{MP2/B}$) and the mean absolute percentage error (MAPE) is calculated based on Eq. 15 (*mol*: molecule, n_{mol} : number of molecules). Similar definitions can be found in Eqs. 17 and 18 for the AC. From Table 2, we can see that the predictions from small basis sets to large ones only introduce an extra 0.5 % of error at most, which shows the stability of the NN functionals among these basis sets. Taken all test cases into consideration, NN4 works the best among all NNs. One may notice that AC works even better, however, it should be pointed out that tested molecules in this subsection are small and have a very similar structure type as the molecules for the training. Hence, a more general test set is employed in the following section to compare these approaches.

$$MAPE(NNx/A \rightarrow B) = \frac{1}{n_{mol}} \sum_{mol} \left| \frac{E_C^{NNx/A \rightarrow B}(mol) - E_C^{MP2/B}(mol)}{E_C^{MP2/B}(mol)} \right|, \quad x = 1 - 4 \quad (15)$$

$$E_C^{NNx/A \rightarrow B}(mol) = \int f^{NNx/B}(\rho^{HF/A}(\mathbf{r}), \rho'^{HF/A}(\mathbf{r}), |\nabla \rho^{HF/A}(\mathbf{r})|, \tau^{HF/A}(\mathbf{r})) d\mathbf{r}, \quad x = 1 - 4 \quad (16)$$

$$MAPE(AC/B) = \frac{1}{n_{mol}} \sum_{mol} \left| \frac{E_C^{AC/B}(mol) - E_C^{MP2/B}(mol)}{E_C^{MP2/B}(mol)} \right| \quad (17)$$

$$E_C^{AC/B}(mol) = \sum_{atom \in mol} E_C^{AC/B}(atom) \quad (18)$$

Table 2: MAPE of the predicted E_C with respect to E_C^{MP2} with NNx (x=1-4) or AC on 32 small molecules. The definition of $A \rightarrow B$ is explained in Eq. 16. AC prediction does not depend on the basis set of the descriptors, so such test cases are omitted.

Test case $A \rightarrow B$	NN1	NN2	NN3	NN4	AC
cc-pVDZ $\rightarrow$ cc-pVDZ	2.9 %	2.4 %	2.4 %	2.3 %	2.3 %
cc-pVDZ $\rightarrow$ cc-pVTZ	2.8 %	2.7 %	2.4 %	2.3 %	–
cc-pVTZ $\rightarrow$ cc-pVTZ	2.6 %	2.2 %	2.4 %	2.1 %	1.7 %
cc-pVDZ $\rightarrow$ cc-pVQZ	2.1 %	1.9 %	2.1 %	2.1 %	–
cc-pVTZ $\rightarrow$ cc-pVQZ	2.0 %	1.9 %	1.8 %	1.7 %	–
cc-pVQZ $\rightarrow$ cc-pVQZ	2.0 %	1.7 %	1.7 %	1.6 %	1.5 %

Transferability across molecular systems

A group of molecules representating different types of structures (see Table 3) is selected to evaluate the transferability of the NN functionals. From Fig. 4, E_C of most molecules can be estimated within 5 % relative error using NN2, NN3, or NN4. All predicted values of C_{20} have very large relative errors, which may be explained by the absence of hydrogen atoms, thus being a more challenging system because its structure is far from the ones used in the training data. NN predicted E_C of glucose has the relative error between +5 % to +8 %, which could be again, as we have pointed out in the last subsection, due to the presence of oxygen atom(s). However, this problem can be avoided by training the NN on selected molecules (see next section). AC for these structures gives relatively large prediction errors, suggesting its limited transferability across structures. We would also like to point out here that NN2 and NN4 are able to recover the E_C difference between the isomers of glycine very well (see Table 4).

Two molecules, chignolin (140 atoms) and β -cyclodextrin (147 atoms), are tested to show the ability of predicting E_C at the MP2 level for larger systems. Training data of **13mol** or **sel** (C_2H_6 , C_2H_4 , CH_3NH_2 , CH_2NH , CH_3OH , and $HCHO$ for chignolin; C_2H_6 and CH_3OH for β -cyclodextrin) is used and the corresponding relative prediction errors are computed. Also, we provide the prediction from the STO-3G¹⁰⁴ and 3-21G¹¹⁸ basis sets to the cc-pVDZ basis set (see Eqs. 16 and 18 for definitions). The relative errors of predictions

Table 3: Structural features of the 14 chosen molecules.

Molecule	Structural features
Benzene	aromatic
Cyclohexane	cyclic
Cubane	cubic
1,3-Butadiene	conjugated
Adamantane	polycyclic; diamondoid
C_{20}	fullerenic
Salicylaldehyde	hydrogen bonded
Camphor	polycyclic; strained
Cortisone	steroidal
Glucose	carbohydratic
Retinol	terpenic
Porphyrin	heterocyclic; macrocyclic
Glycine	
Glycine zwitterion	neutral/zwitterionic

Figure 4: Relative error of the predicted E_C with respect to E_C^{MP2} using NNx (x=1-4) and AC on 14 molecules representing various structures. The cc-pVDZ basis set is adopted in both training and prediction. The shading region indicates the relative error up to 5 %.

Table 4: Dynamic correlation energy difference between glycine and glycine zwitterion $E_C(\text{glycine zwitterion}) - E_C(\text{glycine})$ using MP2, NNx (x=1-4), and AC. The cc-pVDZ basis set is adopted in both training and prediction. Unit: mH

MP2	NN1	NN2	NN3	NN4	AC
2.9	1.0	2.7	4.1	2.7	0.0

are shown in Fig. 5. Using **sel** as training data gives better predictions (with NN2^{sel} , NN3^{sel} , and NN4^{sel}) for β -cyclodextrin. This indicates that selecting specific molecules for training is a good choice if the molecule of interest contains only certain bond types. In these two tests, NN3^{sel} and NN4^{sel} can predict E_C within 2 % of relative error given the descriptors obtained with the cc-pVDZ basis set. Using descriptors from smaller basis sets generally overestimates E_C ; the relative errors are within 5 % using NNx^{sel} ($x=1-4$) with descriptors calculated with the 3-21G basis set. Relative errors using AC and AC^{sel} are around 5 % for these two molecules. Detailed information can be found in Tables S2-S5.

Figure 5: Relative error of the predicted E_C with respect to E_C^{MP2} using NNx ($x=1-4$) and AC with **13mol** and **sel** as training data (noted as NNx^{sel} and AC^{sel} in the figures). Two large molecules, (a) chignolin (**sel**: C_2H_6 , C_2H_4 , CH_3NH_2 , CH_2NH , CH_3OH , HCHO) and (b) β -cyclodextrin (**sel**: C_2H_6 , CH_3OH) are tested for the STO-3G $\rightarrow$ cc-pVDZ, 3-21G $\rightarrow$ cc-pVDZ, and cc-pVDZ $\rightarrow$ cc-pVDZ cases. The shading region indicates the relative error up to 5 %.

Transferability across different nuclear configurations

DFT-MD frames of 6 molecules are extracted and the single point E_C of each frame is predicted and compared with the reference value E_C^{MP2} . In the preparation of descriptors, we select the optimized geometry as the reference geometry, the wavefunction of which is used for the calculation of descriptors of all DFT-MD frames. This assumes that the wavefunctions

do not change much during the DFT-MD simulation. As a consequence, the variation of E_C only depends on the change of MO energy. In practice, we only need to modify ρ' following Eqs. 19 and 20, where $\frac{1}{N_{k,ref}}$ and $\varphi_{i,ref}$ represent the normalization factor and the MO of the reference geometry, respectively. This is less time consuming than using the approach in Eqs. 7–9 which requires the conversion of the wavefunction and the descriptor preparation for each geometry. The NN1 functional does not depend on MO energy so one cannot use it to predict E_C of DFT-MD frames in this way. Similar as before, NNx^{sel} represents the case where the training data are selected molecules. Since relative electronic energies rather than absolute ones during the DFT-MD simulation are of interest, a global shift is added to the predicted E_C when comparing to the reference value E_C^{MP2} . Detailed information can be found in Tables S6-S11.

$$k_i = \frac{1}{N_{k,ref}} \sum_a^{vir} \frac{1}{e_a - e_i} \quad (19)$$

$$\rho'(\mathbf{r}) = 2 \sum_i^{occ} k_i \varphi_{i,ref}^*(\mathbf{r}) \varphi_{i,ref}(\mathbf{r}) \quad (20)$$

As can be seen from Fig. 6, the mean absolute errors (MAEs) are mostly within 1 mH except for HCN which exceeds 2 mH in some cases (see Tables S6-S11 for details). Note that unlike previous studies,^{85,86,88} we do not need many training data points to achieve such accuracy. Only a general data set (**13mol**) with 13 structures or even less (**sel**, one or two structures) has been used to obtain these results. Here, however, **sel** does not necessarily improve the results compared to **13mol**. This might be ameliorated by choosing other molecule(s) for **sel**. It is also essential to check the correlation between the predicted values and the reference ones for the DFT-MD frames, which reveals how close the predictions follow the trend/variation of the exact values. We here provide the E_C^{NN4} and E_C^{MP2} of every frame of benzene and HCN in Figs. 7 and 8 as the "best" and "worst" tests (according to Fig. 6). In the case of benzene, it can be seen that E_C^{NN4} closely follows E_C^{MP2} and the

correlation of them is strong. In the case of HCN, the NN4 functional underestimates the amplitude of the variation though the trend matches. However, one can introduce a scaling factor on the amplitude once the exact value of an extra (or more) data point(s) are known and the results show better agreement with the reference values (as can be seen from the dashline in Fig. 8a). In this case, predictions are mostly located within the range of ± 2 mH (see Fig. 8b, MAE decreases from 1.78 mH to 1.13 mH).

Figure 6: MAE of the predicted E_C with respect to E_C^{MP2} of 100 DFT-MD frames of 6 molecules. The results from NNx and NNx^{sel} (x=2-4) are computed. **sel** of benzene is C₂H₆ and C₂H₄, and **sel** of butane and isobutane is C₂H₆. **sel** of C₂H₆, HCN, and CH₃OH are themselves. Basis set cc-pVDZ is adopted in both training and prediction. The shading region indicates the error up to 2 mH.

Direct application in the DFT context

Here, we present a potential application of our approach to mimic a double-hybrid functional, namely the B2PLYP exchange–correlation functional. NN-based training with **13mol** and the cc-pVDZ basis set are used to predict the correlation energy E_C at the B2PLYP^{22–25,119}/cc-pVDZ level. For this test, all descriptors are calculated from the self-consistent density of B2PLYP without the MP2 contribution (noted as B2PLYP^{DFT}, see Eq. 21 for the for-

Figure 7: Benzene: (a) Variations of E_C predicted with NN4 and the reference E_C^{MP2} along 100 DFT-MD frames; (b) E_C^{NN4} vs. E_C^{MP2} , the shading region indicates the difference up to 2 mH. The cc-pVDZ basis set is adopted in both training and prediction.

Figure 8: HCN: (a) Variations of E_C predicted with NN4 w/, w/o scaling and the reference E_C^{MP2} along 100 DFT-MD frames; (b) E_C^{NN4} w/, w/o scaling vs. E_C^{MP2} . The shading region indicates the difference up to 2 mH. The cc-pVDZ basis set is adopted in both training and prediction. A simple scaling factor of 2.0 is used.

mula of the exchange–correlation energy). Our NN functionals predict E_C^{NNx} on the top of converged B2PLYP^{DFT} and finally give the expression of the predicted correlation energy $E_C^{B2PLYP^{DFT}+NNx}$ (see Eq. 22).

$$E_{XC}^{B2PLYP^{DFT}} = 0.47E_X^{B88} + 0.53E_X^{HF} + 0.73E_C^{LYP} \quad (21)$$

$$E_C^{B2PLYP^{DFT}+NNx} = E_C^{B2PLYP^{DFT}} + 0.27E_C^{NNx}, \quad x = 1 - 4 \quad (22)$$

The results are shown as the relative error with respect to the reference value calculated at B2PLYP level (E_C^{B2PLYP}) in Fig. 9. We have tested on 32 molecules and all errors of the predicted E_C using NNx (x=1-4) are within 5 %. E_C calculated at MP2, GGA (BLYP), and hybrid GGA (B3LYP) level are also listed for comparison. It is worth to mention that we use the data trained on HF calculations for the prediction on DFT ones (B2PLYP^{DFT}, using KS orbital with KS orbital energy) ones here. In principle, one can also train the same systems on DFT calculations, which, however, goes beyond the scope of this work.

General discussion of neural networks

Similar to previous work,⁹¹ the neural network introduced can also be adopted for CC calculations. We have trained NN1 and NN4 on CCSD correlation energy density ε_C^{CCSD} (Eq. 3 in Ref.⁹¹) of **13mol** and tested them on the same 32 small molecules as used in the "Test on small molecules" section. The result is shown in Fig. 10.

To show the features of NNs designed in the present work, we have compared NN4 with the commonly used dense NN. In total, four neural networks are used for the comparison: NN1, NN4, 200/200 which uses 2 layers with 200 neurons per layer (same as the one used in Ref.⁹¹), and NN4+200/200 which has the same dense layers as 200/200 (instead of 48/4). For the test set containing 32 small molecules, MAPE of predicted E_C is shown in Table 5, and the error of reaction energies for 12 reactions studied in Ref.⁹¹ is shown in Table 6 (ΔE_C

Figure 9: Relative error of the predicted E_C with respect to E_C^{B2PLYP} using NNx (x=1-4). Descriptors are from the wavefunction of B2PLYP^{DFT} (B2PLYP without the perturbation step) calculation. For comparison, E_C of the same system calculated at MP2, BLYP, and B3LYP level is also presented. The cc-pVDZ basis set is adopted in both training and prediction. The shading region indicates the relative error up to 5 %.

Figure 10: Relative error of the predicted E_C with respect to E_C^{CCSD} using NN1 and NN4. The cc-pVDZ basis set is used for both training and prediction. The shading region indicates the relative error up to 5 %. MAPE is 2.8 % and 2.5 % for NN1 and NN4, respectively.

represents the difference between correlation energies of reactants and products). It is found that more neurons in each layer increase the accuracy of the dense NN from the comparison between 200/200 and NN1. Nevertheless, more neurons in the dense layer of NN4 do not significantly increase the prediction accuracy. In general, more neurons in NN will increase the time for both training and prediction so we would like to keep NN as simple as possible.

Table 5: MAPE of the predicted E_C with respect to E_C^{MP2} with four NNs on 32 small molecules. The cc-pVDZ basis set is used for both training and prediction.

200/200	NN1	NN4	NN4+200/200
2.5 %	2.9 %	2.3 %	2.2 %

Table 6: MAE and maximum absolute error (MaxAE) of the predicted ΔE_C with respect to ΔE_C^{MP2} with four NNs on 12 reactions in Ref.⁹¹ The cc-pVDZ basis set is used for both training and prediction. Unit: kcal/mol

	200/200	NN1	NN4	NN4+200/200
MAE	5.48	8.01	4.24	3.97
MaxAE	12.79	16.41	8.23	10.62

For the test set consisting of 14 molecules given in Table 3, MAPE of predicted E_C is shown in Table 7. In this case, more neurons in NNs (e.g. 200/200 and NN4+200/200) cause larger error due to overfitting. The relative error of each molecule can be found in Fig. S2.

Table 7: MAPE of the predicted E_C with respect to E_C^{MP2} with four NNs on 14 molecules in Table 3. The cc-pVDZ basis set is used for both training and prediction.

200/200	NN1	NN4	NN4+200/200
5.0 %	3.9 %	2.7 %	4.4 %

The same MD frames as used for Fig. 6 are also tested using NN4+200/200, and Table 8 shows the comparison between NN4 and NN4+200/200. The two NNs have similar accuracy for most of the systems. The large error for benzene using NN4+200/200 might be due to overfitting.

Table 8: MAE of the predicted E_C with respect to E_C^{MP2} of 100 DFT-MD frames of 6 molecules using NN4 and NN4+200/200. The cc-pVDZ basis set is used for both training and prediction. Unit: mH.

	NN4	NN4+200/200
C_2H_6	0.48	0.60
Benzene	0.29	1.95
HCN	1.78	2.12
CH_3OH	0.88	1.09
Butane	0.77	0.70
Isobutane	0.78	0.75

Dependence of prediction accuracy on the number of training data is also tested on 32 small molecules using NN1 and NN4. We have randomly chosen a certain number of grid points from the 70000 integration grid points standardly used in our calculations (see section Method) as training data and MAPE in each case is calculated (see Fig. 11). Densenet type NN1 requires 50000 grid points per atom to achieve 3.0 % MAPE, while NN4 only needs about 1/10 of the training data to obtain the same accuracy.

Figure 11: MAPE of the predicted E_C with respect to E_C^{MP2} using NN1 and NN4 with different number of training data (i.e. grid points). 32 small molecules in Fig. 3 is used as test set. The cc-pVDZ basis set is used for both training and prediction.

Conclusion

We have introduced a grid-based ML approach for the prediction of electron correlation. The descriptors consist of electron density (ρ), the Euclidean norm of the density gradient ($|\nabla\rho|$), the kinetic energy density (τ), and the orbital energy-weighted electron density (ρ'). ρ' , which represents an orbital energy-weighted electron density, was first constructed in this work and was found capable of decreasing the prediction error (e.g. in NN2, NN3, and NN4) compared to only mGGA type descriptors (NN1). Four NNs have been designed considering different amount of ρ and ρ' involved. Various molecular systems have been tested and transferability across basis sets, structures and nuclear configurations has been evaluated using only a few systems (at most 13 molecules) as training data. Most systems, including those with the types of structures outside the training domain and molecules of more than 100 atoms, have the relative prediction error within 5 %. For small molecules, the MAPE can achieve ~ 2 % even in calculations using the cc-pVDZ basis set to predict E_C obtained with the cc-pVQZ basis set. Different nuclear configurations extracted from DFT-MD trajectories of several small molecules have been tested, and the prediction closely follows the variance of E_C at MP2 level. This, however, only requires one or a few structures as training data, those do not even have to be the system itself. It is worth to note that such an approach for DFT-MD frames only works under the assumption that the electron density does not change much across these nuclear configurations. We have also provided an application of using this method to predict the perturbation energy given by B2PLYP, and the relative prediction errors of E_C are all within 5 %. Nevertheless, there might be cases that one may not obtain accurate results, e.g. if targeted systems are far from the training domain (in the aspect of e.g. element type, structures). Though we applied the same idea of correlation energy density as in previous work,⁹¹ the focus of our work is different. In previous work,⁹¹ it was trained against CCSD(T) correlation energies and aimed at a precise prediction of exchange–correlation energy in a relatively small dataset while we train towards the MP2 correlation energy and focus on the prediction of correlation energy with transferability in

chemical and configurational space. The new design of the neural network with orbital energy-weighted electron density is essential for such transferability.

The proposed ML approach is straightforward and efficient. The information required for the prediction are the wavefunction and orbital energies, which are directly available in WFN or MOLDEN file generated from a HF calculation. Compared to MP2 calculation, the ML prediction decreases the computational cost from $\mathcal{O}(N^5)$ to $\mathcal{O}(N_{grid}N)$ and the memory usage/disk storage from $\mathcal{O}(N^4)$ (for two electron integrals) to $\mathcal{O}(N^2)$. Also, only another FCIDUMP file is required if one would like to train on other systems. Though the preparation of ε_C^{MP2} is time consuming, we have developed a code that can be efficiently parallelized on both GPUs and CPUs. Due to the transferability of this approach, one can save time by training on selected small molecules (e.g. similar elements and structures) with relatively small basis sets while predicting the correlation energy of larger molecules at the quality of larger basis sets, which is often hard to accomplish with MP2 at a reasonable time scale. Moreover, it is possible to apply our ML approach on the top of DFT calculations and evaluate the correlation energy which can be used in double-hybrid functionals.

Overall, our ML approach provides an efficient manner to predict electron correlation energies comparable to MP2 at the cost of standard DFT calculations. We have shown the transferability accross chemical and configurational space for organic molecules. Additionally, our ML functionals can be readily used in the DFT context to reduce the cost of the evaluation of double-hybrid exchange-correlation functionals.

Acknowledgment

Funding by the University of Zurich and the Swiss National Science Foundation (grant no: PP00P2_170667) is gratefully acknowledged. We thank the Swiss National Supercomputing Center for computing resources (project ID: s745, s788, and s1001).

Supporting Information Available

Tables in the supporting information provide the AC values used for each atom (Table S1), detailed data for chingnolin and β -cyclodextrin shown in Fig. 5 (Tables S2-S5), and shifted value and MAE of the predicted correlation energy of DFT-MD frames (Tables S6-S11). Figures in the supporting information provide error of the predicted correlation energy of 32 small molecules (Figure S1) and comparison of the prediction error using more neurons in the dense layer of NNs (Figure S2). File coord_all.txt contains Cartesian coordinates of all optimized geometries used in this work.

Appendix

In order to calculate the integration in Eq. 11, we use spherical coordinates for $\frac{1}{|\mathbf{r}-\mathbf{r}_g|}$ and carry out the integration on dr and $d\theta d\phi$ with Legendre quadrature¹⁰⁰ and Lebedev quadrature,¹⁰¹ respectively (see Eq. 23). Note that the integration can be done at arbitrary direction, so we manually set $\theta = 0$ for a given $\mathbf{r}_g$ grid point to simplify the calculation. However, the quadratures above require that the integration part has only one center, but the overlap of primitives $\chi_\mu^*(\mathbf{r})\chi_\nu(\mathbf{r})$ can have two centers under the Gaussian basis (see Eq. 24 in cartesian coordinates, $\mathbf{r} = (x, y, z)$, $\mathbf{r}_a = (x_a, y_a, z_a)$ $\mathbf{r}_b = (x_b, y_b, z_b)$). To circumvent this, we follow the transformation in Ref.¹²⁰ and construct Eq. 25 in the form of a one center Gaussian at $\mathbf{r}_{ab}$. All related variables are defined in Eqs. 26 to 29.

$$\int \frac{\chi_\mu^*(\mathbf{r})\chi_\nu(\mathbf{r})}{|\mathbf{r}-\mathbf{r}_g|}d\mathbf{r} = \int \frac{\chi_\mu^*(\mathbf{r})\chi_\nu(\mathbf{r})}{|\mathbf{r}(r, \theta, \phi) - \mathbf{r}_g(r, 0, \phi)|}drd\theta d\phi \quad (23)$$

$$\begin{aligned} \chi_\mu^*(\mathbf{r})\chi_\nu(\mathbf{r}) = & (x - x_a)^{n_\mu}(y - y_a)^{l_\mu}(z - z_a)^{m_\mu} \exp[-\zeta_\mu(\mathbf{r} - \mathbf{r}_a)^2] \\ & (x - x_b)^{n_\nu}(y - y_b)^{l_\nu}(z - z_b)^{m_\nu} \exp[-\zeta_\nu(\mathbf{r} - \mathbf{r}_b)^2] \end{aligned} \quad (24)$$

$$\begin{aligned} \chi_\mu^*(\mathbf{r})\chi_\nu(\mathbf{r}) = & \exp[-\xi_{\mu\nu}(\mathbf{r}_a - \mathbf{r}_b)^2] \{ (x - x'_a)^{n_\mu} (y - y'_a)^{l_\mu} (z - z'_a)^{m_\mu} \\ & (x - x'_b)^{n_\nu} (y - y'_b)^{l_\nu} (z - z'_b)^{m_\nu} \exp[-\zeta_{\mu\nu}(\mathbf{r} - \mathbf{r}_{ab})^2] \} \end{aligned} \quad (25)$$

$$\zeta_{\mu\nu} = \zeta_\mu + \zeta_\nu \quad (26)$$

$$\xi_{\mu\nu} = \frac{\zeta_\mu \zeta_\nu}{\zeta_{\mu\nu}} \quad (27)$$

$$\mathbf{r}_{ab} = \frac{\zeta_\mu \mathbf{r}_a + \zeta_\nu \mathbf{r}_b}{\zeta_{\mu\nu}} \quad (28)$$

$$\mathbf{r}'_p = \mathbf{r}_p + \frac{1}{2}\mathbf{r}_{ab}, \quad \mathbf{r}'_p = (x'_p, y'_p, z'_p), \quad p \in \{a, b\} \quad (29)$$

References

- (1) Löwdin, P.-O. Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction. *Phys. Rev.* **1955**, *97*, 1474–1489.
- (2) Cioslowski, J. Density-driven self-consistent-field method: Density-constrained correlation energies in the helium series. *Phys. Rev. A* **1991**, *43*, 1223–1228.
- (3) Gottlieb, A. D.; Mauser, N. J. New Measure of Electron Correlation. **2005**, *95*.
- (4) Raeber, A.; Mazziotti, D. A. Large eigenvalue of the cumulant part of the two-electron reduced density matrix as a measure of off-diagonal long-range order. *Phys. Rev. A* **2015**, *92*.

- (5) Benavides-Riveros, C. L.; Lathiotakis, N. N.; Schilling, C.; Marques, M. A. L. Relating correlation measures: The importance of the energy gap. *Phys. Rev. A* **2017**, *95*.
- (6) Ziesche, P. On relations between correlation, fluctuation and localization. *J. Mol. Struct.* **2000**, *527*, 35–50.
- (7) Juhász, T.; Mazziotti, D. A. The cumulant two-particle reduced density matrix as a measure of electron correlation and entanglement. *J. Chem. Phys.* **2006**, *125*, 174105.
- (8) Mazziotti, D. A. Approximate solution for electron correlation through the use of Schwinger probes. *Chem. Phys. Lett.* **1998**, *289*, 419–427.
- (9) Mok, D. K. W.; Neumann, R.; Handy, N. C. Dynamical and Nondynamical Correlation. *J. Chem. Phys.* **1996**, *100*, 6225–6230.
- (10) Valderrama, E.; Ludeña, E. V.; Hinze, J. Analysis of dynamical and nondynamical components of electron correlation energy by means of local-scaling density-functional theory. *J. Chem. Phys.* **1997**, *106*, 9227–9235.
- (11) Valderrama, E.; Ludeña, E. V.; Hinze, J. Assessment of dynamical and nondynamical correlation energy components for the beryllium-atom isoelectronic sequence. *J. Chem. Phys.* **1999**, *110*, 2343–2353.
- (12) Ramos-Cordoba, E.; Salvador, P.; Matito, E. Separation of dynamic and nondynamic correlation. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24015–24023.
- (13) Benavides-Riveros, C. L.; Lathiotakis, N. N.; Marques, M. A. L. Towards a formal definition of static and dynamic electronic correlations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 12655–12664.
- (14) Ramos-Cordoba, E.; Matito, E. Local Descriptors of Dynamic and Nondynamic Correlation. *J. Chem. Theory Comput.* **2017**, *13*, 2705–2711.

- (15) Via-Nadal, M.; Rodríguez-Mayorga, M.; Ramos-Cordoba, E.; Matito, E. Singling Out Dynamic and Nondynamic Correlation. *J. Phys. Chem. Lett.* **2019**, *10*, 4032–4037.
- (16) Henzinger, T. A.; Lukina, A.; Schilling, C. Outside the Box: Abstraction-Based Monitoring of Neural Networks. *arXiv e-prints* **2019**, arXiv:1911.09032.
- (17) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (18) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.
- (19) Sherrill, C. D.; Schaefer, H. F. *Advances in Quantum Chemistry*; Elsevier, 1999; pp 143–269.
- (20) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- (21) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (22) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- (23) Vosko, S. H.; Wilk, L. Influence of an improved local-spin-density correlation-energy functional on the cohesive energy of alkali metals. *Phys. Rev. B* **1980**, *22*, 3812–3815.
- (24) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (25) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

- (26) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (27) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*.
- (28) Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.
- (29) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Chem. Phys.* **1994**, *98*, 11623–11627.
- (30) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (31) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*.
- (32) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*.
- (33) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.
- (34) Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.;

- Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **2013**, *15*, 095003.
- (35) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Phys. Rev. X* **2018**, *8*.
- (36) Bonati, L.; Parrinello, M. Silicon Liquid Structure and Crystal Nucleation from Ab Initio Deep Metadynamics. *Phys. Rev. Lett.* **2018**, *121*.
- (37) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*.
- (38) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **2017**, *95*.
- (39) Hegde, G.; Bowen, R. C. Machine-learned approximations to Density Functional Theory Hamiltonians. *Sci. Rep.* **2017**, *7*.
- (40) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2016**, *121*, 511–522.
- (41) Li, Y.; Li, H.; Pickard, F. C.; Narayanan, B.; Sen, F. G.; Chan, M. K. Y.; Sankaranarayanan, S. K. R. S.; Brooks, B. R.; Roux, B. Machine Learning Force Field Parameters from Ab Initio Data. *J. Chem. Theory Comput.* **2017**, *13*, 4492–4503.
- (42) Jinnouchi, R.; Karsai, F.; Kresse, G. On-the-fly machine learning force field generation: Application to melting points. *Phys. Rev. B* **2019**, *100*.
- (43) Li, Z.; Kermode, J. R.; Vita, A. D. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*.
- (44) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*.

- (45) Fletcher, T. L.; Popelier, P. L. A. Transferable kriging machine learning models for the multipolar electrostatics of helical deca-alanine. *Theor. Chem. Acc.* **2015**, *134*.
- (46) Fletcher, T. L.; Popelier, P. L. A. Toward amino acid typing for proteins in FFLUX. *J. Comput. Chem.* **2016**, *38*, 336–345.
- (47) Fletcher, T. L.; Popelier, P. L. A. FFLUX: Transferability of polarizable machine-learned electrostatics in peptide chains. *J. Comput. Chem.* **2017**, *38*, 1005–1014.
- (48) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
- (49) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, e1603015.
- (50) Chmiela, S.; Sauceda, H. E.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. sGDML: Constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* **2019**, *240*, 38–45.
- (51) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*.
- (52) Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* **2018**, *9*, 2261–2269.
- (53) Lubbers, N.; Smith, J. S.; Barros, K. Hierarchical modeling of molecular energies using a deep neural network. *J. Chem. Phys.* **2018**, *148*, 241715.
- (54) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.

- (55) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.
- (56) Faber, F. A.; Christensen, A. S.; Huang, B.; von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys.* **2018**, *148*, 241717.
- (57) Christensen, A. S.; Bratholm, L. A.; Faber, F. A.; von Lilienfeld, O. A. FCHL revisited: Faster and more accurate quantum machine learning. *J. Chem. Phys.* **2020**, *152*, 044107.
- (58) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, 114102.
- (59) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*.
- (60) Glielmo, A.; Sollich, P.; Vita, A. D. Accurate interatomic force fields via machine learning with covariant kernels. *Phys. Rev. B* **2017**, *95*.
- (61) Glielmo, A.; Zeni, C.; Vita, A. D. Efficient nonparametricn-body force fields from machine learning. *Phys. Rev. B* **2018**, *97*.
- (62) Zeni, C.; Rossi, K.; Glielmo, A.; Fekete, Á.; Gaston, N.; Baletto, F.; Vita, A. D. Building machine learning force fields for nanoclusters. *J. Chem. Phys.* **2018**, *148*, 241739.

- (63) Huan, T. D.; Batra, R.; Chapman, J.; Kim, C.; Chandrasekaran, A.; Ramprasad, R. Iterative-Learning Strategy for the Development of Application-Specific Atomistic Force Fields. *J. Phys. Chem. C* **2019**, *123*, 20715–20722.
- (64) Chapman, J.; Batra, R.; Ramprasad, R. Machine learning models for the prediction of energy, forces, and stresses for Platinum. *Comput. Mater. Sci.* **2020**, *174*, 109483.
- (65) Botu, V.; Chapman, J.; Ramprasad, R. A study of adatom ripening on an Al (1 1 1) surface with machine learning force fields. *Comput. Mater. Sci.* **2017**, *129*, 332–335.
- (66) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2016**, *121*, 511–522.
- (67) Botu, V.; Ramprasad, R. Adaptive machine learning framework to accelerate ab initio molecular dynamics. *Int. J. Quant. Chem.* **2014**, *115*, 1074–1083.
- (68) Huan, T. D.; Batra, R.; Chapman, J.; Krishnan, S.; Chen, L.; Ramprasad, R. A universal strategy for the creation of machine learning-based atomistic force fields. *Npj Comput. Mater.* **2017**, *3*.
- (69) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **2017**, *3*, e1701816.
- (70) Han, R.; Lubber, S. Trajectory-based machine learning method and its application to molecular dynamics. *Mol. Phys.* **2020**, 1–14.
- (71) Zaverkin, V.; Kästner, J. Gaussian Moments as Physically Inspired Molecular Descriptors for Accurate and Scalable Machine Learning Potentials. *J. Chem. Theory Comput.* **2020**,
- (72) Anderson, B.; Hy, T.-S.; Kondor, R. Cormorant: Covariant Molecular Neural Networks. *arXiv e-prints* **2019**, arXiv:1906.04015.

- (73) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, I., Thomas F. OrbNet: Deep Learning for Quantum Chemistry Using Symmetry-Adapted Atomic-Orbital Features. *arXiv e-prints* **2020**, arXiv:2007.08026.
- (74) Goryaeva, A. M.; Maillet, J.-B.; Marinica, M.-C. Towards better efficiency of inter-atomic linear machine learning potentials. *Comput. Mater. Sci.* **2019**, *166*, 200–209.
- (75) Cooper, A. M.; Hallmen, P. P.; Kästner, J. Potential energy surface interpolation with neural networks for instanton rate calculations. *J. Chem. Phys.* **2018**, *148*, 094106.
- (76) Denzel, A.; Kästner, J. Gaussian process regression for geometry optimization. *J. Chem. Phys.* **2018**, *148*, 094114.
- (77) Zaverkin, V.; Kästner, J. Gaussian Moments as Physically Inspired Molecular Descriptors for Accurate and Scalable Machine Learning Potentials. *J. Chem. Theory Comput.* **2020**, *16*, 5410–5421.
- (78) Cooper, A. M.; Kästner, J.; Urban, A.; Artrith, N. Efficient training of ANN potentials by including atomic forces via Taylor expansion and application to water and a transition-metal oxide. *Npj Comput. Mater.* **2020**, *6*.
- (79) Klinting, E. L.; Thomsen, B.; Godtliebsen, I. H.; Christiansen, O. Employing general fit-bases for construction of potential energy surfaces with an adaptive density-guided approach. *J. Chem. Phys.* **2018**, *148*, 064113.
- (80) Schmitz, G.; Christiansen, O. Gaussian process regression to accelerate geometry optimizations relying on numerical differentiation. *J. Chem. Phys.* **2018**, *148*, 241704.
- (81) Schmitz, G.; Godtliebsen, I. H.; Christiansen, O. Machine learning for potential energy surfaces: An extensive database and assessment of methods. *J. Chem. Phys.* **2019**, *150*, 244113.

- (82) Schmitz, G.; Artiukhin, D. G.; Christiansen, O. Approximate high mode coupling potentials using Gaussian process regression and adaptive density guided sampling. *J. Chem. Phys.* **2019**, *150*, 131102.
- (83) Chen, Y.; Zhang, L.; Wang, H.; E, W. Ground state energy functional with Hartree-Fock efficiency and chemical accuracy. *arXiv e-prints* **2020**, arXiv:2005.00169.
- (84) Bogojeski, M.; Vogt-Maranto, L.; Tuckerman, M. E.; Mueller, K.-R.; Burke, K. Density Functionals with Quantum Chemical Accuracy: From Machine Learning to Molecular Dynamics. **2019**,
- (85) Dick, S.; Fernandez-Serra, M. Machine learning accurate exchange and correlation functionals of the electronic density. *Nat. Commun.* **2020**, *11*.
- (86) Welborn, M.; Cheng, L.; Miller, T. F. Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis. *J. Chem. Theory Comput.* **2018**, *14*, 4772–4779.
- (87) Cheng, L.; Kovachki, N. B.; Welborn, M.; Miller, T. F. Regression Clustering for Improved Accuracy and Training Costs with Molecular-Orbital-Based Machine Learning. *J. Chem. Theory Comput.* **2019**, *15*, 6668–6677.
- (88) Cheng, L.; Welborn, M.; Christensen, A. S.; Miller, T. F. A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules. *J. Chem. Phys.* **2019**, *150*, 131103.
- (89) Schmidt, J.; Benavides-Riveros, C. L.; Marques, M. A. L. Machine Learning the Physical Nonlocal Exchange–Correlation Functional of Density-Functional Theory. *J. Phys. Chem. Lett.* **2019**, *10*, 6425–6431.
- (90) Nagai, R.; Akashi, R.; Sugino, O. Completing density functional theory by machine learning hidden messages from molecules. *Npj Comput. Mater.* **2020**, *6*.

- (91) Nudgejima, T.; Iwabata, Y.; Seino, J.; Yoshikawa, T.; Nakai, H. Machine-learned electron correlation model based on correlation energy density at complete basis set limit. *J. Chem. Phys.* **2019**, *151*, 024104.
- (92) Lei, X.; Medford, A. J. Design and analysis of machine learning exchange-correlation functionals via rotationally invariant convolutional descriptors. *Phys. Rev. Mater.* **2019**, *3*.
- (93) Imamura, Y.; Nakai, H. Energy density analysis for second-order Møller-Plesset perturbation theory and coupled-cluster theory with singles and doubles: Application to C₂H₄-CH₄ complexes. *J. Comput. Chem.* **2008**, *29*, 1555–1563.
- (94) Nakai, H. Energy density analysis with Kohn-Sham orbitals. *Chem. Phys. Lett.* **2002**, *363*, 73–79.
- (95) Imamura, Y.; Takahashi, A.; Nakai, H. Grid-based energy density analysis: Implementation and assessment. *J. Chem. Phys.* **2007**, *126*, 034103.
- (96) Vyboishchikov, S. F. A Simple Local Correlation Energy Functional for Spherically Confined Atoms from ab Initio Correlation Energy Density. *Chem. Phys. Chem.* **2017**, *18*, 3478–3484.
- (97) Knowles, P. J.; Handy, N. C. A determinant based full configuration interaction program. *Comput. Phys. Commun.* **1989**, *54*, 75–83.
- (98) Aquilante, F.; Autschbach, J.; Carlson, R. K.; Chibotaru, L. F.; Delcey, M. G.; Vico, L. D.; Galván, I. F.; Ferré, N.; Frutos, L. M.; Gagliardi, L.; Garavelli, M.; Giusani, A.; Hoyer, C. E.; Manni, G. L.; Lischka, H.; Ma, D.; Malmqvist, P. Å.; Müller, T.; Nenov, A.; Olivucci, M.; Pedersen, T. B.; Peng, D.; Plasser, F.; Pritchard, B.; Reiher, M.; Rivalta, I.; Schapiro, I.; Segarra-Martí, J.; Stenrup, M.; Truhlar, D. G.; Ungur, L.; Valentini, A.; Vancoillie, S.; Veryazov, V.; Vysotskiy, V. P.; Weingart, O.;

- Zapata, F.; Lindh, R. Molcas8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.* **2016**, *37*, 506–541.
- (99) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: an open-source ab initio electronic structure program. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *2*, 556–565.
- (100) Golub, G. H.; Welsch, J. H. Calculation of Gauss quadrature rules. *Math. Comput.* **1969**, *23*, 221–221.
- (101) Lebedev, V. Quadratures on a sphere. *USSR Comput. Math. Math. Phys.* **1976**, *16*, 10–24.
- (102) Mayer, I.; Salvador, P. Overlap populations, bond orders and valences for ‘fuzzy’ atoms. *Chemical Physics Letters* **2004**, *383*, 368–375.
- (103) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (104) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules. *J. Chem. Phys.* **1970**, *52*, 2769–2773.
- (105) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.;

- Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.
- (106) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.
- (107) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.
- (108) Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **2005**, *114*, 145–152.
- (109) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (110) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (111) CP2K version 7.0 (Development Version), the CP2K developers group. CP2K is freely available from <https://www.cp2k.org/>.
- (112) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *4*, 15–25.

- (113) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
- (114) Chollet, F., et al. Keras. 2015; <https://github.com/fchollet/keras>.
- (115) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015; <http://tensorflow.org/>, Software available from tensorflow.org.
- (116) Givon, L. E.; Unterthiner, T.; Erichson, N. B.; Chiang, D. W.; Larson, E.; Pfister, L.; Dieleman, S.; Lee, G. R.; van der Walt, S.; Menn, B.; Moldovan, T. M.; Bastien, F.; Shi, X.; Schlüter, J.; Thomas, B.; Capdevila, C.; Rubinsteyn, A.; Forbes, M. M.; Frelinger, J.; Klein, T.; Merry, B.; Merrill, N.; Pastewka, L.; Liu, L. Y.; Clarkson, S.; Rader, M.; Taylor, S.; Bergeron, A.; Ukani, N. H.; Wang, F.; Lee, W.-K.; Zhou, Y. scikit-cuda 0.5.3: a Python interface to GPU-powered libraries. 2019.
- (117) Klöckner, A.; Pinto, N.; Lee, Y.; Catanzaro, B.; Ivanov, P.; Fasih, A. PyCUDA and PyOpenCL: A Scripting-Based Approach to GPU Run-Time Code Generation. *Parallel Comput.* **2012**, *38*, 157–174.
- (118) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102*, 939–947.

- (119) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *The Journal of Chemical Physics* **2006**, *124*, 034108.
- (120) Cioslowski, J.; Liu, G. Fast evaluation of electron intracule and extracule densities on large grids of points. *J. Chem. Phys.* **1996**, *105*, 4151–4158.

Table of Contents

A machine learning based density functional approximation for the MP2 electron correlation energies is proposed. The transferability of the approach across chemical and configurational space is tested and application in the evaluation of double-hybrid exchange–correlation functionals is presented.

Figure 12: For Table of Contents Only.