
DeepDFT: Neural Message Passing Network for Accurate Charge Density Prediction

Peter Bjørn Jørgensen
Energy Conversion and Storage
Technical University of Denmark

Arghya Bhowmik
Energy Conversion and Storage
Technical University of Denmark
arbh@dtu.dk

Abstract

We introduce DeepDFT, a deep learning model for predicting the electronic charge density around atoms $\rho(r)$, the fundamental variable in electronic structure simulations from which all ground state properties can be calculated. The model is formulated as neural message passing on a graph, consisting of interacting atom vertices and special query point vertices for which the charge density is predicted. The accuracy and scalability of the model are demonstrated for molecules, solids and liquids. The trained model achieves lower average prediction errors than the observed variations in charge density obtained from density functional theory simulations using different exchange correlation functionals.

1 Introduction

Machine learning methods have been gaining popularity in materials research community recently [1]. Data driven models are being used to accelerate simulation based materials design through development of faster surrogate models that replace or work with physics based models. In the realm of atomic scale modelling, the focus has been on the mapping of the molecular geometry to a target property, such as the total energy or the band gap [2]. Although more challenging to predict, the electronic charge density around atoms of a given molecule or material can describe the system completely, i.e. all ground state properties can be obtained from the charge density with reasonable accuracy [3]. It is the fundamental variable in Kohn Sham density functional theory (KS-DFT) [4] based quantum simulations - the most popular electronic structure method used for materials research and design. DFT being an $O(n^3)$ method, computational cost inhibits simulating more than few hundred atoms. Computational design of materials from atomistic scale is severely limited by the system size as atomic scale representations of real materials and processes (e.g. battery electro-chemistry at cathode-electrolyte interface) needs 10^5 atoms or more. In this work we aim at obtaining a linear scaling, general and flexible machine learning model that is able to utilise the vast amount of existing simulation results and those being produced daily in supercomputers around the world to predict the charge density for atomic structures of unprecedented size, enabling new avenues of atomic scale materials design.

2 State of the art and scope

The first efforts towards predicting the electron density with machine learning [5, 6] used a kernel ridge regression model, which takes, as input, an artificial Gaussian potential sampled on a 3D grid and outputs coefficients that represents the electron density in a Fourier basis set. The model can accurately predict the electron density in molecular dynamics simulations, however, by design, the model does not transfer across different molecular systems. This limitation was later overcome with a similar kernel model [7, 8], in which the total electron density is decomposed into additive

atom-centered contributions and a symmetry-adapted Gaussian process regression model [9] is used to predict the expansion coefficients of each contribution based on the local environment around the atoms. The locality of the model allows it to transfer across different molecules, but the cubic computational cost (in number of training examples) of Gaussian process regression is prohibitive for applying the model to systems with a large number of atom types, where large number of data points are required to cover the chemical space of interest. Deep learning models generally scale well with the number of training examples and using these models might be the right direction for training with large data sets. Among deep learning approaches - deep convolutional neural networks have been used to solve the charge density prediction problem by posing it as an image to image translation problem. For example, given a low accuracy DFT calculated density a neural network model predicts the output of high accuracy DFT [10]. Such an approach is restricted to a fixed voxel size and is not equivariant to rotations. A different approach is to manually construct fingerprints describing the local environment in 3D space and map the fingerprints to a density value using neural networks [11, 12, 13]. Irrespective of machine learned or hand-engineered, features, when collected from longer distances (i.e. larger cutoff radius), may provide better accuracy. As the number of neighbours grow $O(r^3)$ with the cutoff radius r , a relatively small cutoff radius need to be used. For neural message passing (or graph convolution) models, using a small graph connectivity cutoff in combination with multi-step message passing enables efficient propagation of long-range information in comparison to just increasing atomic environment cutoff distance - computational complexity grows only linearly with the number of message passing steps.

In the previous graph convolution based model [14] a new graph is created for each electron density query point. The atomic environment representation is thus dependent on the choice of query point and therefore the learned atomic environment representations do not directly transfer across different query points. In contrast, our model first obtain a representation of atoms and their local environment and in the next step map these representations to the electron density at the requested query points. This two-step approach allows the model to efficiently reuse the learned atomic environment representation. Both steps are learned simultaneously in an end-to-end fashion. In this work we demonstrate the excellent scalability and accuracy of our DeepDFT model. We train on very large scale datasets and achieve prediction errors lower than the variations in DFT charge density obtained with different exchange correlation (Xc) approximations.

3 Neural Message Passing Network

The DeepDFT density model is framed in the message passing framework devised by [15] and components of the model architecture is inspired by the SchNet model [16]. The input to the model is a graph representation of the molecule or crystal structure and the graph has a vertex for each atom in the molecule or for each atom in the crystal structure unit cell. Edges are defined by a constant cutoff distance, i.e. we draw an edge between vertex v and w if the distance between v and w is less than a certain cutoff distance (4 \AA). The edges may cross periodic boundary conditions as in quotient graphs [17, 18]. Special probe vertices, that only accept incoming edges, are placed at each query point. A simple example graph with four atoms and three query points is illustrated in Figure 1.

Each vertex v has a hidden state h_v^t at "time" step t . They are updated in a number of interaction "time" steps T . We use $T = 6$ interaction steps in all the experiments in this work. The hidden states of the vertices are updated in two steps. First messages from neighbour vertices are computed using the message function $M_t(\cdot)$ and then the vertex state is updated by a state transition function $S_t(\cdot)$.

$$m_v^{t+1} = \sum_{w \in N(v)} M_t(h_v^t, h_w^t, e_{vw}^{t+1}), \quad h_v^{t+1} = S_t(h_v^t, m_v^{t+1}), \quad (1)$$

where $N(v)$ denotes the neighbourhood of the vertex v . For the initial atom vertex state we use the atomic number to look up an embedding vector for each vertex. As edge feature we use the distance between the two corresponding atoms expanded in a series of exponentiated quadratic functions:

$$(e_{vw}^0)_k = \exp\left(-\frac{(\varepsilon_{vw} - (-\mu_{\min} + k\Delta))^2}{2\Delta^2}\right), k = 0 \dots k_{\max} \quad (2)$$

where μ_{\min} , Δ , and k_{\max} are chosen such that the centers of the functions covers the cutoff range. This can be seen as a soft 1-hot-encoding of the distances, which makes it easier for a neural network to learn a function where the input distance is uncorrelated with the output of the network if that is

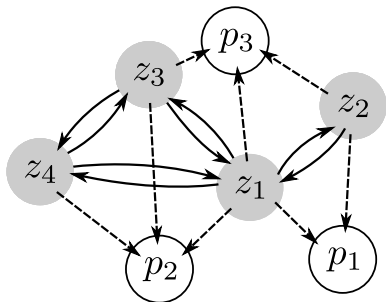


Figure 1: Messages are exchanged between atom vertices z_i in several steps while the probe nodes p_j only receive messages.

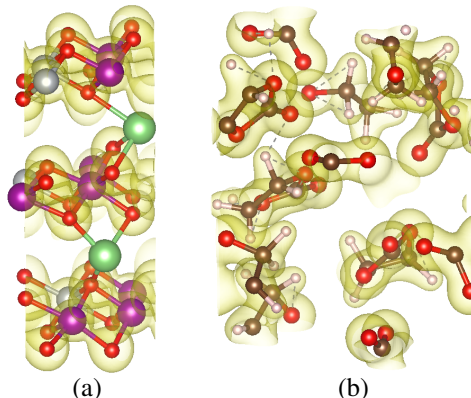


Figure 2: Data set examples from (a) Lithium-Ion battery cathode - mixed transition metal (CO/Mn/Ni) layered oxide and (b) Ethylene-carbonate electrolyte in liquid state.

necessary. The atom-to-atom message function is a function of the sending vertex state h_w , the edge features e_{vw} , and can be written as

$$M_t(h_v^t, h_w^t, e_{vw}^t) = M_t(h_w^t, e_{vw}^t) = (W_1^t h_w^t) \odot c(d_{vw}) g(W_3^t g(W_2^t e_{vw}^t)), \quad (3)$$

where \odot denotes element-wise multiplication, $g(\cdot)$ is the soft-plus activation function, and $c(\cdot)$ is a soft cutoff function $c(x) = 1 - \text{sigmoid}(5 * (x - (d_{\text{cut}} - 1.5)))$ where $d_{\text{cut}} = 4 \text{ \AA}$ is the graph edge cutoff distance. The expression on the right-hand side of the element-wise multiplication in (3) can be seen as a filter generating function [16]. Because the filter depends only on the distance between atoms it becomes a radial filter rather than the directional filters normally used in image processing.

The state transition function is a two layer neural network on the sum of incoming messages and the result is added to the current hidden state as in Residual Networks [19]:

$$S_t(h_v^t, m_v^{t+1}) = h_v^t + W_5^t g(W_4^t m_v^{t+1}), \quad (4)$$

For the special probe vertices the hidden state is initialised as a vector of zeros. We use the same form of message function as for the inter-atom messages (3), but the weight matrices are not shared between the inter-atom message function and probe message function. The state transition function for the probe vertices is also very similar to the atom hidden state transition function (4). The only difference is that we add a “forget” gate term $F_t(\cdot)$, a two-layer neural network, that controls whether and which of the probe state entries that are updated or kept after each interaction step. The state transition functions for the probe vertices are thus given by:

$$\tilde{S}_t(h_v^t, m_v^{t+1}) = F_t(h_v^t) \odot h_v^t + (1 - F_t(h_v^t)) \odot W_5^t g(W_4^t m_v^{t+1}), \quad (5)$$

$$\text{where } F_t(h_v^t) = \text{sigmoid}(W_7^t g(W_6^t m_v^{t+1})), \quad (6)$$

The PyTorch implementation and pretrained model are available on Github.

4 Dataset and Model Training

To assess the model we use the QM9 dataset [20, 21] (134k small molecules with up to nine heavy atoms (CNOF)) that is widely used for benchmarking machine learning models for molecular property prediction. Additionally we also train and test with charge density data from crystalline and liquid state materials. 1: A challenging to model but industrially important multi-transition metal layered oxide lithium ion battery (LIB) cathode. 5000 configurations are generated through Monte Carlo style crystal site occupation for transition metal ions (Ni/Mn/Co) and lithium/vacancy to represent varieties of chemistry and lithiation state followed by structural optimization. 2: Liquid ethylene carbonate - the most used LIB electrolyte. 12000 disordered configurations are generated through high temperature (3000K) accelerated identity preserving molecular dynamics. KS-DFT is used with PBE XC-functional and VASP code [22] for obtaining valence charge density on a volumetric grid. For each gradient step we sample 1000 query points in two training set molecules/crystal-structures and use the mean-squared-error at the query points as the cost function. The model is trained for two weeks on a single Nvidia GPU and a small validation set is used for early stopping.

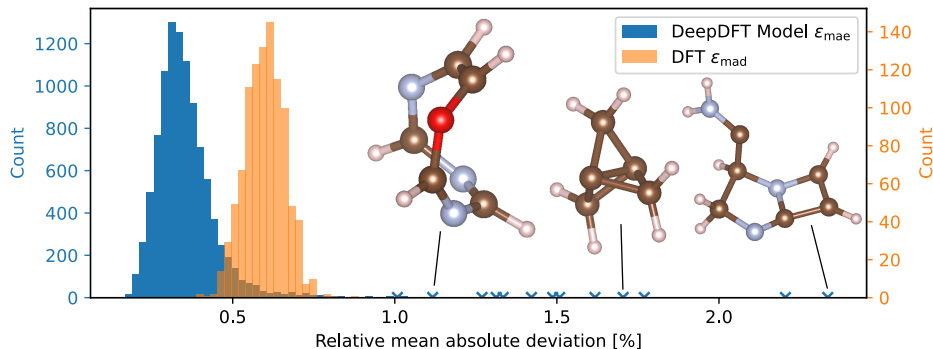


Figure 3: Histogram of relative MAE on QM9 test set. The markers show the outliers.

Table 1: Datasets and prediction errors for DeepDFT model

Dataset	Dataset Splits			Test Set Error (ε_{mae} %)	
	Train	Val.	Test	Average	Max.
QM9	123835	50	10000	0.36	2.35
LIB Cathode	3903	50	1000	0.10	0.29
Ethylene Carbonate	7330	50	4000	0.53	0.69

5 Results

To evaluate the model, we integrate the mean absolute error (MAE) over the whole simulation box normalized by total number of electrons (eq. (7)) following [7, 8]. The DFT calculated density is used as ground truth $\rho(x)$. The distribution of ε_{mae} errors for the QM9 test set is shown in Figure 3. The DFT uncertainties, also included in the histogram, are estimated using eq. (8) from an ensemble of densities $K = 8$ calculated using VASP with eight different XC functionals (BF, PE, 91, CA, PZ, RP, RE, PS) [23, 24, 25, 26, 27, 28, 29, 30] for 1000 molecules. For each grid point we compute the mean absolute deviation (MAD) of the ensemble around the median density. The MAD is then integrated across all grid points and divided by the total number of electrons eq. (8).

$$\varepsilon_{\text{mae}} = \frac{\int_{x \in V} |\rho(x) - \hat{\rho}(x)|}{\int_{x \in V} |\rho(x)|} \quad (7) \quad \varepsilon_{\text{mad}} = \frac{\int_{x \in V} \frac{1}{K} \sum_{k=1}^K |\rho_k(x) - \rho_{\text{median}}(x)|}{\int_{x \in V} |\rho_1(x)|} \quad (8)$$

Predictions errors are mostly below 0.5% with an average of 0.35% and this is below the estimated DFT error which averages at 0.6%. This indicates our model is adept at learning from electronic structure simulations which are more accurate than KS-DFT. For comparison, another baseline, which is the superposition of the single atom electron densities gives an average error of 18% for QM9 molecules (not shown). The DeepDFT model predictions include a number of outliers with very large errors compared to the average error, but they are still an order of magnitude more accurate than the superposition of single atom densities. Manual inspection of the QM9 test set outliers shows that presence of exotic structural units (see Figure 3), which are underrepresented or unrepresented in the training sets, leads to high error. For the other two datasets the average ε_{mae} errors are also low (Table 1) and no outliers are observed as chemical variations are well sampled.

6 Conclusion

The proposed DeepDFT model achieves scalable and accurate prediction of electronic charge densities. We envision that this type of model will allow design of new materials in a wide range of scientific/engineering problems where the size of the problem limits the applicability of conventional methods. For example, in future work, we will investigate the capability of the model to capture charge transfers in reactive dynamical systems. We will also investigate the data efficiency of the method, i.e. if learning the charge density first is more data efficient than directly predicting the properties of interest, such as total energy or band gaps.

Acknowledgments and Disclosure of Funding

Peter Bjørn Jørgensen and Arghya Bhowmik acknowledges financial support from VILLUM FONDEN by a research grant (00023105) for the DeepDFT project.

References

- [1] Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559(7715):547–555, 2018.
- [2] Frank Noé, Alexandre Tkatchenko, Klaus-Robert Müller, and Cecilia Clementi. Machine learning for molecular simulation. *Annual review of physical chemistry*, 71:361–390, 2020.
- [3] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Challenges for density functional theory. *Chemical reviews*, 112(1):289–320, 2012.
- [4] Mike C Payne, Michael P Teter, Douglas C Allan, TA Arias, and ad JD Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Reviews of modern physics*, 64(4):1045, 1992.
- [5] Felix Brockherde, Leslie Vogt, Li Li, Mark E Tuckerman, Kieron Burke, and Klaus-Robert Müller. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.*, 8(1):872, October 2017.
- [6] Mihail Bogojeski, Felix Brockherde, Leslie Vogt-Maranto, Li Li, Mark E. Tuckerman, Kieron Burke, and Klaus-Robert Müller. Efficient prediction of 3d electron densities using machine learning. *arXiv:1811.06255*, 2018.
- [7] Andrea Grisafi, Alberto Fabrizio, Benjamin Meyer, David M Wilkins, Clemence Corminboeuf, and Michele Ceriotti. Transferable Machine-Learning Model of the Electron Density. *ACS Cent Sci*, 5(1):57–64, January 2019.
- [8] Alberto Fabrizio, Andrea Grisafi, Benjamin Meyer, Michele Ceriotti, and Clemence Corminboeuf. Electron density learning of non-covalent systems. *Chemical Science*, 10(41):9424–9432, 2019.
- [9] Andrea Grisafi, David M. Wilkins, Gábor Csányi, and Michele Ceriotti. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.*, 120:036002, Jan 2018.
- [10] Anton V. Sinitskiy and Vijay S. Pande. Deep neural network computes electron densities and energies of a large set of organic molecules faster than density functional theory (DFT). *arXiv:1809.02723*, 2018.
- [11] Eric Schmidt, Andrew T. Fowler, James A. Elliott, and Paul D. Bristowe. Learning models for electron densities with Bayesian regression. *Computational Materials Science*, 149:250–258, June 2018.
- [12] Anand Chandrasekaran, Deepak Kamal, Rohit Batra, Chiho Kim, Lihua Chen, and Rampi Ramprasad. Solving the electronic structure problem with machine learning. *npj Computational Materials*, 5(1):1–7, February 2019.
- [13] Leonardo Zepeda-Núñez, Yixiao Chen, Jiefu Zhang, Weile Jia, Linfeng Zhang, and Lin Lin. Deep Density: Circumventing the Kohn-Sham equations via symmetry preserving neural networks. *arXiv:1912.00775*, November 2019.
- [14] Sheng Gong, Tian Xie, Taishan Zhu, Shuo Wang, Eric R. Fadel, Yawei Li, and Jeffrey C. Grossman. Predicting charge density distribution of materials using a local-environment-based graph convolutional network. *Physical Review B*, 100(18):184103, November 2019.
- [15] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural Message Passing for Quantum Chemistry. In *International Conference on Machine Learning*, pages 1263–1272, July 2017.
- [16] K T Schütt, H E Sauceda, P-J Kindermans, A Tkatchenko, and K-R Müller. SchNet - A deep learning architecture for molecules and materials. *J. Chem. Phys.*, 148(24):241722, June 2018.
- [17] S J Chung, Th Hahn, and W E Klee. Nomenclature and generation of three-periodic nets: The vector method. *Acta Crystallogr. A*, 40(1):42–50, January 1984.
- [18] W E Klee. Crystallographic nets and their quotient graphs. *Cryst. Res. Technol.*, 39(11):959–968, November 2004.
- [19] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [20] Lars Ruddigkeit, Ruud van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.*, 52(11):2864–2875, November 2012.

- [21] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Sci Data*, 1:140022, August 2014.
- [22] Jürgen Hafner. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of computational chemistry*, 29(13):2044–2078, 2008.
- [23] Jess Wellendorff, Keld T Lundgaard, Andreas Møgelhøj, Vivien Petzold, David D Landis, Jens K Nørskov, Thomas Bligaard, and Karsten W Jacobsen. Density functionals for surface science: Exchange-correlation model development with bayesian error estimation. *Physical Review B*, 85(23):235149, 2012.
- [24] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.
- [25] John P Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical review B*, 45(23):13244, 1992.
- [26] David M Ceperley and Berni J Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45(7):566, 1980.
- [27] John P Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048, 1981.
- [28] Bjørk Hammer, Lars Bruno Hansen, and Jens Kehlet Nørskov. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Physical review B*, 59(11):7413, 1999.
- [29] Yingkai Zhang and Weitao Yang. Comment on “generalized gradient approximation made simple”. *Physical Review Letters*, 80(4):890, 1998.
- [30] Gábor I Csonka, John P Perdew, Adrienn Ruzsinszky, Pier HT Philipsen, Sébastien Lebègue, Joachim Paier, Oleg A Vydrov, and János G Ángyán. Assessing the performance of recent density functionals for bulk solids. *Physical Review B*, 79(15):155107, 2009.