



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna | Austria



Dissertation

Machine learning application for DFT exchange functionals

carried out for the purpose of obtaining the degree of Doctor of Natural Sciences,
submitted at TU Wien, Institute of Materials Chemistry, by

Péter Kovács
Mat.Nr.: 11738290

under the supervision of

Univ. Prof. Dr. Georg K. H. Madsen
E165
Institute of Materials Chemistry

Vienna, August 2021



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Acknowledgement

First of all I would like to thank Univ. Prof. Dr. Georg K. H. Madsen who supervised my work and guided me on the path to become a researcher. He encouraged me to follow my interest even in non-strictly related topics, which also heavily influenced the final form of my work. He was always ready to answer my questions, no matter how silly they were sometimes, and teach me or at least point me the right direction whenever i needed. He was also very supportive and patient when it came to the specific parts of my work with which i struggled with, and thanks his guidance I have also improved, I hope, in those areas.

I can not write my acknowledgement, without mentioning Fabien Tran, who provided most of the data I based my research on. He was always eager to help whenever i asked for something more or new and his ideas also progressed my research significantly. He often drew my attention to new publications in the field, which turned out to be very interesting and useful for me.

I would also like to thank Dr. Jesús Carrete for all the support and encouragement I received during the past years. My discussion with him contributed a lot to my understanding and helped me see many things from a different perspective. I am also grateful to him for involving me in the project about the IR spectra of PAHs, which taught me a lot and i really enjoyed working on.

I'm thankful to Dr. Peter Blaha and Dr. Zhao Wang for the fruitful discussions during our collaborations.

I would also like to express my appreciation to all my teachers who sparked and kept my interest in science long before I became a PhD student. Especially Dr. Gergő Pokol and Dr. László Udvardi who guided my work during my bachelor and master studies.

I would also like to thank my friends, both from here and Hungary, for their support, for always being able to cheer me up and for simply being the best friends i could hope for.

Finally, i would like to thank my girlfriend, Andi, for always being there for me, encouraging me, listening to my complaining and so much more. I am also very grateful for my parents, for all the love and support i received from them.

Abstract

Currently density functional theory is one of the most commonly used electronic structure calculation method, thanks to its good accuracy - computational cost ratio. Since its invention the performance of DFT drastically increased, due to improvements both in the numerical approaches and in the exchange-correlation(XC) functional approximations. XC functional approximations are often categorized based on their complexity and these categories are referred to steps on Jacob's ladder of DFT. The lowest step consists of the local density approximations(LDAs), which use only the density to approximate the XC energy at a given point. The higher rungs contain the generalized gradient approximations(GGAs), which also use the gradient of the density, and the meta-generalized gradient approximations (mGGAs), which include the laplacian of the density or the kinetic energy density as well. Since density functional approximations (DFAs) play a crucial role in the success of DFT, they are a focus of research.

In the present thesis, we examine why the SCAN functional results in poor equilibrium lattice parameter predictions for alkali - and alkaline earth metals, while it performs well for the rest of the tested solids. It was found that the exchange part of the XC energy plays the dominant role in the lattice parameter predictions. We identified the semi-core region to be responsible for a push in the direction of larger lattice parameters in all solids, but in materials with more interstitial electrons this effect was suppressed by the much stronger effect of the interstitial region.

Using principal component analysis we also explored how much information does the laplacian of the density carry when it is used alongside the gradient of the density and the kinetic energy density for DFAs. We showed that in most cases the laplacian can be reproduced as a linear combination of the other descriptors, an exception to this was only

found in the middle of covalent bonds. As a result we concluded that while the laplacian might contain some useful information its inclusion in DFAs alongside the kinetic energy density is not expected to cause significant improvements in accuracy.

Employing unsupervised machine learning methods we developed a way to identify groups of materials in databases, which occupy similar regions in the space of mGGA functional descriptors. This method was able to reproduce groups formed based on chemical intuition in a purely data driven way. Using our method databases with strong biases can be identified and rebalanced to produce better benchmarking or training datasets for DFA development.

To find the limits of mGGA functionals and understand how specific changes in their functional form affect their performance we trained 25 mGGAs with different weights on equilibrium lattice parameter, cohesive energy and band gap errors. The training was carried out on a set of 44 solids for lattice parameter and cohesive energy and on 440 materials for band gap. It was found that mGGAs express a similar tradeoff between the accuracy of lattice parameters and cohesive energies as it was seen for GGA functionals, but in the meantime they manage to predict the band gaps significantly better. Compared to other existing functionals the trained ones showed better performance on this three specific errors, hinting that this might be the limit of mGGAs. The functional trained mostly on cohesive energies showed similarities to the mBEEF functional, which also produces low cohesive energy errors. And the functional trained mostly on band gaps resembled the TASK functional, which was created to produce good band gaps. The similarities with existing functionals indicate that our findings could be general rules for functionals which excel at these specific properties.

Finally, we also developed a neural network based model to predict the infrared(IR) spectra of polycyclic aromatic hydrocarbons(PAHs). These molecules are a focus of interest for astronomers since they are abundant in the universe and are suspected to be responsible for the so called "unidentified infrared emission" features in the IR spectra of various interstellar sources. The vast number of possible PAH configurations makes the bruteforce prediction of their IR spectra with DFT impossible. Our solution based on Morgan fingerprints is many magnitudes faster and the accuracy of predictions is on par

with DFT results. We also proposed a way to assess the accuracy of the predictions based on an ensemble of neural networks and in cases of poor accuracy the calculation can fall back to the traditional DFT approach.



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Contents

1	Introduction	1
1.1	Motivation and timeline	1
2	Background	5
2.1	Density functional theory	5
2.1.1	Interacting electron Hamiltonian	5
2.1.2	Hohenberg-Kohn theorems	6
2.1.3	Kohn-Sham equations	6
2.1.4	Exchange-correlation functionals	8
	Local density approximation	8
	Generalized gradient approximations	9
	Meta-generalized gradient approximations	11
	Hybrid functionals	12
	Success of GGAs and mGGAs	13
	Theoretical constraints on exchange functionals	14
2.1.5	DFT accuracy	16
2.2	Machine learning	19
2.2.1	Supervised learning	20
	Neural networks	20
	Decision trees	21
2.2.2	Unsupervised learning	22
	k-means clustering	22

	Affinity propagation	22
	Hierarchical clustering	23
3	Exchange functional training	25
3.1	Material database	26
3.2	Calculation of different properties	27
3.3	Details of training	29
3.3.1	Approximation form	29
3.3.2	Loss function	30
3.3.3	Training steps	30
3.4	Results	31
3.4.1	Lattice parameter - cohesive energy tradeoff	33
3.4.2	Band gap - cohesive energy tradeoff	35
3.4.3	Band gap - lattice parameter tradeoff	36
3.4.4	Comparison with existing functionals	38
3.4.5	mGGA surface	41
4	Similarity clustering for representative sets of solids for density functional testing	43
4.1	Additional similarity metrics	43
4.1.1	Normalized dot-product	44
4.1.2	Normalized Euclidean distance	44
4.1.3	Normalized Manhattan distance	45
4.1.4	Comparison of distance metrics	45
4.2	Additional clustering methods	46
4.2.1	Affinity propagation	47
4.2.2	Hierarchical clustering	47
5	Outlook	51
5.1	Exchange-correlation functionals	51
5.2	Machine learning in IR prediction	52

<i>Contents</i>	xi
<hr/>	
5.3 Similarity clustering of materials	53
Bibliography	55
6 List of publications	63
7 Curriculum vitae	113



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

1 Introduction

1.1 Motivation and timeline

Since its invention density functional(DFT) theory grew to be one of the most used ab initio electronic structure calculation methods thanks to its great accuracy compared to the relatively low computational cost. While there are multiple more accurate methods, those are often limited to a few atoms per calculation, yet with DFT calculations including more than 1 million atoms are still possible. Nowadays DFT is commonly used in drug screening, catalyst development, semiconductor research and many other fields of chemistry.

Aside from the numerical limitations, the only accuracy restricting factor of DFT is the approximation of the exchange-correlation energy. Through the years numerous approximations were developed with different complexities and goals. These different level of approximations are often referred to steps on the Jacob's ladder of DFT. Functionals on higher rungs tend to be more accurate, since they use more information to approximate the exchange-correlation energy, like the electron density gradient or the kinetic energy density. This fact can be viewed from two directions, either the multiple descriptors carry more information about the underlying system, so this extra information can be used to better predict the exchange energy, or simply more input parameters allow the functional to fit the exact functional more accurately. The goal of my work was to understand the behaviour of these functionals, explore their limitations and create a potentially better approximation than the current ones.

The first step of this journey was to get familiar with the two most used functionals for solid state calculations, namely the PBE and SCAN. SCAN being on a higher rung of

Jacob's ladder was expected to outperform PBE, which was true in most cases except for alkali metals. Investigating this artifact led to a better understanding of how the exchange functionals affect the different calculated properties of materials, and helped to create a framework to analyze these discrepancies. This work was published in the "Comparative study of the PBE and SCAN functionals: the particular case of alkali metals" paper, which can be found in chapter 6.

The accuracy of a functional strongly depends on the information carried by its descriptors. In my second work this extra information carried by the laplacian of the density was analyzed to assess its usefulness along the density, density gradient and kinetic energy density descriptors. The findings were presented in the "Orbital-free approximations to the kinetic-energy density in exchange-correlation MGGA functionals: Tests on solids" paper included in chapter 6.

Benchmarking or training a new functional is a complex problem, since the results highly depend on the materials used in the process. To tackle this problem, or at least quantify the biases in the dataset the next step of my work was to develop a method to identify groups of materials which sample similar regions of the descriptor space. Using these groups we also proposed a method to create new balanced or rebalance existing datasets to minimize the effect of the previously mentioned biases. This work has been submitted to Journal of Chemical Theory and Computation and the manuscript is attached in chapter 6. Extensions to the manuscript are also included in chapter 4.

The final goal of my thesis was to explore the space of new functionals, understand their behaviour and possibly create new, more accurate approximations than the currently available. Based on the findings in the previous steps the newly created functional uses the density gradient and the kinetic energy density as descriptors and aims to predict the lattice parameter and cohesive energy of 44 solids, along with the band gap of more than 400 other materials. The results of the functional training are presented in chapter 3.

As a collaboration with the Guangxi University, I've also worked on developing a neural network based approach to predict infrared spectra of polycyclic aromatic hydrocarbons. These molecules are suspected to be responsible for a yet unexplained range of

infrared emission observed in wide variety of interstellar sources. While DFT is a powerful tool to predict these spectra, the vast number of possible PAHs makes it inefficient. Our method is able to predict these spectra orders of magnitudes faster than the "traditional" DFT approach with comparable accuracy. The details are published in the "Machine-learning Prediction of Infrared Spectra of Interstellar Polycyclic Aromatic Hydrocarbons" paper and also included in chapter 6.



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

2 Background

2.1 Density functional theory

Nowadays density functional theory is one of the most widely used electronic structure calculation methods. Since its main accuracy limiting factor is the exchange-correlation approximation, the potential improvements and general understanding of XC functionals is a focus of research. The following sections cover the basics of DFT with an emphasis on the different levels of exchange functional approximations and their accuracy measured on a set of solids for lattice parameter, cohesive energy and band gap.

2.1.1 Interacting electron Hamiltonian

The non-relativistic Hamiltonian of an electronic system can be written the following way:

$$H = -\frac{1}{2} \sum_i^N \nabla_i^2 + \sum_{i<j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_i^N V_{ext}(\mathbf{r}_i), \quad (2.1)$$

where the first term describes the kinetic energy of electrons, the second term the electron-electron interaction and the third term the external potential. In most cases the description of nuclei can be separated from the description of electrons, so the external potential coming from the them takes the form

$$V_{ext}(\mathbf{r}) = \sum_{\alpha}^M \frac{Z_{\alpha}}{|\mathbf{r} - \mathbf{R}_{\alpha}|} \quad (2.2)$$

where \mathbf{R}_α stands for the location and Z_α for the charge of the fixed nuclei. Since the external potential operators are simple multiplicative one-electron operators, the energy coming from this term can be calculated based on the electron density:

$$E_{ext} = \int V_{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r} \quad (2.3)$$

Because of the electron-electron interaction term the solution of this Hamiltonian becomes impossible for larger systems, so approximations have to be made.

2.1.2 Hohenberg-Kohn theorems

The Hohenberg-Kohn theorems[18] serve as the foundation of density functional theory. Both theorems can be derived from the variational principle of wavefunctions, which shows that given the Hamiltonian of a system, the energy of any trial wavefunction is larger than or equal to the ground state energy.

The first theorem states that the total energy, the wavefunction, the external potential and the Hamiltonian of a system are all defined by its ground state electron density. Introducing an energy functional with a fixed external potential:

$$E[n] = I[n] + \int n(\mathbf{r})V_{ext}(\mathbf{r})d^3\mathbf{r}, \quad I[n] = \min_{\Psi \rightarrow n} \langle \Psi | -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \Psi \rangle \quad (2.4)$$

the second theorem proves

$$E[n_0] = E_{GS} < E[n_1] \text{ for every } n_1 \neq n_0 \quad (2.5)$$

where n_0 is the ground state electron density belonging to the fixed V_{ext} .

2.1.3 Kohn-Sham equations

In an effort to utilize the results of the Hohenber-Kohn theorems Walter Kohn and Lu Jeu Sham proposed[21] an approach to calculate the ground state electron density with a fixed

external potential. The problem is mapped to a system of non-interacting electrons in a fictitious potential, for which the total energy is partitioned in four parts:

$$E[n] = T_s[n] + E_H[n] + E_{xc}[n] + E_{ext}[n] \quad (2.6)$$

The wavefunction of this fictitious system is a Slater determinant formed from the lowest energy solutions of its Hamiltonian. The density of the system is simply:

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 \quad (2.7)$$

The kinetic energy of the noninteracting particles is:

$$T_s[n] = \sum_i \langle \psi_i | -\frac{1}{2} \nabla^2 | \psi_i \rangle \quad (2.8)$$

The classical Hartree energy of the electrons is:

$$E_H[n] = \frac{1}{2} \int \int \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (2.9)$$

And the E_{ext} term for the interaction with the external potential was described in Eq. 2.3. The $E_{xc}[n]$ part is called the exchange-correlation energy, which includes all the necessary corrections emerging because of the noninteracting particle model and the classical Hartree-energy. Finding the ground state of this system can be done by minimizing the total energy with respect to the single particle wavefunctions.

$$\frac{\delta E[n]}{\delta \psi_i^*(\mathbf{r})} = -\frac{1}{2} \nabla^2 \psi_i(\mathbf{r}) + \left[\int \frac{n(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' + \frac{E_{xc}[n]}{\delta n(\mathbf{r})} + V_{ext}(\mathbf{r}) \right] \frac{\delta n(\mathbf{r})}{\delta \psi_i^*(\mathbf{r})} \quad (2.10)$$

Adding the constraint of orthonormal wavefunctions using Lagrange multipliers results in:

$$\frac{\delta}{\delta \psi_i^*} \left[E[n] + \sum_i \sum_j \epsilon_{ij} \left(\int \psi_i^* \psi_j d\mathbf{r} - \delta_{ij} \right) \right] = 0 \quad (2.11)$$

where ϵ_{ij} are the Lagrange multipliers associated with the orthonormal condition of ψ_i and ψ_j and δ_{ij} is the Kronecker delta. With unitary transformations the wavefunctions can be transformed in another basis, where the minimization takes the following Schrödinger-equation like form:

$$-\frac{1}{2}\nabla^2\psi_i(\mathbf{r}) + \left[\int \frac{n(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' + \frac{\delta E_{xc}[n](\mathbf{r})}{\delta n(\mathbf{r})} + V_{ext}(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \quad (2.12)$$

the $\frac{\delta E_{xc}}{\delta n}$ part is usually referred as the exchange-correlation potential. Following these steps the original problem has been reduced to a noninteracting electron problem, which can be tackled efficiently even for large structures. The only approximation in these steps was the exchange-correlation functional. Even though with the exact exchange-correlation functional this approach would result in perfectly accurate densities and energies, the form of this functional is unknown, and approximations have to be made.

2.1.4 Exchange-correlation functionals

The only theoretical limit on the accuracy of DFT calculations is the accuracy of the used XC functional. Since the birth of DFT numerous functionals were developed. These functionals are usually grouped based on their complexity and computational cost on different levels of the so called Jacob's ladder.[34]

The first few steps of the ladder are the local and semi-local functionals, since they can be written in the form of $\int \epsilon_{xc}(\mathbf{r})n(\mathbf{r})d^3\mathbf{r}$. Generally the exchange and correlation parts are approximated separately in the form of $\int (\epsilon_x(\mathbf{r}) + \epsilon_c(\mathbf{r}))n(\mathbf{r})d^3\mathbf{r}$.

Local density approximation

The simplest form of XC functionals are the local density approximations (LDAs). These functionals rely only on the local density and are often derived from the homogeneous

electron gas (HEG) model. The exchange energy density of HEG can be calculated exactly:[21]

$$\epsilon_x^{LDA}(n) = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} n^{1/3} \quad (2.13)$$

while for the correlation energy parametrizations[36, 55] of quantum Monte Carlo calculations are available.

Generalized gradient approximations

It is possible that a system contains multiple regions with the same electron density, but in very different chemical environments. LDA functionals are not able to differentiate between these regions, ultimately restricting their accuracy. The gradients of the density also carry important information, so including these in the functional is beneficial.[2] Functionals relying not only on the density, but also its gradient are the family of generalized gradient approximations (GGA).[3, 35] The gradient of the density is not used directly in the construction of these functionals, but its reduced version $\nabla n/n^{4/3}$. In the present treatment we will use the reduced gradient normalized to the scale of the local Fermi wave length.

$$p = s^2 = \frac{|\nabla n|^2}{4(3\pi^2)^{2/3}n^{8/3}} \quad (2.14)$$

In practice exchange functionals are usually defined based on the LDA exchange:

$$\epsilon_x^{GGA}(\mathbf{r}) = \epsilon_x^{LDA}(\mathbf{r})F_x(p), \quad (2.15)$$

where the F_x is called the enhancement factor. One of the most widely used GGA functional is the PBE[35] functional defined as:

$$F_x^{PBE}(p) = 1 + \kappa - \frac{\kappa}{(1 + \mu p/\kappa)}, \quad (2.16)$$

where $\mu = 0.21951$ and $\kappa = 0.804$ are constants set to satisfy constraints known for the exact functional. At $p = 0$ PBE starts at $F_x = 1$ to reproduce the LDA results, then the slope

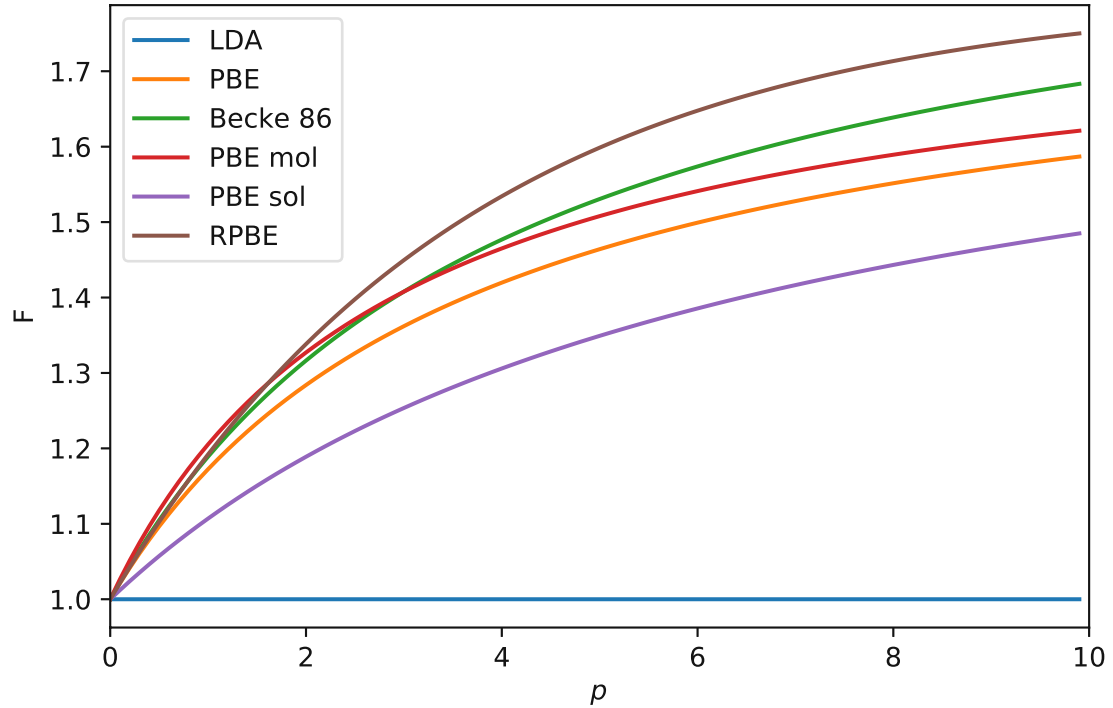


FIGURE 2.1: Enhancement factors of various exchange functionals[2, 9, 16, 21, 35, 39]

is determined based on the linear response function of the homogeneous electron gas[8], finally at $p \rightarrow \infty$ the enhancement factor converges to $F_x = 1.804$ to satisfy the Lieb-Oxford bound.[24] GGA functionals having multiple parameters can also be fine-tuned for specific groups of materials, like molecules[9] or solids [39] or a specific property, like adsorption energies.[16] The enhancement factors of these GGAs are shown on Fig. 2.1, which also shows the general shape of most GGA functionals.

Meta-generalized gradient approximations

Following the logic of extending the information available to the functional, meta-generalized gradient approximations can use also the kinetic energy density (KED) and/or the laplacian of the density as well. Just like in the case of GGAs usually in the formulation of mGGA functionals also the normalized/reduced values are used. Based on the Kohn-Sham kinetic energy density $\tau^{\text{KS}} = (1/2) \sum_{i=1}^N |\nabla \psi_i|^2$ and the kinetic energy density of the HEG [13, 48] $\tau^{\text{TF}} = (3/10)(3\pi^2)^{2/3} n^{5/3}$ the normalized kinetic energy density is:

$$t = \frac{\tau^{\text{KS}}}{\tau^{\text{TF}}} \quad (2.17)$$

In iso-orbital regions where the density is dominated by one or two orbitals of the same shape, the KED is given exactly by the von Weizsäcker limit: [56]

$$\tau^{\text{vW}} = \frac{1}{8} \frac{|\nabla n|^2}{n} \quad (2.18)$$

which leads to another useful descriptor

$$\alpha = \frac{\tau^{\text{KS}} - \tau^{\text{vW}}}{\tau^{\text{TF}}} = t - \frac{5p}{3} \quad (2.19)$$

In iso-orbital regions $\alpha = 0$ [4] and in regions with slowly varying density $\alpha \approx 1$. [45] Low α values can also be an indicator of covalent bonds, while interlayer regions of graphite has been shown to have large values. [27]

An example of how the α dependence can be incorporated in mGGA functionals is shown on the left panel of Fig. 2.2 through the enhancement map of SCAN [44]. Compared to PBE an important difference is that the derivative of the enhancement factor w.r.t. p is negative at high p values, and the SCAN enhancement factor also decreases with increasing α , which is not possible for a GGA. On the right panel the enhancement factor curves are shown for SCAN, and two other successful functionals, the mBEEF [57] and the TASK [1] for low and high α values. These functionals not only differ in their $\alpha = 1$ curves, but the way these curves change when α shifts from 1 to 10. For both

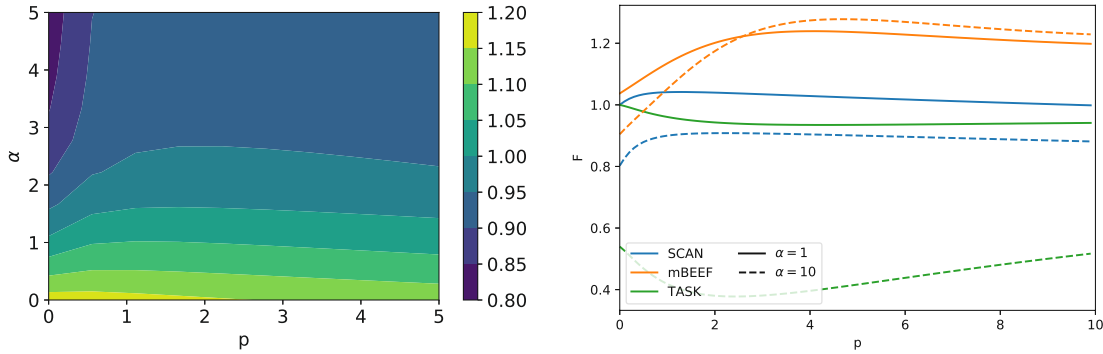


FIGURE 2.2: Enhancement factor of the SCAN functional (left panel) and comparison of three mGGA functionals on small and high α values (right panel).

SCAN and TASK the enhancement factors decrease with increasing α , while mBEEF does not show this kind of behaviour. These three examples already show how versatile the mGGA functionals can be, hinting that they might be able to fit the exact functional more closely resulting in better accuracy for multiple properties.

Hybrid functionals

Solving Eq. 2.12 results in the electronic orbitals which reproduce the real densities. These orbitals then can be used to calculate the exchange energy of this hypothetical system exactly:

$$E_x = -\frac{1}{2} \sum_{ij}^{occ} \delta_{\sigma,\sigma'} \int \int \frac{\psi_{i,\sigma}^*(\mathbf{r}) \psi_{j,\sigma'}^*(\mathbf{r}') \psi_{i,\sigma}(\mathbf{r}') \psi_{j,\sigma}(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \quad (2.20)$$

The idea of hybrid functionals[5] is to mix this exchange with approximations from the previously mentioned levels.

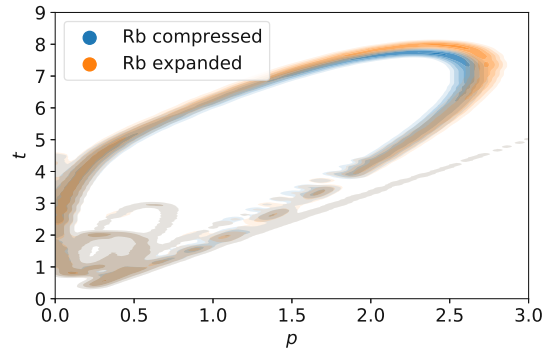


FIGURE 2.3: $p - t$ descriptor map of Rb with a 1% compressed and 1% expanded structure compared to experimental lattice parameter

Success of GGAs and mGGAs

One of the biggest drawbacks of the LDA approximation is the overbinding of solids, which is reflected in the -1.5% mean relative error of lattice parameters in the benchmark of 44 materials.

The GGAs can correct this overbinding, using the normalized density gradient, p , defined in Eq. 2.14. When expanding the unit cell of a solid, the interstitial regions tend to have larger p values compared to the compressed case, an example of this behaviour for Rb is shown Fig. 2.3.

One of the most widely used GGA functionals is the PBE, defined as Eq. 2.16. The PBE enhancement factor as shown on Fig. 2.1 is strictly increasing with respect to p , thus the regions in the expanded structure with larger p values experience higher enhancement factors, reducing the total energy of the stretched structures and shifting the calculated lattice parameter towards larger values.

This behaviour is very well represented by the functionals shown on Fig. 2.1. The mean relative lattice parameter errors of the LDA, PBEsol, PBE and RPBE on the 44 solids set[50] are -1.5%, -0.1%, 1.1% and 2.4% respectively, showing that functionals with larger enhancements predict more expanded structures.

mGGA functionals extending the available information with the kinetic energy density or the laplacian of the density are able to better distinguish between different regions, thus they are able to predict multiple properties more accurately.

Theoretical constraints on exchange functionals

While the exact form of the exchange functional from Eq. 2.6 is unknown some constraints can be derived for it, which helps the development of approximations. It is important to note that these constraints are constraints on the exact functional and the approximations are not obliged to satisfy them. The SCAN functional satisfies most known "constraints" an mGGA level functional can:

- **Negativity:** The exchange energy is negative for every electron density. A sufficient, but not necessary condition for the enhancement factors to satisfy this is: $F > 0$
- **Spin-scaling:**[32] The exchange energy of a spin polarized system can be calculated as:

$$E_x[n_\uparrow, n_\downarrow] = \frac{1}{2}E_x[2n_\uparrow] + \frac{1}{2}E_x[2n_\downarrow] \quad (2.21)$$

This equality does not put any constraint on the enhancement factor.

- **Uniform density scaling:**[23] Introducing a scaled density as:

$$n_\lambda(\mathbf{r}) = \lambda^3 n(\lambda\mathbf{r}) \quad (2.22)$$

The exchange energy of the scaled density scales with λ :

$$E_x[n_\lambda] = \lambda E_x[n] \quad (2.23)$$

Using the previously introduced dimensionless reduced variables p and t to calculate the exchange energy in the form of $E_x[n] = \int e_x^{LDA}(\mathbf{r})n(\mathbf{r})F(p(\mathbf{r}), t(\mathbf{r}))d\mathbf{r}$ automatically satisfies this constraint.

- **Fourth order gradient expansion:**[46] The enhancement factor in a slowly varying density ($s \ll 1$ and $\alpha \approx 1$) can be approximated:

$$F_x \approx 1 + \frac{10}{81}s^2 - \frac{1606}{18255}s^4 + \frac{511}{13500}s^2(1 - \alpha) + \frac{5913}{405000}(1 - \alpha)^2 \quad (2.24)$$

while this expansion sets the derivative of the enhancement factor with respect to $s^2 = p$, other approximations are also available. As an example PBE uses the linear response of the uniform electron gas[8] to set the same gradient to $\mu = 0.21951$.

- **Non-uniform density scaling:**[41] Following the example of Eq. 2.22 one can define:

$$n_\lambda^x(x, y, z) = \lambda n(\lambda x, y, z) \quad (2.25)$$

for this scaled density and its exchange energy:

$$\lim_{\lambda \rightarrow \infty} E_x[n_\lambda^x] > -\infty \quad (2.26)$$

which can be achieved by making the enhancement factor decay with $s^{-\frac{1}{2}}$ as $s \rightarrow \infty$.

- **Tight bound for two-electron densities:**[38] The lower bound of the total exchange energy of a two electron toy system[38] with density n was derived:

$$E_x[n] \geq 1.174 E_x^{LDA}[n] \quad (2.27)$$

since the mentioned system could take any constant p values in its whole region with $\alpha = 0$, the previous equation was translated into $F(p, \alpha = 0) < 1.174$.

- **Lieb-Oxford bound:**[38] A less strict version of the previous bound was derived using a reformulation of the Lieb-Oxford bound[24] in the case of spin-unpolarized electron densities, which results in:

$$F_x(p) < 1.804 \quad (2.28)$$

This bound was used to set the asymptotic limit of the PBE functional in the $p \rightarrow \infty$ case.

- **LDA limit:** The homogeneous electron gas is characterized by $p = 0$ and $\alpha = 1$ and its exchange energy density is known exactly. This sets our last constraint on the enhancement factor:

$$F(p = 0, \alpha = 1) = 1 \quad (2.29)$$

This constraint serves as the starting point of numerous functionals, such as PBE, SCAN, PW91[37], SOGGA[59], TASK[1], TM[47] and many others.

To reiterate, these constraints are set on the exact exchange functional and its enhancement factor, an approximation is not obliged to satisfy any of them. For example in the case of a GGA functional the LDA constraint forces the enhancement factor to be 1 at $p = 0$, regardless of the α value, even in regions with $\alpha \neq 1$ which are not well approximated by the HEG. Because of this for specific systems breaking any of the aforementioned limits could be beneficial for the overall accuracy, while deteriorating their performance for the systems which the limits were derived on.

2.1.5 DFT accuracy

While the accuracy of DFT calculations depends on multiple different factors, in the case of optimal settings it is only limited by the underlying XC functional. Extensive studies[7, 15, 50, 51] have been made to benchmark the vast range of functionals available nowadays. The comparison of these functionals is only meaningful when they are evaluated on the same dataset for the same properties, since some of them would excel on one property or one type of materials, but may completely fail for others. There are also functionals which perform generally well for a wide range of properties, but they can be overshadowed by specialized functionals if compared only on one property.

Notable datasets for molecules are the G2/97[10] and G3/99[11] with 302 and 376 atomization- and ionization energies, reaction barrier heights and proton- and electron affinities respectively. For solids a set[43] of 18 materials and an extended set[50] of 44

strongly bound materials exists for lattice parameters, cohesive energies and bulk moduli. Also for band gap calculations a set of 473 materials is available.[7]

As stepping up the ladder of DFT functionals better accuracy is expected, since higher steps of the ladder rely on more information. The LDA, GGA and mGGA levels with the functionals used in the benchmark of 44 solids[50] are shown on Fig. 2.4 for lattice parameter and cohesive energy. Strictly deciding that a functional is better than the other is complicated, since one functional can give better results for one property and fail for the others. The ME and MRE plots show straight lines on which the GGA and mGGA functionals lie, with line of mGGA functionals being closer to the origin showing their superior accuracy.

As Fig. 2.4 shows the overall best functionals are two mGGAs, closely followed by GGAs. The SCAN mGGA[44] has the smallest MARE for lattice parameter and with 4.9% MARE for cohesive energy it performs only slightly worse than PBEalpha with 4.1%. SCAN also has the smallest MARE for bulk modulus with 6.5%.

The same tendency of improving accuracy was reported for dipole moments[15], where the calculations were done on 152 different molecules showing a best RMSE of 13.67%, 8.85% and 7.56% on the LDA, GGA and mGGA levels. In this benchmark SCAN placed as the 2nd best functional of these three levels, only mBEEF[57] being more accurate.

Another important property of bulk materials is the band gap, which was also extensively studied[7] using 33 XC functionals of various levels (LDA, GGA, mGGA and hybrid). Not surprisingly the modified Becke-Johnson mGGA potential[49] performed the best with 30% MARE and 0.5 eV MAE on the set of 473 materials, since it was created specifically for band gaps. The interesting speciality of the mBJ potential is the inclusion of a non-local descriptor in the otherwise local potential:

$$c = \alpha + \beta \left(\frac{1}{V_{cell}} \int_{cell} \frac{|\nabla n(r)|}{n(r)} d^3r \right)^{1/2} \quad (2.30)$$

As shown on Fig. 2.5 the best semi-local functionals are the HLE17,[54] M06L,[58] SCAN and TASK(mGGA)[1], yet they still result in larger than 32% MAREs. Both the

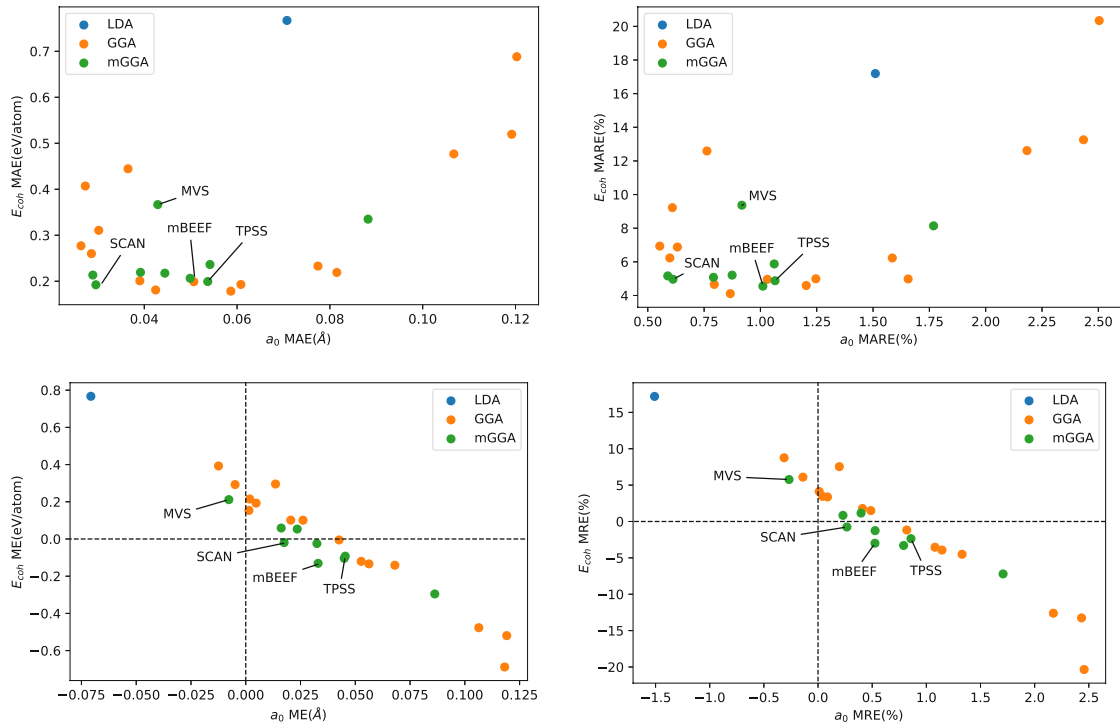


FIGURE 2.4: Lattice parameter and cohesive energy ME, MRE, MAE and MARE values for 44 solids[50] with LDA, GGA and mGGA functionals

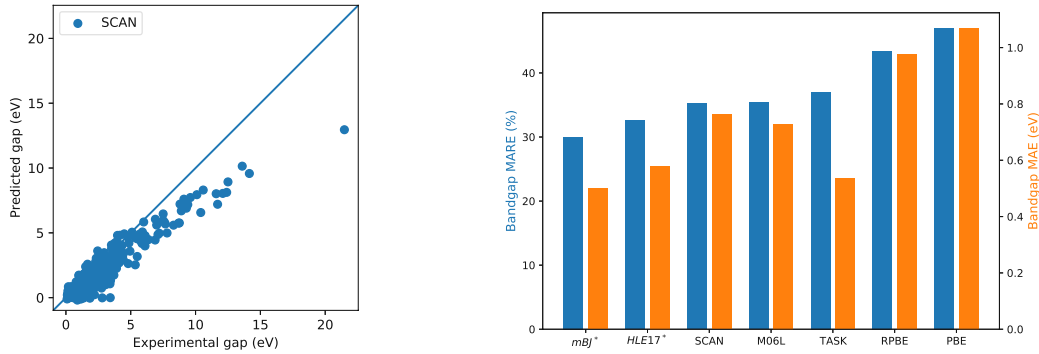


FIGURE 2.5: Predicted band gaps with the SCAN functional compared to experimental values (left panel) and mean absolute relative band gap errors with a selection of mGGA and GGA functionals (right panel). Both plots were calculated using a 440 material subset of the database from the original study[7]

M06L and SCAN underestimate the gaps by around 20% on average, and even the HLE17 produces a MRE of -11%. The TASK functional which was constructed focusing on the derivative discontinuity, contrary to the other functionals overestimates the band gaps by 14% and also results in the lowest MAE right after mBJ. On the higher level the HSE06 hybrid functional[17] performed the best, with slightly larger than 30% MARE. Also in this case the LDA and GGA level approximations resulted in worse accuracy than the more complex ones, except for the HLE16[53] GGA, which shows similar accuracy as the HLE17 mGGA.

2.2 Machine learning

Machine Learning is the study of computer algorithms that improve automatically through experience.[29] Thanks to improvements in machine learning (ML) algorithms and the increase in affordable computational power, nowadays machine learning based approaches are widely used basically everywhere from scientific research to sorting cat photos on social media. ML consists of three main branches: supervised learning, unsupervised

learning and reinforcement learning. During my work I employed various unsupervised learning methods to find and understand connections between groups of materials and also supervised approaches to predict specific properties or optimize functionals. While reinforcement learning is also a very powerful tool for various problems, in the present thesis it has not been utilized, so we will only focus on the other two branches.

2.2.1 Supervised learning

In supervised learning the training samples are labeled, meaning that for every sample input we know the expected output. The goal of the training is to create a mapping from the input(descriptor) space to output space, so that the outputs match with the labels as good as possible. This very general approach can be used on a vast range of different models:

- tables
- decision trees
- Bayesian networks
- support-vector machines
- neural networks (NN)
- ...

In our work we used two of these models, decision trees for their simplicity and explainability, and neural networks for their flexibility.

Neural networks

A neural network is a computational model loosely based on the structure of biological neural networks. It consists of nodes (neurons) and edges (synapses). Every edge is associated with a weight, and every node with an activation function. The nodes take input

from their incoming edges and generate a numeric output which can be passed to multiple other nodes. The output of a node is calculated by applying the activation function to the weighted sum of the incoming edges. According to the universal approximation theorem sufficiently large networks of these structures are able [19] to approximate any "well behaved" function on a compact region.

After fixing the architecture of a neural network, during training the weights of the edges are changed so that the output of the network is as close as possible to the expected output. To change the weights the most often used algorithm is a variant of the gradient descent algorithm. Since we use neural networks in supervised learning, for every input we know the ideal output of the network, so we can define a function describing how much the actual result differs from the expected one. This function is called a loss function, which is the subject of our gradient descent minimization process.

During training, the samples of the training set are shown to the network and the weights are updated the following way:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha \frac{\delta L(\mathbf{x})}{\delta \mathbf{w}} \quad (2.31)$$

where \mathbf{x} is the input vector, \mathbf{w}_t the list of weights at a given step t , L the loss function and α the learning rate. The input samples are shown to the network and the weights are updated until convergence or other stopping criterion is reached.

Decision trees

Just like other models in supervised learning, decision trees are used to predict a value (number or tag) based on one or multiple inputs. They do it by iteratively splitting the training dataset in two parts, based on one of the descriptors at a time, until a maximum depth or the desired accuracy is reached. Then for evaluation a new sample is passed through this decision tree, and the last group it ends up in determines the predicted value or tag of that given sample.

2.2.2 Unsupervised learning

Unsupervised learning algorithms opposed to the supervised ones learn patterns based on unlabeled data. They are often used for clustering, anomaly detection or dimension reduction. In our work we mostly used clustering algorithms, so some of those will be introduced in detail.

k-means clustering

One of the simplest and well known clustering algorithm is the k-means clustering. It is used to split n input samples into k groups, so that every sample belongs to the cluster of the nearest centroid. The algorithm[25] iteratively updates the assigned clusters until convergence. The iteration steps are the following:

- **Initialization:** The k centroid of the clusters are chosen randomly of the n samples.
- **Assigning clusters:** Every sample is assigned to the cluster with the nearest centroid.
- **Recalculating centroids:** For every cluster a new centroid is calculated as the average position of its elements.
- **Repeat until convergence:** Steps 2-4 are repeated until the clusters do not change or other stopping criteria is met.

Because of the random initialization in step 1, the resulting clusters can change from run to run. The "goodness" of a clustering in this case can be characterized with the average distance of the samples from its cluster centers. To overcome the issue of randomness the algorithm is usually ran multiple times with different seeds and the final result is the one with the lowest average distance.

Affinity propagation

Affinity propagation[14] is another example of clustering algorithms. Unlike k-means clustering affinity propagation does not need a preset number of clusters, only a similarity

function and results not only in clusters, but also generates representatives of each cluster. The similarity function $s(i, j)$ represents how similar the samples i and j are and results in larger values for samples more similar to each other. The flow of the algorithm is the following:

- Initialize the responsibility and availability matrices as 0.
- Calculate the new responsibility matrix:

$$r'(i, j) = s(i, j) - \max_{k \neq j} [a_t(i, k) + s(i, k)] \quad (2.32)$$

- Calculate the new availability matrix:

$$a'(i, j) = \min \left(0, r_t(j, j) + \sum_{k \notin \{i, j\}} r_t(k, j) \right) \quad (2.33)$$

- Update the old matrices through mixing with the new matrices:

$$r_{t+1}(i, j) = \lambda r_t(i, j) + (1 - \lambda) r'(i, j) \quad (2.34)$$

$$a_{t+1}(i, j) = \lambda a_t(i, j) + (1 - \lambda) a'(i, j) \quad (2.35)$$

- Repeat steps 2-4 until convergence or a fixed number of steps.
- Materials for which $r(i, i) + a(i, i) > 0$ are chosen as exemplars. The number of formed clusters is the same as the number of exemplars.

Hierarchical clustering

Hierarchical clustering is a method to cluster samples based on some kind of similarity of distance. The difference compared to the previously mentioned algorithms is that this method does not only create one clustering, but a list of different level of clusterings from every sample having its own cluster to every sample belonging in the same cluster. These

lists are most commonly represented as dendograms. The algorithm for the greedy agglomerative approach:

- All samples start as their own cluster.
- A linkage criteria is calculated between all clusters.
- The two clusters with the lowest linkage criteria between each other are merged in one.
- Repeat steps 2-3 until only one cluster remains.

Based on the type of samples one wants to cluster multiple linkage functions are available.

3 Exchange functional training

Since the first LDA approximation the number of exchange-correlation functionals grew rapidly with more than 400 functionals existing today in the LibXC library[22]. In the development of these functionals two general approaches took place, both with comparable success. Two of the already mentioned functionals, SCAN and PBE were developed to obey many of the constraints known to be true for the exact functional, some of which were listed in section 2.1.4. This approach is very attractive from an ab initio point of view, since it relies on exact theoretical results, even though there are still multiple functional forms, which can satisfy the constraints. Since these constraints are universal, functionals created this way are expected to be transferable to various systems. The other approach is to fit a flexible functional[40, 57, 58] to reproduce experimental results or results calculated with a higher level method. In recent years replacing the traditional functional forms neural networks also have been employed[12, 20, 31] as XC functional approximations. These functionals can strongly depend on the database they were trained on and might perform worse on systems far from the training data or for properties not included in the training. On the other hand the previously mentioned constraints are only constraints of the exact functional, and it is possible that a semi-local approximation perform better without them.

As shown on Fig. 2.4 mGGAs have the ability to be more accurate for lattice parameters and cohesive energies than GGA functionals, yet still exhibit the tradeoff between the accuracy for these two properties. As shown in Fig. 2.5 mGGAs furthermore showed a systematic improvement over GGAs w.r.t. band gaps. However, the state of the art functionals were still not able to fully match the mBJ potential in accuracy. It is therefore interesting that Fig 2.2 showed how mGGA functionals with quite distinct functional form

and "construction philosophies" can lead to rather similar results for the lattice parameters and cohesive energy. In this regard, it is an open question whether the extra flexibility of mGGA over GGA functionals coming from the normalized kinetic energy descriptor can be utilized to predict more accurate band gaps and to what degree this results in a deterioration of the lattice parameter and cohesive energy errors.

In this chapter, I present the functionals obtained with the optimization of a flexible mGGA functional form using 44 solids for lattice parameter and cohesive energy predictions and more than 400 materials for band gap calculations as reference. Multiple functionals were trained with different relative weight on the importance of these properties. The trained functionals are used to analyze the tradeoff in detail and to draw the surface of achievable accuracy in the lattice parameter - cohesive energy - band gap error space

3.1 Material database

As the idea of "garbage in garbage out" is clearly true in the case of most machine learning methods, the first important step in an exchange functional training is to find or generate an accurate and big enough dataset. In our case these datasets were available for lattice parameters, cohesive energies, bulk moduli and band gaps by extracting them from previous WIEN2k[6] calculations. The dataset consists of four converged calculations (three on different volumes around the experimental lattice parameter and one for the atomic cases) for materials used for the lattice parameter and cohesive energy prediction and three converged calculations (one for a neutral case and two for the ± 1 electron cases) for the band gap calculations. The data about a converged run contains the densities, density gradients and kinetic energy densities on a dense grid of a unit cell. Since the goal is to train an exchange functional for every converged run the other components of the total energy, namely $T_s[n]$, $E_H[n]$, $E_{ext}[n]$ and E_c in Eq. 2.6 are also stored.

The set used for band gap calculations contains 440 materials, a subset of the materials used in the benchmark[7] mentioned in the introduction. The other set for lattice parameter and cohesive energy calculations contain the same 44 materials as the benchmark[50]

producing Fig. 2.4.

Since during the training only the exchange energy is reevaluated and our approximation follows the idea established in Eq. 2.15 the size of the dataset can be significantly reduced. This reduction can be done by binning sample points with similar p - t values together and replacing them with a single point using their averaged p - t values and the sum of their LDA exchange energy. If the bins are chosen too small, there would be no reduction in the database size, while too large bins can deteriorate the accuracy. In our case the bin size was chosen to be 0.005 in both the p and t directions, which reduced the memory and training time requirements by a factor of 10, while still keeping a good accuracy.

3.2 Calculation of different properties

The calculation of the lattice parameter, cohesive energy and band gap all rely on total energy differences. Since all components of the total energy except the exchange part are stored in the dataset, we only have to care about the latter, which is calculated as

$$E_x = \sum_i E_x^{LDA}(i) F(p_i, t_i) \quad (3.1)$$

where i runs through all the previously defined bins of the specific material, p_i and t_i are the averaged p and t of the bin and $E_x^{LDA}(i)$ is the sum of the LDA exchange energy in bin i . After calculating the exchange part it is added to the rest of the other components to obtain the total energy.

For the lattice parameter calculation the three runs with different volumes are used. The total energies for all volumes are calculated and the minimum of the fitted parabola is used to estimate the lattice parameter and the total energy on this optimal value. We checked that using three volumes is accurate enough for our purposes, therefore using more points would have increased the training time with no significant benefit.

Since the data for the atomic cases are also available the calculation of the cohesive energy is easily carried out by subtracting the previously calculated energy at the optimal

lattice parameter from the sum of the total energy of the constituent atoms. This calculation is also affected by the accuracy of the lattice parameter prediction, but the differences are not significant here either.

In density functional theory, the difference between the ionization energy and electron affinity, the fundamental band gap, is often approximated with the energy difference of the highest occupied and lowest unoccupied orbitals, the Kohn-Sham gap.

$$E_g = I(N) - A(N) = [E_{tot}(N-1) - E_{tot}(N)] - [E_{tot}(N) - E_{tot}(N+1)] \quad (3.2)$$

$$\approx E_g^{KS} = \epsilon_{LUMO}(N) - \epsilon_{HOMO}(N)$$

Yet in the case of the Kohn-Sham framework it was shown that these two quantities differ by the so called derivative discontinuity,

$$E_g = E_g^{KS} + \Delta_{xc} \quad (3.3)$$

which arises from the fact, that the exchange-correlation potential exhibits jumps at integer electron numbers, and finally results in an underestimation of the fundamental gap.

Kinetic energy based exchange mGGA functionals, thanks to their non-multiplicative exchange potential, can result in nonzero derivative discontinuity, thus they have the possibility to lead to larger, and therefore more accurate, band gaps than GGA and LDA functionals.

As it has been shown previously[51] using Eq. 3.2 without approximation can be done for periodic solids relatively easily. The total energy of the system consisting of N_k unit cells, N_k being the number of k-points used in the calculation, is evaluated with the density $\rho(N-1) = \rho(N) - 1/N_k |\psi_{HOMO}|^2$ and kinetic energy density $t(N-1) = t(N) - 1/N_k \nabla \psi_{HOMO}^* \nabla \psi_{HOMO}$ for the $N-1$ case. For the $N+1$ case the same process is done by adding the LUMO orbital. In the case of real solids where $N_k \rightarrow \infty$ adding or removing one electron does not influence the orbitals, thus all three energies can be evaluated using the same orbitals calculated for the neutral system".

3.3 Details of training

This chapter presents the methods and approximations used in the training.

3.3.1 Approximation form

After initial attempts with a densely connected neural network to describe the form of the trained functional, we opted for a less complex solution, the Padé-approximant.[33] This approximation in 1D takes the form of:

$$A(x) = \frac{a_0 + a_1x + a_2x^2 + \dots}{b_0 + b_1x + b_2x^2 + \dots} \quad (3.4)$$

which we extended for the 2D descriptor space for mGGA approximations as:

$$F(p, t) = \frac{c_t + a_t p + b_t t + a_{2t} p^2 + x_t p t + b_{2t} t^2}{c_b + a_b p + b_b t + a_{2b} p^2 + x_b p t + b_{2b} t^2} \quad (3.5)$$

The advantage of this approximation, is that it is expected to produce smooth derivatives in the relevant region of the phase space and contains considerably less parameters than a general neural network, so less prone to overfitting. This functional form is also flexible enough to reproduce the PBE functional presented in Eq. 2.16 exactly, with $c_t = 1$, $a_t = \mu + \mu/\kappa$, $c_b = 1$ and $a_b = \mu/\kappa$ while all other coefficients are zero. For more complicated functionals when the exact representation is not possible, the Padé form can still result in good approximations. On Fig. 3.1 the SCAN functional and the 2nd order Padé approximant fitted to reproduce the SCAN enhancement factors in the $\alpha, p \in [0, 5]$ region with minimizing the squared errors are shown, along with the relative error of the approximation for every point. While the approximation has less than 4% error for every point of the relevant phase space area the calculated lattice parameter, cohesive energy and band gap errors still differ significantly from the SCAN results.

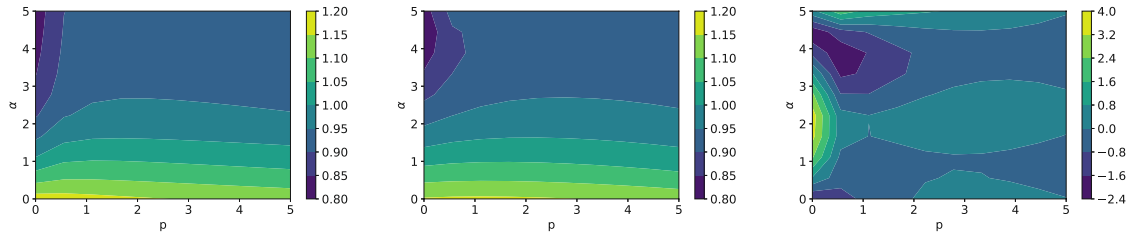


FIGURE 3.1: SCAN and its Padé approximation enhancement factors on left and middle. The relative error of the approximation on the right.

3.3.2 Loss function

As mentioned in the introduction deciding which functional is better is not straightforward, so defining the loss function is also a non-trivial task. The most common error metrics are the mean absolute error(MAE) and mean absolute relative error(MARE), but other, more complex descriptions like the deviation from a specific density distribution calculated with a higher level approximation are also possible.[31] In our case the gradient descend training method restricts the loss function only to differentiable functions with respect to the parameters.

To sample the whole space of possible functionals we described our loss as a mixture of lattice parameter, cohesive energy and band gap errors, and trained a functional for multiple different mixing of these errors.

$$\mathcal{L} = \sum_{prop}^{[a_0, E_{coh}, gap]} w_{prop} MARE_{prop} \quad (3.6)$$

3.3.3 Training steps

The training was carried out on a 2nd order Padé approximation as shown in Eq. 3.5 for 25 different mixing of errors in the loss function using the SCAN correlation energies. The starting parameters of the first functional were $c_t = c_b = 1$, $a_t = 0.05$ and everything else set to 0. Setting the a_t coefficient helped to avoid divergences in the early stages of the

training. All following functionals used the result of the previous training as the starting configuration. The training was considered finished for a specific error mixing, when the loss did not decrease for 40 consecutive epochs.

To investigate the effect of the LDA limit two constraining approaches were employed. The first was to enforce the LDA limit only approximately by adding the $\alpha(F(0,1) - 1)^2$ term to the loss function and continue training the functional, where α can be used to tune the importance of this term. Functionals trained this way with $\alpha = 5$ are the softly constrained functionals. The second approach was to set $F(p = 0, t = 1)$ exactly to 1, by fixing the c_b as $c_b = c_t + b_t + b_{2t} - b_b - b_{2b}$ and continue training in a similar fashion.

During the training the whole database is used, so no test or validation sets are available. This ensures that the resulting functionals are as good as possible for the given dataset at the cost of generalization and transferability. Since our goal was to explore the limits of the functionals and not necessarily create a new general purpose one, the overfitting caused by this approach is actually beneficial for us. It is also important to note that our functional form has only 12 free parameters and the training set contains 528 experimental values to fit for so the effect of overfitting is not expected to be strong.

3.4 Results

The results of the previously described training process are shown on Fig. 3.2 and Fig. 3.3. As seen on Fig. 3.3 the mGGA functionals outperform the GGAs in almost every case. GGAs only show similar accuracy in the 0.8-1.1% lattice parameter MARE region on the left panel of Fig. 3.3, but even for those functionals the mGGAs results in better band gap errors. This confirms that the kinetic energy carries useful information for functionals.

On the right panel of Fig. 3.2 only the mGGA functionals are shown with no/soft/exact LDA constraint. The fact that all three set of points lie on the same surface shows that even when the LDA limit was not enforced the trained functionals did not severely violate this limit. Since obeying this limit had no considerable effect on the accuracy, in the following only the exact LDA constrained mGGA functionals are considered.

The tradeoff between the accuracy of lattice parameters and cohesive energies shown on Fig. 2.4 can also be seen on the right panel of Fig. 3.2. The points with small cohesive energy errors have large lattice parameter error and the other way around. The same effect can be seen between the band gap and cohesive energy. The relationship of the lattice parameter and band gap errors is harder to see, since most of the trained functionals are at least partially trained on cohesive energy errors as well. Including the cohesive energy in the training even with a small weight deteriorates the accuracy for the other two properties, so for most of the functionals the lattice parameter errors are correlated with the band gap errors, yet this correlation is only caused by the inclusion of cohesive energy in the loss function. The true tradeoff between these two properties can only be seen for functionals not trained on cohesive energy at all.

In the following the mGGA results will be presented in more detail, along with a more accurate description of the tradeoffs between the three properties.

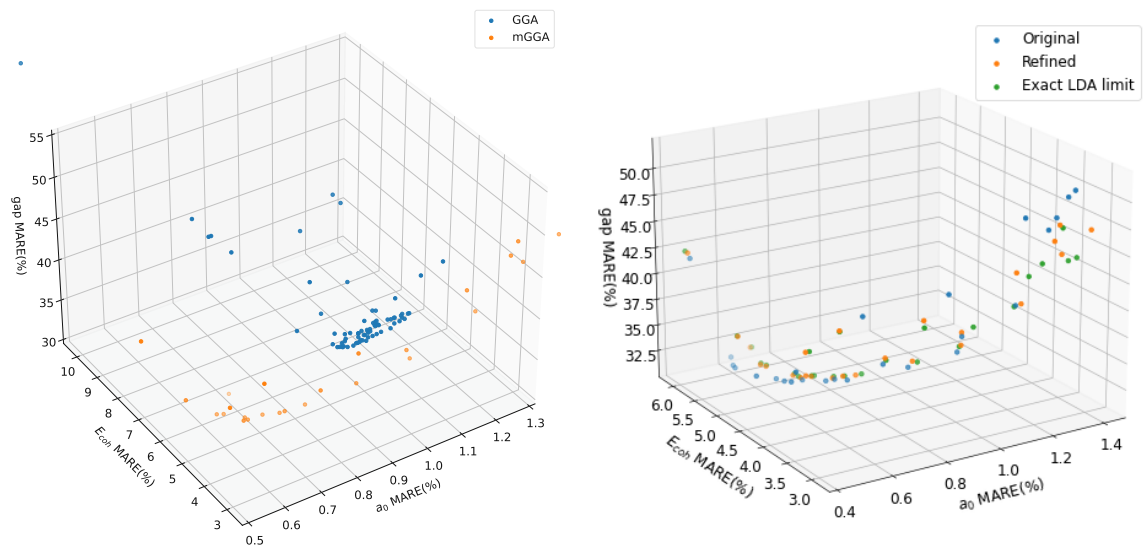


FIGURE 3.2: Lattice parameter, cohesive energy and band gap MAREs of the trained GGA and mGGA functionals (left panel) and the mGGA functionals with no/soft/exact LDA constraint (right panel).

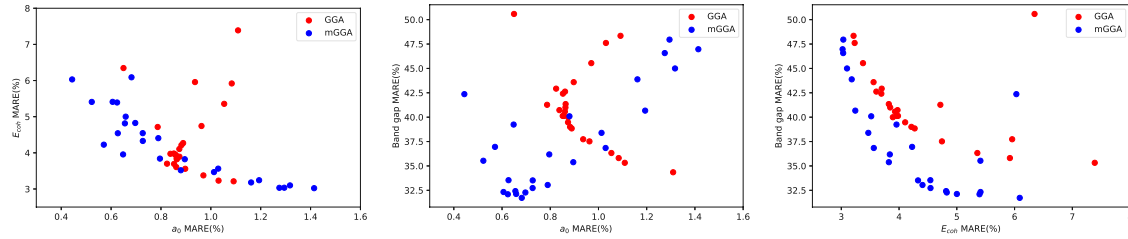


FIGURE 3.3: Comparison of GGA and mGGA functionals on lattice parameter, cohesive energy and band gap errors.

3.4.1 Lattice parameter - cohesive energy tradeoff

The tradeoff between the lattice parameter and cohesive energy is shown on Fig. 3.4. The best functional for lattice parameter results in 0.44% MARE, but at the same time with more than 6% MARE for cohesive energies. On the other side of the plot the best cohesive energy functional has 3% error, with more than 1.3% error for lattice parameter.

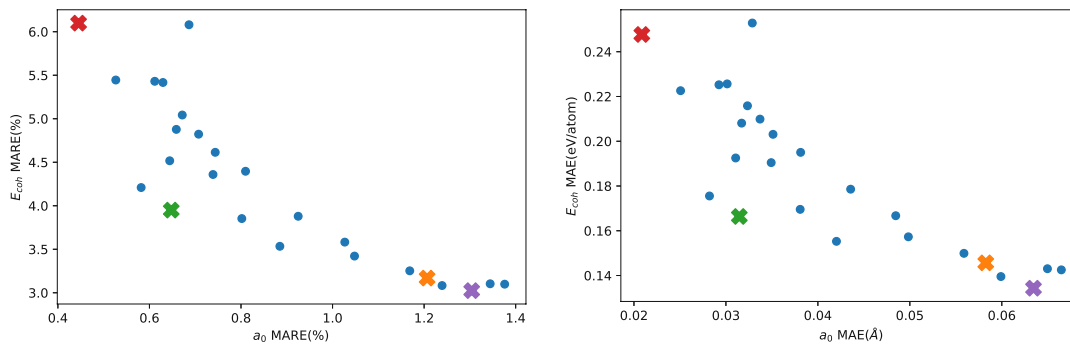


FIGURE 3.4: Lattice parameter and cohesive energy errors of the 25 trained functionals. The enhancement factor curves of the highlighted functionals are shown on Fig. 3.5.

This behaviour can be explained by comparing the enhancement factor curves of the functionals represented by colored points on Fig. 3.4. As it was shown[26] previously in

GGA functionals a steeper slope and a larger enhancement factor at large p values pushes materials to larger equilibrium lattice constants.

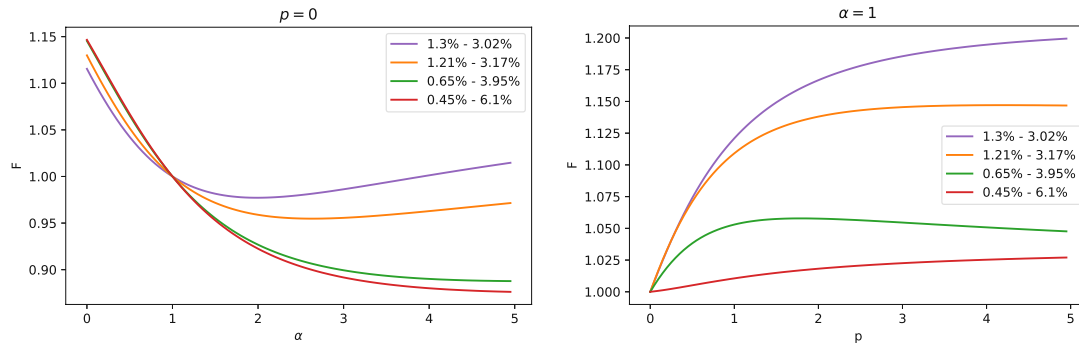


FIGURE 3.5: Enhancement factor plots of the functionals with fixed α or p trained on mostly lattice parameter and cohesive energy as the weights shifts from the cohesive energy to the lattice parameter. The legends contain the cohesive energy MARE and lattice parameter MARE for that specific functional.

For most of the materials the functional weighted towards cohesive energy, purple curve on Fig. 3.5, overestimated the lattice parameter, but for the alkali- and alkaline-earth metals there is an underestimation. The orange functional with the overall slightly decreased enhancement factors partially fixes the overestimations, but the red and green functionals with the drastically reduced enhancements also manage to correct the overestimated lattice parameters. The $\alpha = 1$ purple curve with the large initial slope and reaching a plateau around $F = 1.2$ shows some similarity with the mBEEF functional, shown on Fig. 2.2, which performs very well for cohesive energies. On the other end the green functional with a smaller initial slope with respect to p , a negative gradient on high p values and relatively large negative $\frac{\delta F}{\delta \alpha}$ values resembles the SCAN functional, also presented on Fig. 2.2, which was shown to give accurate lattice parameter results.

3.4.2 Band gap - cohesive energy tradeoff

As shown on Fig. 3.6, the relationship between the band gap MAREs and cohesive energy MAREs is very similar to the one between the lattice parameter and cohesive energy.

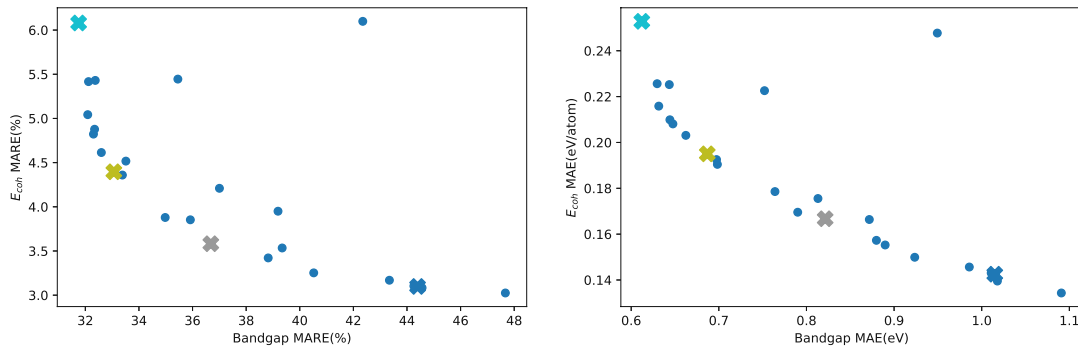


FIGURE 3.6: Band gap and cohesive energy errors of the 25 trained functionals. The enhancement factor curves of the highlighted functionals are shown on Fig. 3.7.

While the explanation of this behaviour is more complicated than the previous one, one can expect the enhancement factor curves to show some similarity based on the similarity of the error curves.

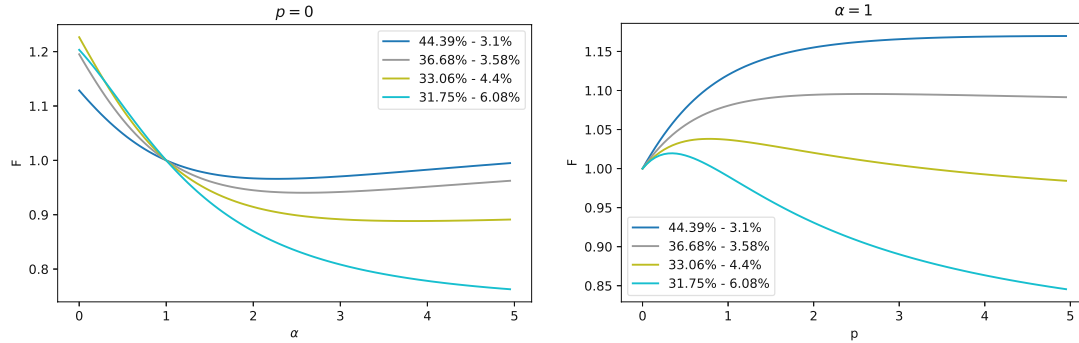


FIGURE 3.7: Enhancement factor plots of the functionals with fixed α or p trained on mostly cohesive energy and band gaps as the weights shift from the cohesive energy to the band gap. The legends contain the cohesive energy MARE and band gap MARE for that specific functional.

Just like in the previous case, reducing the importance of the cohesive energy errors results in generally smaller enhancement factors and smaller $\frac{\delta F}{\delta p}$ and $\frac{\delta F}{\delta \alpha}$ derivatives. But in the case of band gaps these effects are stronger. While the best functional for lattice parameters changes from $F(0,0) = 1.15$ to $F(0,5) = 0.87$ the best one for band gaps changes from $F(0,0) = 1.2$ to $F(0,5) = 0.76$, also in the fix $\alpha = 1$ case the best band gap functional takes significantly smaller than 1 values. This kind of behaviour is characteristic for the TASK functional, which is one of the most accurate functional for band gaps. In the case of TASK the strong negative derivative w.r.t. α was an intentional choice to obtain a large derivative discontinuity, and arriving to a similar result in a purely data driven approach also confirms this idea.

3.4.3 Band gap - lattice parameter tradeoff

When compared to cohesive energy, the band gap and lattice parameter errors show anticorrelation. Above we showed that functionals with larger cohesive energy error tend to do better for the other properties. The case of lattice parameter - band gap errors is more complex as shown on Fig. 3.8; the functionals on the right side of both plots are

mostly trained on cohesive energies, and as the weight decreases both the gap and lattice parameter errors improve up to some point. Weighting the lattice parameter even more in the training causes not only the cohesive energy, but the band gap errors to deteriorate as well.

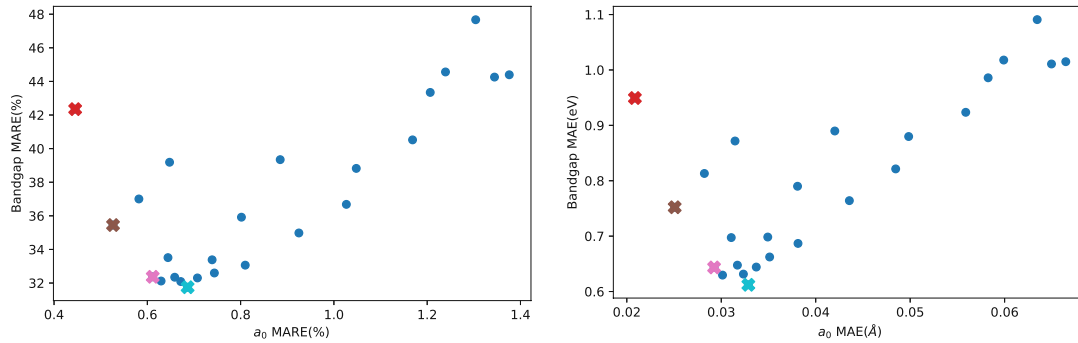


FIGURE 3.8: Band gap and lattice parameter errors of the 25 trained functionals. The enhancement factor curves of the highlighted functionals are shown on Fig. 3.9.

For the analysis of the tradeoff, the functionals which were trained only on lattice parameters and band gaps are shown in detail on Fig. 3.9, since the inclusion of cohesive energy worsens the accuracy for both these properties and its effect was already analyzed in previous two sections. Shifting the focus of the training from the band gaps to the lattice parameter, from the light blue functional to the red one, the overall enhancement factor increases, but still remains at much lower values than it was seen in the case purple functional which produced the best cohesive energies. One interesting property of the light blue functional can be found in the low p region, where a slight positive gradient of the enhancement factor can be observed w.r.t. p .

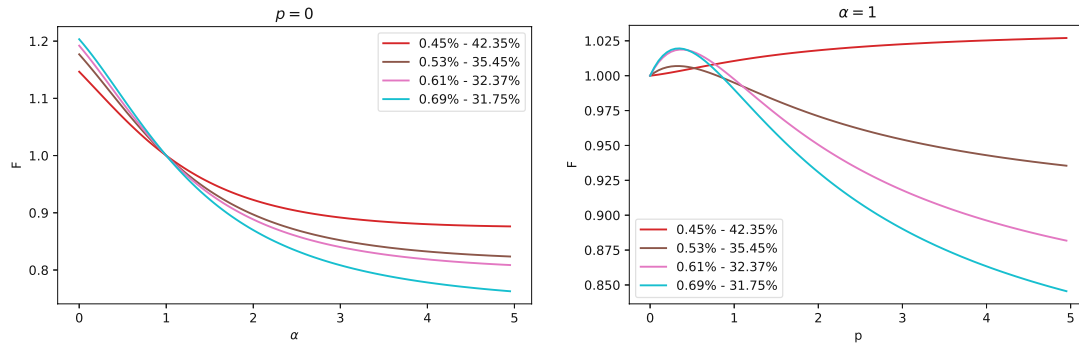


FIGURE 3.9: Enhancement factor plots of the functionals with fixed α or p trained on mostly lattice parameter and band gaps as the weights shifts from the lattice parameter to the band gap. The legends contain the lattice parameter MARE and band gap MARE for that specific functional.

3.4.4 Comparison with existing functionals

The MARE and MAE values of the colored functionals, notable general purpose functionals (PBE, SCAN) and band gap specific functionals (TASK, HLE17) are shown in Table 3.1.

The trained functionals are expected to perform at least as good as any previous exchange functional. First of all, the Padé approximation form is very flexible and the same database was used for evaluation and training. Secondly, the only constraint forced on the functionals was the LDA limit and this was shown not to impact the accuracy.

Name	Lattice constant		Cohesive energy		Band gap	
	MARE(%)	MAE(Å)	MARE(%)	MAE(eV/atom)	MARE(%)	MAE(eV)
Trained	1.38	0.07	3.1	0.14	44.39	1.01
Trained	1.21	0.06	3.17	0.15	43.34	0.99
Trained	0.65	0.03	3.95	0.17	39.19	0.87
Trained	0.45	0.02	6.1	0.25	42.35	0.95
Trained	1.3	0.06	3.02	0.13	47.67	1.09
Trained	0.53	0.03	5.45	0.22	35.45	0.75
Trained	0.61	0.03	5.43	0.23	32.37	0.64
Trained	1.03	0.05	3.58	0.17	36.68	0.82
Trained	0.81	0.04	4.4	0.2	33.06	0.69
Trained	0.69	0.03	6.08	0.25	31.75	0.61
PBE	1.2	0.06	5.0	0.19	47.03	1.07
SCAN	0.6	0.03	4.9	0.19	35.22	0.76
HLE17*	2.9	0.12	14.6	0.57	31.42	0.6
TASK*	4.6	0.23	35.2	1.34	37.68	0.54

TABLE 3.1: Lattice parameter, cohesive energy and band gap errors of the colored trained functionals and four reference ones. The color coding of the trained functionals correspond to the colors on Fig. 3.5, 3.7 and 3.9. *Band gap results taken from a benchmark[7] based on 473 materials.

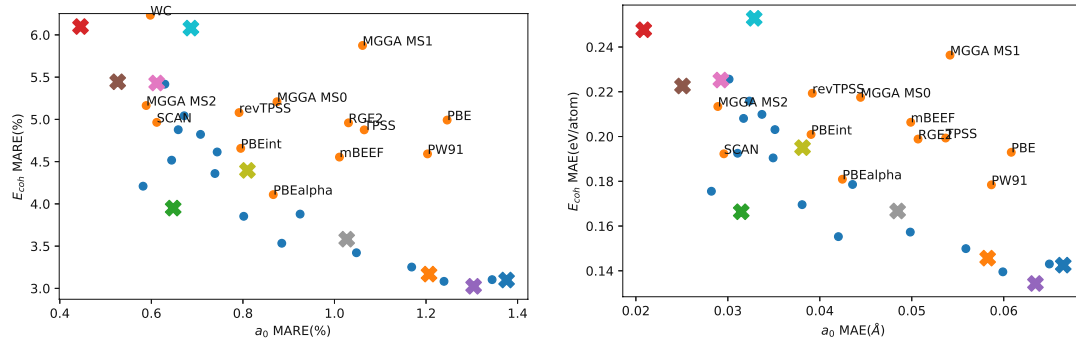


FIGURE 3.10: Lattice parameter and cohesive energy errors for some notable functionals and the 25 trained ones. The functionals from Table 3.1 are also marked.

Fig. 3.10 shows how the trained functionals perform compared to a few known functionals for lattice parameters and cohesive energies. As expected, for all the existing functionals there is at least one trained functional which is not worse in any way, meaning both the lattice parameter and cohesive energy MAREs and MAEs are not larger. If only these two properties are considered the new functionals show a significant improvement in accuracy. Even if we include the band gaps there is a new functional which performs better than the SCAN, the best general purpose functional in our tests. The new functional results in only 0.03% worse MARE for lattice parameters, but improves on SCAN by 0.45% and 1.72% in cohesive energy and band gap MAREs respectively.

The other interesting result comes from the functional trained more for the band gap, marked by light blue in Fig. 3.9. This functional performs on par with HLE17, with only 0.3% worse band gap MARE, but improves both the lattice parameter and cohesive energy errors significantly. Compared to TASK it has 0.07eV larger band gap MAE, but performs significantly better in every other error metric.

3.4.5 mGGA surface

Showing all the trained functionals on Fig. 3.2 draws a well defined surface of possible mGGAs in the 2nd order Padé approximation form on the given dataset. The fact that none of the considered already existing functionals performed strictly better than the trained ones might suggest that this surface is the limit of "well behaved" mGGAs. It's important to note that for the training of mGGAs the SCAN correlation functional was used, so in that part there might still be room for improvement. Also in the formulation of these mGGAs the Laplacian of the density was not used after early results showed no significant improvement,[52] but including it also might increase the accuracy slightly.



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4 Similarity clustering for representative sets of solids for density functional testing

The following chapter aims to extend the work presented in the "Similarity clustering for representative sets of solids for density functional testing" paper. In this paper our goal was to create a method which can be used to efficiently create small datasets of materials to represents various regions of the descriptor space of most mGGA functionals, rendering these datasets useful for functional training and low-cost functional evaluation. As a secondary goal the elements of the small set were chosen in a way that the errors calculated with the larger and smaller sets differ as little as possible.

4.1 Additional similarity metrics

As mentioned in the paper, the classical Euclidean-distance between the maps of materials is unusable in our case, since even when there are no overlapping regions this distance would depend on the exact shape of the occupied density regions.

4.1.1 Normalized dot-product

Our choice to calculate the similarities between materials was the normalized dot product, which is very similar to the correlation of the materials and is defined the following way:

$$S(A, B) = \frac{\sum_{i,j} A[i, j]B[i, j]}{N(A)N(B)} \quad (4.1)$$

with A and B being the $p - t$ maps of the materials, and i, j indexing the bins of the mappings in the two directions. The N normalization function is:

$$N(A) = \sqrt{\sum_{i,j} A[i, j]^2} \quad (4.2)$$

Having the similarity, the distance of materials is $1-S(A, B)$.

4.1.2 Normalized Euclidean distance

It is also possible to fix the Euclidean-distance to satisfy the necessary conditions with proper normalization. Treating the $p - t$ maps as vectors and renormalizing them to unit length maximizes the distance of any two non-overlapping materials to $\sqrt{2}$, regardless of their shape on the maps. The definition of this distance is:

$$d_2(A, B) = \sqrt{\sum_{i,j} \left(\frac{A[i, j]}{N(A)} - \frac{B[i, j]}{N(B)} \right)^2} \quad (4.3)$$

with the normalization factors:

$$N(A) = \sqrt{\sum_{i,j} A[i, j]^2} \quad (4.4)$$

The similarity based on this error metric is defined as:

$$S(A, B) = 1 - \frac{1}{\sqrt{2}} d_2(A, B) \quad (4.5)$$

4.1.3 Normalized Manhattan distance

Following the logic of the normalized Euclidean distance, one can also try to modify the Manhattan distance of the two maps to find a usable distance definition.

$$d_1(A, B) = \sum_{i,j} \left| \frac{A[i,j]}{N(A)} - \frac{B[i,j]}{N(B)} \right| \quad (4.6)$$

with the normalization:

$$N(A) = \sum_{i,j} |A[i,j]| \quad (4.7)$$

In this case the similarity is:

$$S(A, B) = 1 - \frac{1}{2} d_1(A, B) \quad (4.8)$$

4.1.4 Comparison of distance metrics

In the paper the similarity matrix of the 44 materials already helped to identify clusters of materials which are alike. These matrices with all three distance definitions are shown on Fig. 4.1.

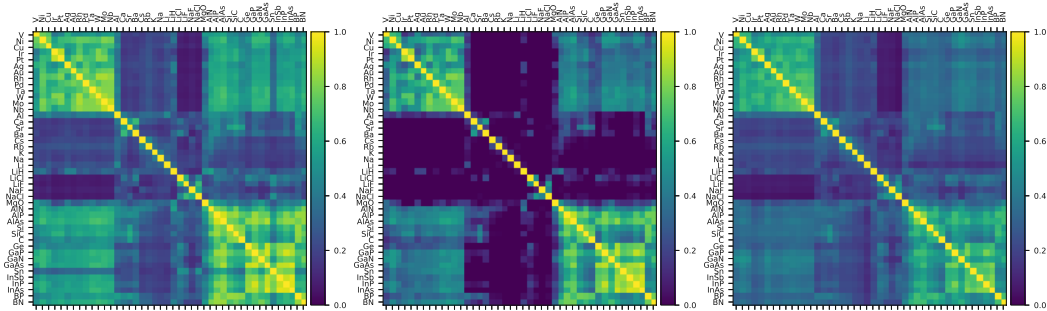


FIGURE 4.1: Similarity matrices of the 44 materials with different distance metrics. Normalized dot product on the left, normalized Euclidean distance in the middle and the normalized Manhattan distance on the right.

With all three metrics the two large groups of materials, the metals and the semiconductors, are easily identifiable in the high similarity regions. The dissimilarity of the ionic materials, the alkali- and earth alkali metals to everything else is also visible with all metrics, but in the case of the Euclidean distance it is much more pronounced. The smaller but still visible groups of the ionic materials and earth alkali metals are also noticeable on all three plots.

Doing the k-means clustering using these similarity maps results in only small differences amongst the formed clusters. One of the differences is the LiH moving between the [MgO, Al, Rb, Cs] and [Li, Na, K] clusters, where the [LiH, MgO, Al, Rb, Cs] cluster was already identified as the least stable one. The only other difference is the AlAs and GaN materials jumping between the two semiconductor clusters.

4.2 Additional clustering methods

If the results of the clustering would significantly change depending on the used clustering method, that could indicate that the formed clusters are unstable or are simply an artifact of the specific algorithm. To rule out this possibility, we carried out the clustering with affinity propagation and two types of hierarchical clustering as well.

4.2.1 Affinity propagation

The affinity propagation, described in section 2.2.2, can be directly applied to the similarity matrix, so the errors coming from the MDS can be avoided. Tweaking the preference parameter the number of clusters can be set to seven, resulting in the following grouping: [C, Si, SiC, BN, BP, AlN, AlP, AlAs, LiH, MgO, Al, V], [Ge, Sn, GaN, GaP, GaAs, InP, InAs, InSb], [LiF, LiCl, NaF, NaCl], [Li, Na, K], [Ca, Sr, Ba], [Rb, Cs], [Ni, Cu, Nb, Mo, Rh, Pd, Ag, Ta, W, Ir, Pt, Au].

The only large difference compared to the k-means result is the [LiH, MgO, Al, Rb, Cs] group splitting up, removing the [Rb, Cs] alkali metals, but the rest merging with one of the semiconductor groups. Another small difference is [V] moving from the metals to the previously mentioned semiconductor group.

4.2.2 Hierarchical clustering

Using an agglomerative clustering method, gives insight not only into the formed clusters, but also into the order and stability. Carrying out the clustering using the Ward linkage results in the same clusters as the k-means algorithm, except moving the [Rb, Cs] materials from the [LiH, MgO, Al] to the [Li, Na, K] group, collecting all the alkali metals in one group.

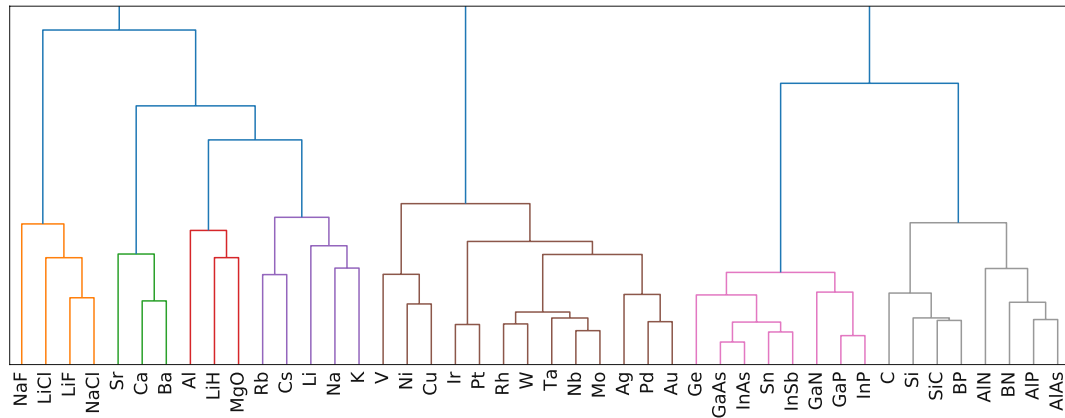


FIGURE 4.2: Dendrogram of the agglomerative clustering using Ward linkage. For better visibility the top part of the dendrogram is discarded.

Looking at the structure of the dendrogram on Fig. 4.2 the three distinct groups of materials can be noticed, which link only on the later stages of the clustering. The clusters are also stable with respect to the used linkage method, using the "maximum" linkage, shown on Fig. 4.3, results in the same cluster, except the two semiconductor clusters being merged when 6 clusters are used. Going to seven only splits down the [NaF] from the ionic cluster, yet in the case of eight clusters the semiconductors are also broken up in two clusters.

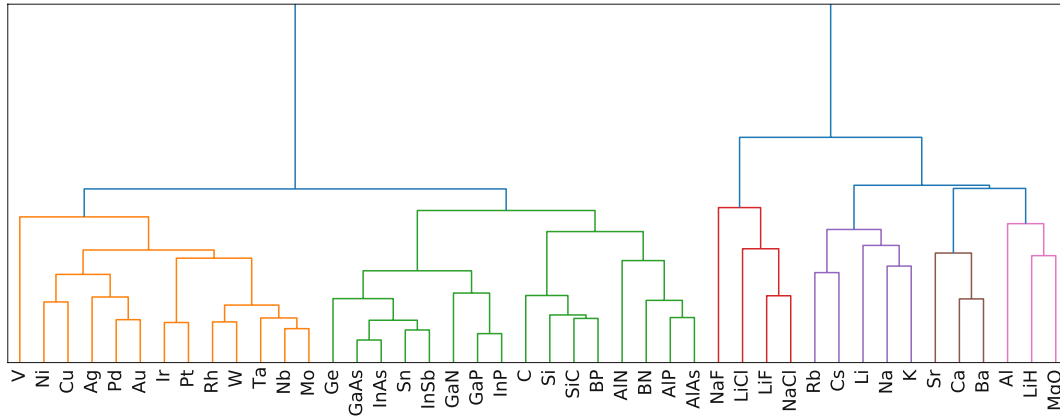


FIGURE 4.3: Dendrogram of the agglomerative clustering using "maximum" linkage. For better visibility the top part of the dendrogram is discarded.



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

5 Outlook

My thesis consists of three distinct lines of work, all of which can be improved upon.

5.1 Exchange-correlation functionals

With the development of new ML methods and the increasing application of the newest discoveries in quantum chemistry, ML applications have a real potential to revolutionize electronic structure calculations. Recent examples in this direction are:

- Good approximation of electron correlation through neural network wavefunction approximations[42]
- Significant speedup in molecular dynamics calculations using machine learned force field on par with DFT accuracy[30]
- Complete replacement of the XC functional by a neural network trained on only a small number of molecules[31]

Yet these applications are not completely replacing traditional functionals or DFT calculations, however with the fast progress of the field it could be possible in the near future.

As a direct extension of the method presented in the current thesis, the ML functional could be used to predict both correlation and exchange energies, removing the requirement of relying on correlation energies calculated with previous approximations. While correlation is the smaller part of the exchange-correlation energy, using it in the fitting could make significant improvements in the accuracy.

Other than extending the functional for correlation, one can include more descriptors as inputs for the functional. The first candidate for this could be the laplacian of the density, but as it was shown previously this descriptor contains only a small amount of extra information when it is used alongside the kinetic energy density. The more interesting extensions would be non-local descriptors, which could help to incorporate long range interactions.

5.2 Machine learning in IR prediction

Our work described in the "Machine-learning Prediction of Infrared Spectra of Interstellar Polycyclic Aromatic Hydrocarbons" study focuses on predictions for polycyclic aromatic hydrocarbons, yet none of the steps in the process requires the materials to come from this specific group. A straightforward improvement of our current model could be to simply include larger databases of different materials in the training. Additionally to the obvious advantages of being able to predict for different material groups, this would allow us to investigate how transferable our model is and if the extra information from training on other materials groups improves or deteriorates the accuracy on PAHs.

While an extended connectivity fingerprint with sufficiently large radius contains all the information of a molecular graph, it might not be to most efficient descriptor for IR spectra prediction. The usage of message passing on a molecular graph to create the input of feed forward neural network for IR spectra prediction was also shown[28] to be a successful approach, which could be investigated as the replacement of the ECFP in our process.

Another interesting application would be the reverse prediction, meaning that the NN takes an IR spectra and reconstructs the ECFP or directly the molecular graph. While the ECFP prediction for fixed set of bits can be done with a traditional feed forward neural network, reconstructing the molecular graph is more complicated.

5.3 Similarity clustering of materials

The methodology presented in the "Similarity clustering for representative sets of solids for density functional testing" study is used to cluster materials which sample similar regions of p-t space of mGGA functionals. This has been used to investigate how well balanced the 44 solids dataset used for the lattice parameter and cohesive energy calculation of the exchange functional training is and to present a possibly better balanced "representative set".

The natural extension to this study could be to simply apply the methodology to other notable datasets, like the one used for bandgap benchmarks[7] with 473 solids or the G2/97[10] and G3/99[11] with 302 and 376 molecules respectively.

A probably more interesting line of research could be extending the type of materials used. Currently the methodology was only tested on bulk solids, but testing on solids with defects, surfaces or organic molecules could reveal new knowledge which could help to tackle these kind of materials more accurately with DFT.



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Bibliography

- [1] Thilo Aschebrock and Stephan Kümmel. “Ultranonlocality and accurate band gaps from a meta-generalized gradient approximation”. In: *Physical Review Research* 1.3 (Nov. 2019). DOI: [10.1103/physrevresearch.1.033082](https://doi.org/10.1103/physrevresearch.1.033082). URL: <https://doi.org/10.1103/physrevresearch.1.033082>.
- [2] A. D. Becke. “Density functional calculations of molecular bond energies”. In: *The Journal of Chemical Physics* 84.8 (Apr. 1986), pp. 4524–4529. DOI: [10.1063/1.450025](https://doi.org/10.1063/1.450025). URL: <https://doi.org/10.1063/1.450025>.
- [3] A. D. Becke. “Density-functional exchange-energy approximation with correct asymptotic behavior”. In: *Physical Review A* 38.6 (Sept. 1988), pp. 3098–3100. DOI: [10.1103/physreva.38.3098](https://doi.org/10.1103/physreva.38.3098). URL: <https://doi.org/10.1103/physreva.38.3098>.
- [4] A. D. Becke and K. E. Edgecombe. “A simple measure of electron localization in atomic and molecular systems”. In: *The Journal of Chemical Physics* 92.9 (May 1990), pp. 5397–5403. DOI: [10.1063/1.458517](https://doi.org/10.1063/1.458517). URL: <https://doi.org/10.1063/1.458517>.
- [5] Axel D. Becke. “Density-functional thermochemistry. III. The role of exact exchange”. In: *The Journal of Chemical Physics* 98.7 (Apr. 1993), pp. 5648–5652. DOI: [10.1063/1.464913](https://doi.org/10.1063/1.464913). URL: <https://doi.org/10.1063/1.464913>.
- [6] P. Blaha et al. *WIEN2k: An Augmented Plane Wave plus Local Orbitals Program for Calculating Crystal Properties*. ISBN 3-9501031-1-2. Vienna University of Technology, Austria, 2018.
- [7] Pedro Borlido et al. “Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning”. In: *npj Computational Materials* 6.1

- (July 2020). DOI: [10.1038/s41524-020-00360-0](https://doi.org/10.1038/s41524-020-00360-0). URL: <https://doi.org/10.1038/s41524-020-00360-0>.
- [8] C. Bowen, G. Sugiyama, and B. J. Alder. “Static dielectric response of the electron gas”. In: *Physical Review B* 50.20 (Nov. 1994), pp. 14838–14848. DOI: [10.1103/physrevb.50.14838](https://doi.org/10.1103/physrevb.50.14838). URL: <https://doi.org/10.1103/physrevb.50.14838>.
- [9] Jorge M. del Campo et al. “Non-empirical improvement of PBE and its hybrid PBE0 for general description of molecular properties”. In: *The Journal of Chemical Physics* 136.10 (Mar. 2012), p. 104108. DOI: [10.1063/1.3691197](https://doi.org/10.1063/1.3691197). URL: <https://doi.org/10.1063/1.3691197>.
- [10] Larry A. Curtiss et al. “Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation”. In: *The Journal of Chemical Physics* 106.3 (Jan. 1997), pp. 1063–1079. DOI: [10.1063/1.473182](https://doi.org/10.1063/1.473182). URL: <https://doi.org/10.1063/1.473182>.
- [11] Larry A. Curtiss et al. “Assessment of Gaussian-3 and density functional theories for a larger experimental test set”. In: *The Journal of Chemical Physics* 112.17 (May 2000), pp. 7374–7383. DOI: [10.1063/1.481336](https://doi.org/10.1063/1.481336). URL: <https://doi.org/10.1063/1.481336>.
- [12] Sebastian Dick and Marivi Fernandez-Serra. “Machine learning accurate exchange and correlation functionals of the electronic density”. In: *Nature Communications* 11.1 (July 2020). DOI: [10.1038/s41467-020-17265-7](https://doi.org/10.1038/s41467-020-17265-7). URL: <https://doi.org/10.1038/s41467-020-17265-7>.
- [13] E. Fermi. In: *Rend. Accad. Naz. Lincei* 6 (1927), p. 602.
- [14] B. J. Frey and D. Dueck. “Clustering by Passing Messages Between Data Points”. In: *Science* 315.5814 (Feb. 2007), pp. 972–976. DOI: [10.1126/science.1136800](https://doi.org/10.1126/science.1136800). URL: <https://doi.org/10.1126/science.1136800>.
- [15] Diptarka Hait and Martin Head-Gordon. “How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values”. In: *Journal of Chemical Theory and Computation* 14.4 (Mar. 2018),

- pp. 1969–1981. DOI: [10.1021/acs.jctc.7b01252](https://doi.org/10.1021/acs.jctc.7b01252). URL: <https://doi.org/10.1021/acs.jctc.7b01252>.
- [16] B. Hammer, L. B. Hansen, and J. K. Nørskov. “Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals”. In: *Physical Review B* 59.11 (Mar. 1999), pp. 7413–7421. DOI: [10.1103/physrevb.59.7413](https://doi.org/10.1103/physrevb.59.7413). URL: <https://doi.org/10.1103/physrevb.59.7413>.
- [17] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. “Hybrid functionals based on a screened Coulomb potential”. In: *The Journal of Chemical Physics* 118.18 (May 2003), pp. 8207–8215. DOI: [10.1063/1.1564060](https://doi.org/10.1063/1.1564060). URL: <https://doi.org/10.1063/1.1564060>.
- [18] P. Hohenberg and W. Kohn. “Inhomogeneous Electron Gas”. In: *Physical Review* 136.3B (Nov. 1964), B864–B871. DOI: [10.1103/physrev.136.b864](https://doi.org/10.1103/physrev.136.b864). URL: <https://doi.org/10.1103/physrev.136.b864>.
- [19] Kurt Hornik. “Approximation capabilities of multilayer feedforward networks”. In: *Neural Networks* 4.2 (1991), pp. 251–257. DOI: [10.1016/0893-6080\(91\)90009-t](https://doi.org/10.1016/0893-6080(91)90009-t). URL: [https://doi.org/10.1016/0893-6080\(91\)90009-t](https://doi.org/10.1016/0893-6080(91)90009-t).
- [20] Muhammad F. Kasim and Sam M. Vinko. *Learning the exchange-correlation functional from nature with fully differentiable density functional theory*. 2021. eprint: [arXiv:2102.04229](https://arxiv.org/abs/2102.04229).
- [21] W. Kohn and L. J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Physical Review* 140.4A (Nov. 1965), A1133–A1138. DOI: [10.1103/physrev.140.a1133](https://doi.org/10.1103/physrev.140.a1133). URL: <https://doi.org/10.1103/physrev.140.a1133>.
- [22] Susi Lehtola et al. “Recent developments in libxc — A comprehensive library of functionals for density functional theory”. In: *SoftwareX* 7 (Jan. 2018), pp. 1–5. DOI: [10.1016/j.softx.2017.11.002](https://doi.org/10.1016/j.softx.2017.11.002). URL: <https://doi.org/10.1016/j.softx.2017.11.002>.

- [23] Mel Levy and John P. Perdew. "Hellmann-Feynman, virial, and scaling requisites for the exact universal density functionals. Shape of the correlation potential and diamagnetic susceptibility for atoms". In: *Physical Review A* 32.4 (Oct. 1985), pp. 2010–2021. DOI: [10.1103/physreva.32.2010](https://doi.org/10.1103/physreva.32.2010). URL: <https://doi.org/10.1103/physreva.32.2010>.
- [24] Elliott H. Lieb and Stephen Oxford. "Improved lower bound on the indirect Coulomb energy". In: *International Journal of Quantum Chemistry* 19.3 (Mar. 1981), pp. 427–439. DOI: [10.1002/qua.560190306](https://doi.org/10.1002/qua.560190306). URL: <https://doi.org/10.1002/qua.560190306>.
- [25] S. Lloyd. "Least squares quantization in PCM". In: *IEEE Transactions on Information Theory* 28.2 (Mar. 1982), pp. 129–137. DOI: [10.1109/tit.1982.1056489](https://doi.org/10.1109/tit.1982.1056489). URL: <https://doi.org/10.1109/tit.1982.1056489>.
- [26] Georg K. H. Madsen. "Functional form of the generalized gradient approximation for exchange: ThePBEfunctional". In: *Physical Review B* 75.19 (May 2007). DOI: [10.1103/physrevb.75.195108](https://doi.org/10.1103/physrevb.75.195108). URL: <https://doi.org/10.1103/physrevb.75.195108>.
- [27] Georg K. H. Madsen, Lara Ferrighi, and Bjørk Hammer. "Treatment of Layered Structures Using a Semilocal meta-GGA Density Functional". In: *The Journal of Physical Chemistry Letters* 1.2 (Dec. 2009), pp. 515–519. DOI: [10.1021/jz9002422](https://doi.org/10.1021/jz9002422). URL: <https://doi.org/10.1021/jz9002422>.
- [28] Charles McGill et al. "Predicting Infrared Spectra with Message Passing Neural Networks". In: *Journal of Chemical Information and Modeling* 61.6 (May 2021), pp. 2594–2609. DOI: [10.1021/acs.jcim.1c00055](https://doi.org/10.1021/acs.jcim.1c00055). URL: <https://doi.org/10.1021/acs.jcim.1c00055>.
- [29] Tom Mitchell. *Machine Learning*. New York: McGraw-Hill, 1997. ISBN: 0-07-042807-7.
- [30] Hadrián Montes-Campos et al. *A Differentiable Neural-Network Force Field for Ionic Liquids*. 2021. eprint: [arXiv:2106.16220](https://arxiv.org/abs/2106.16220).
- [31] Ryo Nagai, Ryosuke Akashi, and Osamu Sugino. "Completing density functional theory by machine learning hidden messages from molecules". In: *npj Computational*

- Materials* 6.1 (May 2020). DOI: 10.1038/s41524-020-0310-0. URL: <https://doi.org/10.1038/s41524-020-0310-0>.
- [32] G. L. Oliver and J. P. Perdew. "Spin-density gradient expansion for the kinetic energy". In: *Physical Review A* 20.2 (Aug. 1979), pp. 397–403. DOI: 10.1103/physreva.20.397. URL: <https://doi.org/10.1103/physreva.20.397>.
- [33] H. Padé. "Sur la représentation approchée d'une fonction par des fractions rationnelles". In: *Annales scientifiques de l'École normale supérieure* 9 (1892), pp. 3–93. DOI: 10.24033/asens.378. URL: <https://doi.org/10.24033/asens.378>.
- [34] John P. Perdew. "Jacob's ladder of density functional approximations for the exchange-correlation energy". In: *AIP Conference Proceedings*. AIP, 2001. DOI: 10.1063/1.1390175. URL: <https://doi.org/10.1063/1.1390175>.
- [35] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. "Generalized Gradient Approximation Made Simple". In: *Physical Review Letters* 77.18 (Oct. 1996), pp. 3865–3868. DOI: 10.1103/physrevlett.77.3865. URL: <https://doi.org/10.1103/physrevlett.77.3865>.
- [36] John P. Perdew and Yue Wang. "Accurate and simple analytic representation of the electron-gas correlation energy". In: *Physical Review B* 45.23 (June 1992), pp. 13244–13249. DOI: 10.1103/physrevb.45.13244. URL: <https://doi.org/10.1103/physrevb.45.13244>.
- [37] John P. Perdew et al. "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation". In: *Physical Review B* 46.11 (Sept. 1992), pp. 6671–6687. DOI: 10.1103/physrevb.46.6671. URL: <https://doi.org/10.1103/physrevb.46.6671>.
- [38] John P. Perdew et al. "Gedanken densities and exact constraints in density functional theory". In: *The Journal of Chemical Physics* 140.18 (May 2014), 18A533. DOI: 10.1063/1.4870763. URL: <https://doi.org/10.1063/1.4870763>.

- [39] John P. Perdew et al. "Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces". In: *Physical Review Letters* 100.13 (Apr. 2008). DOI: [10.1103/physrevlett.100.136406](https://doi.org/10.1103/physrevlett.100.136406). URL: <https://doi.org/10.1103/physrevlett.100.136406>.
- [40] Roberto Peverati and Donald G. Truhlar. "M11-L: A Local Density Functional That Provides Improved Accuracy for Electronic Structure Calculations in Chemistry and Physics". In: *The Journal of Physical Chemistry Letters* 3.1 (Dec. 2011), pp. 117–124. DOI: [10.1021/jz201525m](https://doi.org/10.1021/jz201525m). URL: <https://doi.org/10.1021/jz201525m>.
- [41] L Pollack and J P Perdew. In: *Journal of Physics: Condensed Matter* 12.7 (Feb. 2000), pp. 1239–1252. DOI: [10.1088/0953-8984/12/7/308](https://doi.org/10.1088/0953-8984/12/7/308). URL: <https://doi.org/10.1088/0953-8984/12/7/308>.
- [42] James Spencer. "Learning many-electron wavefunctions with deep neural networks". In: *Nature Reviews Physics* 3.7 (June 2021), pp. 458–458. DOI: [10.1038/s42254-021-00330-5](https://doi.org/10.1038/s42254-021-00330-5). URL: <https://doi.org/10.1038/s42254-021-00330-5>.
- [43] Viktor N. Staroverov et al. "Tests of a ladder of density functionals for bulk solids and surfaces". In: *Physical Review B* 69.7 (Feb. 2004). DOI: [10.1103/physrevb.69.075102](https://doi.org/10.1103/physrevb.69.075102). URL: <https://doi.org/10.1103/physrevb.69.075102>.
- [44] Jianwei Sun, Adrienn Ruzsinszky, and John P. Perdew. "Strongly Constrained and Appropriately Normed Semilocal Density Functional". In: *Physical Review Letters* 115.3 (July 2015). DOI: [10.1103/physrevlett.115.036402](https://doi.org/10.1103/physrevlett.115.036402). URL: <https://doi.org/10.1103/physrevlett.115.036402>.
- [45] Jianwei Sun et al. "Density Functionals that Recognize Covalent, Metallic, and Weak Bonds". In: *Physical Review Letters* 111.10 (Sept. 2013). DOI: [10.1103/physrevlett.111.106401](https://doi.org/10.1103/physrevlett.111.106401). URL: <https://doi.org/10.1103/physrevlett.111.106401>.
- [46] P. S. Svendsen and U. von Barth. "Gradient expansion of the exchange energy from second-order density response theory". In: *Physical Review B* 54.24 (Dec. 1996), pp. 17402–17413. DOI: [10.1103/physrevb.54.17402](https://doi.org/10.1103/physrevb.54.17402). URL: <https://doi.org/10.1103/physrevb.54.17402>.

- [47] Jianmin Tao and Yuxiang Mo. “Accurate Semilocal Density Functional for Condensed-Matter Physics and Quantum Chemistry”. In: *Physical Review Letters* 117.7 (Aug. 2016). DOI: [10.1103/physrevlett.117.073001](https://doi.org/10.1103/physrevlett.117.073001). URL: <https://doi.org/10.1103/physrevlett.117.073001>.
- [48] L. H. Thomas. “The calculation of atomic fields”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 23.5 (Jan. 1927), pp. 542–548. DOI: [10.1017/s0305004100011683](https://doi.org/10.1017/s0305004100011683). URL: <https://doi.org/10.1017/s0305004100011683>.
- [49] Fabien Tran and Peter Blaha. “Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential”. In: *Physical Review Letters* 102.22 (June 2009). DOI: [10.1103/physrevlett.102.226401](https://doi.org/10.1103/physrevlett.102.226401). URL: <https://doi.org/10.1103/physrevlett.102.226401>.
- [50] Fabien Tran, Julia Stelzl, and Peter Blaha. “Rungs 1 to 4 of DFT Jacob’s ladder: Extensive test on the lattice constant, bulk modulus, and cohesive energy of solids”. In: *The Journal of Chemical Physics* 144.20 (May 2016), p. 204120. DOI: [10.1063/1.4948636](https://doi.org/10.1063/1.4948636). URL: <https://doi.org/10.1063/1.4948636>.
- [51] Fabien Tran et al. “On the calculation of the bandgap of periodic solids with MGGA functionals using the total energy”. In: *The Journal of Chemical Physics* 151.16 (Oct. 2019), p. 161102. DOI: [10.1063/1.5126393](https://doi.org/10.1063/1.5126393). URL: <https://doi.org/10.1063/1.5126393>.
- [52] Fabien Tran et al. “Orbital-free approximations to the kinetic-energy density in exchange-correlation MGGA functionals: Tests on solids”. In: *The Journal of Chemical Physics* 149.14 (Oct. 2018), p. 144105. DOI: [10.1063/1.5048907](https://doi.org/10.1063/1.5048907). URL: <https://doi.org/10.1063/1.5048907>.
- [53] Pragya Verma and Donald G. Truhlar. “HLE16: A Local Kohn–Sham Gradient Approximation with Good Performance for Semiconductor Band Gaps and Molecular Excitation Energies”. In: *The Journal of Physical Chemistry Letters* 8.2 (Jan. 2017), pp. 380–387. DOI: [10.1021/acs.jpcllett.6b02757](https://doi.org/10.1021/acs.jpcllett.6b02757). URL: <https://doi.org/10.1021/acs.jpcllett.6b02757>.

- [54] Pragya Verma and Donald G. Truhlar. "HLE17: An Improved Local Exchange–Correlation Functional for Computing Semiconductor Band Gaps and Molecular Excitation Energies". In: *The Journal of Physical Chemistry C* 121.13 (Mar. 2017), pp. 7144–7154. DOI: [10.1021/acs.jpcc.7b01066](https://doi.org/10.1021/acs.jpcc.7b01066). URL: <https://doi.org/10.1021/acs.jpcc.7b01066>.
- [55] S. H. Vosko, L. Wilk, and M. Nusair. "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis". In: *Canadian Journal of Physics* 58.8 (Aug. 1980), pp. 1200–1211. DOI: [10.1139/p80-159](https://doi.org/10.1139/p80-159). URL: <https://doi.org/10.1139/p80-159>.
- [56] C. F. von Weizsäcker. In: *Z. Phys.* 96 (1935), p. 431.
- [57] Jess Wellendorff et al. "mBEEF: An accurate semi-local Bayesian error estimation density functional". In: *The Journal of Chemical Physics* 140.14 (Apr. 2014), p. 144107. DOI: [10.1063/1.4870397](https://doi.org/10.1063/1.4870397). URL: <https://doi.org/10.1063/1.4870397>.
- [58] Yan Zhao and Donald G. Truhlar. "A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions". In: *The Journal of Chemical Physics* 125.19 (Nov. 2006), p. 194101. DOI: [10.1063/1.2370993](https://doi.org/10.1063/1.2370993). URL: <https://doi.org/10.1063/1.2370993>.
- [59] Yan Zhao and Donald G. Truhlar. "Construction of a generalized gradient approximation by restoring the density-gradient expansion and enforcing a tight Lieb–Oxford bound". In: *The Journal of Chemical Physics* 128.18 (May 2008), p. 184109. DOI: [10.1063/1.2912068](https://doi.org/10.1063/1.2912068). URL: <https://doi.org/10.1063/1.2912068>.

6 List of publications

Comparative study of the PBE and SCAN functionals: The particular case of alkali metals

Cite as: J. Chem. Phys. **150**, 164119 (2019); <https://doi.org/10.1063/1.5092748>

Submitted: 14 February 2019 . Accepted: 04 April 2019 . Published Online: 25 April 2019

Péter Kovács,  Fabien Tran,  Peter Blaha, and  Georg K. H. Madsen



ARTICLES YOU MAY BE INTERESTED IN

Regularized SCAN functional

The Journal of Chemical Physics **150**, 161101 (2019); <https://doi.org/10.1063/1.5094646>

A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu

The Journal of Chemical Physics **132**, 154104 (2010); <https://doi.org/10.1063/1.3382344>

Rungs 1 to 4 of DFT Jacob's ladder: Extensive test on the lattice constant, bulk modulus, and cohesive energy of solids

The Journal of Chemical Physics **144**, 204120 (2016); <https://doi.org/10.1063/1.4948636>

Challenge us.

What are your needs for periodic signal detection?

 Watch



 Zurich Instruments

Comparative study of the PBE and SCAN functionals: The particular case of alkali metals

Cite as: J. Chem. Phys. 150, 164119 (2019); doi: 10.1063/1.5092748

Submitted: 14 February 2019 • Accepted: 4 April 2019 •

Published Online: 25 April 2019



View Online



Export Citation



CrossMark

Péter Kovács, Fabien Tran, Peter Blaha, and Georg K. H. Madsen^{a)}

AFFILIATIONS

Institute of Materials Chemistry, Vienna University of Technology, Getreidemarkt 9/165-TC, A-1060 Vienna, Austria

^{a)} Author to whom correspondence should be addressed: georg.madsen@tuwien.ac.at

ABSTRACT

The SCAN meta-generalized gradient approximation (GGA) functional is known to describe multiple properties of various materials with different types of bonds with greater accuracy, compared to the widely used PBE GGA functional. Yet, for alkali metals, SCAN shows worse agreement with experimental results than PBE despite using more information about the system. In the current study, this behavior for alkali metals is explained by identifying an inner semicore region which, within SCAN, contributes to an underbinding. The inner semicore push toward larger lattice constants is a general feature but is particularly important for very soft materials, such as the alkali metals, while for harder materials the valence region dominates.

© 2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5092748>

I. INTRODUCTION

Presently, the most common theoretical approach to calculating the properties of solids and molecules is Kohn-Sham density functional theory (KS-DFT).¹ The main accuracy restricting factor of this method is the functional form of the exchange-correlation energy, E_{xc} . For local and semilocal functionals, E_{xc} is given by

$$E_{xc} = \int e_{xc}(\mathbf{r}) d\mathbf{r}, \quad (1)$$

where e_{xc} is the exchange-correlation energy per volume unit and is a function of local electronic properties, such as the electron density, electron density gradient, or kinetic-energy density (KED). The simplest approximation of e_{xc} is the local density approximation (LDA).¹ The next step of functional development was to add a functional dependence on the gradient of the density. This led to the generalized gradient approximations (GGAs),^{2,3} which have better accuracy in multiple cases. In the meta-GGA (MGGAs) functionals, the KED and/or the Laplacian of the density are also used in the parameterization of e_{xc} . Several MGGAs with different constraints and goals have been developed (see Ref. 4 for a review), and benchmarks of these different functionals have shown how MGGAs can improve the overall accuracy compared to GGAs.^{5–8} The improved performance can, depending on the point of view, be related to

the MGGAs being able to distinguish more bonding situations,^{9–11} better fit reference data,¹² or satisfy more exact constraints.¹³

One of the recent MGGAs, which has gained considerable attention, is the SCAN functional.¹³ For instance, it has shown successes in calculating the formation enthalpy of various solids¹⁴ or the structural and energetic properties of ferroelectric materials.¹⁵ On the other hand, SCAN performs poorly for the magnetic properties of transition metals.^{16–19} It is natural to compare SCAN with the PBE GGA.³ First of all, they are constructed following a similar philosophy of constraint satisfaction. Furthermore, PBE can be considered as a good functional for solids since it gives reasonable equilibrium lattice constants, a_0 , and cohesive energies, E_{coh} (see, e.g., Ref. 5). While it is possible to construct GGA functionals which give better results than PBE for the lattice constants,^{20–23} these will tend to overestimate the cohesive energies of solids.⁵ Thus, for SCAN to be a systematic improvement on PBE for solids, one requirement would be that it simultaneously improves on the lattice constants and the cohesive energy. Numerical tests have shown that on average the SCAN does exactly this for a wide range of solids.^{5,24}

The improvement of SCAN over PBE is, however, not universal. A close look at the results in Ref. 5 also reveals how SCAN performs disappointingly for most alkali metals. This is illustrated in Fig. 1 where SCAN is compared to LDA and PBE for a few selected materials. Considering first Si and Ge, which we use for illustrative

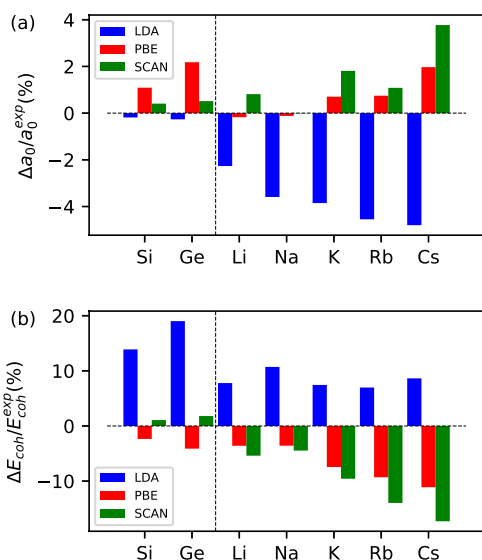


FIG. 1. Relative error (in %) in the lattice constant (a) and cohesive energy (b) obtained with the LDA, PBE, and SCAN functionals for Si, Ge, and the alkali metals.

purposes as representatives for a large group of systems (covalent semiconductors), the well known tendencies of LDA to overestimate the cohesive energies, thereby underestimating the lattice constant, and PBE to overestimate the lattice constants, thereby underestimating the cohesive energies, can be seen. For Si and Ge, SCAN systematically improves both the lattice constant and cohesive energies over PBE. For the alkali metals on the other hand, we can see that SCAN underbinds even more, i.e., gives even larger lattice constants (except for Na) and even smaller cohesive energies, than PBE (Fig. 1).

In the present study, we aim at a detailed understanding of how the poor performance of SCAN for the alkali metals is related to its functional form. Apart from the obvious interest in the alkali metals themselves, understanding the disappointing performance of SCAN for this class of materials is also important for developing more accurate density functionals in general. While SCAN may perform well in statistical studies, where the focus is on average errors for databases containing a large number of strongly ionic and covalently bonded materials, these studies may somewhat hide a systematic problem for the more weakly bonded alkali metals because these systems only make up a small subset of the database. Actually, the density distribution in the alkali metals is rather particular. The bonding region is characterized by both the density and reduced gradients being low. This means that the correlation energy becomes comparable to the exchange and that regions of e_{xc} that are otherwise not sampled are probed.

II. METHODOLOGY

As will be discussed below, we will focus on the exchange energy in the present analysis. To describe the analytical form of

an exchange functional, it is common to define the enhancement factor

$$F_x(\mathbf{r}) = \frac{e_x(\mathbf{r})}{e_x^{\text{LDA}}(\mathbf{r})}, \quad (2)$$

where $e_x^{\text{LDA}} = -C_x n^{4/3}$ [$C_x = (3/4)(3/\pi)^{1/3}$, atomic units are used throughout this work] is the LDA exchange-energy density for the homogeneous electron gas¹ and $n = \sum_{i=1}^N |\psi_i|^2$ is the electron density. For GGA functionals, F_x depends on the gradient of the density, ∇n , while for MGGAs functionals it also depends on the noninteracting KS KED $\tau^{\text{KS}} = (1/2) \sum_{i=1}^N |\nabla \psi_i|^2$ (in the present work, we are not concerned with Laplacian-dependent MGGAs). In the following, we will use dimensionless expressions to characterize the density, namely, the reduced density gradient

$$p = \frac{|\nabla n|^2}{4(3\pi^2)^{2/3} n^{8/3}} \quad (3)$$

and reduced KED

$$t = \frac{\tau^{\text{KS}}}{\tau^{\text{TF}}}, \quad (4)$$

where $\tau^{\text{TF}} = (3/10)(3\pi^2)^{2/3} n^{5/3}$ is the Thomas-Fermi (TF) KED^{25,26} which is exact for the homogeneous electron gas. Here, we note that in our previous studies^{8,10} and others,^{11,12} $\tau^{\text{KS}}/\tau^{\text{TF}}$ was instead labeled as t^{-1} . In iso-orbital regions where the density is dominated by one orbital, the KED is given exactly by the von Weizsäcker form²⁷

$$\tau^{\text{vW}} = \frac{1}{8} \frac{|\nabla n|^2}{n}, \quad (5)$$

which makes

$$\alpha = \frac{\tau^{\text{KS}} - \tau^{\text{vW}}}{\tau^{\text{TF}}} \quad (6)$$

a convenient measure of how much the density n at a point of space is dominated by a single orbital.⁹ Since one can write $\tau^{\text{vW}}/\tau^{\text{TF}} = 5p/3$, then

$$\alpha = t - \frac{5}{3}p. \quad (7)$$

Note that τ^{vW} is a strict lower bound to the KED^{9,28-30} so that $5p/3$ is a lower bound to t .

As mentioned in Sec. I, the goal of the present work is to rationalize the SCAN results on the alkali metals and to understand the worsening in the performance compared to PBE. For this purpose, potassium is the case study that will be considered in Sec. III. The analysis will consist of a careful comparison of the PBE and SCAN enhancement factors F_x . The calculations were carried out with the WIEN2k code.³¹ The SCAN calculations were done non-self-consistently using the PBE densities and orbitals^{5,32} so that the only difference in the total energy stems from the functional form of e_{xc} . Note that in our previous work,⁵ the self-consistent effects were estimated to be quite small, below 0.01 Å in most cases, except for the van der Waals systems where they could be larger. For the spatial distribution analysis of the exchange-correlation energy, the sampling of the Voronoi cell of one atom was done on an equidistant radial mesh for 400 different directions from the atom. The distance between the sample points is the same for both volumes, resulting in a larger number of samples for the expanded structure.

III. ANALYSIS

We start by showing in Fig. 2 the difference between PBE and SCAN of the exchange-correlation energy E_{xc} as a function of lattice constant a for the alkali metal potassium. Since all terms in the total energy except E_{xc} are the same for the PBE and SCAN calculations and the SCAN energy is evaluated with PBE density, the slope of $E_{xc}^{\text{SCAN}} - E_{xc}^{\text{PBE}}$ is directly related to the difference between the equilibrium lattice constants a_0^{PBE} and a_0^{SCAN} . As seen in Fig. 2, the slope of the exchange-energy difference $E_x^{\text{SCAN}} - E_x^{\text{PBE}}$ is negative. As a direct consequence, the SCAN equilibrium lattice constant a_0 is “pushed” toward a larger value than the one obtained with PBE. As PBE already overestimates a_0 of potassium (as well as Rb and Cs, see Fig. 1), then SCAN worsens the agreement. Thus, it is the exchange component of E_{xc}^{SCAN} that is responsible for the overestimated lattice constant of K. An interesting feature about the alkali metals is that the low density in the bonding region means that the correlation energy density is comparable to the exchange energy density. Figure 2 shows that the correlation energy exhibits the opposite trend and somewhat compensates for the “push” toward larger volume by the SCAN exchange. However, the compensation is only partial and the slope of the total exchange-correlation energy curve remains negative. In the following, we will thus focus on a detailed analysis of the exchange energy, which is the driving force behind the overestimated lattice constant.

The variation of E_x with respect to the lattice constant a can be explained in terms of changes in the density n and enhancement factor F_x [Eq. (2)]. We separate these two effects by expanding the exchange energy shifts and keeping only terms that are first order in the perturbation

$$\begin{aligned} \delta e_x &\approx -C_x(n + \delta n)^{4/3}(F_x + \delta F_x) + C_x n^{4/3} F_x \\ &\approx -C_x \left(n^{4/3} \delta F_x + \frac{4}{3} n^{1/3} F_x \delta n \right). \end{aligned} \quad (8)$$

The first part, $\delta e_x^{\text{enha}} = -C_x n^{4/3} \delta F_x$, corresponds to the changes in the enhancement factor and the second part, $\delta e_x^{\text{dens}} = -C_x (4/3) n^{1/3} F_x \delta n$, to the changes in the density upon volume change. Two unit cell volumes V (or equivalently two different lattice constants a) were used to obtain δe_x . The smaller and bigger volumes correspond to $V^{\text{small}} = 0.97V^{\text{exp}}$ and $V^{\text{large}} = 1.03V^{\text{exp}}$,

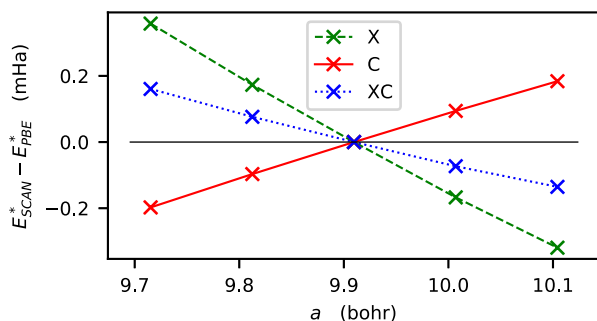


FIG. 2. Energy differences between SCAN and PBE in potassium. A shift, $E^*(a) = E(a) - E(a_0^{\text{exp}})$, is added so that $E(a_0^{\text{exp}}) = 0$. X, C, and XC denote the exchange, correlation, and exchange-correlation energies, respectively.

respectively. This choice of volumes for calculating δe_x is somehow arbitrary; however, the linear monotonic behavior of E_x seen in Fig. 2 shows that it is unimportant for the conclusion. For the sampling of δe_x , we have chosen grids of equidistant points starting at the atomic positions. Since the grid contains more points for the larger volume, an additional contribution, $\delta e_x^{\text{new}} = -C_x n^{4/3} F_x$, representing the new sample points has to be taken into account and added to δe_x^{enha} and δe_x^{dens} to get the full δe_x . Figure 3(a) illustrates these components to the difference $\delta e_x^{\text{SCAN}} - \delta e_x^{\text{PBE}}$ as functions of distance from the potassium atom. δe_x^{new} is small and only appears for distances larger than 4 bohrs because of the shape of the Voronoi cell. In the valence region that we define as the distance beyond 2 bohrs, the density and enhancement terms tend to cancel each other. Actually, $\delta e_x^{\text{enha,SCAN}} - \delta e_x^{\text{enha,PBE}}$ is positive, which indicates that the SCAN exchange enhancement factor is less sensitive to a change in the volume. On the other hand, $\delta e_x^{\text{dens,SCAN}} - \delta e_x^{\text{dens,PBE}}$ is negative reflecting that F_x^{PBE} is larger than F_x^{SCAN} .

In Fig. 3(a), one can also identify a region, between 1 and 2 bohrs, where the total exchange energy density difference $\delta e_x^{\text{SCAN}} - \delta e_x^{\text{PBE}}$ is negative, thus forcing SCAN lattice constant to be larger than the PBE one. This region of negative values is clearly due to the component δe_x^{enha} , i.e., a faster increase in the magnitude of F_x^{SCAN} in a region with a high density when the volume gets bigger, which is particularly important for the lattice constant [see Fig. 3(b)].

Actually, a strong influence on the equilibrium lattice constant coming from the region between 1 and 2 bohrs is at first sight somewhat surprising as one would associate it with an inner semicore region. To understand its origin, we first show $5p/3$ and t , Eqs. (3) and (4), and the normalized orbital densities of a free potassium atom as functions of the distance to the atom in Fig. 4. It is seen that the 1–2 bohr region is indeed dominated by the 3s and 3p semicore orbitals. In this region, the electron density n is very large compared to the valence region such that even small changes δF_x in the enhancement factor have a large impact on the exchange energy [since δF_x is multiplied by $n^{4/3}$, Eq. (8)] and thereby lead to large values of δe_x^{enha} . From Fig. 4, it is also possible to understand why the 4s shell also contributes to $\delta e_x^{\text{SCAN}} - \delta e_x^{\text{PBE}}$ below 2 bohrs. Indeed, at a distance around 1.6 bohrs from the atom, the outer lobe of the 4s shell starts to become important. Since the 4s shell is strongly perturbed by the chemical bonding, then its influence on δF_x in the 1.6–2.0 bohr region should be important.

To obtain insight into the individual contributions³³ to δF_x in δe_x^{enha} , we proceed by expanding it as

$$\delta F_x = \left. \frac{\partial F_x}{\partial p} \right|_{a_0} \delta p + \left. \frac{\partial F_x}{\partial t} \right|_{a_0} \delta t. \quad (9)$$

We first consider the variations δp and δt due to an expansion of the volume. These are depicted for the inner semicore region in Fig. 5(a). As expected, the reduced density gradient p gets larger when the volume increases, i.e., $\delta p > 0$, especially for a distance larger than 1.6 bohrs. The reduced KED, t , on the other hand, shows an interesting behavior. δt is negative up to about 1.9 bohrs, where it changes sign. This illustrates that the KED contains important information

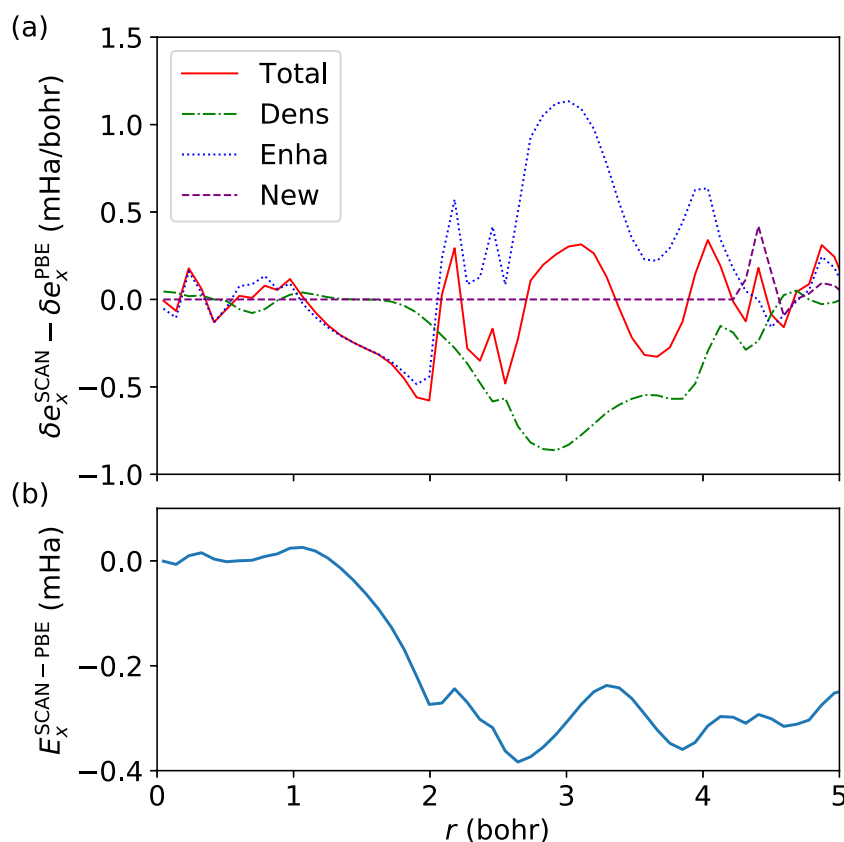


FIG. 3. The differences in δe_x between SCAN and PBE in potassium. (a) The differences are integrated over shells centered at the atomic positions $e_x(r) = r^2 \int e_x(\mathbf{r}) d\Omega$. The integration is done in the Voronoi cell of one atom. (b) Integrated energy differences, $E_x^{\text{SCAN-PBE}}(R) = \int_0^R \delta e_x^{\text{SCAN}}(r) - \delta e_x^{\text{PBE}}(r) dr$.

that is not available in the gradient of the electron density, something that is a premise for the development of MGGA. The behavior of t can be understood from Fig. 4. As observed earlier,^{33,34} one can clearly identify peaks in the reduced density gradient p that are

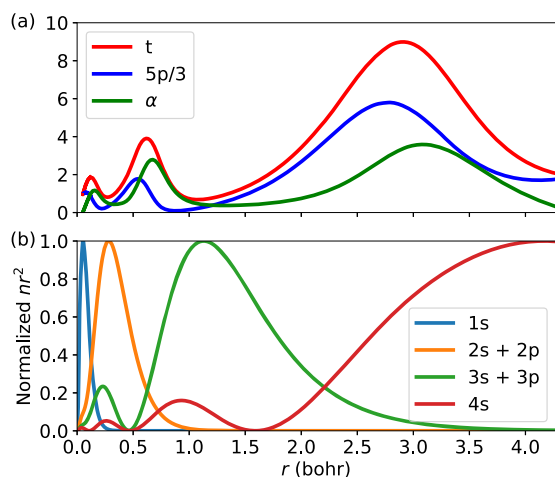


FIG. 4. (a) $5p/3$ [Eq. (3)], α [Eq. (7)], and t [Eq. (4)] for the free potassium atom plotted as functions of the distance from the atom. (b) Normalized densities of the different shells, where the maximum of every curve is set to 1.

located in transition regions where the dominating shell is switching from one to another. In these inter-shell regions, t is substantially larger than $5p/3$ (see Fig. 4) so that α is, as expected from Eq. (7), larger in such regions with contributions coming from different shells.⁹ In the inner semicore (1–2 bohrs) region, α is small, reflecting how it is dominated by orbitals of similar shape ($n = 3$). In the solid, the inner semicore region becomes increasingly dominated by the 3s and 3p orbitals as the unit cell expands, thereby becoming more “atomiclike.” As p hardly changes [$\delta p \approx 0$ for $r < 1.6$ bohrs, Fig. 5(a)], the smaller values of α are the result of smaller values of t ($\delta t < 0$).

The partial derivatives in Eq. (9), depicted in Fig. 5(b), reflect the dependence of the functional on changes in p and t around their values at the equilibrium lattice constant a_0 . Figure 5(b) also shows that $\partial F_x^{\text{SCAN}}/\partial p$ is approximately twice larger than $\partial F_x^{\text{PBE}}/\partial p$. Actually, the large derivatives of SCAN are somewhat surprising because earlier illustrations (see, e.g., Fig. 1 of Ref. 13) give more the impression of a smooth and subdued functional form. However, in Fig. 6(a), we show that the smooth behavior is mainly along lines of constant values of α . Perpendicular to these lines, F_x^{SCAN} shows a somewhat more “bumpy” behavior. Such bumps lead to an erratic behavior of the derivatives, as shown in Figs. 6(b) and 6(c).

The large positive $\partial F_x^{\text{SCAN}}/\partial p$ will, when multiplied by the negative $-C_x n^{4/3}$, Eq. (8), and the positive δp in the inner semicore region, Fig. 5(a), contribute to the negative slope of $E_x^{\text{SCAN}} - E_x^{\text{PBE}}$

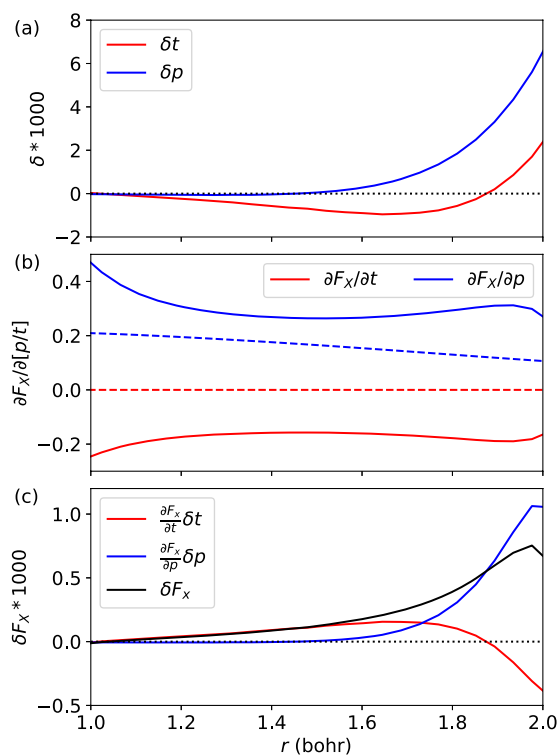


FIG. 5. Various functions in potassium plotted as functions of the distance from the atom. (a) $\delta p = p(V^{\text{large}}) - p(V^{\text{small}})$ and $\delta t = t(V^{\text{large}}) - t(V^{\text{small}})$. (b) $\partial F_x / \partial p$ and $\partial F_x / \partial t$ for PBE (dashed) and SCAN (solid). $\partial F_x^{\text{PBE}} / \partial t = 0$, as a GGA functional has no dependence on t . (c) $\delta F_x^{\text{SCAN}} - \delta F_x^{\text{PBE}}$ [Eq. (9)] and its two components.

observed in Fig. 2. We have already discussed how δt shows a different behavior than δp in the inner semicore region and, in principle, a MGGA could compensate for this contribution in its dependence on the KED. However, as both δp and the corresponding partial derivative $\partial F_x^{\text{SCAN}} / \partial p$ have opposite signs of δt and $\partial F_x^{\text{SCAN}} / \partial t$, their contributions to δF_x^{SCAN} add up instead of canceling [see Fig. 5(c)]. Thereby, both contribute to a too large value for a_0 .

Equation (9) underlines how the partial derivatives are an important factor in determining energy differences and thereby the performance of a MGGA. This would suggest that they should be routinely shown when reporting a new functional. We should also point out that the partial derivatives in Eq. (9) are part of the analytical expression of the MGGA potential for self-consistent calculations,³⁵ and the behavior observed in Figs. 6(b) and 6(c) could thus be responsible for SCAN resulting in a large overestimation of the magnetic moment in itinerant transition metals.^{16–19} In this context, it is interesting to note that a fixed-spin moment calculation, which involved only the SCAN energy and used the PBE potential, resulted in the same overestimation of the magnetic moment as a self-consistent calculation.¹⁹ The expansion with respect to volume, Eq. (8), highlights how features of the analytic form of the energy functional are directly related to the potential. Similar to the expansion with respect to volume, the exchange-

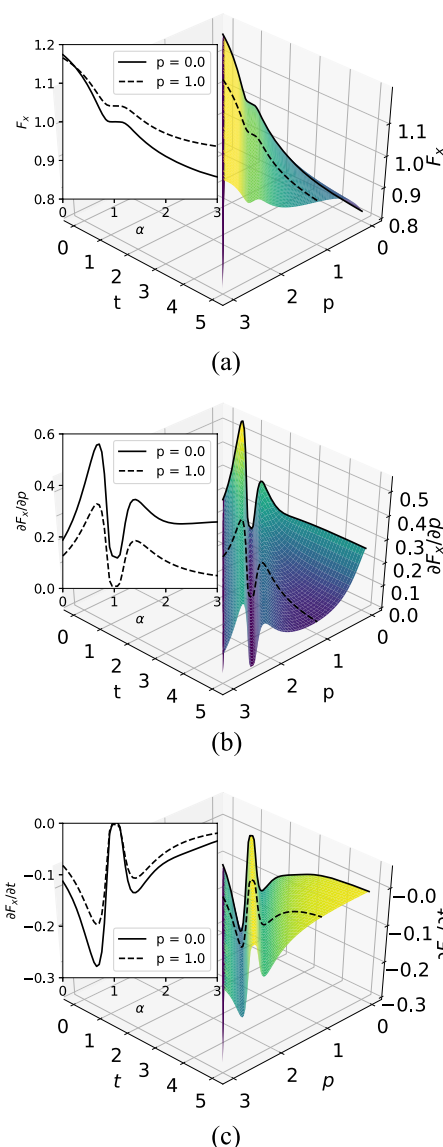


FIG. 6. Maps of the SCAN exchange enhancement factor F_x and its partial derivatives. On the 2D plots, slices of these maps are shown for constant p values as a function of α , where $\alpha = t - 5p/3$. The view is set so that lines of constant values of α are perpendicular to the paper.

correlation energy can also be expanded with respect to the magnetic moment.

The observation of an erratic behavior of the functional also falls in line with recent observations of relative strong grid dependence of SCAN results.^{36,37} Such a grid dependence has also been analyzed previously in Ref. 38 for other MGGA functionals, where it was also pointed out that poor convergence with grid size can lead to unintended contributions to the energy differences.

It is noteworthy that among the 44 solids tested in Ref. 5, the alkali and alkaline earth metals are the only ones for which lattice

constants obtained with SCAN are larger than those obtained with PBE. The analysis above raises the question why the SCAN inner semicore push toward larger volumes is not observed for more systems. To answer this, we will use the closed packed metal Al, where the semicore region, as in the alkali metals, constitutes a significant part of the volume. Despite this, SCAN actually predicts a smaller lattice constant, $a_0 = 4.012 \text{ \AA}$, than both PBE $a_0 = 4.041 \text{ \AA}$ and experiment $a_0 = 4.022 \text{ \AA}$ (see the supplementary material of Ref. 5). The plot of $\delta e_x^{\text{SCAN}} - \delta e_x^{\text{PBE}}$ for Al is shown in Fig. 7. The inner semicore region can be identified between 0.5 and 1.0 bohrs and does indeed have a negative total $\delta e_x^{\text{SCAN}} - \delta e_x^{\text{PBE}}$ due to the δF_x contribution. It will therefore push SCAN to have a larger lattice constant compared to PBE, similar to what was observed for potassium (Fig. 3). However, contrary to potassium, the influence of the valence region is much larger than the inner semicore region. The valence contribution is mainly positive which results overall in a smaller SCAN equilibrium lattice constant than with PBE. We have performed the same analysis for FCC-Ca (not shown) Si in the diamond lattice. Also here, an inner semicore push toward larger lattice constants due to δe_x^{enha} can be identified. However, this is compensated by the valence region, which means that SCAN and PBE lead to very similar lattice constants for Ca.

Finally, one could also argue that the SCAN underbinding of the alkali metals should be cured by explicitly adding contributions for the long-range dispersion interactions.³⁶ Such corrections will however universally strengthen the bonding and thus lead to a worse performance in cases such as Ca and Al where SCAN already tends to overbind. Thus, two strategies could be followed to cure the problem of SCAN for the alkali metals: either by modifying the functional form such that the results for the alkali metals are improved or by adding a term that explicitly accounts for the dispersion term.

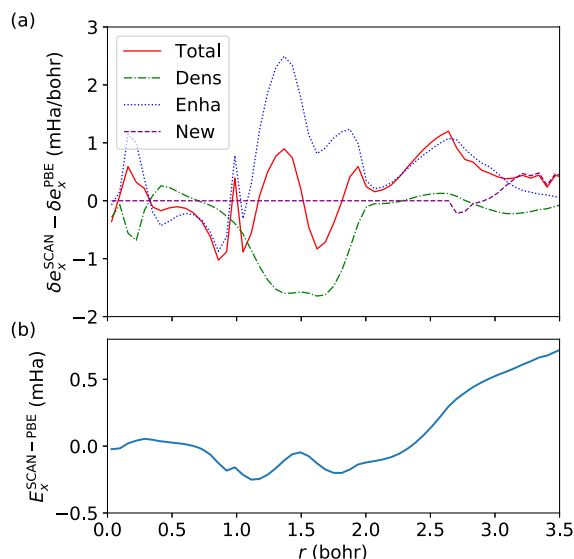


FIG. 7. The differences in δe_x between SCAN and PBE in Al. (a) The differences are integrated over shells centered at the atomic positions $e_x(r) = r^2 \int e_x(\mathbf{r}) d\Omega$. The integration is done in the Voronoi cell of one atom. (b) Integrated energy differences, $E_x^{\text{SCAN-PBE}}(R) = \int_0^R (\delta e_x^{\text{SCAN}}(r) - \delta e_x^{\text{PBE}}(r)) dr$.

However, in the latter case, the functional form of SCAN should also be modified in order to avoid an overbinding for other systems like Ca or Al.

IV. SUMMARY

In the current study, we have analyzed in detail the results obtained with the MGGA functional SCAN for the alkali metals. For these systems, SCAN is less accurate than the more simple GGA functional PBE. SCAN has a clear tendency to underbind the alkali metals; i.e., the equilibrium lattice constants are too large and the cohesive energies are too small. We have shown that this behavior of SCAN is attributed to an inner semicore push toward larger lattice constants, which was revealed by a careful comparison of the PBE and SCAN enhancement factors. Such an inner semicore push toward larger lattice constants can probably be identified for many materials; however, it is the most important mechanism for soft materials such as alkali metals, while for harder materials (e.g., semiconductors and ionic insulators) the valence region dominates (as shown for Al).

A detailed analysis, such as the one that we have presented, leads to a clear understanding of the failures or unexpected results that a functional produces. A functional may have an analytical form that is inappropriate within a particular regime, e.g., for low densities or high density gradients, and the precise identification of the problem in the functional form may give a clue of how to modify the functional form to cure the problem. Our study furthermore highlights the importance of the partial derivatives in determining energy differences and suggests that these should be routinely shown when reporting a new functional.

ACKNOWLEDGMENTS

Support from the Austrian Science Fund (FWF) via Project Nos. F41 (SFB ViCoM), P27738-N28, and CODIS (I 3576-N36) is acknowledged.


REFERENCES

- W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996); Erratum, **78**, 1396 (1997).
- F. Della Sala, E. Fabiano, and L. A. Constantin, *Int. J. Quantum Chem.* **116**, 1641 (2016).
- F. Tran, J. Stelzl, and P. Blaha, *J. Chem. Phys.* **144**, 204120 (2016).
- G. I. Csonka, J. P. Perdew, A. Ruzsinszky, P. H. T. Philipsen, S. Lebègue, J. Paier, O. A. Vydrov, and J. G. Ángyán, *Phys. Rev. B* **79**, 155107 (2009).
- Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.* **41**, 157 (2008).
- L. Ferrighi, G. K. H. Madsen, and B. Hammer, *J. Chem. Phys.* **135**, 084704 (2011).
- A. D. Becke and K. E. Edgecombe, *J. Chem. Phys.* **92**, 5397 (1990).
- G. K. H. Madsen, L. Ferrighi, and B. Hammer, *J. Phys. Chem. Lett.* **1**, 515 (2010).
- J. Sun, B. Xiao, Y. Fang, R. Haunschild, P. Hao, A. Ruzsinszky, G. I. Csonka, G. E. Scuseria, and J. P. Perdew, *Phys. Rev. Lett.* **111**, 106401 (2013).
- Y. Zhao and D. G. Truhlar, *J. Chem. Phys.* **125**, 194101 (2006).
- J. Sun, A. Ruzsinszky, and J. P. Perdew, *Phys. Rev. Lett.* **115**, 036402 (2015).
- Y. Zhang, D. A. Kitchaev, J. Yang, T. Chen, S. T. Dacek, R. A. Sarmiento-Pérez, M. A. L. Marques, H. Peng, G. Ceder, J. P. Perdew, and J. Sun, *npj Comput. Mater.* **4**, 9 (2018).

- ¹⁵Y. Zhang, J. Sun, J. P. Perdew, and X. Wu, *Phys. Rev. B* **96**, 035143 (2017).
- ¹⁶E. B. Isaacs and C. Wolverton, *Phys. Rev. Materials* **2**, 063801 (2018).
- ¹⁷S. Jana, A. Patra, and P. Samal, *J. Chem. Phys.* **149**, 044120 (2018).
- ¹⁸M. Ekholm, D. Gambino, H. J. M. Jönsson, F. Tasnádi, B. Alling, and I. A. Abrikosov, *Phys. Rev. B* **98**, 094413 (2018).
- ¹⁹Y. Fu and D. J. Singh, *Phys. Rev. Lett.* **121**, 207201 (2018).
- ²⁰R. Armiento and A. E. Mattsson, *Phys. Rev. B* **72**, 085108 (2005).
- ²¹Z. Wu and R. E. Cohen, *Phys. Rev. B* **73**, 235116 (2006).
- ²²G. K. H. Madsen, *Phys. Rev. B* **75**, 195108 (2007).
- ²³J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, *Phys. Rev. Lett.* **100**, 136406 (2008); Erratum, **102**, 039902 (2009).
- ²⁴G.-X. Zhang, A. M. Reilly, A. Tkatchenko, and M. Scheffler, *New J. Phys.* **20**, 063020 (2018).
- ²⁵L. H. Thomas, *Proc. Cambridge Philos. Soc.* **23**, 542 (1927).
- ²⁶E. Fermi, *Rend. Accad. Naz. Lincei* **6**, 602 (1927).
- ²⁷C. F. von Weizsäcker, *Z. Phys.* **96**, 431 (1935).
- ²⁸M. Hoffmann-Ostenhof and T. Hoffmann-Ostenhof, *Phys. Rev. A* **16**, 1782 (1977).
- ²⁹Y. Tal and R. F. W. Bader, *Int. J. Quantum Chem., Quantum Chem. Symp.* **14**, 153 (1978).
- ³⁰S. Kurth, J. P. Perdew, and P. Blaha, *Int. J. Quantum Chem.* **75**, 889 (1999).
- ³¹P. Blaha, K. Schwarz, G. K. H. Madsen, D. Kvasnicka, J. Luitz, R. Laskowski, F. Tran, and L. D. Marks, *WIEN2k: An Augmented Plane Wave Plus Local Orbitals Program for Calculating Crystal Properties* (Vienna University of Technology, Austria, 2018), ISBN: 3-9501031-1-2.
- ³²F. Tran, P. Kovács, L. Kalantari, G. K. H. Madsen, and P. Blaha, *J. Chem. Phys.* **149**, 144105 (2018).
- ³³P. Haas, F. Tran, P. Blaha, K. Schwarz, and R. Laskowski, *Phys. Rev. B* **80**, 195109 (2009).
- ³⁴D. C. Langreth and M. J. Mehl, *Phys. Rev. B* **28**, 1809 (1983); Erratum, **29**, 2310 (1984).
- ³⁵R. Neumann, R. H. Nobes, and N. C. Handy, *Mol. Phys.* **87**, 1 (1996).
- ³⁶J. G. Brandenburg, J. E. Bates, J. Sun, and J. P. Perdew, *Phys. Rev. B* **94**, 115144 (2016).
- ³⁷Y. Yao and Y. Kanai, *J. Chem. Phys.* **146**, 224105 (2017).
- ³⁸E. R. Johnson, A. D. Becke, C. D. Sherrill, and G. A. DiLabio, *J. Chem. Phys.* **131**, 034111 (2009).



Machine-learning Prediction of Infrared Spectra of Interstellar Polycyclic Aromatic Hydrocarbons

Péter Kovács¹, Xiaosi Zhu², Jesús Carrete¹, Georg K. H. Madsen¹, and Zhao Wang^{1,2} 

¹Institute of Materials Chemistry, TU Wien, A-1060 Vienna, Austria; zw@gxu.edu.cn

²Laboratory for Relativistic Astrophysics, Department of Physics, Guangxi University, 530004 Nanning, People's Republic of China

Received 2020 July 16; revised 2020 September 2; accepted 2020 September 5; published 2020 October 16

Abstract

We design and train a neural network (NN) model to efficiently predict the infrared spectra of interstellar polycyclic aromatic hydrocarbons with a computational cost many orders of magnitude lower than what a first-principles calculation would demand. The input to the NN is based on the Morgan fingerprints extracted from the skeletal formulas of the molecules and does not require precise geometrical information such as interatomic distances. The model shows excellent predictive skill for out-of-sample inputs, making it suitable for improving the mixture models currently used for understanding the chemical composition and evolution of the interstellar medium. We also identify the constraints to its applicability caused by the limited diversity of the training data and estimate the prediction errors using an ensemble of NNs trained on subsets of the data. With help from other machine-learning methods like random forests, we dissect the role of different chemical features in this prediction. The power of these topological descriptors is demonstrated by the limited effect of including detailed geometrical information in the form of Coulomb matrix eigenvalues.

Unified Astronomy Thesaurus concepts: Polycyclic aromatic hydrocarbons (1280); Interstellar molecules (849); Infrared astronomy (786); Neural networks (1933)

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are among the most widely studied organic compounds in the fields of astronomy (Herbst & van Dishoeck 2009), chemistry (Zhang et al. 2015), biology, and environmental science (Ravindra et al. 2008; Moorthy et al. 2015). As some of the most abundant molecules in the universe (Snow & Witt 1995), PAHs are understood to play an essential role in the evolution of the interstellar medium (ISM; Tielens 2008; Hardegree-Ullman et al. 2014; McGuire et al. 2018; Qi et al. 2018; Hanine et al. 2020). They are also thought to have acted as elemental building blocks of complex organic molecules related to the origin of life (Ehrenfreund & Charnley 2000). Since the infrared (IR) spectrum of a molecule contains valuable information of the molecular bonding configuration (Meier 2003; Neubrech et al. 2017), IR spectroscopy has become an indispensable tool in many observatory projects such as the Stratospheric Observatory for Infrared Astronomy and the Spitzer Space Telescope (Young et al. 2012; Deming & Knutson 2020).

Due to their complex structure–property relationship, identifying PAHs from their IR spectra is anything but straightforward. Since IR activity is related to changes in the molecular dipole moment, a good characterization requires knowledge of both the dynamics of the atomic nuclei, specifically in the form of a set of normal modes, and of the electronic charge distribution. In many cases, the best option available at a reasonable computational cost is density functional theory (DFT). Unfortunately, the number of possible existing PAH species in the ISM is so vast that a brute-force application of DFT is unlikely to be successful in interpreting experimentally measured IR spectra from mixtures of arbitrary molecules (Andrews et al. 2015; Croiset et al. 2016; Shannon et al. 2018). Indeed, while the “unidentified” infrared emission (UIE) features dominating the mid-IR spectra of a wide variety of interstellar sources has been linked to PAHs (Allamandola et al. 1999; Maltseva et al. 2015; Bouwman et al. 2019) the exact

chemical species responsible for UIE are still under debate (Kwok & Zhang 2011; Li & Draine 2012; Kwok & Zhang 2013). Therefore, developing efficient approaches to the prediction of IR spectra of interstellar PAHs remains an important goal with a view to the accurate identification of the UIE band carriers among other sources.

Recently, the rapid development of machine-learning (ML) methods has opened new and reliable ways of investigating molecular structure–property relationships (Gastegger et al. 2017; Butler et al. 2018; Marquez-Neila et al. 2018; Ghosh et al. 2019). However, vibrational spectra are a challenging property for any ML method as they cannot be explained in terms of global composition or local bonding, but depend on hybridizations involving many atoms. In the present study we aim at developing a neural-network (NN) based accelerated model to predict the IR spectra of PAHs using just their skeletal formula as an input. Such formulas encode the topology of the molecule without reference to the exact coordinates of the atoms and, despite their abstraction of the geometric details, are central in any discussion of the structure of organic molecules and exhibit a large amount of predictive power. They thus provide an ideal starting point for the kind of accelerated model that we develop here and do not require computationally intensive electronic structure calculations to determine optimized geometries.

We present an efficient, data-driven approach to the prediction of the IR spectra of PAHs that combines a NN and inputs extracted from the NASA Ames PAH IR spectroscopic database. The potential of NNs to predict the IR spectra of organic molecules was explored more than 20 yr ago (Weigel & Herges 1996; Selzer et al. 2000). However, the discriminatory power of these pioneering attempts was not particularly convincing. Moreover, recent developments in deep learning have led to large improvements in the effectiveness of NN. The results of a NN depend intricately on the descriptors and penalty function used for training. In the present paper we demonstrate that it is possible to obtain good

predictive power from a multilayer NN trained on Morgan descriptors that represented skeletal formulas. We discuss the success of these descriptors in detail and show how including the Coulomb matrix (Rupp et al. 2012; Schütt et al. 2014), which has otherwise been very successful for encoding molecular structure, does not improve the predictive power of the model.

We furthermore train random forests (RFs) on the same data. Generally speaking, RFs have lower quantitative predictive skill than well-trained NNs. However, they can be easier to interpret. For instance, RFs have an intrinsic metric for the importance of each feature that can be computed simply by reverting all the choices based on that feature. Moreover, they are naturally resistant to overfitting and work well with correlated inputs. In this study we use NNs to provide quantitative predictions and RFs to take a closer look at the effect of adding and removing information from the input.

2. Methods

2.1. Data Set

The NASA Ames Research Center has assembled computed and experimental IR spectra of PAHs into a public database, available since 2010 (Bauschlicher et al. 2010). This database comprises more than 3000 spectra, and has undergone two major updates (Boersma et al. 2014; Bauschlicher et al. 2018). It includes PAH IR spectra obtained using DFT, as well as a number of spectra from experimental measurements. This database is an important tool for determining the IR spectra of PAHs in order to develop and test hypotheses regarding astronomical PAHs (Allamandola et al. 1989; Draine & Li 2007; Tielens 2008; Peeters 2011).

For this work, we take the structures and IR spectra of all PAHs in the computational data set version 3.00. As detailed in the next subsection, we use topological descriptors, so we discard those cases in which several geometries exist that are compatible with the same topology. Such cases include, but are not limited to, topologically equivalent structures with different charge states. That leaves us with 2670 molecules from the 3129 in the database. We then turn the set of discrete lines of the IR spectrum into a histogram with a bin width of 21.39 cm^{-1} determined using Knuth's Bayesian rule (Knuth 2006). Each histogram consists of 252 bins covering the range from 6.95 to 5376 cm^{-1} . Bins beyond the 176th (i.e., beyond 3751 cm^{-1}) are discarded since there is a single compound in the whole database contributing to this region of the spectrum with a single (and possibly spurious) peak. We then split each of the truncated histograms into a low-frequency and a high-frequency part. The splitting is done at the 106th bin (2253 cm^{-1}) because it lies in the middle of a gap without contributions from any compound in the database. Frequencies above this cutoff typically correspond to localized vibrations involving hydrogen atoms. Finally, each of the two subhistograms is normalized with the obvious exception of histograms composed entirely of zeros. At the end of this round of preprocessing, the IR spectrum of each compound in the database is represented by two vectors, one for the low-frequency part of the histogram and the other for the high-frequency part, and the components of each add up to one. Those vectors are the targets for the prediction.

2.2. Loss Function

A key piece of a good ML model is a suitable loss function, i.e., a target to be minimized during the training process. A common way to build such a function is to introduce a notion

of distance between output values and then sum the distances between the known and predicted values of the output over the training set. In the context of the current application, each of those values comes in the form of an array representing a normalized histogram. Therefore, it is of critical importance to define a sensible idea of distance between two histograms that takes into account the nature of the elements in those arrays. Among the requirements for that distance is that a slight misprediction of the position of a line should contribute less to the distance than a significantly larger error in that prediction. General-purpose distances like the Euclidean norm of the difference between histograms do not fulfill this criterion, since they do not take the distance between bins into account. Therefore we opt for a more specialized function, in particular a version of the earth mover's distance (EMD; Monge 1781; Dobrushin 1970). Introduced in 1781 in the context of the literal transport of dirt between two sites and now known to be a special case of the Wasserstein metric, the EMD measures the minimal cost of transforming a histogram into another when the cost of moving a unit of mass from bin i to bin j is set to a fixed nonnegative value c_{ij} . We specifically make c_{ij} proportional to the distance between the center of the bins, $|i - j|$. With this choice, if one takes the spectrum of a molecule and introduces a random perturbation in all frequencies of the order of the bin width, the distance between the two histograms will be rather small, and in particular much smaller than the distance to another arbitrary molecule. In contrast, big errors in the placement of lines will increase the distance much more significantly. Moreover, this particular choice of costs allows for a simple and efficient implementation of the EMD. Let $\mathbf{a} = (a_i)_{i=1}^N$ and $\mathbf{b} = (b_i)_{i=1}^N$ be two normalized histograms with the same set of bins, and $(A_i)_{i=1}^N$ and $(B_i)_{i=1}^N$ the corresponding cumulative histograms, with $A_i = \sum_{j=1}^i a_j$ and a similar expression for B_i . The distance between the histograms is computed as

$$\text{EMD}(\mathbf{a}, \mathbf{b}) = \sum_i |A_i - B_i| = \sum_i \left| \sum_{j \leq i} (a_j - b_j) \right|. \quad (1)$$

There is a clear connection with other measures of differences between distributions like the Kolmogorov–Smirnov statistic (Smirnov 1944), in which the distance is the maximum value of $|A_i - B_i|$.

In addition to its role in building the loss function, we also employ the EMD to evaluate the quality of a prediction and to quantify the similarity between two spectra.

To illustrate the distribution of EMD values, the IR spectra of perylene is calculated with two different hybrid functionals and basis sets using the NWChem software package (Valiev et al. 2010). The spectra are then scaled according to the prescription for version 3.0 of the database (Bauschlicher et al. 2018), which splits the peaks in three different regions and scales the frequencies for these regions individually to get better agreement with experimental results. Figure 1 shows the EMD values between these calculated spectra. As expected, the (B3LYP, 4-31g) calculation matches the data included in the database. A change in the functional introduces only small differences, whereas changing the basis set causes larger EMD values between the calculated spectra. The worst agreement can be found between the (B3LYP, 4-31g) and (PBE0, cc-pVDZ) calculations, as illustrated in the bottom half of Figure 1. The EMD in that case is 2.79, which we will use as

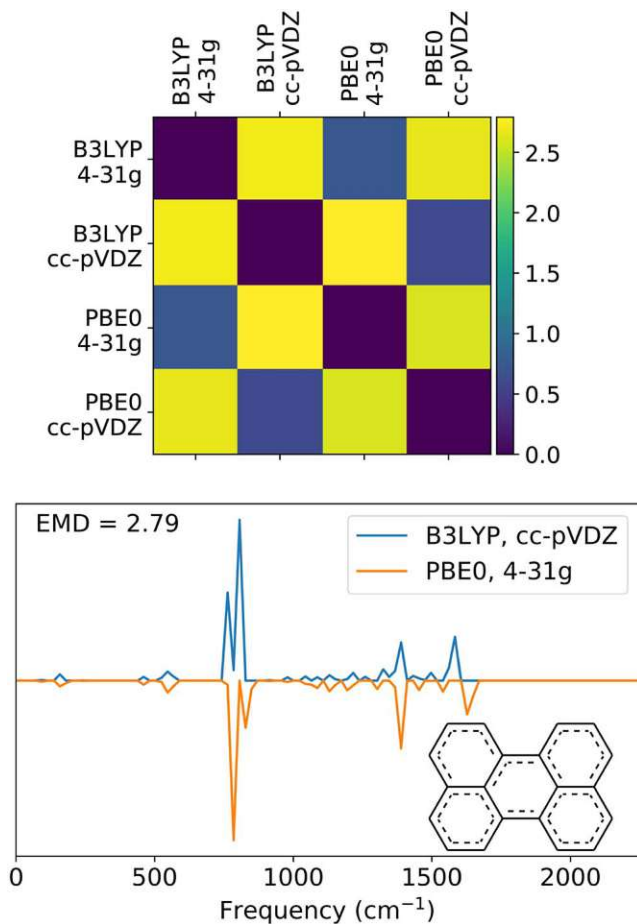


Figure 1. Top: EMD between the DFT-calculated IR spectra obtained using different functionals and basis sets for perylene. The database contains the spectrum from a B3LYP, 4-31g calculation. Bottom: direct comparison of the most dissimilar IR spectra predicted by DFT calculations with different functionals and basis sets in the case of perylene.

a reference value for “good” predictions since they are comparable to the variance among DFT calculations.

2.3. Descriptors

We focus on topological fingerprints as our primary descriptors of the skeletal formula. Often used for substructure and similarity searching, such fingerprints express whether a molecular graph contains particular subgraphs and how many copies of those subgraphs it contains. We specifically use the implementation of extended connectivity fingerprints (ECFPs; Rogers & Mathew 2010) in RDKit³. These fingerprints are calculated using a modified version of the Morgan algorithm (Morgan 1965), originally designed to create a canonical numbering scheme for atoms in molecules. Each nonhydrogen atom is initially assigned a 32-bit integer identifier derived from the properties used in the Daylight atomic invariants rule (Weininger et al. 1989). The algorithm then proceeds for a predefined number of iterations, replacing that identifier with the hash of an array formed by the identifiers of the atom and its first neighbors listed in a deterministic order. In Figure 2 four examples of substructures are shown, generated by 0 (red, blue), 1 (green), and 2 (yellow) iterations. The results of all iterations are put together and the occurrences of each

substructure counted to create the final fingerprint. To be able to detect large substructures of potential relevance for the low-frequency portions of the spectrum, we perform 11 iterations of the algorithm.

The structures are extracted from the database in the form of a set of atomic coordinates, which are then converted to the SDF format (Dalby et al. 1992) using Open Babel (O’Boyle et al. 2011) and finally to a simplified molecular-input line-entry system (SMILES) string from which the aforementioned descriptors are extracted. Since the XYZ to SDF conversion involves the use of bond detection heuristics not guaranteed to work in every case, the conversion fails for 18 compounds and those are dropped at this stage.

As mentioned in the previous section, topologically equivalent structures with different charge states were removed from the data set. The remaining charged molecules with unique descriptors make up around 7% of the data set. The average EMD for these molecules (5.3) is around 2.9 times larger than that for the neutral ones (1.8). We conclude that charged molecules are not accurately described by the SMILES in our processing pipeline. To check if this has an impact on the results, we also train NNs only on neutral molecules. However, the average EMD remains the same and the only significant difference was a thinner tail in the EMD histogram. We therefore keep those charged molecules.

The topological fingerprints cannot encode geometric information, and we aim at developing a model that does not rely on such information. To assess whether this choice influences the result we also train models based on the eigenvalues of the Coulomb matrix (Rupp et al. 2012).

2.4. Machine-learning Models

2.4.1. Neural Networks

NNs are one of the most powerful families of ML techniques in use today (Bishop 1996), and also one of the most widespread, partly because of the existence of high-performance implementations for both CPUs and GPUs. An NN is a graph consisting of layers of nodes, or neurons. In the cases of interest for this discussion, each node produces an output based on a linear combination of the nodes from the previous layer plus a constant. Specifically, the output from a neuron is computed as

$$Y = f\left(\sum_{i=1}^N w_i X_i + b\right), \quad (2)$$

where X_i is the output of the i th neuron in the previous layer, w_i is the connection weight to the current neuron, b is the bias and f is the activation function, responsible for the nonlinearity in the network. During training, all the weights and biases of the neural network are fitted so as to minimize a target penalty, which in the present case is the EMD, described by Equation (1).

Due to their many and diverse applications, different NN architectures have been devised, such as convolutional NNs (where all neurons in a layer share their weights, but their sparse connections to the previous layer are displaced in a systematic way) or recurrent neural networks (where the output of the NN is fed to it again as an input). However, in our case we do not need to capture the type of features those architectures were designed for, so we opt for an archetypal fully connected multilayer NN, where each neuron is connected

³ <http://www.rdkit.org>

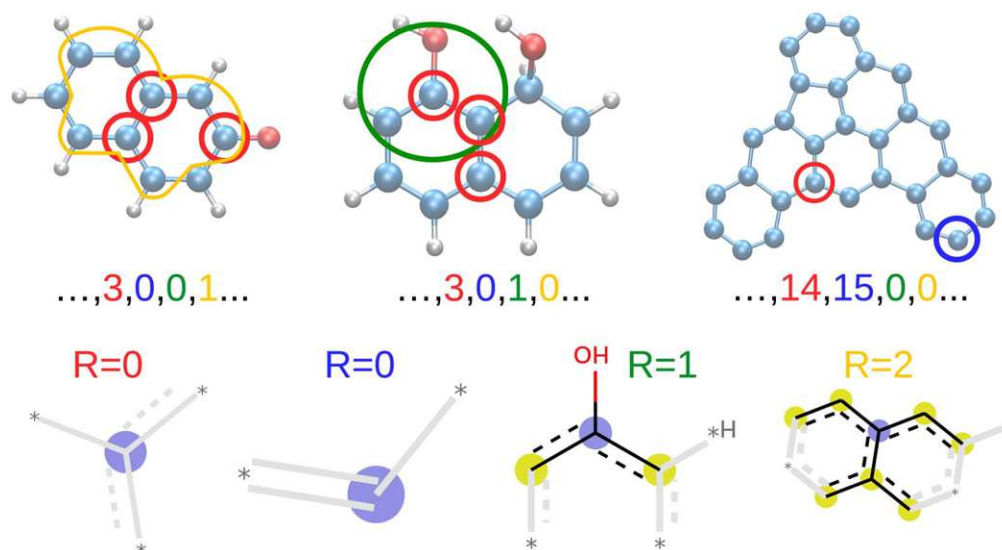


Figure 2. Illustration of how the topological descriptors are built based on the presence of different molecular fragments. Example molecules are shown on top, with marks showing the positions of the specific fragments depicted below. The middle line shows the corresponding region of the generated counting fingerprints. The red and blue fragments are generated by the first iteration of the fingerprint generation, so they only contain information about their base atom. In the current case, the red items represent carbon atoms with three nonhydrogen neighbors and no hydrogens connected to them, while the blue items represent carbon atoms with two nonhydrogen bonds and, likewise, with no hydrogen neighbors. The green and yellow circles show fragments generated by the second and third iterations, respectively. During training we use more than 9200 unique fragments generated by up to 11 iterations of the algorithm.

to every node of the previous and the next layers. Aside from the input and the output layers, our network has four hidden layers with 1500, 1000, 850, and 600 neurons, respectively. As activation functions we use rectified linear units (Lecun et al. 2015) with an extra linear layer and absolute value function before the output.

The input data set is randomly split in training, test, and validation subsets containing 70%, 15%, and 15% of the data, respectively. The inputs to the NN are the descriptors defined above, after removing all elements that did not appear in the training set. The feature vectors so constructed contain 9231 ± 15 elements. The weights of the NN are initialized using the Glorot algorithm (Glorot & Bengio 2010) and all the biases are initialized to zero. The training is then carried out with an Adam optimizer (Kingma & Ba 2015) with the EMD as the target, until the validation error fails to decrease for 50 consecutive epochs. The model is implemented, trained and evaluated using TensorFlow (Abadi et al. 2015). The results presented are calculated with an ensemble of 40 individual NNs for whose training the training/validation/test split was performed in 40 different ways, always according to the proportions quoted above.

2.4.2. Random Forests

The second ML technique that we use is RFs (Breiman 2001). Each RF regressor (or classifier, as the case might be) consists in an ensemble of classification trees, each of them trained on a random subset of the observations (a technique known as “bagging”) and performing splits at each level of the tree based on a random subset of the variables. The result of the RF regression for a new structure is obtained by running the set of descriptors down each tree, obtaining the corresponding individual predictions, and averaging them. Since the prediction of each tree so built is always a value from the training set, a RF regressor of this type has a strong centralizing tendency (Carrete et al. 2014).

We use the implementation of RF in scikit-learn (Pedregosa et al. 2011). Our forests contain 1000 trees each. A tree stops growing when leaf nodes contain just one element or when the depth (i.e., the number of splits) equals 15. A maximum of four features are considered when looking for an optimal split. The splitting criterion is the minimization of the mean square error instead of the EMD. The RFs are mainly used as an interpretative tool and an EMD splitting criterion resulted in training times that were too long.

3. Results and Discussion

In the following we present the results of our NN and RF models.

3.1. NN Performance Metrics

As described in the previous section, we train the NN architecture separately for the low- and high-frequency parts of the spectrum. Figures 3 and 4 provide, respectively, a quantitative and a qualitative window into the performance of the NN for the low-frequency part. This is often labeled the fingerprint region (Smith 2011), is used to identify the molecule, and comprises most of the mass of the histograms. More specifically, Figure 3 (top panel) shows how the model for the low-frequency part performs, by way of the distribution of the EMD between each database record of the test set and the corresponding prediction. Most of the EMD values are found well below the 2.8 reference value extracted from DFT calculations on perylene with different parameters (Figure 1).

A reasonable criterion for what values of the EMD can be considered as good is the discriminatory power, i.e., to be useful, a predicted spectrum for a molecule A should be significantly closer to the “real” spectrum of A than the spectrum of some other molecule B. Therefore, Figure 3 also shows, in the bottom panel, the distribution of EMDs between pairs of structures in the database. The distribution is color coded according to the 25%, 50%, and 75% percentiles. Most

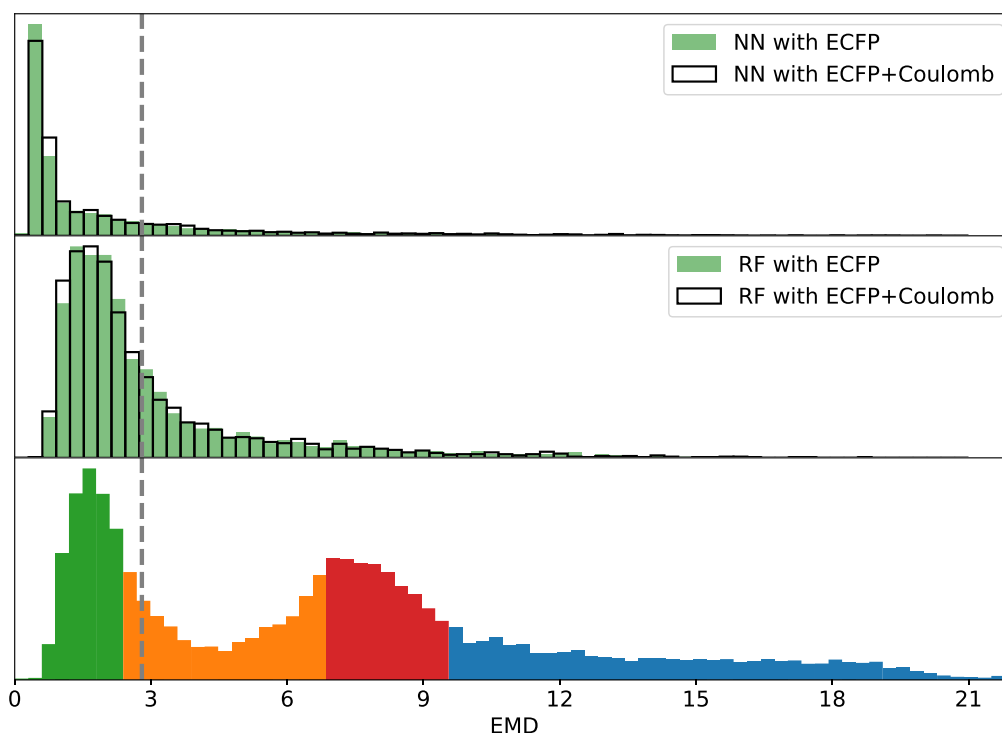


Figure 3. Distribution of EMD values between each individual NN (top) and RF (middle) predictions and the database spectra. The green bars give the results for models trained using only the topological fingerprints and the empty bars the results for models trained also using the 10 largest eigenvalues of the Coulomb matrix. The bottom panel shows the EMDs between random pairs of structures from the database. As a reference, the 25%, 50%, and 75% percentiles of the EMDs between random pairs are shown as changes in color in the bottom panel. The gray vertical line shows the reference 2.8 EMD from Figure 1. All EMDs shown here correspond to the low-frequency parts of the spectra.

of the NN-predicted spectra yield EMDs well below those marks, with 73%, 92%, and 95.7% below the first, second, and third quartiles, respectively, indicating a good predictive skill of the model. Another baseline is the average EMD of 5.79 between random samples of the database, which could indicate if the model has any predictive power at all.

An interesting feature of the baseline histogram in Figure 3 (bottom) is its bimodal structure, with a clear divisory line around an EMD of 4.5. Both of the peaks are well populated, with $\sim 20\%$ of the compound pairs in the central region of each. One possible explanation of this behavior could be that the database contains groups of molecules with relatively small intragroup EMDs, thus creating the first peak in Figure 3 (bottom), and significantly larger intergroup EMDs, forming the second peak and the long tail that contains a further $\sim 25\%$ of the compound pairs. We put this hypothesis to the test using a clustering algorithm, specifically *k*-medoids (Hastie et al. 2001) because it allows us to use a custom metric for clustering, which we set to the low-frequency EMD depicted in the histogram. We select the optimum number of clusters as that maximizing the average silhouette score (Rousseeuw 1987), i.e., the average over all structures of $(b - a) / \max(a, b)$, where a is the mean intracluster distance for that particular structure and b is the distance from the structure to the nearest cluster other than its own. This optimum number turns out to be two; the clusters contain 67% and 33% of the structures, and their median intracluster EMDs are 2.2 and 7.3, respectively. Therefore, the initial hypothesis is false: as a matter of fact, the database consists of a core of closely related compounds (the first cluster) and a second cluster of more loosely similar ones. Each of the peaks in the baseline histogram contains the intracluster distances of one of these two clusters, and the tail of the distribution comes mostly from intercluster distances.

An additional indicator of the good performance of the NN model is the fact that this bimodal structure is absent from the top panel of Figure 3. This goes to show that the model does not act as a mere nearest-neighbor interpolant, looking for similar molecules whose spectrum to copy, but is actually able to pick up different structural features from each molecule and build an accurate prediction based on them.

As a more qualitative illustration, Figure 4 shows four example spectra to provide the reader with an idea of what can be considered a good or a poor prediction in the context of our model. The four structures are chosen at random from each of the quartiles of the NN EMD distribution. The quartiles for the NN results are 0.49, 0.82, 2.54, and 23. This means that 75% of the predicted spectra have an EMD < 2.54 , which is comparable to the EMD between DFT predictions obtained with different DFT parameterizations (Figure 1). Figure 4 also illustrates how even relatively large EMDs are qualitatively informative, which underlines the suitability of the EMDs as a tailored distance metric and the modified Morgan fingerprints as a molecular descriptor.

The NN for the high frequencies yields a median EMD of 0.15. While this value is remarkably small in the context of Figure 3, the median EMD between the high-frequency parts of two random histograms is only 0.33. In fact, as will be seen in more detail below, the high-frequency parts of the spectra have a narrow unimodal distribution and contain far less detail that can or needs to be predicted. We trained a binary classifier to try and predict which molecular structures have a high-frequency part in their histograms at all. After trying both NNs and RFs for this task, we find that it is easy to build a classifier with perfect precision and almost perfect recall, that is, free of false positives and with a single false negative. The reason for this unusually high level of

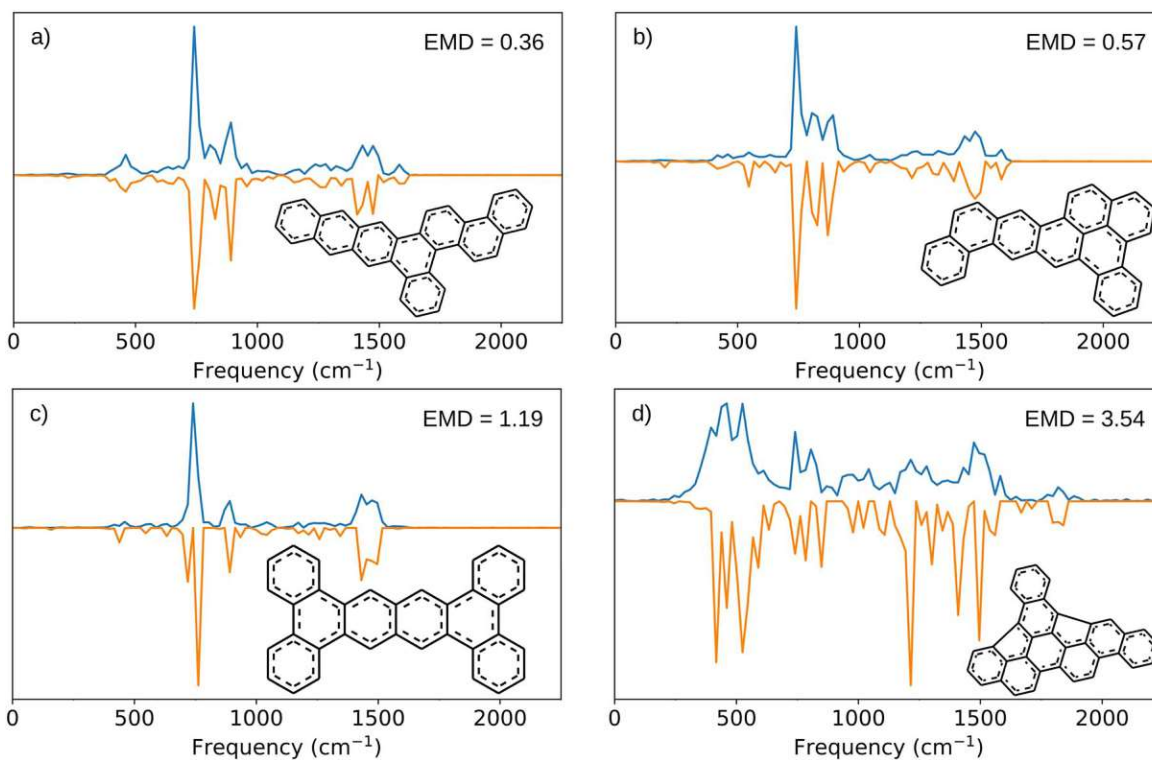


Figure 4. Comparisons between the database result (blue lines) and the NN prediction (orange lines) for four compounds in the test set drawn at random from the regions between (a) 0% and 25%, (b) 25% and 50%, (c) 50% and 75%, and (d) 75% and 100% of the distribution of EMDs. In other words, the four panels provide examples of what a prediction looks like for different levels of quality, from good to poor. All histograms shown here correspond to the low-frequency parts of the spectra.

skill can be analyzed by looking at the importance of each feature in the RF model or by systematically pruning the input features in the case of the NN. Interestingly, in both cases it is revealed that nearly perfect classification can be obtained by using just a single feature, specifically the fingerprint bit shown as blue in Figure 2, representing an unsaturated carbon atom on the edge of an aromatic ring. In our data set, the molecules containing multiple instances of the mentioned fragment have no hydrogen atoms. This points to C–H bonds as responsible for the localized, high-frequency vibrations, in agreement with physicochemical intuition. The finding provides an example of how ML techniques can replicate domain knowledge and, in particular, well known “rules of thumb”, without any specific guidance from specialists.

3.2. Feature Importance

Our next step consists in training an RF model based on the same data set as the NN. As expected from the discussion in the previous section, the predictive power of the RFs is lower, partly because of the flexibility of the model and partly because the RF was trained to minimize the mean square error and not the EMD. This can be seen in the center panel of Figure 3, which shows the distribution of EMD in the test set. On the other hand, an advantage of RFs is the intrinsic feature importance metric they provide. In the left panel of Figure 5 we show the 10 most important features of the low-frequency RF models trained on the topological fingerprints. The values in each list have been renormalized to assign an importance of one to the most important feature.

A remarkable result is that four features are important for both low- and high-frequency predictions (not shown): 1088 and 1089, 1200, and 1358. The substructures that those features

represent are depicted in Figure 6. Their complexity reveals that ML models of vibrational behavior must consider sequences of many bonds to achieve good predictive skill. For the low frequencies this is intuitively obvious, since those normal modes arise from the hybridization of many individual vibrations and involve many atoms.

3.3. Coulomb Matrix

We then test whether the predictive power can be improved by adding information to the input that the topological fingerprints cannot encode, namely descriptors of the molecular geometry. A priori there are situations where IR spectra depend on their stereochemistry, for instance if the effect of nonbonded interactions between atoms far away in the molecular graph causes large changes in the vibrational frequencies of the structure. It is clear that nothing in the topological descriptors can directly address those situations. However, the real question is whether the connection between topology and geometry is strong enough for PAHs in interstellar space that the former can be used as a proxy for the latter.

To answer this question, the histograms in the top and center panels of Figure 3 show the results of NN and RF regression models based on topological information only (filled) and models obtained when the topological descriptors are supplemented with the 10 largest eigenvalues of the Coulomb matrix (unfilled contours). Comparing each pair of histograms reveals very little improvement in model performance coming from the addition of geometry, especially in comparison with the large differences introduced by switching the underlying model or the loss function. The lower importance of the Coulomb matrix eigenvalues compared to the fingerprints might be due to the

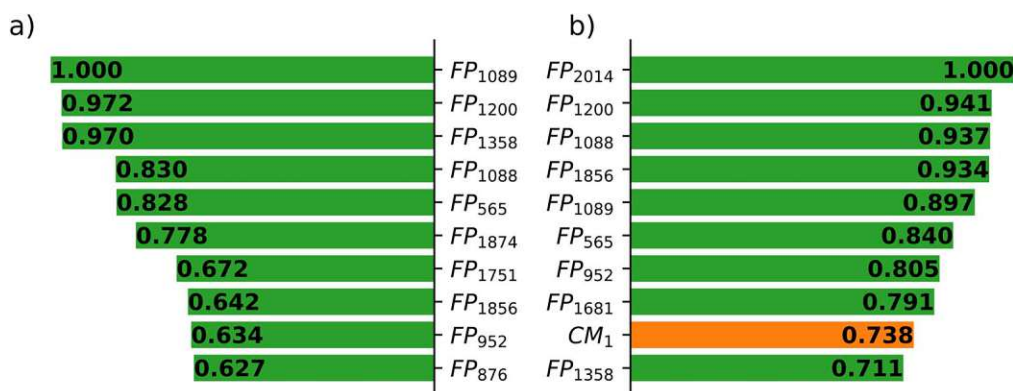


Figure 5. Most important features according to the RF models for the low-frequency part, when only the topological fingerprints are included in the input (left-hand-side panel) and when those are supplemented with the eigenvalues of the Coulomb matrix (right-hand-side panel). Green bars denote topological features, orange bars represent eigenvalues of the Coulomb matrix. Importances have been renormalized so that the first feature in each list has an importance of unity.

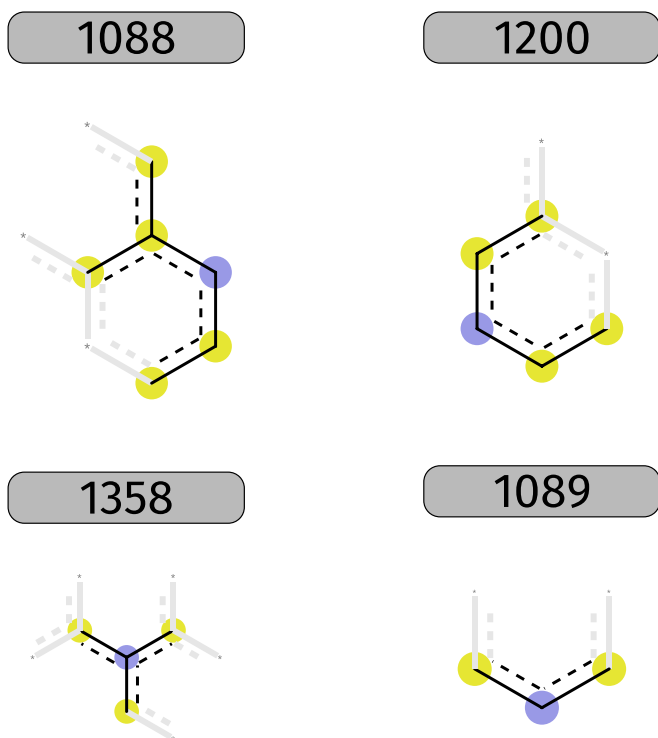


Figure 6. Most important molecular fragments and their unique identifiers for the low-frequency prediction according to the RF model.

loss of information inherent in using a few eigenvalues of a matrix with $4n - 6$ independent elements that encode all structural information of an n -atomic molecule, but also point to part of the geometric structure being predictable from the topology itself, as expected.

The inclusion of the Coulomb matrix eigenvalues does have a discernible effect on the structure of the RFs. This is evidenced by a comparison between the left- and right-hand-side panels of Figure 5, which list the 10 most important features in the models built without and with that information, respectively. However, the four features discussed above remain among the most 10 most important and only one eigenvalue enters this group. It does thus not seem probable that including more structural information would improve the predictive power, and, at least for the PAHs, the skeletal structure is a sufficient descriptor.

3.4. Application to Specific PAHs

Finally, we test the predictive power of the NNs in detail for three PAHs which have recently been discussed in terms of their presence in the interstellar medium (McGuire et al. 2018; Bouwman et al. 2019). Two of these, perylene and peropyrene (Bouwman et al. 2019), are present in the NASA spectroscopic database, while one, benzonitrile (McGuire et al. 2018), is not. Benzonitrile is furthermore very different from the other molecules in the data set from a structural point of view: first of all, it contains a $C\equiv N$ triple bond which is not present in any molecule in the data set and, second, the database is focused on PAHs so there is only another single aromatic ring in the data set (phenol). To show how our method performs for those compounds, we trained 20 additional NNs where perylene and peropyrene were explicitly included in the test set and thus excluded from the training material. Furthermore, we calculated the vibrational spectrum of benzonitrile with DFT using the same prescription employed for the compounds in the database and described in the methodological section.

The EMDs for perylene and peropyrene are 1.44 and 2.29 respectively, both in the same range as the EMD caused by using a different basis set for perylene, Figure 1. The good quality of these predictions is illustrated in Figure 7(a) where it can be seen that the main peaks are found at the right frequencies. As expected, a much larger EMD of 9.56 is found for benzonitrile. This would indicate a poor agreement between the calculated spectra and the model spectra, also illustrated in Figure 7(b). As a regression model, the trained NN regression is inadequate for use on materials that differ substantially from the training set.

3.5. Error Estimation

It is important that the model also be able to identify such cases. As a measure of uncertainty we have defined the cross-NN EMD as the average EMD between every prediction provided by an ensemble of NNs for a given molecule

$$\frac{2}{N(N-1)} \sum_{i \neq j} \text{EMD}(\text{NN}_i, \text{NN}_j),$$

where NN_i is the spectrum predicted by the i th neural network in the ensemble for the molecule and N is the number of NNs in the ensemble.

The poor predictive power for benzonitrile is clearly reflected in these cross-validation EMD. The average and

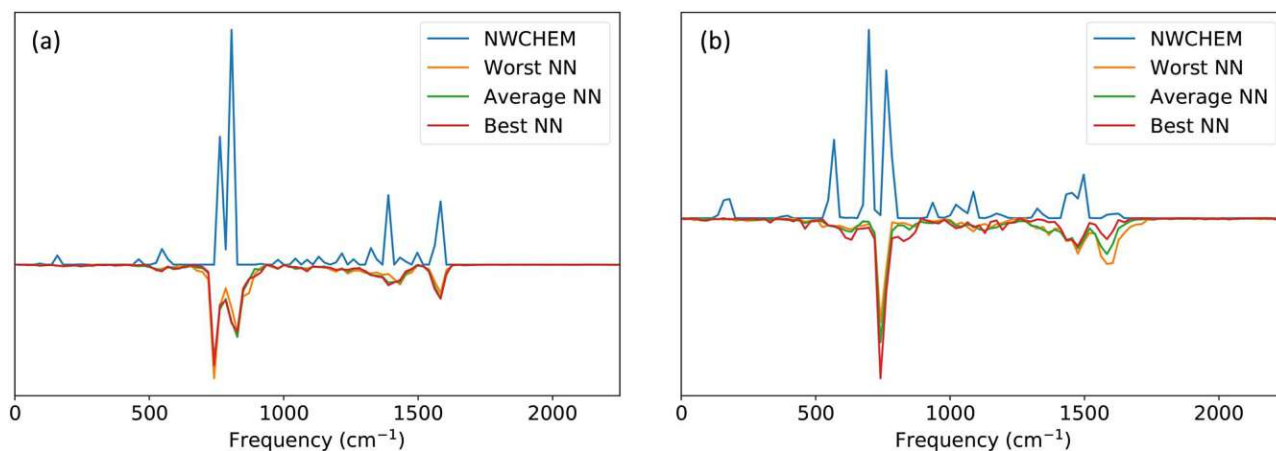


Figure 7. Worst, average, and best neural network predictions for perylene and benzonitrile compared to the database spectra.

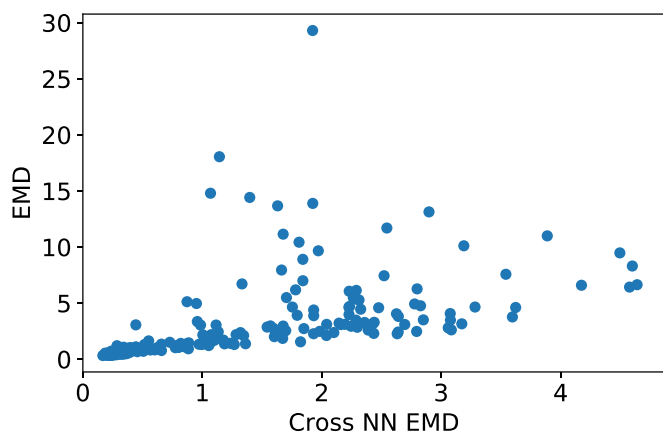


Figure 8. Average EMD from the NN predictions to the database spectrum vs. the average cross EMD between the ensemble of NN predictions.

largest cross-NN EMD are 2.54 and 9.30, respectively. In contrast, for perylene and peropyrene the average/largest cross-NN EMDs are 0.66/1.53 and 0.76/1.60, respectively. In general, this tendency can also be shown for the test set of the original database. In Figure 8 a clear correlation between the average cross EMD and the error compared to the true value can be appreciated. Roughly speaking, we can identify three areas: an average cross EMD below 1 points to a large probability that the predicted spectrum is reliable, an average cross EMD between 1 and 2.5 indicates that the model prediction might still be correct, and an average cross EMD larger than 2.5 is correlated with a model prediction that is probably incorrect.

4. Conclusions

We extract the set of molecular structures from the NASA Ames PAH IR repository and translate them into topological fingerprints identifying the abundance of different chemical fragments in their molecular graph. We also extract the infrared spectra from the database, codify them into histograms, and split them into a low- and a high-frequency part. Using the EMD as a metric we design and train a multilayer NN model to predict each of those parts of the spectrum based on the fingerprints. The resulting models show excellent predictive power for out-of-sample IR spectra, making them suitable for predicting the spectra of larger libraries of PAHs that better

support more accurate interpretations of astronomical IR observations. Moreover, NNs are able to recover identifiable pieces of knowledge, like the role of hydrogen in high-frequency vibrations. This will be helpful for answering the puzzling questions raised by astronomical observations on the chemical composition of the ISM (Li 2020).

We compare the NN predictions with DFT calculations for three different compounds and show that the average error of the NN predictions falls in the range of errors caused by choosing different basis sets for the DFT calculation. We also find that the NN is only applicable to compounds similar to the training set and use multiple NNs to give an approximation of the expected error of the predictions.

We complement this analysis using RF regression. While the accuracy of RFs is lower than what can be achieved with NNs, they allow us to explore which molecular features are most relevant for determining the molecular spectra. We identify four substructures of high importance for both the low- and high-frequency parts of the spectrum. At the same time, however, the results point to a high degree of fungibility among descriptors, whereby similar levels of performance can be achieved using different combinations of those. We also check whether any important information about the molecule is left out by the topological descriptors by supplementing them with geometric information in the form of the largest eigenvalues of the Coulomb matrix. The models do not improve to any significant degree, showing that the topology of the molecular graph alone is enough to satisfactorily characterize the vibrational dynamics of these structures.

This study shows that NNs can be efficiently trained to bypass expensive first-principles calculations, offering useful levels of accuracy and incomparably lower computational cost even for demanding properties like the vibrational spectra. Moreover, it points to the possibility of extracting simple, intuitive rules from trained models that replicate or supplement existing specialist knowledge.

5. Data Access

An example model, data set, and training code are available on Zenodo at doi:10.5281/zenodo.3979217.

P. Blaha and A. Li are acknowledged for helpful discussion. Partial financial support from the National Natural Science Foundation of China (11964002), the Guangxi Science Foundation

(2018GXNSFAA138179), and the Scientific Research Foundation of Guangxi University (XTZ160532) is acknowledged.

ORCID iDs

Zhao Wang  <https://orcid.org/0000-0003-1887-223X>

References

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-scale Machine Learning on Heterogeneous Systems, v1.14. <https://www.tensorflow.org/>
- Allamandola, L. J., Hudgins, D. M., & Sandford, S. A. 1999, *ApJL*, **511**, L115
- Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1989, *ApJS*, **71**, 733
- Andrews, H., Boersma, C., Werner, M. W., et al. 2015, *ApJ*, **807**, 99
- Bauschlicher, C. W. J., Boersma, C., Ricca, A., et al. 2010, *ApJS*, **189**, 341
- Bauschlicher, C. W. J., Ricca, A., Boersma, C., & Allamandola, L. J. 2018, *ApJS*, **234**, 32
- Bishop, C. M. 1996, *Neural Networks for Pattern Recognition* (New York: Oxford Univ. Press)
- Boersma, C., Bauschlicher, C. W. J., Ricca, A., et al. 2014, *ApJS*, **211**, 8
- Bouwman, J., Castellanos, P., Bulak, M., et al. 2019, *A&A*, **321**, A80
- Breiman, L. 2001, *Mach. Learn.*, **45**, 5
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. 2018, *Natur*, **559**, 547
- Carrete, J., Li, W., Mingo, N., Wang, S., & Curtarolo, S. 2014, *PhRvX*, **4**, 011019
- Croiset, B. A., Candian, A., Berne, O., & Tielens, A. G. G. M. 2016, *A&A*, **590**, A26
- Dalby, A., Nourse, J. G., Hounshell, W. D., et al. 1992, *J. Chem. Inf. Comput. Sci.*, **32**, 244
- Deming, D., & Knutson, H. A. 2020, *NatAs*, **4**, 453
- Dobrushin, R. L. 1970, *Theory Probab. Appl.*, **15**, 458
- Draine, B. T., & Li, A. 2007, *ApJ*, **657**, 810
- Ehrenfreund, P., & Charnley, S. B. 2000, *ARA&A*, **38**, 427
- Gastegger, M., Behler, J., & Marquetand, P. 2017, *Chem. Sci.*, **8**, 6924
- Ghosh, K., Stuke, A., Todorovič, M., et al. 2019, *Adv. Sci.*, **6**, 1801367
- Glorot, X., & Bengio, Y. 2010, in *Proc. Machine Learning Research 9*, Proc. Thirteenth Int. Conf. on Artificial Intelligence and Statistics, ed. Y. W. Teh & M. Titterton (MLResearchPress), 249
- Hanine, M., Meng, Z., Lu, S., et al. 2020, *ApJ*, **900**, 188
- Hardegree-Ullman, E. E., Gudipati, M. S., Boogert, A. C. A., et al. 2014, *ApJ*, **784**, 172
- Hastie, T., Tibshirani, R., & Friedman, J. 2001, *The Elements of Statistical Learning* (New York: Springer)
- Herbst, E., & van Dishoeck, E. F. 2009, *ARA&A*, **47**, 427
- Kingma, D. P., & Ba, J. 2015, in 3rd Int. Conf. on Learning Representations, ICLR 2015, ed. B. Bengio & Y. LeCun (San Diego, CA: Conf. Track Proc.), 1, <https://dblp.org/rec/journals/corr/KingmaB14>
- Knuth, K. H. 2006, arXiv:physics/0605197
- Kwok, S., & Zhang, Y. 2011, *Natur*, **479**, 80
- Kwok, S., & Zhang, Y. 2013, *ApJ*, **771**, 5
- Lecun, Y., Bengio, Y., & Hinton, G. 2015, *Natur*, **521**, 436
- Li, A. 2020, *NatAs*, **4**, 339
- Li, A., & Draine, B. T. 2012, *ApJL*, **760**, L35
- Maltseva, E., Petrigiani, A., Candian, A., et al. 2015, *ApJ*, **814**, 23
- Marquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. 2018, *NatAs*, **2**, 719
- McGuire, B. A., Burkhardt, A. M., Kalenskii, S., et al. 2018, *Sci*, **359**, 202
- Meier, R. 2003, in *Handbook of Vibrational Spectroscopy*, ed. J. Chalmers & P. R. Griffiths (Chichester: Wiley)
- Monge, G. 1781, *HARSB*, 666
- Moorthy, B., Chu, C., & Carlin, D. J. 2015, *Toxicol. Sci.*, **145**, 5
- Morgan, H. L. 1965, *J. Chem. Doc.*, **5**, 107
- Neubrech, F., Huck, C., Weber, K., Pucci, A., & Giessen, H. 2017, *ChRv*, **117**, 5110
- O'Boyle, N. M., Banck, M., James, C. A., et al. 2011, *J. Cheminformatics*, **3**, 33
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825, <https://www.jmlr.org/papers/v12/pedregosa11a.html>
- Peeters, E. 2011, in *IAU Symp. 280, The Molecular Universe* (Cambridge: Cambridge Univ. Press), 149
- Qi, H., Picaud, S., Devel, M., Liang, E., & Wang, Z. 2018, *ApJ*, **867**, 133
- Ravindra, K., Sokhi, R., & van Grieken, R. 2008, *AtmEn*, **42**, 2895
- Rogers, D., & Mathew, H. 2010, *J. Chem. Inf. Model.*, **50**, 742
- Rousseuw, P. J. 1987, *JCoAM*, **20**, 53
- Rupp, M., Tkatchenko, A., Müller, K. R., & von Lilienfeld, O. A. 2012, *PhRvL*, **108**, 058301
- Schütt, K. T., Glawe, H., Brockherde, F., et al. 2014, *PhRvB*, **89**, 205118
- Selzer, P., Gasteiger, J., Thomas, H., & Salzer, R. 2000, *CEJ*, **6**, 920
- Shannon, M. J., Peeters, E., Cami, J., & Blommaert, J. A. D. L. 2018, *ApJ*, **855**, 32
- Smirnov, N. V. 1944, *Uspekhi Mat. Nauk*, **10**, 179, <http://mi.mathnet.ru/eng/umn8798>
- Smith, J. G. 2011, *Mass Spectrometry and Infrared Spectroscopy* (3rd edn.; New York: McGraw-Hill), 463
- Snow, T. P., & Witt, A. N. 1995, *Sci*, **270**, 1455
- Tielens, A. G. G. M. 2008, *ARA&A*, **46**, 289
- Valiev, M., Bylaska, E. J., Govind, N., Kowalski, K., & Straatsma, T. 2010, *CoPhC*, **181**, 1477
- Weigel, U. M., & Herges, R. 1996, *Anal. Chim. Acta*, **331**, 63
- Weininger, D., Weininger, A., & Weininger, J. L. 1989, *J. Chem. Inf. Comput. Sci.*, **29**, 97
- Young, E. T., Becklin, E. E., Marcum, P. M., et al. 2012, *ApJL*, **749**, L17
- Zhang, L., Cao, Y., Colella, N. S., et al. 2015, *AcChR*, **48**, 500

Orbital-free approximations to the kinetic-energy density in exchange-correlation MGGA functionals: Tests on solids

Cite as: J. Chem. Phys. **149**, 144105 (2018); <https://doi.org/10.1063/1.5048907>

Submitted: 18 July 2018 . Accepted: 14 September 2018 . Published Online: 08 October 2018

 Fabien Tran, Péter Kovács, Leila Kalantari, Georg K. H. Madsen, and Peter Blaha



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Rungs 1 to 4 of DFT Jacob's ladder: Extensive test on the lattice constant, bulk modulus, and cohesive energy of solids](#)

The Journal of Chemical Physics **144**, 204120 (2016); <https://doi.org/10.1063/1.4948636>

[Bethe-Salpeter correlation energies of atoms and molecules](#)

The Journal of Chemical Physics **149**, 144106 (2018); <https://doi.org/10.1063/1.5047030>

[A consistent and accurate ab initio parametrization of density functional dispersion correction \(DFT-D\) for the 94 elements H-Pu](#)

The Journal of Chemical Physics **132**, 154104 (2010); <https://doi.org/10.1063/1.3382344>



Challenge us.
What are your needs for periodic signal detection? 

 Zurich Instruments

Orbital-free approximations to the kinetic-energy density in exchange-correlation MGGA functionals: Tests on solids

Fabien Tran, Péter Kovács, Leila Kalantari, Georg K. H. Madsen, and Peter Blaha
*Institute of Materials Chemistry, Vienna University of Technology, Getreidemarkt 9/165-TC,
 A-1060 Vienna, Austria*

(Received 18 July 2018; accepted 14 September 2018; published online 8 October 2018)

A recent study of Mejia-Rodriguez and Trickey [Phys. Rev. A **96**, 052512 (2017)] showed that the deorbitalization procedure (replacing the exact Kohn-Sham kinetic-energy density by an approximate orbital-free expression) applied to exchange-correlation functionals of the meta-generalized gradient approximation (MGGA) can lead to important changes in the results for molecular properties. For the present work, the deorbitalization of MGGA functionals is further investigated by considering various properties of solids. It is shown that depending on the MGGA, common orbital-free approximations to the kinetic-energy density can be sufficiently accurate for the lattice constant, bulk modulus, and cohesive energy. For the bandgap, calculated with the modified Becke-Johnson MGGA potential, the deorbitalization has a larger impact on the results. *Published by AIP Publishing.*
<https://doi.org/10.1063/1.5048907>

I. INTRODUCTION

Kohn-Sham density functional theory^{1,2} (KS-DFT) is a computationally efficient quantum method, which allows the treatment of molecules, surfaces, and solids containing up to several thousands of atoms. KS-DFT is particularly fast when the exchange and correlation (xc) effects are treated at the semilocal level of approximation. The drawback is, however, that there can be some degree of uncertainty in the results with semilocal methods.^{3,4} The most simple semilocal functional E_{xc} is the local density approximation (LDA),^{2,5,6} which is purely a functional of the electron density $\rho = \sum_{i=1}^N |\psi_i|^2$. Higher accuracy can be obtained by using functionals of the generalized gradient approximation (GGA)^{7–10} which depend additionally on the first derivative of ρ ($\nabla\rho$). Nowadays, the most advanced and accurate semilocal functionals are the so-called meta-GGA (MGGA),¹¹ which, in addition of ρ and $\nabla\rho$, depend also on the positive-definite KS kinetic-energy density (KED)

$$t^{\text{KS}}(\mathbf{r}) = \frac{1}{2} \sum_{i=1}^N \nabla\psi_i^*(\mathbf{r}) \cdot \nabla\psi_i(\mathbf{r}) \quad (1)$$

and/or the second derivative of ρ ($\nabla^2\rho$),

$$E_{xc}^{\text{MGGA}} = \int \varepsilon_{xc}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}), \nabla^2\rho(\mathbf{r}), t^{\text{KS}}(\mathbf{r})) d^3r. \quad (2)$$

Considering how constructing a functional using more ingredients brings more flexibility to it, MGGA functionals should be universally more accurate than LDA and GGA functionals. As with GGA functionals, a plethora of MGGA functionals have been proposed (see Ref. 11 for an exhaustive list) and among the recent ones, SCAN¹² and TM¹³ for instance, have shown to be accurate for many types of systems and properties.^{14–22}

As discussed in detail in Ref. 11, most MGGA functionals depend only on the KED t^{KS} , while only very few use

also (or only) $\nabla^2\rho$. One of the main reasons for not using $\nabla^2\rho$ in E_{xc} are the difficulties encountered when calculating the potential (i.e., the functional derivative of E_{xc}) for self-consistent calculations. Indeed, the presence of $\nabla^2\rho$ in E_{xc} means that the potential contains a term, $\nabla^2(\partial\varepsilon_{xc}/\partial(\nabla^2\rho))$, that involves the third and fourth derivatives of ρ (see Ref. 23) which may lead to numerical problems like a greater sensitivity to the integration grid.^{23–26} (To our knowledge, only Ref. 27 reports an implementation of $\nabla^2\rho$ -MGGA with integration by part of the relevant Hamiltonian matrix elements²⁸ to avoid the third and fourth derivatives of ρ .) As a comparison, a GGA potential involves only the first and second derivatives of ρ (or only the first if integration by part in the Hamiltonian matrix²⁹ is done), and a t^{KS} -dependency in a MGGA functional leads to an additional (non-multiplicative) term in the potential, $-(1/2)\nabla \cdot ((\partial\varepsilon_{xc}/\partial t^{\text{KS}})\nabla\psi_i)$, that involves the derivatives of ψ_i up to the second order (or only the first if integration by part in the Hamiltonian matrix²⁸ is done). Therefore, MGGA calculations have been done using mostly t^{KS} -MGGAs and are becoming increasingly popular (see Refs. 30–33 for recent studies reporting self-consistent implementations for periodic solids). Furthermore, from the theoretical point of view, a benefit of using t^{KS} is that regions of space dominated by a single orbital can be detected (see, e.g., Ref. 34), which may be necessary to satisfy known exact constraints.¹²

On the other hand, $\nabla^2\rho$ -MGGAs have the advantage to be explicit functionals of ρ such that the functional derivative leads to a true KS (i.e., multiplicative) potential, which is not the case with t^{KS} -MGGAs. Also, except for the problems with the high derivatives of ρ mentioned above, a new self-consistent implementation of MGGA should be easier for $\nabla^2\rho$ -MGGAs. Furthermore, according to Refs. 35 and 36, $\nabla^2\rho$ -MGGAs lead to faster calculations compared to t^{KS} -MGGAs, which may be of importance for large systems.

Thus, from the fundamental and practical point of views, $\nabla^2\rho$ -MGGAs are still of interest and worth to be further considered as done in recent studies.^{25,26,35,36}

In particular, Mejia-Rodriguez and Trickey²⁶ investigated the effect of replacing the exact orbital-dependent t^{KS} in existing t^{KS} -MGGAs functionals by some orbital-free (OF) approximations t^{OF} . They called this procedure *deorbitalization*, meaning that a t^{KS} -MGGAs is transformed into an explicit density functional $\nabla^2\rho$ -MGGAs. The properties that they considered are the heat of formation, bond lengths, and vibration frequencies of molecules. This study showed that the replacement $t^{\text{KS}} \rightarrow t^{\text{OF}}$ can have some impact on the results depending on the xc-MGGAs or the OF KED. For instance, the average error for the heat of formation is in some cases only slightly modified, while in some other cases it is increased by one order of magnitude. Also, it seems that none of the OF KED they considered, including the three new ones proposed by Mejia-Rodriguez and Trickey, lead to reasonably small changes in all cases.

For the present study, we pursue the investigations on the deorbitalization procedure by considering properties of solids. Several t^{KS} -MGGAs energy functionals will be deorbitalized and tested on the lattice constant, bulk modulus, and cohesive energy, while the deorbitalization of the modified Becke-Johnson potential³⁷ will be considered for the electronic structure. We note that, in a subsequent work of Mejia-Rodriguez and Trickey,³⁶ that was made available just after completion of our work, solids were also considered and basically the same properties were calculated.

The structure of the paper is the following. Section II provides a brief description of the theory and the computational details. In Sec. III, the results obtained with the deorbitalized MGGAs are presented and discussed, while Sec. IV provides some analysis, and Sec. V gives the summary of this work.

II. THEORY AND COMPUTATIONAL DETAILS

A. Orbital-free kinetic energy densities

In the KS-DFT method,² the noninteracting kinetic energy component of the total energy is given by $T_s^{\text{KS}} = \int t^{\text{KS}} d^3r$, where t^{KS} is given by Eq. (1). Note that another common expression for the integrand in T_s^{KS} is $t^{\text{KS}'} = -(1/2) \sum_{i=1}^N \psi_i^* \nabla^2 \psi_i$ which is related to t^{KS} by $t^{\text{KS}'} = t^{\text{KS}} - (1/4) \nabla^2 \rho$ and leads to the same value of T_s^{KS} since the integral of $\nabla^2 \rho$ is zero. For the development of fully OF DFT methods³⁸⁻⁴⁰ or in the framework of embedding schemes,⁴¹⁻⁴⁴ expressions for T_s which are explicit functionals of ρ have been proposed, and as for xc-functionals, the majority of them are of semilocal type. The most simple is the LDA of Thomas and Fermi^{45,46} (TF) which is the exact expression for the homogeneous electron gas and reads

$$T_s^{\text{TF}} = C_{\text{TF}} \int \rho^{5/3}(\mathbf{r}) d^3r, \quad (3)$$

where $C_{\text{TF}} = (3/10)(3\pi^2)^{2/3}$. With respect to the exact values (T_s^{KS}), the TF functional leads to underestimations for atoms⁴⁷ and molecules⁴⁸⁻⁵¹ of about 10%. Since the kinetic energy is

a major component of the total energy E_{tot} (from the virial theorem $T_s \approx -E_{\text{tot}}$), such errors are extremely large. Much better values for T_s can be obtained with gradient-corrected type (GGA) functionals (errors below 0.5% for the best ones⁴⁸⁻⁵³),

$$T_s^{\text{GGA}} = C_{\text{TF}} \int \rho^{5/3}(\mathbf{r}) F_s(s(\mathbf{r})) d^3r, \quad (4)$$

where $s = |\nabla\rho| / \left(2(3\pi^2)^{1/3} \rho^{4/3}\right)$ is the reduced density gradient and F_s is the kinetic enhancement factor for which many forms have been proposed in the literature (see Refs. 51 and 53-55 for compilations) like, for instance, those that were obtained using the *conjointness conjecture* between the exchange and kinetic energy functionals.^{52,56,57} While GGAs can lead to rather accurate (albeit far from enough for an useful OF DFT method) values of T_s , the GGA KEDs defined as the integrand of Eq. (4) show absolutely no resemblance to Eq. (1).⁵⁸⁻⁶¹ This can be understood by considering the density-gradient expansion approximation (GEA) of Eq. (1) which, at the second order, is given by^{62,63} (L in GEA2L indicates the presence of $\nabla^2\rho$)

$$t^{\text{GEA2L}}(\mathbf{r}) = t^{\text{TF}}(\mathbf{r}) + \frac{1}{9} t^{\text{W}}(\mathbf{r}) + \frac{1}{6} \nabla^2 \rho(\mathbf{r}), \quad (5)$$

where $t^{\text{TF}} = C_{\text{TF}} \rho^{5/3}$ [the integrand of Eq. (3)] and $t^{\text{W}} = |\nabla\rho|^2 / (8\rho)$ is the von Weizsäcker⁶⁴ KED. It is only by considering $\nabla^2\rho$ in an OF KED t^{OF} that the shape of t^{OF} can be made reasonably close to t^{KS} (see Refs. 58, 59, 65, and 66), and despite some attempts,⁶⁰ it is most likely hopeless to construct a GGA KED that looks similar to t^{KS} .

Thus, one has to consider $\nabla^2\rho$ -dependent OF KED t^{OF} for a replacement of t^{KS} in a t^{KS} -MGGAs xc-functional with the hope of not changing much the results. As mentioned above, a term $c\nabla^2\rho$ (c is a constant) in the KED [like in Eq. (5)] integrates to zero, but would also not contribute to the kinetic potential $\delta T_s / \delta \rho$ in a OF or embedding scheme since the contribution is $\nabla^2(c\nabla^2\rho) / \partial(\nabla^2\rho) = \nabla^2 c = 0$. However, MGGAs xc-functionals depend on the KED in a more complicated way such that $c\nabla^2\rho$ cannot be discarded.

The $\nabla^2\rho$ -dependent KED t^{OF} that we will consider for a replacement of t^{KS} in xc-MGGAs are now listed (more detail can be found in the respective references). GEA2L^{62,63} as given by Eq. (5). TW02L, which consists of the GGA TW02 proposed in Ref. 52 (a reparametrization of the GGA exchange of Perdew, Burke, and Ernzerhof (PBE)⁹ with $\kappa = 0.8438$ and $\mu = 0.2319$) augmented with $(1/6)\nabla^2\rho$. PC from Perdew and Constantin,⁶⁵ which was constructed to recover the fourth-order GEA in the slowly varying density limit and t^{W} in the rapidly varying limit, as well as to satisfy $t^{\text{W}} \leq t^{\text{PC}}$. CR from Cancio and Redd⁶⁷ [Eqs. (20) and (21) in Ref. 67 with $\alpha = 4$], which was constructed in a rather similar way as PC. GEALoc from Cancio and Redd⁶⁷ [Eq. (37) in Ref. 67], which has the same form as Eq. (5) but with different (optimized) parameters in front of t^{W} and $\nabla^2\rho$. PCopt and CROpt from Mejia-Rodriguez and Trickey²⁶ that are reoptimized versions of PC and CR, respectively. Many other expressions for t^{OF} could also be considered, e.g., any of the integrand (augmented by $c\nabla^2\rho$) of the

numerous proposed T_s^{GGA} or those proposed recently in Refs. 68–71. Nevertheless, our selection of seven different OF KED should be good enough to give us a general idea of the change in the performance of a xc-MGGA when it is deorbitalized.

It is important to mention that, as done, e.g., in Ref. 43, for all considered OF KED, we chose to enforce the lower bound $t^{\text{W}} \leq t$.^{72,73} Thus, it is in fact

$$t^{\text{OF}'}(\mathbf{r}) = \max(t^{\text{OF}}(\mathbf{r}), t^{\text{W}}(\mathbf{r})) \quad (6)$$

that replaces t^{KS} in the MGGA xc-functionals, which is also a way to locally reduce the error in t^{OF} . Note that depending on the MGGA xc-functional, Eq. (6) may be anyway necessary to apply if negative values of $t^{\text{OF}} - t^{\text{W}}$ or t^{OF} lead to problems like, for instance, under a square root. Note that Eq. (6) may introduce a kink in the KED, and therefore also in ε_{xc} , which may translate into a discontinuity in the potential.

We also mention that the generalization of the OF KED formulas for spin-polarized systems is trivially given by⁷⁴ $t[\rho_{\uparrow}, \rho_{\downarrow}] = t_{\uparrow}[\rho_{\uparrow}] + t_{\downarrow}[\rho_{\downarrow}]$, where $t_{\sigma}[\rho_{\sigma}] = (1/2)t[2\rho_{\sigma}]$ with $t[2\rho_{\sigma}]$ being the non-spin-polarized formula in which ρ is replaced by $2\rho_{\sigma}$.

B. MGGA exchange-correlation functionals

The MGGA xc-energy functionals that we will consider to test the accuracy of OF KED are MVS⁷⁵ and SCAN,¹² that were used by Mejia-Rodriguez and Trickey²⁶ for their molecular tests, as well as TM that was proposed by Tao and Mo.¹³ The recent SCAN and TM functionals have been shown to be accurate in many circumstances (see, e.g., Refs. 14–22). Additionally, the modified Becke-Johnson MGGA potential³⁷ (mBJLDA, combined with LDA correlation⁶) will also be used to test the accuracy of OF KED by considering the bandgap. The mBJLDA potential, which is based on the BJ potential,^{76,77} was shown to be much more reliable than the standard LDA and GGA methods for bandgap calculations and to lead to values that are in very good agreement with experiment in most cases.^{37,78–82}

With an energy functional (MVS, SCAN, or TM), the closeness between OF KEDs and the exact KS KED is quantified by considering properties that depend on the total energy (lattice constant, bulk modulus, and cohesive energy). With the mBJLDA potential, properties like band structure or electron density are more interesting to look at.

C. Computational details

The calculations were done with WIEN2k,⁸³ which is an all-electron code based on the linearized augmented plane-wave method.^{84,85} Very good parameters were chosen such that the results are well converged. As in our previous work,¹⁴ the lattice constant, bulk modulus, and cohesive energy obtained with MGGAs were calculated using the GGA PBE⁹ orbitals and density since in WIEN2k there is no implementation of the potential for MGGAs (neither of the non-multiplicative type for t^{KS} -MGGAs nor of the multiplicative type for $\nabla^2\rho$ -MGGAs). As discussed in Ref. 14, the effect of self-consistency on the results should be very small for strongly

bound (i.e., covalent, ionic, metallic) solids. However, self-consistency is expected to affect more the results for weakly bound van der Waals solids. Therefore, this is only via the energy functional that the replacement $t^{\text{KS}} \rightarrow t^{\text{OF}}$ will produce changes in the lattice constant, bulk modulus, and cohesive energy. The calculations of the bandgap with the multiplicative mBJLDA potential were done self-consistently.

III. RESULTS

A. Lattice constant, bulk modulus, and binding energy

We start with the results for the equilibrium lattice constant a_0 , bulk modulus B_0 , and cohesive energy E_{coh} of 44 strongly bound solids (listed in Table S1 of the [supplementary material](#)). Table I shows the mean error (ME), mean absolute error (MAE), mean relative error (MRE), and mean absolute relative error (MARE) with respect to the experiment. The values of a_0 , B_0 , and E_{coh} can be found in Tables S1–S9 and Figs. S1–S24 of the [supplementary material](#). The errors obtained with the parent t^{KS} -MGGA, namely, MVS, SCAN, or TM, are considered as the reference that should be reproduced at best by an OF t^{OF} -MGGA [denoted MGGA(X), where X is one of the OF approximations t^{OF} mentioned in Sec. II A]. Since the amount of results shown in Table I is rather substantial and would make a detailed discussion rather lengthy and tedious, a concise discussion, only in terms of MAE and ME, of the most interesting observations is provided.

In the case of the SCAN and TM xc-functionals, the deorbitalization procedure leads to changes in the MAE and ME that are the smallest if t^{KS} is replaced by t^{GEA2L} , t^{TW02L} , t^{PC} , or t^{CR} . The change in the MAE is in most cases below 0.003 Å for a_0 , 2.5 GPa for B_0 , and 0.03 eV/atom for E_{coh} , such that it is reasonable to consider the overall (in)accuracy of the xc-functional as unaffected by its deorbitalization. t^{PCopt} also belongs to the group of the accurate OF KED in the case of SCAN, but not TM especially for the bulk modulus and cohesive energy. If the deorbitalization of SCAN or TM is done with t^{GEAloc} or t^{CRopt} , then larger changes in the MAE and ME can sometimes, but not systematically, be observed. This seems to be particularly the case with t^{CRopt} , which, for instance, leads for SCAN to changes of 0.023 Å and 3.8 GPa in the MAE of a_0 and B_0 , respectively. t^{CRopt} also leads to the largest change in the MAE of a_0 and E_{coh} for TM. Thus, replacing t^{KS} by t^{GEAloc} or t^{CRopt} , in particular, affects more the accuracy of a functional and would probably change the position of the xc-functional in some performance ranking (see Ref. 14).

Compared to SCAN and TM, the deorbitalization procedure of MVS leads to changes in the MAE that are in general clearly larger. This is due to the analytical form of MVS that depends more strongly on the KED. For instance, for B_0 there is a decrease in the MAE that is in the range 3.2–5.9 GPa, while for E_{coh} the MAE of the t^{OF} -MVS can be decreased by 0.15 eV/atom [with MVS(GEAloc)] or increased by 0.07 eV/atom [with MVS(CRopt)]. Concerning the ME of MVS, t^{GEA2L} , t^{TW02L} , t^{PC} , and t^{CR} are more efficient than t^{GEAloc} , t^{PCopt} , and t^{CRopt} for reproducing the values of MVS. Note that, in terms

TABLE I. The ME, MAE, MRE, and MARE of the parent t^{KS} -MGGA functionals (MVS, SCAN, and TM) with respect to experiment^{86,87} on the testing set of 44 strongly bound solids for the lattice constant a_0 , bulk modulus B_0 , and cohesive energy E_{coh} . The values for the t^{OF} -MGGA functionals are also with respect to experiment, but with the value of the parent functional subtracted, e.g., $\text{ME}(t^{\text{OF}}\text{-MGGA})-\text{ME}(t^{\text{KS}}\text{-MGGA})$. The units of the ME and MAE are Å, GPa, and eV/atom for a_0 , B_0 , and E_{coh} , respectively, and % for the MRE and MARE. The large differences with respect to the parent t^{KS} -MGGA are italic. All results were obtained non-self-consistently using PBE orbitals/density.

Functional	a_0				B_0				E_{coh}			
	ME	MAE	MRE	MARE	ME	MAE	MRE	MARE	ME	MAE	MRE	MARE
MVS	-0.008	0.043	-0.3	0.9	12.2	13.3	8.2	12.7	0.21	0.37	5.8	9.3
MVS(GEA2L)	<i>-0.016</i>	<i>-0.007</i>	<i>-0.3</i>	<i>-0.1</i>	<i>-4.0</i>	<i>-3.4</i>	<i>-1.1</i>	<i>-3.3</i>	<i>-0.03</i>	<i>-0.13</i>	<i>-1.2</i>	<i>-3.0</i>
MVS(TW02L)	<i>-0.007</i>	<i>-0.009</i>	<i>-0.1</i>	<i>-0.2</i>	<i>-4.7</i>	<i>-3.6</i>	<i>-2.5</i>	<i>-4.0</i>	<i>-0.13</i>	<i>-0.13</i>	<i>-3.9</i>	<i>-2.6</i>
MVS(PC)	<i>-0.014</i>	<i>-0.008</i>	<i>-0.2</i>	<i>-0.2</i>	<i>-4.6</i>	<i>-3.2</i>	<i>-1.5</i>	<i>-3.4</i>	<i>-0.08</i>	<i>-0.13</i>	<i>-2.3</i>	<i>-3.0</i>
MVS(CR)	<i>-0.016</i>	<i>-0.007</i>	<i>-0.3</i>	<i>-0.1</i>	<i>-3.9</i>	<i>-3.4</i>	<i>-1.1</i>	<i>-3.3</i>	<i>-0.02</i>	<i>-0.12</i>	<i>-0.8</i>	<i>-2.9</i>
MVS(GEAloc)	0.006	<i>-0.007</i>	0.2	<i>-0.1</i>	<i>-9.3</i>	<i>-5.9</i>	<i>-4.6</i>	<i>-5.2</i>	<i>-0.29</i>	<i>-0.15</i>	<i>-6.9</i>	<i>-3.4</i>
MVS(PCopt)	<i>-0.011</i>	0.001	<i>-0.2</i>	0.0	<i>-8.4</i>	<i>-3.8</i>	<i>-3.0</i>	<i>-3.2</i>	<i>-0.25</i>	<i>-0.08</i>	<i>-5.3</i>	<i>-2.6</i>
MVS(CRopt)	<i>0.045</i>	0.007	<i>1.0</i>	0.1	<i>-17.1</i>	<i>-3.2</i>	<i>-11.8</i>	<i>-3.7</i>	<i>-0.59</i>	0.07	<i>-14.1</i>	1.4
SCAN	0.018	0.030	0.3	0.6	3.5	7.4	-0.5	6.5	-0.02	0.19	-0.7	4.9
SCAN(GEA2L)	<i>-0.012</i>	<i>-0.002</i>	<i>-0.2</i>	0.0	<i>-4.5</i>	2.4	<i>-0.7</i>	1.3	0.05	<i>-0.01</i>	1.0	<i>-0.3</i>
SCAN(TW02L)	<i>-0.007</i>	<i>-0.001</i>	<i>-0.1</i>	0.0	<i>-5.2</i>	2.5	<i>-1.6</i>	1.5	<i>-0.00</i>	0.00	<i>-0.5</i>	0.1
SCAN(PC)	<i>-0.010</i>	<i>-0.001</i>	<i>-0.2</i>	0.0	<i>-5.0</i>	2.7	<i>-1.0</i>	1.4	0.02	0.00	0.3	0.0
SCAN(CR)	<i>-0.012</i>	<i>-0.003</i>	<i>-0.2</i>	0.0	<i>-4.5</i>	2.3	<i>-0.7</i>	1.3	0.06	<i>-0.01</i>	1.1	<i>-0.3</i>
SCAN(GEAloc)	<i>0.016</i>	0.010	<i>0.4</i>	0.2	<i>-10.4</i>	3.4	<i>-3.8</i>	2.4	<i>-0.20</i>	0.06	<i>-4.5</i>	1.3
SCAN(PCopt)	<i>-0.004</i>	<i>-0.002</i>	0.0	0.0	<i>-6.4</i>	0.3	<i>-1.8</i>	0.2	<i>-0.07</i>	<i>-0.02</i>	<i>-1.6</i>	<i>-0.1</i>
SCAN(CRopt)	<i>0.034</i>	<i>0.023</i>	<i>0.8</i>	<i>0.5</i>	<i>-11.7</i>	3.8	<i>-6.2</i>	3.4	<i>-0.28</i>	0.12	<i>-6.6</i>	3.0
TM	-0.006	0.023	-0.2	0.5	2.4	6.6	2.1	6.2	0.24	0.27	6.4	7.0
TM(GEA2L)	<i>-0.005</i>	0.002	<i>-0.1</i>	0.0	<i>-0.9</i>	0.9	<i>-0.5</i>	0.4	<i>-0.01</i>	0.01	<i>-0.3</i>	0.2
TM(TW02L)	<i>-0.003</i>	0.001	<i>-0.1</i>	0.0	<i>-0.9</i>	0.9	<i>-0.8</i>	0.3	<i>-0.02</i>	0.01	<i>-0.7</i>	0.0
TM(PC)	<i>-0.006</i>	0.003	<i>-0.1</i>	0.1	<i>-0.7</i>	1.0	<i>-0.1</i>	0.6	<i>-0.02</i>	0.02	<i>-0.5</i>	0.4
TM(CR)	<i>-0.005</i>	0.002	<i>-0.1</i>	0.0	<i>-0.8</i>	0.9	<i>-0.5</i>	0.4	<i>-0.00</i>	0.01	<i>-0.1</i>	0.2
TM(GEAloc)	<i>-0.010</i>	0.003	<i>-0.2</i>	0.1	<i>-0.4</i>	1.6	0.9	1.2	<i>-0.01</i>	0.03	0.2	1.1
TM(PCopt)	0.004	0.004	0.1	0.1	<i>-2.9</i>	1.7	<i>-1.5</i>	0.9	<i>-0.09</i>	<i>-0.03</i>	<i>-2.0</i>	<i>-0.6</i>
TM(CRopt)	0.009	0.004	0.2	0.1	<i>-3.6</i>	1.0	<i>-2.4</i>	0.5	<i>-0.13</i>	<i>-0.05</i>	<i>-2.8</i>	<i>-1.0</i>

of MAE, MVS(CRopt) seems to be the closest to MVS, but this is fortuitous since the ME are completely different and of opposite sign.

Figure 1 shows for each solid the relative error in the lattice constant and cohesive energy obtained with the parent SCAN and four of its deorbitalized versions. We can see that the results with SCAN(GEA2L) and SCAN(PC), which are basically the same, are very or fairly close to SCAN results in most cases. The most visible exceptions are the alkali and alkaline earth metals for which the SCAN(CRopt) values follow very closely those obtained with SCAN, in particular, for a_0 . We also note some large differences in E_{coh} between SCAN(GEA2L/PC) and SCAN for some of the 3d and 4d transition metals and the ionic compounds. Except for the aforementioned alkali and alkaline earth metals, the lattice constants and cohesive energies obtained with SCAN(CRopt) differ noticeably from SCAN. SCAN(PCopt) leads to results that are intermediate between SCAN(GEA2L/PC) and SCAN(CRopt).

Thus, in summary the performance of a xc-MGGA functional for strongly bound solids is modified the least when

t^{KS} is replaced by t^{GEA2L} , t^{TW02L} , t^{PC} , t^{CR} , or t^{PCopt} . For SCAN and TM, the performance is overall barely changed by the deorbitalization using one of these OF KED, but more for MVS.

Although the goal of replacing t^{KS} by t^{OF} in a xc-MGGA was not to improve the agreement with experiment, we mention that it is sometimes the case. By looking at the MA(R)E in Table I, we can see that, for instance, the deorbitalization of MVS reduces the values for a_0 , B_0 , and E_{coh} .

In their work, Mejia-Rodriguez and Trickey²⁶ reported changes (due to the deorbitalization) in the MAE for bond lengths of molecules that are below 0.002 Å with MVS, which is small. The change in the ME can be larger in some cases since while the ME is -0.0016 Å with MVS, it increases to 0.0069 Å with MVS(PC), but is rather similar, -0.0025 Å, with MVS(PCopt). The deorbitalisation of SCAN leads to larger changes in the MAE of bond lengths (up to ~0.01 Å), but not for the ME since the largest change is ~0.016 Å, which is barely larger than for MVS. From these results on molecular bond lengths, t^{PCopt} seems to be a more accurate OF KED than the others. This is in line with our observation that t^{PCopt}

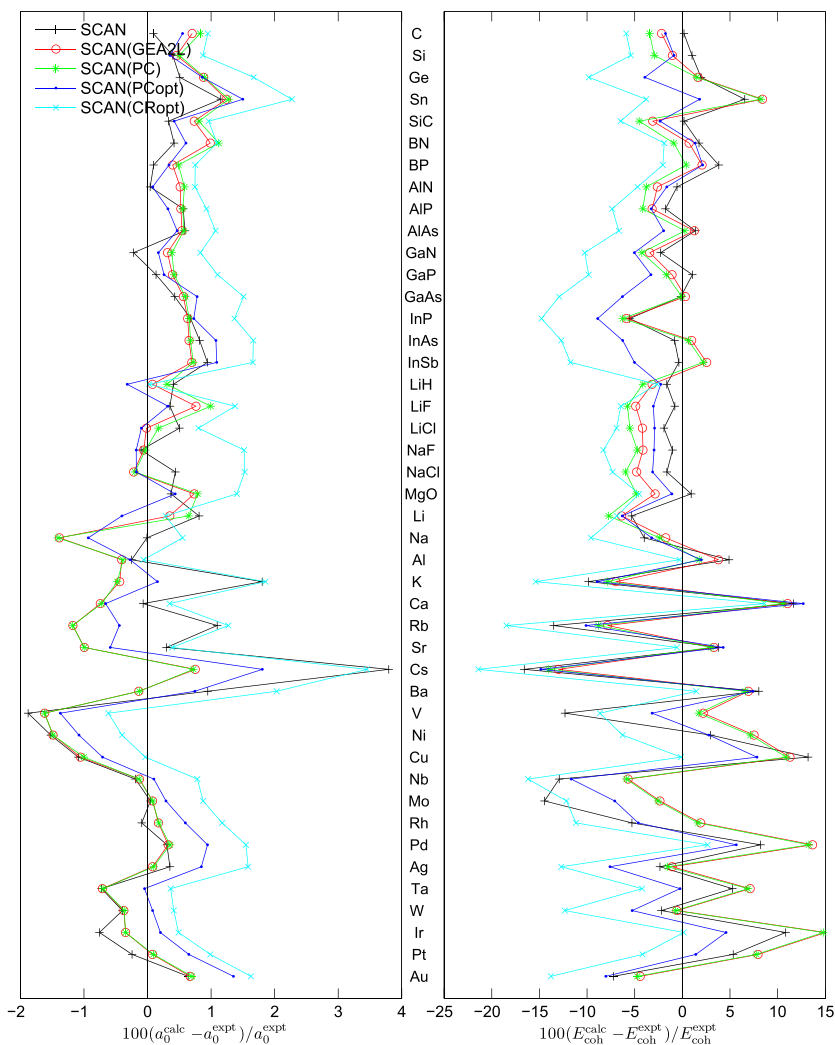


FIG. 1. Relative error (in %) with respect to experiment^{86,87} in the calculated lattice constant (left panel) and cohesive energy (right panel) for the 44 strongly bound solids.

is among the most accurate OF KED for the lattice constants of solids. Concerning the heat of formation,²⁶ the changes in the MAE and ME seem to be in many cases the smallest with t^{PCopt} , as well. For instance, the deorbitalization of SCAN leads to a change in the MAE of +15 and +0.5 kcal/mol with t^{PC} and t^{PCopt} , respectively, and +21 and +6 kcal/mol for the ME. We also mention that from the results of Mejia-Rodriguez and Trickey, we cannot observe a change in the results due to the deorbitalization that is larger in the case of MVS as we did. We should also mention the work of Bienvenu and Knizia³⁵ who observed that the deorbitalization of the MGGA of Tao *et al.*⁹⁰ (TPSS) with the KED PC induces rather large changes in the reaction energy of molecules.

In their more recent work on a similar test set of solids, Mejia-Rodriguez and Trickey³⁶ also investigated the lattice constant, bulk modulus, and cohesive energy, but only one xc-functional (SCAN) and OF KED (PCopt) were considered. Their calculations were done self-consistently and within the projected-augmented wave method. Without entering into detail, their results are similar to ours and therefore also concluded that PCopt is an accurate OF KED in the case of SCAN.

Turning now to weakly bound van der Waals systems, Tables II and III show the results for rare-gas (Ne, Ar, and Kr)

and layered hexagonal solids (graphite, h-BN, TiS_2 , MoTe_2 , and WSe_2), respectively.

The range of errors in the lattice constant obtained in typical performance tests of DFT functionals on van der Waals systems (see, e.g., Refs. 14 and 91–94) is by far much larger than for covalent or ionic solids. Hence, for our systems it should be fair to consider that the performance of a t^{KS} -MGGA (with respect to other functionals) is not really modified by its deorbitalization if the change in the lattice constant is, let us say, below something like (this may be a matter of personal taste) $\sim 0.1\text{--}0.15$ Å for the rare-gas (a_0) and layered solids (c_0). With this criterion, the results show that the replacement $t^{\text{KS}} \rightarrow t^{\text{OF}}$ in SCAN and TM leads to acceptable changes in the lattice constant in most cases except maybe Kr. With MVS, however, the changes are two or three times larger and unacceptable since they may affect the performance of MVS with respect to other functionals.

By choosing, again arbitrarily, $\sim 20\%$ of the reference CCSD(T) (coupled-cluster singles, doubles, and perturbative triples) or RPA (random-phase approximation) binding energy as the largest change that can be accepted when a functional is deorbitalized, then too large variations in E_{coh} or E_{b} are usually observed for MVS, especially for the rare gases. The deorbitalization of SCAN or TM affects less the results, but

TABLE II. Equilibrium lattice constant a_0 (in Å) and cohesive energy E_{coh} (in meV/atom) of rare-gas solids. The values for the $t^{\text{OF-MGGA}}$ functionals are the difference from those obtained with the parent $t^{\text{KS-MGGA}}$, e.g., $a_0(t^{\text{OF-MGGA}}) - a_0(t^{\text{KS-MGGA}})$. The reference CCSD(T) results, which agree closely with experiment,⁸⁸ are also shown. The large differences with respect to the parent $t^{\text{KS-MGGA}}$ are italic. All results were obtained non-self-consistently using the PBE orbitals/density.

Method	Ne		Ar		Kr	
	a_0	E_{coh}	a_0	E_{coh}	a_0	E_{coh}
MVS	4.02	59	5.41	56	5.79	69
MVS(GEA2L)	-0.14	41	-0.34	70	-0.30	80
MVS(TW02L)	-0.03	0	-0.21	29	-0.17	33
MVS(PC)	-0.15	47	-0.34	75	-0.31	85
MVS(CR)	-0.14	41	-0.34	70	-0.30	80
MVS(GEAl _{loc})	-0.11	31	-0.26	54	-0.19	55
MVS(PCopt)	-0.13	38	-0.31	66	-0.23	63
MVS(CRopt)	0.85	-53	1.03	-48	0.75	-53
SCAN	4.03	54	5.31	61	5.74	72
SCAN(GEA2L)	-0.02	11	-0.15	32	-0.20	50
SCAN(TW02L)	0.03	-6	-0.08	9	-0.15	23
SCAN(PC)	-0.03	15	-0.15	36	-0.20	54
SCAN(CR)	-0.02	12	-0.15	32	-0.20	50
SCAN(GEAl _{loc})	0.02	5	-0.06	19	-0.11	33
SCAN(PCopt)	-0.03	12	-0.11	28	-0.14	40
SCAN(CRopt)	0.63	-48	0.25	-37	0.26	-37
TM	4.05	47	5.23	62	5.60	82
TM(GEA2L)	-0.00	7	-0.08	22	-0.08	27
TM(TW02L)	0.03	-5	-0.05	9	-0.05	13
TM(PC)	-0.03	-8	-0.14	12	-0.14	22
TM(CR)	-0.00	7	-0.08	22	-0.08	27
TM(GEAl _{loc})	-0.10	32	-0.17	56	-0.16	67
TM(PCopt)	-0.01	-10	-0.11	7	-0.11	15
TM(CRopt)	0.05	-4	-0.01	7	-0.01	9
Reference	4.30	26	5.25	88	5.60	122

nevertheless the change for the rare gases is in most cases also too large according to our criterion. Interestingly, note that the deorbitalization of the SCAN and TM functionals leads in many cases to a better agreement with CCSD(T) for the binding energy.

For the rare gases, the OF KED that leads overall to the smallest perturbations for the deorbitalization of the xc-MGGAs seems to be t^{TW02L} . Note that, t^{CRopt} shows rather strange results since it is the worst when used in MVS and SCAN, while it is the best for TM. In the case of the layered solids, a good choice for t^{OF} is $t^{\text{GEAl_{loc}}}$ for MVS and SCAN, while with TM all t^{OF} except $t^{\text{GEAl_{loc}}}$ are of similar accuracy.

B. Bandgaps

Turning to the electronic structure, Table IV and Fig. 2 (for selected methods) show the results obtained for the fundamental bandgap E_g calculated with the mBJLDA potential and its deorbitalized versions. The testing set, which was used in our previous studies,^{80,82} consists of 76 solids (listed in Table

S10 of the [supplementary material](#)) of various types: ionic insulators, *sp*-semiconductors, rare gases, and strongly correlated solids. As shown in Refs. 80 and 82, the mBJLDA potential is on average more accurate for the bandgap than all other semilocal potentials and hybrid functionals that were considered for comparison (the PBE⁹ and HSE06^{105,106} results are also shown in Table IV and Fig. 2).

From the statistics shown in Table IV, the first observation is that deorbitalizing the mBJLDA potential leads to an increase of the MAE and MARE, no matter what OF KED is used. The deterioration is the smallest when t^{KS} is replaced by t^{PCopt} , and in this case the MAE increases from 0.47 to 0.67 eV and the MARE from 15% to 16%. This increase in the MARE is clearly negligible, but also quite acceptable for the MAE considering that most other potentials lead to larger MAE for this test set.^{80,82} With mBJLDA(CRopt), a small increase of 4% for the MARE is obtained, while the MAE increases to 0.75 eV, which is now on the verge of being acceptable since other potentials, e.g., AK13,¹⁰⁷ B3PW91,¹⁰⁸ or HSE06^{105,106} lead to similar MAE.^{80,82} Substituting t^{KS} by any of the other OF KED leads to a clearly larger MAE (around 1 eV) and MARE (above 30%, except with $t^{\text{GEAl_{loc}}}$).

Looking into more detail at the results (see Table S11 and Figs. S25–S32 of the [supplementary material](#) and Fig. 2), we can see that an inaccurate OF KED like t^{GEA2L} leads to bandgaps which are in most cases about halfway between the mBJLDA and PBE values, such that a rather clear underestimation is obtained on average (see ME and MRE in Table IV). The mBJLDA bandgaps are in general reproduced more accurately by mBJLDA(PCopt) and/or mBJLDA(CRopt) except for the rare gases for which mBJLDA(GEA2L) is the closest to mBJLDA.

We note that a reoptimization of the parameters α and β in a OF mBJLDA potential [see Ref. 37 for details] may possibly lead to a (partial) recovery of the performance of the original mBJLDA potential. However, we have not made any attempts since this is beyond the scope of this work.

Finally, we mention that Mejia-Rodriguez and Trickey³⁶ compared the bandgaps obtained from self-consistent SCAN and SCAN(PCopt) calculations. On a test set of 21 solids, they observed that the deorbitalization of SCAN leads to an increase in the MAE from 1.26 to 1.58 eV. This is rather similar to the difference between mBJLDA (0.47 eV) and mBJLDA(PCopt) (0.67 eV).

IV. FURTHER DISCUSSION

Thanks to their additional dependency on t^{KS} , $t^{\text{KS-MGGA}}$ are more flexible than GGAs and, therefore, have the possibility to be universally more accurate. As shown above, a $t^{\text{KS-MGGA}}$ can be replaced rather efficiently (albeit not systematically) by a corresponding $\nabla^2\rho$ -MGGA, and in order to shed some light on the relation between t^{KS} and $\nabla^2\rho$, a principal component analysis^{109,110} (PCA) of t^{TF} , t^{W} , $\nabla^2\rho$, and t^{KS} has been carried out. From the PCA, an approximation for t^{KS} that consists of a linear combination of t^{TF} , t^{W} , and $\nabla^2\rho$ is obtained, and its accuracy reveals to which extent t^{KS} can be represented by ρ and its first two derivatives.

TABLE III. Equilibrium lattice constant c_0 (in Å) and interlayer binding energy E_b (in meV/atom) of layered solids. The values for the t^{OF} -MGGA functionals are the difference from those obtained with the parent t^{KS} -MGGA, e.g., $c_0(t^{\text{OF}}\text{-MGGA}) - c_0(t^{\text{KS}}\text{-MGGA})$. The intralayer constant a was not optimized, but kept fixed at the experimental value.⁸⁹ Reference results⁸⁹ from the experiment for c_0 and from RPA for E_b are also shown. The large differences with respect to the parent t^{KS} -MGGA are italic. All results were obtained non-self-consistently using PBE orbitals/density.

Method	Graphite		h-BN		TiS ₂		MoTe ₂		WSe ₂	
	c_0	E_b	c_0	E_b	c_0	E_b	c_0	E_b	c_0	E_b
MVS	6.60	32	6.43	38	5.79	30	14.66	34	13.48	19
MVS(GEA2L)	<i>-0.24</i>	<i>13</i>	<i>-0.21</i>	10	<i>-0.19</i>	<i>18</i>	<i>-0.25</i>	6	<i>-0.22</i>	<i>12</i>
MVS(TW02L)	<i>-0.22</i>	<i>11</i>	<i>-0.19</i>	8	<i>-0.12</i>	9	<i>-0.13</i>	0	<i>-0.09</i>	5
MVS(PC)	<i>-0.24</i>	<i>13</i>	<i>-0.20</i>	10	<i>-0.14</i>	<i>17</i>	<i>-0.25</i>	7	<i>-0.22</i>	<i>12</i>
MVS(CR)	<i>-0.24</i>	<i>13</i>	<i>-0.21</i>	10	<i>-0.19</i>	<i>18</i>	<i>-0.25</i>	6	<i>-0.22</i>	<i>12</i>
MVS(GEAl _{loc})	<i>-0.14</i>	10	<i>-0.13</i>	7	0.02	7	0.12	-2	0.07	4
MVS(PC _{opt})	<i>-0.13</i>	10	<i>-0.12</i>	7	<i>-0.07</i>	<i>11</i>	0.01	2	0.02	7
MVS(CR _{opt})	0.02	-1	<i>0.14</i>	-7	<i>0.28</i>	<i>-11</i>	<i>0.37</i>	<i>-13</i>	<i>0.59</i>	-8
SCAN	6.94	20	6.79	21	5.93	21	14.75	30	13.68	17
SCAN(GEA2L)	<i>-0.13</i>	4	<i>-0.10</i>	5	<i>-0.12</i>	<i>12</i>	<i>-0.33</i>	8	<i>-0.26</i>	10
SCAN(TW02L)	<i>-0.10</i>	2	<i>-0.08</i>	3	<i>-0.09</i>	8	<i>-0.31</i>	5	<i>-0.23</i>	7
SCAN(PC)	<i>-0.13</i>	4	<i>-0.10</i>	5	<i>-0.09</i>	<i>11</i>	<i>-0.33</i>	8	<i>-0.26</i>	10
SCAN(CR)	<i>-0.13</i>	4	<i>-0.10</i>	5	<i>-0.12</i>	<i>12</i>	<i>-0.33</i>	8	<i>-0.26</i>	10
SCAN(GEAl _{loc})	<i>-0.09</i>	3	<i>-0.06</i>	4	0.03	5	0.13	-1	0.05	3
SCAN(PC _{opt})	<i>-0.12</i>	3	<i>-0.08</i>	4	<i>-0.04</i>	8	0.07	0	0.01	4
SCAN(CR _{opt})	0.03	-1	0.05	-2	<i>0.16</i>	-5	<i>0.41</i>	-9	<i>0.36</i>	-5
TM	6.63	29	6.49	32	5.73	44	14.17	50	13.21	35
TM(GEA2L)	<i>-0.08</i>	4	<i>-0.06</i>	3	<i>-0.08</i>	7	<i>-0.16</i>	7	<i>-0.11</i>	6
TM(TW02L)	<i>-0.09</i>	4	<i>-0.07</i>	3	<i>-0.08</i>	6	<i>-0.16</i>	7	<i>-0.11</i>	5
TM(PC)	<i>-0.15</i>	4	<i>-0.11</i>	4	<i>-0.07</i>	6	<i>-0.17</i>	8	<i>-0.14</i>	6
TM(CR)	<i>-0.08</i>	4	<i>-0.06</i>	3	<i>-0.08</i>	7	<i>-0.16</i>	7	<i>-0.11</i>	6
TM(GEAl _{loc})	<i>-0.23</i>	<i>17</i>	<i>-0.21</i>	<i>15</i>	<i>-0.07</i>	<i>15</i>	<i>-0.14</i>	<i>14</i>	<i>-0.15</i>	<i>13</i>
TM(PC _{opt})	<i>-0.12</i>	3	<i>-0.08</i>	2	0.02	3	0.02	2	0.02	2
TM(CR _{opt})	<i>-0.10</i>	7	<i>-0.07</i>	5	0.02	3	0.05	2	0.03	2
Reference	6.70	48	6.69	40	5.71	95	13.97	111	12.96	93

The 4×4 covariance matrix was calculated using uniformly sampled data from one representative of metallic (Cu), layered (graphite), and covalently bound (Si) systems, and

TABLE IV. The ME, MAE, MRE, and MARE (with respect to experiment^{86,95–104}) on the testing set of 76 solids (listed in Table S10 of the [supplementary material](#)) for the fundamental bandgap E_g obtained with mBJLDA and its deorbitalized versions, as well as PBE and HSE06. The units are eV for the ME and MAE and % for the MRE and MARE.

	ME	MAE	MRE	MARE
mBJLDA	-0.30	0.47	-5	15
mBJLDA(GEA2L)	-0.95	0.97	-32	32
mBJLDA(TW02L)	-1.03	1.03	-33	33
mBJLDA(PC)	-1.17	1.18	-32	33
mBJLDA(CR)	-0.94	0.96	-31	32
mBJLDA(GEAl _{loc})	0.39	0.92	6	21
mBJLDA(PC _{opt})	-0.54	0.67	-10	16
mBJLDA(CR _{opt})	-0.08	0.75	-10	19
PBE	-1.99	1.99	-53	53
HSE06	-0.68	0.82	-7	17

diagonalized in order to get the eigenvalues and corresponding eigenvectors spanning the four-dimensional space of t^{TF} , t^{W} , $\nabla^2\rho$, and t^{KS} . In the next step, we neglect the eigenvector with the smallest eigenvalue, thereby obtaining the three-dimensional representation which explains most of the variance in the data. Now, assuming that all points are on this three dimensional hyperplane, one can reconstruct an OF KED from the values of ρ (i.e., t^{TF}), $\nabla\rho$ (i.e., t^{W}), and $\nabla^2\rho$, and the resulting linear combination is given by

$$t^{\text{PCA}}(\mathbf{r}) = 1.069t^{\text{TF}}(\mathbf{r}) - 0.244t^{\text{W}}(\mathbf{r}) + 0.438\nabla^2\rho(\mathbf{r}). \quad (7)$$

The coefficient in front of t^{TF} is close to 1 as it should be in order to recover the homogeneous electron gas limit, while those in front of t^{W} and $\nabla^2\rho$ show big differences from GEA2L [Eq. (5)]. However, it is worth mentioning that a negative coefficient in front of t^{W} is found also in GEAl_{loc}⁶⁷ (-0.165) and in a KED expression derived for the Airy gas⁶⁸ (-1/9 \approx -0.111).

Figure 3 shows for the three selected solids the accuracy of the GEA2L and our PCA approximation with the Weizsäcker lower bound enforced [Eq. (6)]. We can see that

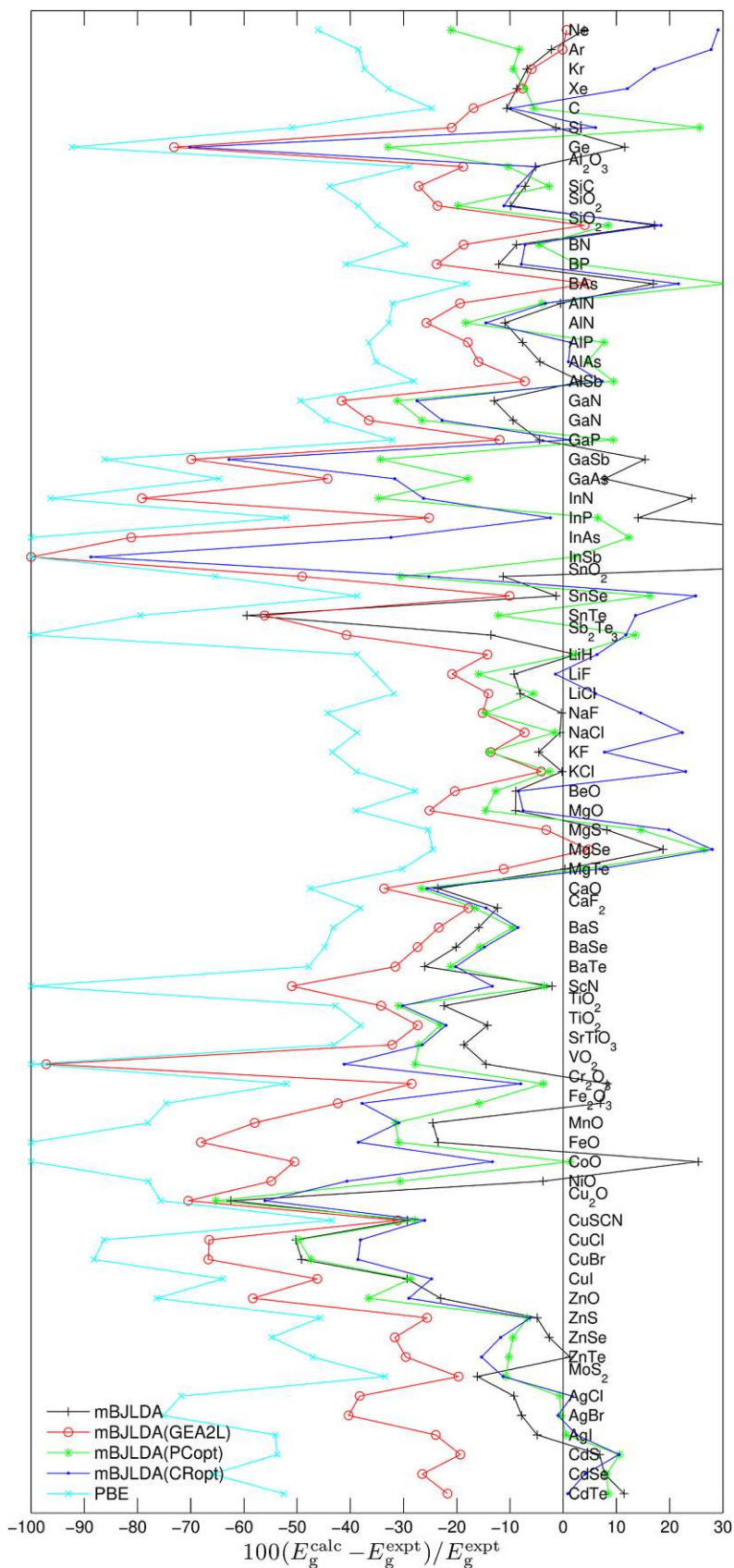


FIG. 2. Relative error (in %) with respect to the experiment^{86,95–104} in the bandgap E_g .

the PCA approximation shows better agreement with the KS KED than GEA2L, similarly obtained by Seino *et al.*⁵¹ for atoms and small organic molecules using a machine learning algorithm. It is also important to note that for both

approximations, there are two regions where one can find larger errors. These two lumps are from Si and graphite, where GEA2L systematically overestimates the KED, while in the PCA approximation these errors are still there but largely

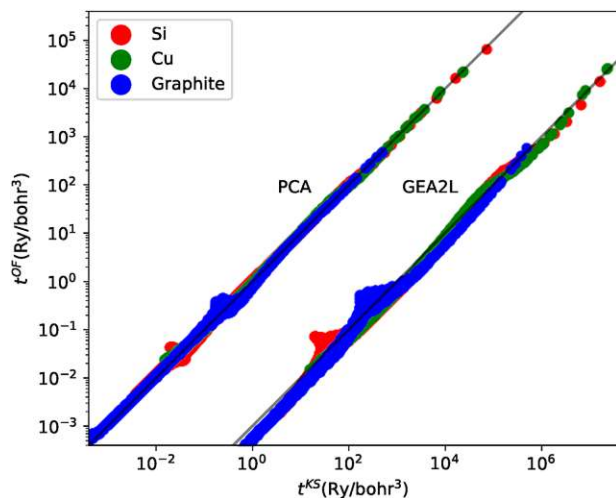


FIG. 3. Comparison between the KS KED and the GEA2L and PCA approximations for different solids. For clarity (no overlap between the GEA2L and PCA data), the t^{KS} values for GEA2L are multiplied by 1000 (i.e., right shifted). A perfect approximation should coincide with the diagonal solid black line.

reduced. Actually, the errors for graphite can be found in the same KED region as the errors for organic molecules.⁵¹

In Fig. 4, the erroneous points from these two regions are shown in real space, where we can see that the bigger errors occur in the middle of covalent bonds. If, for instance, for graphite the same PCA method is applied using only the points in the bonding regions, much better accuracy (in these bonding regions) can be reached, and the resulting linear combination is given by

$$t_{\text{bond}}^{\text{PCA}}(\mathbf{r}) = 0.389t^{\text{TF}}(\mathbf{r}) + 0.635t^{\text{W}}(\mathbf{r}) + 0.084\nabla^2\rho(\mathbf{r}). \quad (8)$$

While this is obviously not useful as a general KED approximation, it is interesting to note that t^{W} has now a small positive coefficient, in agreement with the fact that the covalent σ -bonding in graphite and silicon should be dominated by a single molecular orbital. As shown by Seino *et al.*,⁵¹ considering also the third derivative of ρ further improves the accuracy of OF KED. However, as discussed below, the bonding regions highlighted in Fig. 4 are not necessarily those which are the most relevant for explaining the differences observed in the results for the lattice constant.

In order to provide some insight into the results presented in Sec. III, Fig. 5 compares the energy density of SCAN(GEA2L) and SCAN(CRopt) in Si. For simplicity, only the exchange component, which is much larger than correlation, is considered. SCAN(GEA2L) and SCAN(CRopt) lead to rather different equilibrium lattice constants a_0 for Si, namely, 5.437 and 5.460 Å, respectively, and the following analysis provides details about the regions of space that are involved to explain these different values of a_0 . Figure 5(a) shows $\Delta\varepsilon_x^{\text{F1-F2}}$, which is defined as

$$\Delta\varepsilon_x^{\text{F1-F2}}(r) = r^2 \int \left[\left(\varepsilon_x^{\text{F1},a_{\text{large}}}(\mathbf{r}) - \varepsilon_x^{\text{F1},a_{\text{small}}}(\mathbf{r}) \right) - \left(\varepsilon_x^{\text{F2},a_{\text{large}}}(\mathbf{r}) - \varepsilon_x^{\text{F2},a_{\text{small}}}(\mathbf{r}) \right) \right] d\Omega, \quad (9)$$

where $\varepsilon_x^{\text{F},a}$ is the exchange energy density [defined by Eq. (2)] of functional F [F1 and F2 designate SCAN(CRopt) and

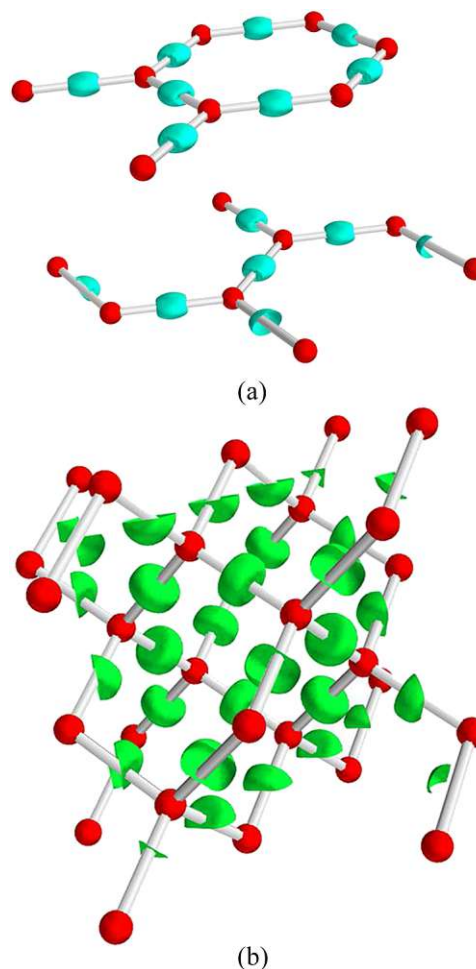


FIG. 4. Real space position of the lumps of Fig. 3. The atoms are represented by red spheres, while the erroneous points for (a) graphite (isosurface corresponding to $t^{\text{GEA2L}}/t^{\text{KS}} = 2.25$) and (b) silicon (isosurface corresponding to $t^{\text{GEA2L}}/t^{\text{KS}} = 1.9$) are in turquoise and green, respectively.

SCAN(GEA2L), respectively] calculated at a given lattice constant (a_{small} or a_{large}). The integration in Eq. (9) is over the spherical angles and r is the distance from one Si atom. As discussed in detail in Refs. 111 and 112, the equilibrium lattice constant a_0 is determined by the slope of the xc-energy E_{xc} , i.e., the variation of E_{xc} with respect to a , and this is basically what Fig. 5 shows since the difference between two values of a (a_{small} and a_{large}) is considered. Figure 5(b) shows the radial integration of $\Delta\varepsilon_x^{\text{F1-F2}}$ up to a given value of r . As already discussed in Ref. 111 for Si but in the case of GGA functionals, two different regions contribute significantly to the variation of E_{xc} with respect to a . The first one, located around 0.5 Å [see the fast variations of the curves in Figs. 5(a) and 5(b)] corresponds to the core-valence separation. The second region extends from 1.2 to 1.7 Å and corresponds to the valence/interstitial region which is rather large since Si has an open structure. Additionally, Fig. 6 shows the isosurface of $|F_x^{\text{SCAN(CRopt)}} - F_x^{\text{SCAN(GEA2L)}}|$ that delimits values larger than 0.03 (where actually $F_x^{\text{SCAN(CRopt)}} > F_x^{\text{SCAN(GEA2L)}}$) and highlights the two types of regions just mentioned above.

The case of graphite was also discussed in Ref. 111, where the electron density and reduced density gradient s in the region

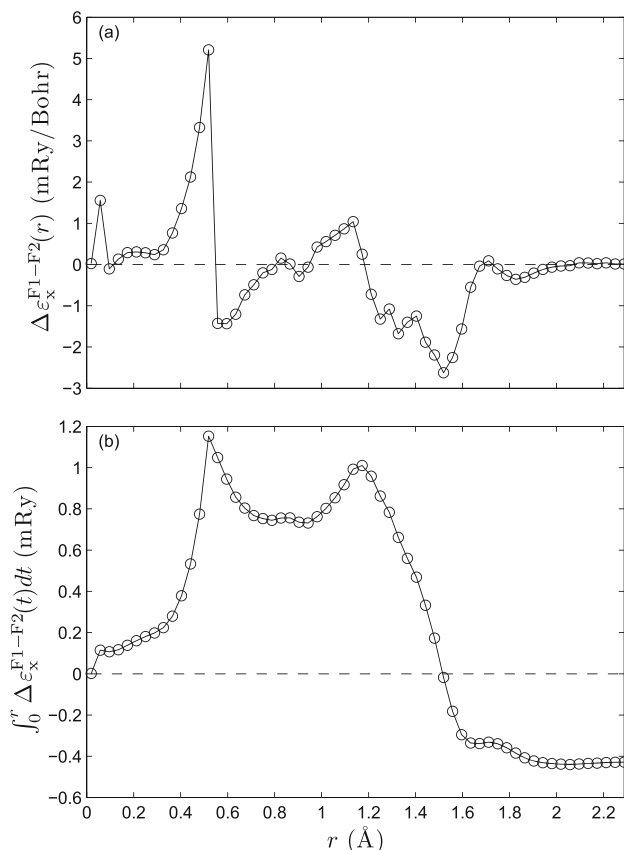


FIG. 5. Difference between the exchange components of SCAN(CRopt) (F1) and SCAN(GEA2L) (F2) in Si plotted as a function of the distance r from an Si atom. Panel (a) shows the angular average of $\Delta\epsilon_x^{F1-F2}$ (see the text for definition), while panel (b) shows the radial integration of $\Delta\epsilon_x^{F1-F2}$ from the atom until r .

between the layers were studied in detail. It was shown that an increase of the interlayer distance leads to a rather large increase of s overall, thus explaining the overestimation of

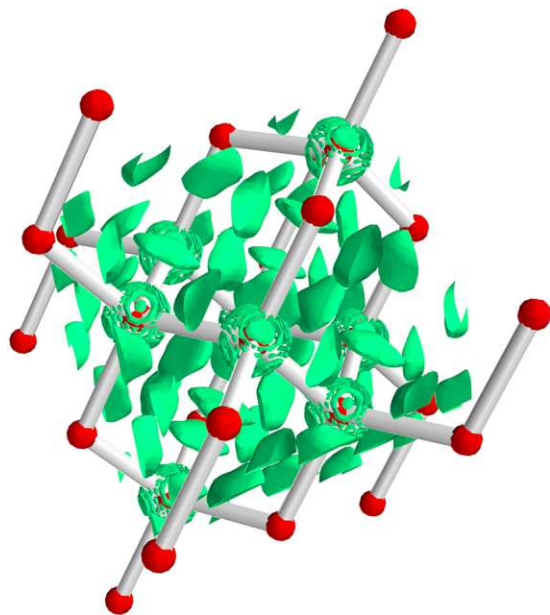


FIG. 6. Isosurface of the absolute value of $F_x^{\text{SCAN(CRopt)}} - F_x^{\text{SCAN(GEA2L)}}$ corresponding to 0.03.

the interlayer distance for GGA functionals with a too strong enhancement factor. Figure 7 shows the ratio $t^{\text{GEA2L}}/t^{\text{KS}}$ with a ratio that is smaller than the one used in Fig. 4(a), such that the isosurface encloses a larger region. We can see that aside from the middle of the short covalent bonds within the layers (not relevant for the interlayer distance), also a non-negligible portion of the space between the layers has a ratio ($t^{\text{KS}}/t^{\text{GEA2L}}$) bigger than 1.9.

In Sec. III, we also observed that an OF KED that is among the most accurate for a property calculated with the total energy, may be among the most inaccurate for the bandgap, or vice versa. For instance, while t^{PCopt} and t^{CROpt} are not among the best KEDs for total-energy related properties of strongly bound solids, they are the most accurate for the bandgap. Such contradictory results could seem quite puzzling at first sight, however this should be rather simple to explain in most cases.

Taking LiH as an example, Fig. 8 compares the xc-energy calculated with SCAN and selected deorbitalized SCANS by showing the difference $E_{xc}^{t^{\text{OF}}-\text{SCAN}} - E_{xc}^{t^{\text{KS}}-\text{SCAN}}$ as a function of the lattice constant a (this is the same kind of analysis as the one used for Si in Fig. 5). Figures 8(a)–8(c) show the contributions from the Li atom, H atom, and interstitial region, respectively, while the sum of them (the total value in the unit cell) is shown in Fig. 8(d). As expected, the SCAN equilibrium lattice constants a_0 of LiH (see Table S2 of the supplementary material) show the same ordering as the curves in Fig. 8(d) [the uppermost (lowest) curve correspond to the smallest (largest) lattice constant]. Thus, in the present case where the same functional is evaluated with different KED, the change in a_0 due to deorbitalization depends on the variation with a of the difference between t^{KS} and t^{OF} . From Fig. 8, we can also see that for all functionals, $E_{xc}^{t^{\text{OF}}-\text{SCAN}} - E_{xc}^{t^{\text{KS}}-\text{SCAN}}$ decreases in the H atom, but increases in the Li atom and interstitial region such that in total an increase is obtained. We also note that with t^{GEA2L} and t^{PC} , there is a very large cancellation of the errors coming from the H atom and interstitial region.

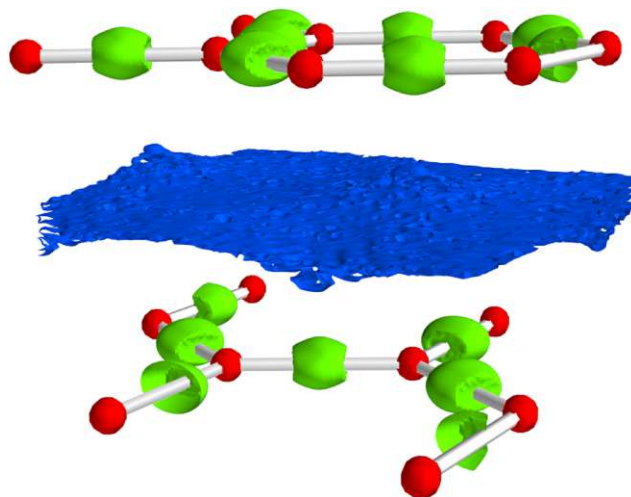


FIG. 7. The regions of space in graphite where $t^{\text{GEA2L}}/t^{\text{KS}}$ and $t^{\text{KS}}/t^{\text{GEA2L}}$ are larger than 1.9 are delimited by the isosurfaces in green and blue, respectively.

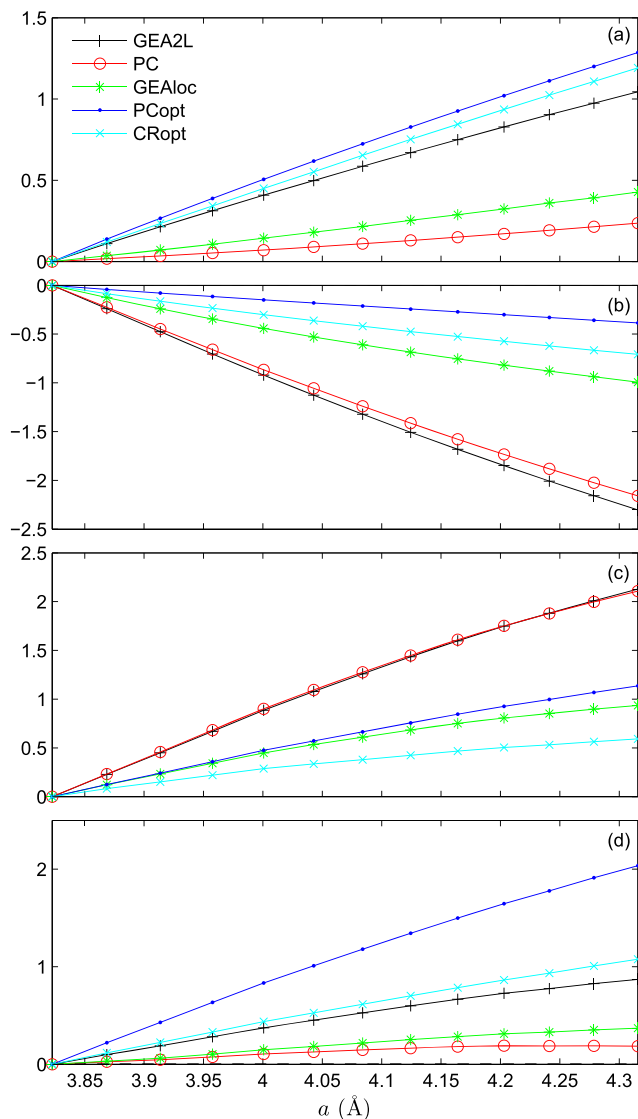


FIG. 8. Difference $E_{xc}^{t^{OF},SCAN} - E_{xc}^{t^{KS},SCAN}$ (in mRy) between the xc-energies of LiH obtained with SCAN and its deorbitalized versions plotted as a function of the lattice constant a . Panels (a), (b), and (c) show the contributions from the Li atom, H atom, and interstitial region, respectively, while panel (d) shows the sum of all contributions (i.e., the whole unit cell). The atomic muffin-tin spheres of the Li and H atoms are 1.7 Bohr. Each curve is vertically shifted such that the zero is at the smallest volume.

As discussed in previous studies,^{77,113,114} the magnitude of the bandgap is determined by the inhomogeneities in the potential, such that, roughly speaking, large inhomogeneities favor larger values of the bandgap. Actually, in most systems, the valence band maximum and conduction band minimum are located close to an atom and in the interstitial region, respectively, which means that the difference in the magnitudes of a potential between these two regions determines the bandgap. Again for LiH, Fig. 9 compares v_{xc} of mBJLDA and its OF variants. The LiH bandgap (see Table S11 of the [supplementary material](#)) with mBJLDA is 5.06 eV and is reproduced at best by mBJLDA(PCopt) (5.03 eV), while mBJLDA(GEALoc) with 6.69 eV leads to the worst agreement. This is in accordance with Fig. 9, where we can see that the mBJLDA(PCopt) potential is the closest to mBJLDA, while the mBJLDA(GEALoc) potential is much higher in the interstitial region (where

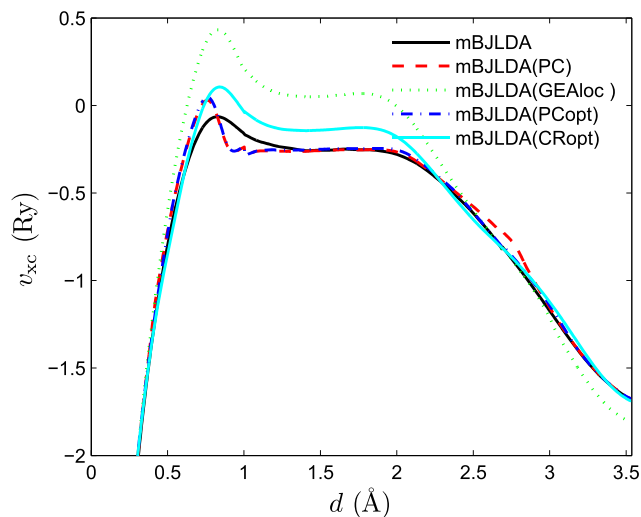


FIG. 9. mBJLDA xc-potential and a few selected of its deorbitalized versions plotted in LiH from the Li atom at $(0, 0, 0)$ to the H atom at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

the conduction band minimum is located) and lower close to the H atom (where the valence band maximum is located).

Thus, from this detailed discussion about LiH, it is rather clear that different mechanisms have to be invoked in order to explain the trends observed for the lattice constant (a total-energy related property) and bandgap, such that opposite conclusions for these two types of properties can be obtained.

V. SUMMARY

In this work, the deorbitalization of several xc-MGGA methods, three energy functionals and one potential, has been investigated by considering properties of solids. The replacement $t^{KS} \rightarrow t^{OF}$ in xc-MGGAs affects the results to some degree which depends on both the xc-MGGA under investigation and the used approximation for the OF KED t^{OF} .

Concerning the energy functionals for the calculation of the lattice constant, bulk modulus, and binding energy, we have shown that the results are in general more sensitive with MVS than with SCAN and TM, which should just be the direct consequence of the analytical form of the functionals that depends more strongly on the KED in the case of MVS. With SCAN and TM, the replacement $t^{KS} \rightarrow t^{OF}$ with most OF KED does not change much the results for strongly bound solids, such that the performance of a xc-MGGA remains pretty much the same. For the weakly bound rare gases, the change in the cohesive energy is usually rather large, while for the layered solids large changes in the interlayer distance are obtained with MVS.

The deorbitalization of the mBJLDA xc-potential leads to appreciable changes in the bandgap and only the OF KED t^{PCopt} can be considered as a somehow reasonable replacement of t^{KS} .

Similarly to Mejia-Rodriguez and Trickey,²⁶ we were not able to identify a OF KED that leads to reasonably small change in the results in most circumstances.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for the detailed results for the lattice constant, bulk modulus, cohesive energy, and bandgap.

ACKNOWLEDGMENTS

This work was supported by the project F41 (SFB ViCoM) of the Austrian Science Fund (FWF) and by the TU-D doctoral college (TU Wien). F.T. acknowledges discussions with J. P. Perdew and L. A. Constantin.

- ¹P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- ²W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- ³A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Chem. Rev.* **112**, 289 (2012).
- ⁴A. D. Becke, *J. Chem. Phys.* **140**, 18A301 (2014).
- ⁵S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).
- ⁶J. P. Perdew and Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).
- ⁷A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- ⁸C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ⁹J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996); **78**, 1396(E) (1997).
- ¹⁰J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, *Phys. Rev. Lett.* **100**, 136406 (2008); **102**, 039902(E) (2009).
- ¹¹F. Della Sala, E. Fabiano, and L. A. Constantin, *Int. J. Quantum Chem.* **116**, 1641 (2016).
- ¹²J. Sun, A. Ruzsinszky, and J. P. Perdew, *Phys. Rev. Lett.* **115**, 036402 (2015).
- ¹³J. Tao and Y. Mo, *Phys. Rev. Lett.* **117**, 073001 (2016).
- ¹⁴F. Tran, J. Stelzl, and P. Blaha, *J. Chem. Phys.* **144**, 204120 (2016).
- ¹⁵H. Peng, Z.-H. Yang, J. P. Perdew, and J. Sun, *Phys. Rev. X* **6**, 041005 (2016).
- ¹⁶Y. Mo, G. Tian, R. Car, V. N. Staroverov, G. E. Scuseria, and J. Tao, *J. Chem. Phys.* **145**, 234306 (2016).
- ¹⁷Y. Mo, R. Car, V. N. Staroverov, G. E. Scuseria, and J. Tao, *Phys. Rev. B* **95**, 035118 (2017).
- ¹⁸Y. Hinuma, H. Hayashi, Y. Kumagai, I. Tanaka, and F. Oba, *Phys. Rev. B* **96**, 094102 (2017).
- ¹⁹S. Jana, A. Patra, and P. Samal, *J. Chem. Phys.* **149**, 044120 (2018).
- ²⁰C. Shahi, J. Sun, and J. P. Perdew, *Phys. Rev. B* **97**, 094111 (2018).
- ²¹Y. Zhang, D. A. Kitchaev, J. Yang, T. Chen, S. T. Dacek, R. A. Sarmiento-Pérez, M. A. L. Marques, H. Peng, G. Ceder, J. P. Perdew, and J. Sun, *npj Comput. Mater.* **4**, UNSP9 (2018).
- ²²Y. Mo, H. Tang, A. Bansil, and J. Tao, *AIP Adv.* **8**, 095209 (2018).
- ²³P. Jemmer and P. J. Knowles, *Phys. Rev. A* **51**, 3571 (1995).
- ²⁴R. Neumann and N. C. Handy, *Chem. Phys. Lett.* **266**, 16 (1997).
- ²⁵A. C. Cancio, C. E. Wagner, and S. A. Wood, *Int. J. Quantum Chem.* **112**, 3796 (2012).
- ²⁶D. Mejia-Rodriguez and S. B. Trickey, *Phys. Rev. A* **96**, 052512 (2017).
- ²⁷S. Laricchia, L. A. Constantin, E. Fabiano, and F. Della Sala, *J. Chem. Theory Comput.* **10**, 164 (2014).
- ²⁸R. Neumann, R. H. Nobes, and N. C. Handy, *Mol. Phys.* **87**, 1 (1996).
- ²⁹J. A. Pople, P. M. W. Gill, and B. G. Johnson, *Chem. Phys. Lett.* **199**, 557 (1992).
- ³⁰L. Ferrighi, G. K. H. Madsen, and B. Hammer, *J. Chem. Phys.* **135**, 084704 (2011).
- ³¹J. Sun, M. Marsman, G. I. Csonka, A. Ruzsinszky, P. Hao, Y.-S. Kim, G. Kresse, and J. P. Perdew, *Phys. Rev. B* **84**, 035117 (2011).
- ³²J. C. Womack, N. Mardirossian, M. Head-Gordon, and C.-K. Skylaris, *J. Chem. Phys.* **145**, 204114 (2016).
- ³³Y. Yao and Y. Kanai, *J. Chem. Phys.* **146**, 224105 (2017).
- ³⁴A. D. Becke and K. E. Edgecombe, *J. Chem. Phys.* **92**, 5397 (1990).
- ³⁵A. V. Biennu and G. Knizia, *J. Chem. Theory Comput.* **14**, 1297 (2018).
- ³⁶D. Mejia-Rodriguez and S. B. Trickey, *Phys. Rev. B* **98**, 115161 (2018).
- ³⁷F. Tran and P. Blaha, *Phys. Rev. Lett.* **102**, 226401 (2009).
- ³⁸V. L. Lignères and E. A. Carter, in *Handbook of Materials Modeling*, edited by S. Yip (Springer, Dordrecht, 2005), p. 137.
- ³⁹*Recent Progress in Orbital-Free Density Functional Theory*, edited by T. A. Wesolowski and Y. A. Wang (World Scientific, Singapore, 2013).
- ⁴⁰V. V. Karasiev and S. B. Trickey, *Adv. Quantum Chem.* **71**, 221 (2015).
- ⁴¹C. R. Jacob and J. Neugebauer, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **4**, 325 (2014).
- ⁴²T. A. Wesolowski, S. Shedge, and X. Zhou, *Chem. Rev.* **115**, 5891 (2015).
- ⁴³S. Šmiga, E. Fabiano, L. A. Constantin, and F. Della Sala, *J. Chem. Phys.* **146**, 064105 (2017).
- ⁴⁴K. Jiang, J. Nafziger, and A. Wasserman, *J. Chem. Phys.* **148**, 104113 (2018).
- ⁴⁵L. H. Thomas, *Math. Proc. Cambridge Philos. Soc.* **23**, 542 (1927).
- ⁴⁶E. Fermi, *Rend. Accad. Naz. Lincei* **6**, 602 (1927).
- ⁴⁷R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
- ⁴⁸A. J. Thakkar, *Phys. Rev. A* **46**, 6920 (1992).
- ⁴⁹S. S. Iyengar, M. Ernzerhof, S. N. Maximoff, and G. E. Scuseria, *Phys. Rev. A* **63**, 052508 (2001).
- ⁵⁰F. Tran and T. A. Wesolowski, *Chem. Phys. Lett.* **360**, 209 (2002).
- ⁵¹J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, *J. Chem. Phys.* **148**, 241705 (2018).
- ⁵²F. Tran and T. A. Wesolowski, *Int. J. Quantum Chem.* **89**, 441 (2002).
- ⁵³D. García-Aldea and J. E. Alvarillo, *J. Chem. Phys.* **127**, 144109 (2007).
- ⁵⁴V. V. Karasiev, S. B. Trickey, and F. E. Harris, *J. Comput.-Aided Mater. Des.* **13**, 111 (2006).
- ⁵⁵F. Tran and T. A. Wesolowski, *Recent Progress in Orbital-Free Density Functional Theory* (World Scientific, Singapore, 2013), p. 429.
- ⁵⁶N. H. March and R. Santamaria, *Int. J. Quantum Chem.* **39**, 585 (1991).
- ⁵⁷H. Lee, C. Lee, and R. G. Parr, *Phys. Rev. A* **44**, 768 (1991).
- ⁵⁸J. A. Alonso and L. A. Girifalco, *Chem. Phys. Lett.* **53**, 190 (1978).
- ⁵⁹W. Yang, R. G. Parr, and C. Lee, *Phys. Rev. A* **34**, 4586 (1986).
- ⁶⁰K. Finzel, *Theor. Chem. Acc.* **134**, 106 (2015).
- ⁶¹A. C. Cancio, D. Stewart, and A. Kuna, *J. Chem. Phys.* **144**, 084107 (2016).
- ⁶²D. A. Kirzhnits, *Sov. Phys. JETP* **5**, 64 (1957).
- ⁶³M. Brack, B. K. Jennings, and Y. H. Chu, *Phys. Lett. B* **65**, 1 (1976).
- ⁶⁴C. F. von Weizsäcker, *Z. Phys.* **96**, 431 (1935).
- ⁶⁵J. P. Perdew and L. A. Constantin, *Phys. Rev. B* **75**, 155109 (2007).
- ⁶⁶E. X. Salazar, P. F. Guarderas, E. V. Ludeña, M. H. Cornejo, and V. V. Karasiev, *Int. J. Quantum Chem.* **116**, 1313 (2016).
- ⁶⁷A. C. Cancio and J. J. Redd, *Mol. Phys.* **115**, 618 (2017).
- ⁶⁸A. Lindmaa, A. E. Mattsson, and R. Armiento, *Phys. Rev. B* **90**, 075139 (2014); **95**, 079902(E) (2017).
- ⁶⁹S. Šmiga, E. Fabiano, S. Laricchia, L. A. Constantin, and F. Della Sala, *J. Chem. Phys.* **142**, 154121 (2015).
- ⁷⁰A. A. Astakhov, A. I. Stash, and V. G. Tsirelson, *Int. J. Quantum Chem.* **116**, 237 (2016).
- ⁷¹L. A. Constantin, E. Fabiano, and F. Della Sala, *J. Phys. Chem. Lett.* **9**, 4385 (2018).
- ⁷²M. Hoffmann-Ostenhof and T. Hoffmann-Ostenhof, *Phys. Rev. A* **16**, 1782 (1977).
- ⁷³S. Kurth, J. P. Perdew, and P. Blaha, *Int. J. Quantum Chem.* **75**, 889 (1999).
- ⁷⁴G. L. Oliver and J. P. Perdew, *Phys. Rev. A* **20**, 397 (1979).
- ⁷⁵J. Sun, J. P. Perdew, and A. Ruzsinszky, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 685 (2015).
- ⁷⁶A. D. Becke and E. R. Johnson, *J. Chem. Phys.* **124**, 221101 (2006).
- ⁷⁷F. Tran, P. Blaha, and K. Schwarz, *J. Phys.: Condens. Matter* **19**, 196208 (2007).
- ⁷⁸D. J. Singh, *Phys. Rev. B* **82**, 205102 (2010).
- ⁷⁹H. Jiang, *J. Chem. Phys.* **138**, 134115 (2013).
- ⁸⁰F. Tran and P. Blaha, *J. Phys. Chem. A* **121**, 3318 (2017).
- ⁸¹K. Nakano and T. Sakai, *J. Appl. Phys.* **123**, 015104 (2018).
- ⁸²F. Tran, S. Ehsan, and P. Blaha, *Phys. Rev. Mater.* **2**, 023802 (2018).
- ⁸³P. Blaha, K. Schwarz, G. K. H. Madsen, D. Kvasnicka, J. Luitz, R. Laskowski, F. Tran, and L. D. Marks, *WIEN2K: An Augmented Plane Wave Plus Local Orbitals Program for Calculating Crystal Properties* (Vienna University of Technology, Austria, 2018).
- ⁸⁴O. K. Andersen, *Phys. Rev. B* **12**, 3060 (1975).
- ⁸⁵D. J. Singh and L. Nordström, *Planewaves, Pseudopotentials, and the LAPW Method*, 2nd ed. (Springer, New York, 2006).
- ⁸⁶L. Schimka, J. Harl, and G. Kresse, *J. Chem. Phys.* **134**, 024116 (2011).
- ⁸⁷K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, *Crit. Rev. Solid State Mater. Sci.* **39**, 1 (2014).
- ⁸⁸K. Rościszewski, B. Paulus, P. Fulde, and H. Stoll, *Phys. Rev. B* **62**, 5482 (2000).
- ⁸⁹T. Björkman, *Phys. Rev. B* **86**, 165109 (2012).
- ⁹⁰J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.* **91**, 146401 (2003).

- ⁹¹E. R. Johnson, R. A. Wolkow, and G. A. DiLabio, *Chem. Phys. Lett.* **394**, 334 (2004).
- ⁹²Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A* **110**, 5121 (2006).
- ⁹³F. Tran and J. Hutter, *J. Chem. Phys.* **138**, 204103 (2013); **139**, 039903 (2013).
- ⁹⁴C. R. C. Rêgo, L. N. Oliveira, P. Tereshchuk, and J. L. F. Da Silva, *J. Phys.: Condens. Matter* **27**, 415502 (2015); **28**, 129501 (2016).
- ⁹⁵J. M. Crowley, J. Tahir-Kheli, and W. A. Goddard III, *J. Phys. Chem. Lett.* **7**, 1198 (2016).
- ⁹⁶M. J. Lucero, T. M. Henderson, and G. E. Scuseria, *J. Phys.: Condens. Matter* **24**, 145504 (2012).
- ⁹⁷S. Bernstorff and V. Saile, *Opt. Commun.* **58**, 181 (1986).
- ⁹⁸R. Gillen and J. Robertson, *J. Phys.: Condens. Matter* **25**, 165502 (2013).
- ⁹⁹D. Koller, P. Blaha, and F. Tran, *J. Phys.: Condens. Matter* **25**, 435503 (2013).
- ¹⁰⁰J. H. Skone, M. Govoni, and G. Galli, *Phys. Rev. B* **89**, 195112 (2014).
- ¹⁰¹H. Shi, R. I. Eglitis, and G. Borstel, *Phys. Rev. B* **72**, 045109 (2005).
- ¹⁰²J. Lee, A. Seko, K. Shitara, K. Nakayama, and I. Tanaka, *Phys. Rev. B* **93**, 115104 (2016).
- ¹⁰³A. M. Ganose and D. O. Scanlon, *J. Mater. Chem. C* **4**, 1467 (2016).
- ¹⁰⁴D. Groh, R. Pandey, M. B. Sahariah, E. Amzallag, I. Baraille, and M. Rérat, *J. Phys. Chem. Solids* **70**, 789 (2009).
- ¹⁰⁵J. Heyd, G. E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **118**, 8207 (2003); **124**, 219906 (2006).
- ¹⁰⁶A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, *J. Chem. Phys.* **125**, 224106 (2006).
- ¹⁰⁷R. Armiento and S. Kümmel, *Phys. Rev. Lett.* **111**, 036402 (2013).
- ¹⁰⁸A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- ¹⁰⁹K. Pearson, *London Edinb. Dublin Philos. Mag. J. Sci.* **2**, 559 (1901).
- ¹¹⁰I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. (Springer, New York, 2002).
- ¹¹¹P. Haas, F. Tran, P. Blaha, K. Schwarz, and R. Laskowski, *Phys. Rev. B* **80**, 195109 (2009).
- ¹¹²H. Levämäki, M. P. J. Punkkinen, K. Kokko, and L. Vitos, *Phys. Rev. B* **89**, 115107 (2014).
- ¹¹³M. Städele, J. A. Majewski, P. Vogl, and A. Görling, *Phys. Rev. Lett.* **79**, 2089 (1997).
- ¹¹⁴F. Tran, P. Blaha, and K. Schwarz, *J. Chem. Theory Comput.* **11**, 4717 (2015).

Similarity clustering for representative sets of solids for density functional testing

Péter Kovács,[†] Fabien Tran,[†] Allan Hanbury,[‡] and Georg K. H. Madsen^{*,¶}

[†]*Institute of Materials Chemistry, Technical University of Vienna, Getreidemarkt
9/165-TC, A-1060 Vienna, Austria*

[‡]*Institute for Information Systems Engineering, Technical University of Vienna,
Favoritenstrasse 9-11/194, A-1040 Vienna, Austria*

[¶]*Institute of Materials Chemistry, Technical University of Vienna, Getreidemarkt
9/165-TC, A-1060 Vienna, Austria*

E-mail: georg.madsen@tuwien.ac.at

Abstract

Benchmarking DFT functionals is complicated since the results highly depend on which properties and materials were used in the process. Unwanted biases can be introduced if a dataset contains too many examples of very similar materials. We show that a clustering based on the distribution of density gradient and kinetic energy density is able to identify groups of chemically distinct solids. We then propose a method to create smaller datasets or rebalance existing datasets in a way that no region of the meta-GGA descriptor space is overrepresented, yet the new dataset reproduces average errors of the original set as closely as possible. We apply the method to an existing set of 44 solids and suggest a representative set of seven solids. The representative sets generated with this method can be used to make more general benchmarks or to train new functionals.

1 Introduction

Currently the most widely used theoretical method to predict the different properties of materials is Kohn-Sham density functional theory (KS-DFT).¹ The accuracy of this approach mainly depends on the underlying functional for the exchange-correlation energy, E_{xc} . To compare and rank these functionals various benchmarks were done on different datasets and properties. Notable datasets for molecules are the G2/97² and G3/99³ containing 302 and 376 energies (atomization- and ionization energies, proton- and electron affinities and reaction barrier heights) respectively. Similar databases are used to benchmark functionals for solids as well, like a set⁴ of 18 solids of different types (main group metals, ionic solids, semiconductors and transition metals), an extension of this set containing 44 strongly bound solids⁵ or a set of more than 300 materials used to benchmark the SCAN functional.⁶ Yet these benchmark datasets are often based on "what is available". This can potentially introduce biases for types of materials which are either over- or underrepresented. Unbalanced datasets are problematic and test results can depend on the chosen set in a way that is not transparent. Furthermore, compounds which are very similar and provide little new information lead to unnecessary computational effort.

To avoid or make bias more transparent and for computational efficiency, it would be appealing to create smaller representative benchmark datasets. Still, the literature on this is surprisingly scarce. One approach created two datasets for molecules containing six representative atomization energies and barrier heights respectively.⁷ The results are quite appealing. Obviously from the point of view of computational effort, but also because the representative molecules are both diverse and make sense as representatives of the much larger original datasets. As such, finding representative molecules is also interesting as a data-driven approach to developing a chemical intuition. On the other hand, the representative molecules were chosen to best possible reproduce the average errors obtained for the complete datasets.⁷ Thereby bias in the original dataset will tend to be reflected in the representative set. A group of compounds that are strongly represented in the original dataset, will also tend to

be in the representative set. Even worse, small, but unique, groups of compounds could be left out, thereby potentially covering up problems of a given functional. In this respect, we recently analyzed how the SCAN functional, that generally performs well for lattice parameter calculations, fails for alkali metals⁸. As there are only a limited number of alkali metals, large errors for this small group is not punished in the benchmarks.

In the present study we aim to find materials which are both representative in terms of the electron density distributions sampled and in terms of the errors. Our approach is based on clustering materials according to their density distribution. The idea being that the materials are clustered according to what part of parameter space, in this case density gradients and kinetic energy densities, they occupy. Then representative materials are chosen according to their errors.

2 Methodology

2.1 Density representation and metric

To achieve the clustering we need a descriptor for the materials on which we can define a similarity function. Since the differences between the functionals arise from the different functional forms for E_{xc} energy it seems natural to base our descriptors on the quantities which enter these. The most common functionals for solids are semilocal, where E_{xc} is given as a functional of the density, n , the magnitude of density gradient, $|\nabla n|$ and sometimes the Laplacian of the density and the kinetic energy density (KED) τ defined as

$$\tau(\mathbf{r}) = \frac{1}{2} \sum_i \nabla \psi_i^*(\mathbf{r}) \cdot \nabla \psi_i(\mathbf{r}) \quad (1)$$

The different level of approximations use different arguments. The local density approximation (LDA) uses only the density, the generalized gradient approximations (GGAs) use the gradients as well and meta-GGAs (mGGAs) can use all four parameters. In the present

study we focus on functionals and descriptors based on n , $|\nabla n|$ and τ . We also tested including the Laplacian in our descriptors, but in agreement with our earlier findings⁹, we did not find important differences in the results and it is left out in the following discussion.

Semilocal functions are typically written in terms of the LDA and an enhancement factor which depends on normalized dimensionless, or reduced, values of the mentioned quantities. It is in this enhancement factor that the functionals typically differ. We use the reduced quantities

$$p = \frac{|\nabla n|^2}{4(3\pi^2)^{2/3}n^{8/3}} \quad (2)$$

$$t = \frac{\tau}{\tau_{\text{TF}}} \quad (3)$$

as the descriptors. Here $\tau_{\text{TF}} = (3/10)(3\pi^2)^{2/3}n^{5/3}$ is the Thomas-Fermi KED.

We consider the 44 solids in a previously published dataset.⁵ We use the all-electron KS-DFT code WIEN2k^{10,11} to calculate the density values at every point of the unit cells of these solids. Based on these data, $p - t$ maps for each material are created by binning the densities in a mesh of $p - t$ combinations with a bin width of 0.02 in both directions. The core regions of the atoms contain a large number of points with large values of electron density and low values of the reduced quantities, Eqs. (2) and (3). To avoid that these chemically inactive regions dominate the descriptors, the mesh was subsequently turned into an indicator function being 1 if there was at least one point at the given p, t value and 0 otherwise. After this a Gaussian smearing was applied to the map with a standard deviation of 0.06.

The choice of the similarity/distance metric is essential to achieve a good clustering. Since our goal is to find materials which cover the same region of the $p - t$ space, if two materials cover overlapping regions their distance should be close to zero. The more specific requirement when defining the distance is that it should have a maximum of one, when the materials have no overlap, and should not diverge based on the exact shapes of occupied

regions. Therefore simple Euclidean distances between the matrices are not usable in this case. A choice for similarity which obeys the mentioned requirements is the normalized dot product of the maps, defined the following way:

$$S(A, B) = \frac{\sum_{i,j} A[i, j]B[i, j]}{N(A)N(B)} \quad (4)$$

where A and B represent the $p - t$ maps of two given materials and i, j index bins of p and t . The N normalization function is

$$N(A) = \sqrt{\sum_{i,j} A[i, j]^2} \quad (5)$$

The values of S are always between 0 and 1, being 0 when there is no overlap in the density maps, and 1 when the maps match exactly. Using this, we can define a distance function simply as $1 - S(A, B)$.

2.2 Clustering method

The clustering is done using k -means clustering, more specifically Lloyd's algorithm¹². Given N samples, every sample being a d dimensional vector, and the desired number of clusters, C , the algorithm chooses C samples randomly as cluster centers. Then two steps are iterated until convergence. First, every sample is assigned to the cluster which has the closest centroid. Secondly, the positions of the centroids are updated to the mean of the samples of the given cluster. With this setup the algorithm is guaranteed to converge to a minimum sum of squared distances between the samples and their cluster centers.

Since the basic k -means algorithm works in Euclidean spaces, our distance matrix has to be embedded in a d -dimensional Euclidean space. For this the multidimensional scaling (MDS)¹³ technique is used, which places the materials in a d -dimensional space based on the distance matrix in a way, that the Euclidean distances between their locations fit the

distance matrix as well as possible. The dimensionality of the embedding space limits the achievable accuracy of the MDS, so we opted to use 43 dimensions to represent our data. This embedding method resulted in 0.02 average absolute error between the distance matrix based on the similarity defined in Eq. (4) and the Euclidean distance matrix of the embedded materials.

Because both the MDS and the k -means algorithm involves some randomness we evaluated multiple different embedding and ran the k -means algorithm 10000 times with random starting centroids for every embedding. We will later focus on seven clusters ($C = 7$). These clusters and especially the representative sets based on these were very stable across multiple runs. The small differences in the loosely connected clusters are discussed later. These clusters were also compared to results from affinity propagation or k -means clusterings on the L1 distances of normalized density maps and the resulting clusters are not only consistent with respect to the random seeds, but also across different clustering methods.

2.3 Error based representative sets

We also apply the method that was used to generate the AE6 and BH6 sets.⁷ The method aims to choose a smaller subset of the original data, which reproduces the mean signed error (MSE), mean unsigned error (MUE) and root mean squared error (RMSE) as well as possible. If we denote the difference between e.g. the MSE of the entire database and the representative set when using functional i as $\Delta_{MSE}(i)$, then the aim is the minimalization of the root mean squared deviation (RMSD), defined as:

$$RMSD = \sqrt{\frac{\sum_i \Delta_{MSE}(i)^2 + \Delta_{MUE}(i)^2 + \Delta_{RMSE}(i)^2}{3M}} \quad (6)$$

where M represents the number of different functionals. To evaluate how good a representative set is, the percent error in representation (PEIR) was used:

$$PEIR = 100\% \frac{RMSD}{ME} \quad (7)$$

where ME is the mean error:

$$ME = \frac{\sum_i |\Delta_{MSE}(i)| + |\Delta_{MUE}(i)| + |\Delta_{RMSE}(i)|}{3M} \quad (8)$$

with the errors calculated on the whole dataset. When the whole database is used as representative set, then the PEIR value is zero.

In our case the database consists of 44 materials and we have 24 different GGA and mGGA functionals for three different properties (lattice parameter, bulk modulus and cohesive energy). To find a representative set with N materials the PEIRs for the three properties are calculated for all $\binom{44}{N}$ combinations and the one with the lowest average PEIR is chosen. A direct minimization of the PEIR by choosing seven compounds from the entire 44 compounds results in the group of:



with a PEIR of 15%. This set inherently carries the imbalances of the full set. Six of the seven compounds belong to the transition metals and diamond-lattice semiconductors. It only contains one representative of the alkali metals, and none of the ionic materials nor the earth-alkali metals which are chemically distinct groups and should be present in a small set. As will be discussed later, setPEIR fails to sample a variety of densities, and can, even if it reproduces average errors well, somewhat misrepresent the error for a specific functional.

3 Results and discussion

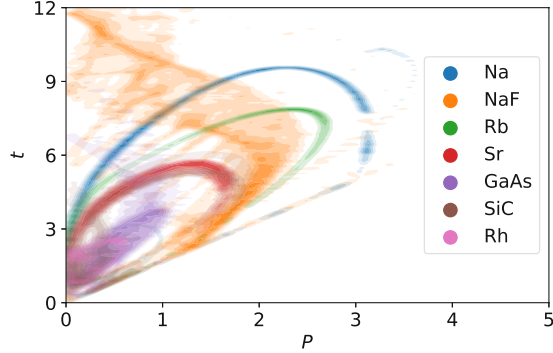


Figure 1: $p-t$ maps of seven representative solids. The clear difference between the different colored regions show that chemically different materials sample distinct regions of the $p-t$ space.

Considering first the $p-t$ maps as descriptors on which the clustering should be based, they are shown for seven different compounds in Fig. 1. It can be seen that these chemically distinct compounds also sample different regions of the $p-t$ maps. Changing e.g. the dependence of the E_{xc} functional on the high p - high t region would mainly influence the results obtained for Na, NaF and similar materials, whereas it would hardly influence the results obtained for the close-packed metal Rh or the semiconductor GaAs. This difference between alkali metals and d -metals or semiconductors falls in line with earlier studies. It has previously been noticed for the atomic electron densities where the maximum value of p (not counting the diverging tail far from the nuclei) decreases along the rows and also along the columns of the periodic table.¹⁴ Furthermore, in case of solid Si and LiF, regions around the outer shell of Li were found to have twice as large p values as in Si.¹⁵ The empty space of the bottom right part of Fig. 1 illustrates the von Weizsäcker limit ($t > 5p/3$). The distance on the y axis from this limit is called $\alpha = t - 5p/3$ and has been shown to carry important information about the bonding properties. In regions occupied by a single orbital $\alpha = 0$,¹⁶ while in regions with slowly varying density $\alpha \approx 1$.¹⁷ α has been also shown to take low values in the covalent bonds of graphite, while being much larger in the interlayer region.¹⁸

The similarity matrix, Eq. (4), of the 44 materials considered here is shown on Fig. 2.

The materials are in an ad-hoc order based on intuition. However, we can still identify multiple groups of similar compounds. These are the close-packed metals, top left, and the semiconductors, bottom right. Some similarity can also be seen between some of the ionic bound materials.

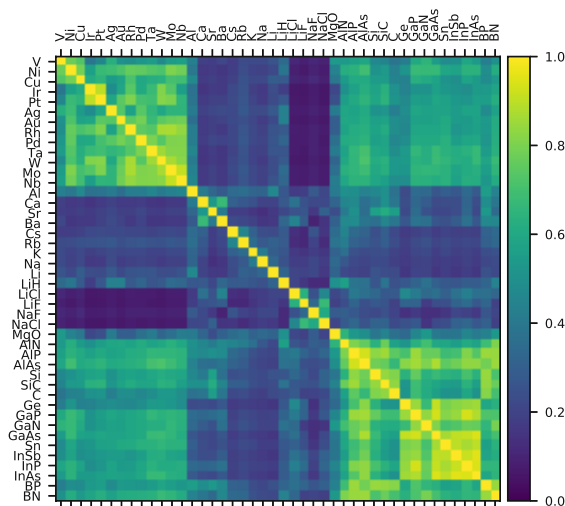


Figure 2: Similarity matrix between materials, with the metal cluster on top left, the semiconductor cluster on bottom right and the less similar groups of ionic materials an alkali- and alkaline earth metals in the middle.

To find the optimal number of clusters we ran the k -means clustering for up to ten clusters. The derivative of the average squared intracluster distances are shown on Fig. 3. It can be seen that making more than seven clusters does not improve the grouping by much. The seven clusters formed this way are: [V, Ni, Cu, Nb, Mo, Rh, Pd, Ag, Ta, W, Ir, Pt, Au], [C, Si, SiC, BN, BP, AlN, AlP], [Ge, Sn, AlAs, GaN, GaP, GaAs, InP, InAs, InSb], [LiH, MgO, Al, Rb, Cs], [LiF, LiCl, NaF, NaCl], [Ca, Sr, Ba], [Li, Na, K]. The intuitive groups that could be recognized by visual inspection of Fig. 2 can be found in this clustering as well. It is pleasing that the transition metals form one large cluster. The diamond-lattice semiconductors are split in two relatively large clusters. Fig. 3 shows how the semiconductors would be grouped into one cluster if only five clusters should be made. The improvement in

mean-square distance between five and seven clusters is however substantial and the splitting is also systematic in the sense that one diamond-lattice cluster tends to contain the atoms from the early periods of the periodic table and the other cluster the atoms from the later periods. There are further smaller clusters of ionic, alkali- and alkaline earth metals. One cluster contains a mixture of ionic compounds and metals, which is also the most unstable cluster, splitting in [LiH, MgO, Al] and [Rb, Cs] groups when 8 instead of 7 clusters are formed.

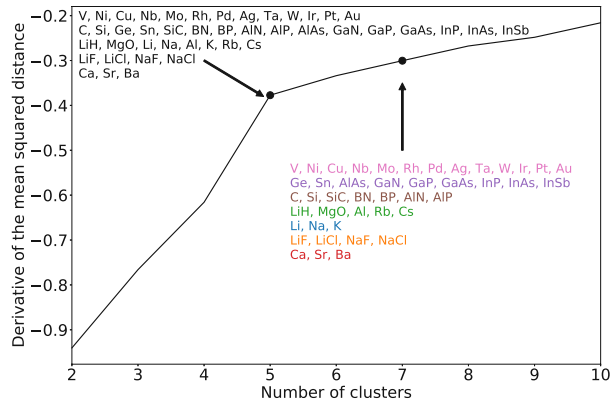


Figure 3: Derivative of the average squared intracluster distances with respect to number of clusters. The colors correspond to the cluster colors of Fig. 1 and 4.

The right panel of Fig.4 shows the 2D representation of the similarity matrix generated by the MDS algorithm. The materials are colored according to the clustering in the 43D space. Each cluster is labelled by one solid, which will later be identified as its first representative. The illustration highlights the strong similarity inside the metal (pink) and semiconductor (purple, brown) clusters and the lower similarity of the clusters containing ionic compounds and alkali and alkaline earth metals. Using only the 2D representation introduces some artifacts mostly around the Na and Rb clusters which results in some of their elements to be seemingly assigned to the wrong cluster. This is only caused by the mismatch of the 2D and 43D representations. The left part of Fig. 4 shows the $p-t$ map obtained by averaging over the solids in each of the seven groups thereby illustrating the most significant regions of $p-t$ values for every cluster. These "average materials" high-

light different regions of mGGA functionals sampled by the materials. If one would use the representative set predicted by the naive PEIR minimization method, setPEIR, the blue, orange and red regions would be unsampled and six of the seven materials would come from the pink, brown and purple areas. These three areas include only the semiconductor and metal clusters and are constrained to the relatively low $p - t$ regions.

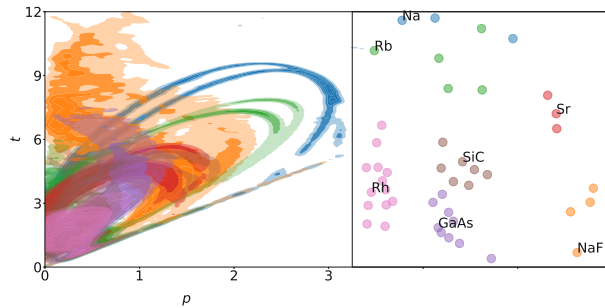
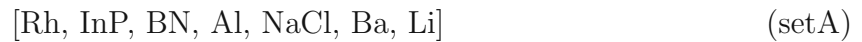


Figure 4: The 2D representation of the distance matrix colored according to the clusters formed by the k -means clustering in the 43D space on the right and $p-t$ maps of the "average materials" of the seven clusters on the left. The labeled materials are the representatives of each cluster and their $p - t$ maps are shown on Fig. 1.

Having the seven clusters, we tested two approaches to find representative sets. The first approach was to calculate the PEIR for every possible combination of seven materials where each material must be from a separate cluster and choose the set with lowest PEIR. This constrained optimization results in the set



with a PEIR of 21%. While this PEIR value is obviously higher than the value of 15% obtained for setPEIR, setA seems more representative. Not only in terms of the $p - t$ maps but also intuitively, in that it is much more diverse in terms of chemistry.

The second approach avoids optimizing the PEIR with respect to the entire dataset and instead chooses from each cluster the material which represents its own cluster best, i.e. the material from each cluster which gives the smallest PEIR with respect to its own cluster.

The set formed this way is

$$[\text{Rh}, \text{GaAs}, \text{SiC}, \text{Rb}, \text{Na}, \text{NaF}, \text{Sr}] \quad (\text{set1})$$

Again this set is representative in terms of $p-t$ and chemical intuition. This set has not been chosen to minimize the PEIR, and the resulting PEIR of 38% is substantially higher than for the sets setPEIR and setA formed by minimizing the total PEIR. However, our goal is not to reproduce the average errors of the full set exactly, but to sample as vast regions of the phase space as possible without unreasonably deviating from the average errors. In the end set1 is preferred since the optimization minimizes the impact of the imbalances of the original dataset. These seven materials are the ones used to label the clusters in Fig. 4 and they were used to exemplify $p-t$ maps in Fig. 1. The strong similarity between Fig. 1 and Fig. 4 shows that the representatives indeed sample the same region as the "average materials" of the given clusters.

Even with representative sets optimized to best possible reproduce an error averaged over functionals and properties according to Eq. (6), it is an open question how well the error for a given property and for a given functional is represented. In Fig. 5 we have chosen the three functionals SCAN¹⁹, TPSS,²⁰ and mBEEF²¹ and show the specific RMSE of setPEIR, setA, and set1, for the three properties. As expected none of the representative sets exactly reproduces the average errors of the entire set. It is worth noting that setPEIR, that was optimized to minimize PEIR without constraints, can result in errors which differ substantially from the full set, e.g. for the cohesive energies obtained with mBEEF or SCAN. It is also noticeable that both sets based on $p-t$ clustering almost always give a lower RMSE than the full set. This is partially caused by the balancing of the dataset. The cluster of close packed metals has the highest RMSE for the lattice constants and cohesive energies of all the clusters for all three of the evaluated functionals. This cluster is down-weighted, in accordance with all these compounds sampling only a small part of $p-t$ space, and therefore

the overall error in the representative sets are reduced.

These results also illustrate that picking one representative material for each cluster may not always be adequate. For both setA and set1 Rh was picked to represent the metal cluster. However, Rh has an error in E_{coh} of 0.3 eV/atom when using the SCAN functional, whereas the RMSE for cohesive energies of the transition metal cluster is 0.54 eV/atom for SCAN. So while Rh is the best material to represent the average error of multiple different functionals, in the sense of Eq. (6), it is somewhat misleading for the E_{coh} error of SCAN. Consequently both set1 and setA give to some degree artificially low error for the SCAN cohesive energy, see Fig. 5. The sets formed by choosing from the representative clustering can, however, be systematically improved by extending the groups of representative materials with additional elements of the clusters. If we choose one additional solid from each cluster by minimize the RMSD with respect to that cluster we obtain



Using set1 and set2 as representative materials a systematic improvement can be observed, see Fig. 5. This can be continued by extending with a third set



The three clusters containing just three compounds, Fig. 3, are then fully present. If the computational cost of the functional evaluation is not a concern, our approach can be still useful to balance the dataset, simply by weighting the different materials based on their cluster size. As an example, the error bar on E_{coh} using TPSS seems to be overestimated due to the strong weight of the transition metal cluster which only samples a rather small part of $p - t$ space.

Irrespective of the average errors of the original set and a representative set, the ranking of the functionals in terms of accuracy is also of importance. Fig. 6 shows the RMSE of 24

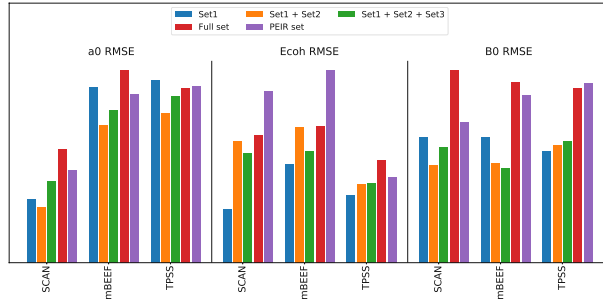


Figure 5: RMSE of three mGGAs for the lattice parameter (left), cohesive energy (middle) and bulk modulus (right), calculated on the original full database, the seven materials set minimising the PEIR and three different size representative sets. The larger representative sets are extended versions of the smaller ones, including two and three materials from every cluster. The errors are scaled, so for every property maximum errors have the same height.

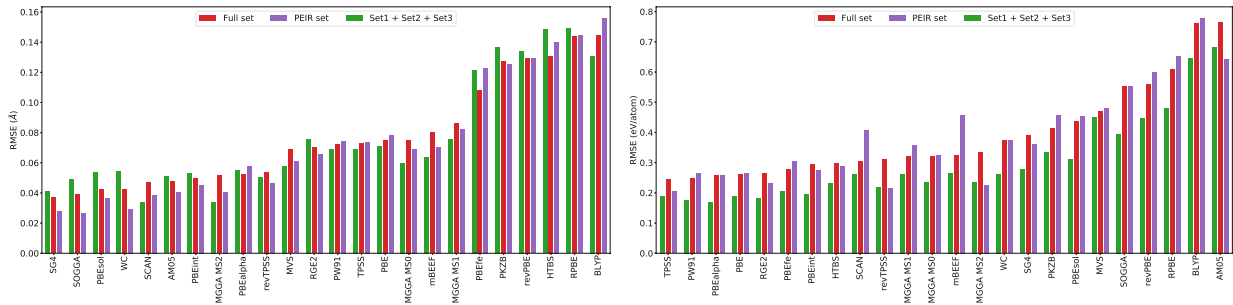


Figure 6: RMSE of 24 GGAs and mGGAs for the lattice parameter (upper panel) and cohesive energy (lower panel) calculated on the original full database, the seven materials set minimising the PEIR and the larger representative set that includes three materials from every cluster. The functionals are ordered according to the RMSE obtained with the original full database. The references for all functionals can be found in Ref. 5.

GGA and mGGA functionals for the lattice parameter and cohesive energy. The functionals are ordered according to the RMSE of the original full set. The ranking of the functionals with two other sets (setPEIR and the set including three materials from every cluster) shows similarities with the original set. By splitting the functionals into three groups, the most accurate, the least accurate, and the middle ones, the groups remain more or less the same independently of the set. There can be inversions within a group of functionals compared to the original database. As discussed above, the error on the cohesive energy seems overestimated for mBEEF and SCAN when using the setPEIR. The representative sets on the other hand give a lower average error for the lattice constants with SCAN and MS2 functionals, mainly due to the down weighting of the transition metals.

4 Summary and conclusions

In the current study we presented a way to group different compounds based on their electron density, allowing us to identify solids which are sampling the same regions of $p - t$ density space. To achieve the grouping we defined a distance metric, which is bound between zero and one, and represents the dissimilarities of the previously mentioned descriptors of different materials. Using multi-dimensional scaling and k -means clustering we formed clusters of similar materials. These are not a pure mathematical construction, but also reflect basic chemical properties. Based on the clustering a small representative set of bulk solid materials is constructed, which not only samples as big regions of the $p - t$ space as possible, but also aims to reproduce the average errors of the original dataset for multiple GGA and mGGA functionals.

The smaller representative sets of the original database, allow for faster evaluation of GGA and mGGA functionals. As the method is able to identify materials which occupy similar regions of the $p - t$ space, thus down weighting highly populated areas can lead to a more general evaluation or functional training.

More recently it has become possible to create test databases based on higher level ab-initio methods.²² An important advantage of the clustering is that it allows for a screening of compounds based just on the DFT descriptors, before computationally heavy calculations are performed.

References

- (1) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (2) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *The Journal of Chemical Physics* **1997**, *106*, 1063–1079.
- (3) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-3 and density functional theories for a larger experimental test set. *The Journal of Chemical Physics* **2000**, *112*, 7374–7383.
- (4) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Tests of a ladder of density functionals for bulk solids and surfaces. *Physical Review B* **2004**, *69*.
- (5) Tran, F.; Stelzl, J.; Blaha, P. Rungs 1 to 4 of DFT Jacob’s ladder: Extensive test on the lattice constant, bulk modulus, and cohesive energy of solids. *J. Chem. Phys.* **2016**, *144*, 204120.
- (6) Zhang, Y.; Kitchaev, D. A.; Yang, J.; Chen, T.; Dacek, S. T.; Sarmiento-Pérez, R. A.; Marques, M. A. L.; Peng, H.; Ceder, G.; Perdew, J. P.; Sun, J. Efficient first-principles prediction of solid stability: Towards chemical accuracy. *npj Comput. Mater.* **2018**, *4*, 9.

- (7) Lynch, B. J.; Truhlar, D. G. Small Representative Benchmarks for Thermochemical Calculations. *The Journal of Physical Chemistry A* **2003**, *107*, 8996–8999.
- (8) Kovács, P.; Tran, F.; Blaha, P.; Madsen, G. K. H. Comparative study of the PBE and SCAN functionals: The particular case of alkali metals. *The Journal of Chemical Physics* **2019**, *150*, 164119.
- (9) Tran, F.; Kovács, P.; Kalantari, L.; Madsen, G. K. H.; Blaha, P. Orbital-free approximations to the kinetic-energy density in exchange-correlation MGGA functionals: Tests on solids. *The Journal of Chemical Physics* **2018**, *149*, 144105.
- (10) Blaha, P.; Schwarz, K.; Madsen, G. K. H.; Kvasnicka, D.; Luitz, J.; Laskowski, R.; Tran, F.; Marks, L. D. WIEN2k: An Augmented Plane Wave plus Local Orbitals Program for Calculating Crystal Properties. ISBN 3-9501031-1-2. 2018.
- (11) Blaha, P.; Schwarz, K.; Tran, F.; Laskowski, R.; Madsen, G. K. H.; Marks, L. D. WIEN2k: An APW+lo program for calculating the properties of solids. *The Journal of Chemical Physics* **2020**, *152*, 074101.
- (12) Lloyd, S. Least squares quantization in PCM. *IEEE Transactions on Information Theory* **1982**, *28*, 129–137.
- (13) Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **1964**, *29*, 1–27.
- (14) del Campo, J. M.; Gázquez, J. L.; Alvarez-Mendez, R. J.; Vela, A. The reduced density gradient in atoms. *International Journal of Quantum Chemistry* **2012**, *112*, 3594–3598.
- (15) Haas, P.; Tran, F.; Blaha, P.; Schwarz, K.; Laskowski, R. Insight into the performance of GGA functionals for solid-state calculations. *Phys. Rev. B* **2009**, *80*, 195109.
- (16) Becke, A. D.; Edgecombe, K. E. *J. Chem. Phys.* **1990**, *92*, 5397–5403.

- (17) Sun, J.; Xiao, B.; Fang, Y.; Haunschild, R.; Hao, P.; Ruzsinszky, A.; Csonka, G. I.; Scuseria, G. E.; Perdew, J. P. Density Functionals that Recognize Covalent, Metallic, and Weak Bonds. *Phys. Rev. Lett.* **2013**, *111*, 106401.
- (18) Madsen, G. K. H.; Ferrighi, L.; Hammer, B. Treatment of Layered Structures Using a Semilocal meta-GGA Density Functional. *J. Phys. Chem. Lett.* **2010**, *1*, 515–519.
- (19) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.
- (20) 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*, 146401.
- (21) Wellendorff, J.; Lundgaard, K. T.; Jacobsen, K. W.; Bligaard, T. mBEEF: An accurate semi-local Bayesian error estimation density functional. *J. Chem. Phys.* **2014**, *140*, 144107.
- (22) Schmidt, P. S.; Thygesen, K. S. Benchmark Database of Transition Metal Surface and Adsorption Energies from Many-Body Perturbation Theory. *The Journal of Physical Chemistry C* **2018**, *122*, 4381–4390.

7 Curriculum vitae

Péter Kovács

PhD student
TU Wien

Untere Augartenstraße 19/2/14
1020 Vienna
Austria
☎ +36 (70) 334 5737
✉ kovpet0330@mail.com

Education

- 2017–2021 **PhD, Theoretical Materials Chemistry**, TU Wien, Vienna, Austria.
2015–2017 **MSc, Physics**, Budapest University of Technology and Economics, Budapest, Hungary.
2012–2015 **BSc, Physics**, Budapest University of Technology and Economics, Budapest, Hungary.

PhD thesis

- title Machine learning application for DFT exchange functionals
supervisor Univ. Prof. Dr. Georg K. H. Madsen
description Developing a method to assess and correct biases in a database of materials w.r.t. the normalized density gradient and normalized kinetic energy density descriptors. Training 25 new exchange functionals to explore the limits of mGGA functionals and understand the differences between various approximations. Developing a neural network based model to predict the infrared spectra of polycyclic aromatic hydrocarbons.

Master thesis

- title Computer simulation of magnetic alloys
supervisor Dr. László Udvardi
description Developing a Monte-Carlo simulation for multi-component Heisenberg-model calculations. Investigating the concentration dependence of the Curie temperature in Fe-Co alloys.

Publications

- "Orbital-free approximations to the kinetic-energy density in exchange-correlation MGGA functionals: Tests on solids"; F Tran, **P Kovács**, L Kalantari, GKH Madsen, and P Blaha; J. Chem. Phys. 149, 144105 [2018]
- "Comparative study of the PBE and SCAN functionals: The particular case of alkali metals"; **P Kovács**, F Tran, P Blaha, and GKH Madsen; J. Chem. Phys. 150, 164119 [2019]

- "Machine-learning Prediction of Infrared Spectra of Interstellar Polycyclic Aromatic Hydrocarbons"; **P Kovács**, X Zhu, J Carrete, GKH Madsen, and Z Wang; *ApJ* 902 100 [2020]
- "Similarity clustering for representative sets of solids for density functional testing"; **P Kovács**, F Tran, A Hanbury, and GKH Madsen; (submitted)
- "Machine Learning interpretation of the correlation between infrared emission features of interstellar polycyclic aromatic hydrocarbons "; Z Meng, X Zhu, **P Kovács**, E Liang, Z Wang; (submitted)

Experience

2017–2021 **University assistant**, TU Wien, Institute of Materials Chemistry, Vienna.

2013–2015 **Market Risk Analyst**, *Morgan Stanley*, Budapest.

Skills

Codes WIEN2k, Tensorflow

Languages C, C#, Python

Languages

English Fluent

German Basic communication skills

Hungarian Native

Mother tongue