Versatile Density Functionals for Computational Surface Science

Wellendorff, Jess; Jacobsen, Karsten Wedel; Bligaard, Thomas

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Versatile Density Functionals for Computational Surface Science

Dissertation for the degree of Doctor Philosophiae

Jess Wellendorff
August 2012

Center for Atomic-scale Materials Design
Department of Physics
Technical University of Denmark

Technical University of Denmark
Department of Physics
Center for Atomic-scale Materials Design
Building 311, DK-2800 Kongens Lyngby, Denmark
www.camd.dtu.dk

Preface

This thesis is submitted in candidacy for the Ph.D. degree in physics from the Technical University of Denmark (DTU). The work has been carried out from August 2009 to Juli 2012 at the Center for Atomic-scale Materials Design (CAMD), Department of Physics, DTU. Supervisors have been Professor Karsten W. Jacobsen, Director of CAMD, and Associate Professor Thomas Bligaard, now Senior Staff Scientist at the SUNCAT Center for Interface Science and Catalysis at SLAC National Accelerator Laboratory, Menlo Park, California. This thesis contains results obtained in collaboration with other researchers at CAMD and SUNCAT.

Lyngby, August 1, 2012
Jess Wellendorff

Abstract in English

Density functional theory (DFT) emerged almost 50 years ago. Since then DFT has established itself as the central electronic structure methodology for simulating atomic-scale systems from a few atoms to a few hundred atoms. This success of DFT is due to a very favorable accuracy-to-computational cost ratio. In a wide range of applications, in fields as diverse as condensed matter physics over geophysics, chemistry, and chemical engineering to molecular biology, DFT now continuously delivers the theoretical base for experimental interpretation and forms an essential component for establishing reliable insights and generation of knowledge. In spite of the great successes of DFT, fundamental challenges still exist for the theory. Systematic improvement of the central ingredient in the theory—the approximate exchange–correlation functional—is by no means a straightforward task: Some materials and chemical properties are simply not treated acceptably well, errors are unknown and often difficult to control, and it is difficult to evaluate whether a highly optimized, empirically fitted, functional is truly a "good" functional or whether it is merely over-parametrized and overfitted. In this thesis I address these problems systematically. It is here analyzed carefully which ingredients must be included to establish a functional which simultaneously performs well for a range of important materials and chemical properties. A general methodology is defined for the multi-objective fitting of exchange–correlation functionals in a language that naturally allows for error prediction within DFT. The optimization methodology includes as a central ingredient establishment of a compromise between high transferability (measured by the smoothness of the resulting functional) and low prediction error using statistical resampling techniques, thereby systematically avoiding problems with overfitting. The first ever density functional presenting both reliable accuracy and convincing error estimation is generated. The methodology is general enough to be applied to more complex functional forms with higher-dimensional fitting and resampling. This is illustrated by searching for meta-GGA type functionals that outperform current meta-GGAs while allowing for error estimation.

Resumé på dansk

Tætheds-funktional teori (DFT) opstod for næsten 50 år siden. Siden da er DFT blevet den vigtigste elektron-struktur metode til simulering af atomar-skala systemer fra et par atomer til få hundrede atomer. Baggrunden for dette er et meget favorabelt forhold mellem nøjagtighed og beregningsmæssig omkostning. I en bred vifte af anvendelser, indenfor felter så forskellige som faststoffysik over geofysik, kemi og kemiteknik til molekylær biologi, leverer DFT nu kontinuerligt det teoretiske grundlag for fortolkning af eksperimentelle resultater, og danner et essentielt grundlag for etablering af pålidelige indsigter og frembringelse af ny viden. På trods at denne store succes for DFT har teorien stadig fundamentelle udfordringer. Systematisk forbedring af den helt centrale komponent i teorien—det approximative exchange–korrelations-funktional—er på ingen måde en simpel opgave: Visse typer af materialer og kemiske egenskaber behandles simpelthen ikke nøjagtigt nok, usikkerheder på beregninger er ukendte og ofte svære at kontrollere, og det er vanskeligt at bedømme om et stærkt optimeret, empirisk tilpasset, funktional virkelig er et "godt" funktional, eller om det blot er over-parametriseret og "overfittet". I denne afhandling adresserer jeg systematisk disse problemer. Der udføres en omhyggelig analyse af hvilke ingredienser der bør inkluderes for at frembringe et funktional som virker godt for en hel række af vigtige egenskaber i materiale-fysik og kemi. En generel metodik defineres for multi-objektiv optimering af exchange–korrelations-funktionaler i en formulering som helt naturligt leder til usikkerheds-estimater i DFT. Et centralt element i optimerings-metoden er bestemmelse, ved hjælp af resamplings-teknikker, af et kompromis mellem høj transferabilitet (målt på blødheden af det resulterende funktional) og lav fejl på forudsigelser. Herved undgås systematisk problemer med overfitning. Det første tætheds-funktional med både pålidelig akkuratesse og overbevisende usikkerheds-estimater fremstilles. Metodikken er generel nok til at blive anvendt på mere komplicerede funktionelle former med højere-dimensional optimering og resampling. Dette illustreres med en søgning efter nye funktionaler af meta-GGA-typen, som overgår nuværende meta-GGA funktionaler, samtidig med at metoden tilbyder muligheden for usikkerheds-estimater.

Acknowledgements

Many people at CAMD and SUNCAT should be acknowledged for their contributions to the successful completeion of this work. First of all I would like to thank Karsten W. Jacobsen and Thomas Bligaard for excellent supervision of my project. It has been a great pleasure working together with such committed and knowledgeable people for the past three years. I have really learned a lot from this. A special thanks also goes to Keld T. Lundgaard, my partner in crime throughout the second half of my time as a Ph.D. student, which was probably also the most productive time. Keld has a large share in this work, and I think we have come a long way in the business of developing density functionals, while also taking the time to enjoy ourselves and discuss completely unrelated matters of the "real world". This was especially so during our several trips abroad for conferences and visits in the SUNCAT group at SLAC.

I would furthermore like to thank all other co-authors of the manuscripts included in this thesis for their contributions; Andreas Møgelhøj, Vivien Petzold, David D. Landis, Jens K. Nørskov, André Kelkkanen, Jens Jørgen Mortensen, and Bengt I. Lundqvist. Frank Abild-Pedersen and Felix Studt from SUNCAT are gratefully acknowledged for several interesting discussions relating to the BEEF-vdW development presented here.

My every-day work for the last three years would have been far less productive without invaluable help by CAMD computer staff Ole Holm Nielsen and Marcin Dułak, and Head of Administration Marianne Ærsøe, who have all made my life much easier. The same thanks goes to Helle Wellejus, Cheryl Johnson, and Christopher O'Grady, who eased my path the two times I visited SLAC.

Finally, I am especially thankful to Nina, who was careless enough to marry me in January 2012. My Ph.D. study has been her project as much as mine, particularly for the last six months, where the heat really went on. Her endless patience with my preoccupied mind has been truly invaluable.

List of Included Papers

Paper I
RPBE-vdW Description of Benzene Adsorption on Au(111)
J. Wellendorff, A. Kelkkanen, J. J. Mortensen, B. I. Lundqvist, and T. Bligaard
Topics in Catalysis **53**, 378 (2010)

Paper II
On the Importance of Gradient-Corrected Correlation for van der Waals Density Functionals
J. Wellendorff and T. Bligaard
Topics in Catalysis **54**, 1143 (2011)

Paper III
Density Functionals for Surface Science: Exchange-correlation Model Development with Bayesian Error Estimation
J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, and K. W. Jacobsen
Physical Review B **85**, 235149 (2012)

Contents

Chapter 1

Introduction

Modern materials science is an ever growing effort to understand the phenomena determining the properties of materials and processes on the atomic level. Though significant progress has been made in experimental techniques to unravel the secrets of condensed matter physics and chemistry, the interactions between nuclei and electrons are ultimately governed by the laws of quantum mechanics and the electronic structure of matter. Computational approaches to materials science are therefore becoming increasingly important, and density functional theory (DFT) in particular. Such first principles calculations offer not only a framework for fundamental understanding of processes in nanoscale physics and chemistry, but also computer-aided *design* of new materials for future technologies. [1,2]

Catalysis and surface chemistry in general is one branch of materials science, and is of immence importance to modern day chemical and petrochemical industry. A catalyst is a substance which accelerates a chemical reaction without itself being consumed in the process. In heterogeneous catalysis the catalyst material is ususally a solid and the molecular reactants and reaction products may be in the gas phase or liquid phase. Modern DFT methods have proven highly valuable for elucidating mechanisms and fundamental trends in enzymatic and heterogeneous catalysis, [3–7] and computational design of chemically active materials by "electronic structure engineering" has been demonstrated. [8–11] The catalyst takes part in a continuous cycle, where gas phase molecules approach the chemically active surface and bonds to it (adsorbs), whereafter one of more chemical transformations (most often bond rupture and formation) leads to the reaction product(s), which may then leave the surface in a desorption process. The active site is (ideally) unaltered after a catalytic cycle. [12,13]

The first catalytic step is adsorption of reactants. This is an exothermic process; it is energetically favorable for the reactant molecules to bond to the surface (and possible dissociate), and the total energy of the combined system of surface plus reactant is lowered upon adsorption. Interactions develop when the molecule approaches the surface, leading first to physical adsorption (physisorption) mediated by long-range van der Waals (vdW) type interactions. However, no actual chemical bonds are formed. As the adsorbate gets closer, significant overlaps develop between the surface and adsorbate electron densities. Hybridization occurs and a surface–reactant chemical bond is formed. This is chemical adsorption (chemisorption), the first step in most catalytic processes of industrial relevance.[13] Note also that an energy barrier often separates the physisorbed and chemisorbed states (a transition state), which must be surmounted for the reaction to proceed to products. Reaction barriers often separate intermediate reaction steps as well, even in the gas phase. Accurate calculation of adsorption energies and reaction barriers for virtually any thinkable (and unthinkable) combination of adsorbate and surface substrate are important targets of DFT for surface science applications, including catalysis.

The quest: Versatile density functionals

Density functional theory relies in practice on approximations to the exact density functional, which is supposed to account for exchange–correlation (XC) effects between electrons. Several levels of XC approximation may be applied, and the challenge is to obtain the maximum computational accuracy with the computing resources available. The generalized gradient approximation (GGA) is very popular in surface science due to a high accuracy-to-cost ratio for many applications, but suffer from a range of shortcomings. Thus, common GGA functionals are well-suited for computing many important quantities in chemistry and condensed matter physics, but appear to be fundamentally unable to accurately describe the physics and chemistry of a surface at the same time. Moreover, the van der Waals dispersion interactions responsible for physisorption phenomena are not accounted for by GGAs. Surface science is an interdisciplinary field, straddling from solid state physics to chemistry and even biology. Furthermore, the interest in applying DFT to more and increasingly complex problems in is not likely to decrease in the years to come. Computationally tractable general-purpose density functionals are therefore a highly active field of research. The work presented here constitute the first few steps in a novel approach to development of versatile density functionals for surface science.

Outline

This thesis is organized "bottom-up", the fundamentals first and applications last, and chapters are expected to be read succesively. Chapter 2 introduces the basics of electronic structure calculations with a focus on the principle and application of density functional theory, but wavefunction methods are also briefly discussed. Chapter 3 introduces common levels of exchange–correlation approximations from the point of view of application, while Chapter 4 considers mainly a range of benchmark datasets from high-level theory or experiments. These are used for evaluating XC approximations throughout.

In Chapter 5 the vdW-DF approach to dispersion forces in DFT is introduced, and several variants of this method are evaluated in terms of general applicability within theoretical chemistry and solid state physics. Moreover, a potential improvement of the vdW-DF method by inclusion of semilocal correlation contributions to the otherwise purely local part of the correlation functional is proposed. Calculations of benzene physisorption on the noble (111) surfaces of Cu, Ag, and Au support this construction as compared to the original vdW-DF.

Chapter 6 introduces some basic principles of Bayesian statistics and machine learning methods for efficient but controlled parametrizations of mathematical models. These tools are applied in Chapter 7, where a methodology for parametrization of general-purpose density functionals is developed. As a bonus, an ensemble of functionals designed for systematic error estimation in DFT comes out naturally. This machinery is applied for generating the BEEF-vdW, a reasonably versatile GGA+vdW density functional for surface science. Finally, in Chapter 8, the latest developments in our group towards even better XC models are highlighted. This includes the so-called meta-GGA XC model space, and promising results have been obtained. This is, however, work in progress.

Chapter 2

Electronic Structure Calculations

Successful use of electronic structure calculations rests heavily on useful approximations for solving the Schrödinger equation for mutually interacting electrons and nuclei. The challenge is therefore to determine, as accurately as possible given the computational ressources at hand, the ground state many-body wavefunction for atomistic systems. This chapter introduces density functional and wavefunction-based methods for tackling the interacting many-body problem, with emphasis on exchange–correlation (XC) approximations for density functional theory (DFT). Methodologies for practical implementation of DFT in computer codes are also discussed. Atomic units are used throughout, meaning $e = \hbar = a_0 = m_e = 1$ for the unit charge, reduced Planck constant, Bohr radius, and electron mass, respectively.

2.1 Many-body problem

A fundamental postulate of quantum mechanics is that all information about a system of particles is contained in the many-body wavefunction $|\Psi\rangle$ for the particles.[14] In electronic structure theory these particles are the electrons and nuclei constituting matter,[15] and the wavefunction is in general an eigenstate of the time-independent and non-relativistic many-body Hamiltonian $\widehat{H}$,

$$\widehat{H}|\Psi_n\rangle = \varepsilon_n|\Psi_n\rangle \tag{2.1}$$

where $|\Psi_n\rangle$ is labeled by the state label n and ε_n is the corresponding energy eigenvalue. The ground state wavefunction $|\Psi_0\rangle$ yields the ground state total energy ε_0 for the fully interacting quantum system.

2.1.1 Electronic Hamiltonian

The many-body Hamiltonian in the Schrödinger equation (2.1) consists of the kinetic energy of all particles (electrons and nuclei) as well as inter-particle interactions. All particles are thus correlated, and the ground state position of each depends on the positions (and spins) of all others in a highly complicated manner. A common and very reliable simplification is applying the Born–Oppenheimer (BO) approximation:[16] Since the nuclei (or ions) are much heavier than electrons, the former may be treated as infinitely heavy. This corresponds to neglecting the nuclear kinetic energy and any other quantum interactions involving nuclei. Applying this static-ion approximation we consider an electronic Hamiltonian,

$$\begin{aligned}\widehat{H} &= \widehat{T} + \widehat{V}_{\text{int}} + \widehat{V}_{\text{ext}} \\ &= -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \sum_{i,I}\frac{Z_i}{|\boldsymbol{r}_i - \boldsymbol{R}_I|},\end{aligned} \tag{2.2}$$

where lowercase (uppercase) indices denote electrons (ions). The first operator, $\widehat{T}$, sums up the the electron kinetic energies expressed in terms of the Laplacian with respect to electron coordinates, $\widehat{V}_{\text{int}}$ sums all electron–electron interactions, and $\widehat{V}_{\text{ext}}$ electron–ion interactions. The latter is the electrostatic potential of the fixed ions acting on the electrons, which is "external" to them.

The electronic Hamiltonian depends only parametrically on the ion positions $\boldsymbol{R}_I$ and looks rather friendly at first sight, but is far from trivial. It contains many-body quantum mechanical interactions between all electrons, and is not solvable analytically except in very simple cases, such as the hydrogen atom. The N-electron wavefunction $|\Psi\rangle$ depends on the $3N$ spatial coordinates and N spin coordinates, and must be antisymmetric under electron permutation in order to satisfy the Pauli exclusion principle for fermions. However complicated, it is the fundamental Hamiltonian in electronic structure theory. In practice, the vast majority of electronic structure calculations aims at approximately solving (2.1) for the electronic Hamiltonian (2.2).

2.1.2 Ground state quantities

Solution of the Schrödinger equation (2.1) is an eigenvalue problem for the many-electron eigenstates. Once these are found, any physical observable O for some state of the system is the expectation value of the corresponding operator $\widehat{O}$ with the wavefunction in question,

$$\langle\widehat{O}\rangle = \langle\Psi|\widehat{O}|\Psi\rangle, \tag{2.3}$$

where $|\Psi\rangle$ is assumed normalized, $\langle\Psi|\Psi\rangle = 1$.

The ground state total electronic energy E_0 of a system of electrons and static ions is one primary object of electronic structure calculations. It is the expectation value of the Hamiltonian operator with the ground state (the N-electron state yielding the lowest energy), which is the zero-temperature equilibrium state for the interacting electrons in the potential of the static ions. [17]

The total electron density $n(\boldsymbol{r})$ is another important quantity. It is defined such that $n(\boldsymbol{r})d\boldsymbol{r}$ is the probaility of finding an electron in the spatial volume element $d\boldsymbol{r}$ around $\boldsymbol{r}$. The density operator $\hat{n}(\boldsymbol{r})$ may be defined in terms of the Dirac delta function,

$$\hat{n}(\boldsymbol{r}) = \sum_{i=1}^{N} \delta(\boldsymbol{r} - \boldsymbol{r}_i), \tag{2.4}$$

such that the electron density is an integral over the coordinates of $N-1$ electrons,

$$n(\boldsymbol{r}) = \sum_{i=1}^{N} \langle\Psi|\delta(\boldsymbol{r} - \boldsymbol{r}_i)|\Psi\rangle = N \int d_2 \dots d_N |\Psi(\boldsymbol{r}, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N), \tag{2.5}$$

where we have implicitly summed over electron spins, $n(\boldsymbol{r}) = n_\uparrow(\boldsymbol{r}) + n_\downarrow(\boldsymbol{r})$.

The expectation value of the external potential operator in (2.2) is simply the action of the external potential $v_{\text{ext}}(\boldsymbol{r})$ on the electron density,

$$\langle \widehat{V}_{\text{ext}} \rangle = \langle\Psi| \sum_{i=1}^{N} v_{\text{ext}}(\boldsymbol{r}_i) |\Psi\rangle = \int v_{\text{ext}}(\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r}, \tag{2.6}$$

where the sum is over electrons. The total energy of the electronic Hamiltonian (2.2) may therefore be written

$$E = \langle \widehat{H} \rangle = \langle\Psi|\widehat{H}|\Psi\rangle = \langle \widehat{T} \rangle + \langle \widehat{V}_{\text{int}} \rangle + \int v_{\text{ext}}(\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r}, \tag{2.7}$$

for ground and excited states of the atomistic system.

2.1.3 Hartree–Fock approximation

The Hartree–Fock (HF) approximation is a standard method in electronic structure theory. Each N-electron eigenstate (ground and excited states) is assumed describable by a single Slater determinant, an antisymmetrized product of one-electron wavefunctions (orbitals). Each electron i is assumed to move in an effective potential $\widehat{V}^i_{\text{eff}}(\boldsymbol{r})$, set up by the ions and all other electrons, [15]

$$\widehat{V}^i_{\text{eff}}(\boldsymbol{r}) = \widehat{V}_{\text{ext}}(\boldsymbol{r}) + U_{\text{H}}(\boldsymbol{r}) + \widehat{V}^i_x(\boldsymbol{r}), \tag{2.8}$$

where $U_{\mathrm{H}}(\boldsymbol{r})$ the Hartree potential describing the classical electrostatic interaction of the electron density with itself,

$$U_{\mathrm{H}}(\boldsymbol{r}) = \frac{1}{2}\int \frac{n(\boldsymbol{r})n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}d\boldsymbol{r}d\boldsymbol{r}'. \tag{2.9}$$

The last term in (2.8), $\widehat{V}_x^i$, is the HF exchange interaction accounting for repulsion between electrons of equal spin due to the Pauli exclusion principle.

The HF Hamiltonian is then written

$$\widehat{H}_{\mathrm{HF}}^i = \widehat{T} + \widehat{V}_{\mathrm{eff}}^i = \widehat{T} + U_{\mathrm{H}} + \widehat{V}_{\mathrm{ext}} + \widehat{V}_x^i. \tag{2.10}$$

If compared to (2.7), it is seen that the sum $U_{\mathrm{H}} + \widehat{V}_x^i$ represents electron–electron interactions in an orbital dependent fashion, since the exchange potential depends explicitly on each particular orbital i. We note in passing that the (unphysical) self-interaction of an electron interacting with itself contained in U_{H} is cancelled out be an equal but oppositely signed term in $\widehat{V}_x^i$ when taking the expectation value of $\widehat{H}_{\mathrm{HF}}^i$.

One may apply the variational principle

$$E_0 \leq \langle\Psi|\widehat{H}|\Psi\rangle, \tag{2.11}$$

in search for the normalized single-determinant HF wavefunction $|\mathrm{HF}\rangle$ that minimizes the expectation value of the HF Hamiltonian. Starting from a trial wavefunction, the self-consistent field method (SCF) is an iterative improvement of $|\Psi\rangle$ towards $|\mathrm{HF}\rangle$ by minimizing the total energy with respect to all degreees of freedom in the wavefunction. We return to the SCF method in Section 2.3.

In the HF approximation the true ground state is, however, almost never found. The HF Hamiltonian neglects all electron–electron correlations except exchange repulsion between equal-spin electrons (Pauli repulsion). The neglected interactions are in wavefunction theory referred to as correlation, which may even be further grouped into dynamical and static correlation.[18] However, the distinction between exchange and correlation becomes more fuzzy in approximate density functional approaches to electronic structure (more about this later).

Lastly, it is important to note that a wealth of high-level quantum chemical approaches to solving the N-electron Schrödinger equation take the HF solution as a starting point for elaborate approximations to the neglected correlation interactions. These are not really the topic of this thesis, but several of the most accurate wavefunction methods provide quality benchmark data, which are used in the present work for assessing the abilities of more approximate DFT approaches. Some of the most relevant post-HF quantum chemical methods are[18]

- **Configuration-interaction (CI) method:** The exact N-electron ground state is approximated by a linear combination of Slater determinants rather than a

single one as in HF. The CI ground state may be thought of as arising from increasingly higher-order electron excitations out of the $|\mathrm{HF}\rangle$ solution, that is, single, double, triple, etc. excitations.

- **Coupled-cluster (CC) method:** The linear CI expansion to infinite excitation order (full CI) is recast as a product wavefunction, and the CC solution is then a linear combination of excited configurations relative to $|\mathrm{HF}\rangle$. This method converges faster towards the true wavefunction for increasing excitation order than CI does. The coupled-cluster method with single, double, and perturbative triple excitations, CCSD(T), is today a standard high-level tool in quantum chemistry.

- **Møller–Plesset (MP) pertubation theory:** Simply applying pertubation theory to second order or higher on the HF ground state leads to estimates of the missing correlation energy at the MP2, MP3, MP4, etc. levels of theory. Especially the MP2 method provides a favorable accuracy-to-cost ratio.

The particularly interested reader is referred to the authoritative textbook of Helgaker *et al.* (Ref. 18).

2.2 Density functional theory

Density functional theory[19–21] takes a rather different approach to the interacting N-electron problem than wavefunction-based methods attempting to refine the HF solution. The fundamental principle of DFT is that any property of a fully interacting system of electrons can be viewed as a functional* of the ground state electron density $n_0(\boldsymbol{r})$. It was thus proved in the pioneering work of Hohenberg and Kohn,[19] almost 50 years ago, that the scalar function $n_0(\boldsymbol{r})$ in principle determines all ground- and excited-state properties of the system. In two theorems, now known as the Hohenberg–Kohn (HK) theorems, they basically showed that there is a one-to-one correspondence between a given external (ionic) potential and the ground state electron density. That is, the ground state density of N electrons in some external potential is unique in the sense that it is not producible by any other external potential. It follows that the N-electron Hamiltonian is also determined by $n_0(\boldsymbol{r})$ (except for a trivial constant), and in principle also any N-electron wavefunction of the system.[15]

HK-DFT is therefore nothing less than a simple but exact theory of correlated many-body systems, including *any* system of interacting particles in an external potential. It leads to the total energy functional $E_{\mathrm{HK}}[n(\boldsymbol{r})]$,

$$E_{\mathrm{HK}}[n] = T[n] + E_{\mathrm{int}}[n] + \int v_{\mathrm{ext}}(\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r}, \tag{2.12}$$

* If the function $y(\boldsymbol{x})$ maps the vector argument $\boldsymbol{x}$ to the scalar function value y, the *functional* $F[y(\boldsymbol{x})]$ is a function of the function $y(\boldsymbol{x})$, and maps $\boldsymbol{x}$ to F through y.

where $n = n(\boldsymbol{r})$ for ease of notation. Again, $T[n]$ is the electron kinetic energy, $E_{\text{int}}[n]$ electron–electron interaction energies, and the last term the energy of the ionic potential acting on the electrons. Though not a necessary condition, (2.12) is clearly written assuming validity of the BO approximation. Since $T[n]$ and $E_{\text{int}}[n]$ are functionals of the electron density only, they constitute a universal functional $F[n]$, valid for *any* system of electrons independently of the external potential. We may therefore write

$$E_{\text{HK}}[n] = F[n] + \int v_{\text{ext}}(\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r}, \tag{2.13}$$

$$F[n] = T[n] + E_{\text{int}}[n]. \tag{2.14}$$

The HK approach boils down to knowing $F[n]$ and minimizing the total energy (2.13) with respect to all variations in $n(\boldsymbol{r})$ (the variational principle). The density $n_0(\boldsymbol{r})$ that minimizes the total energy functional is the ground state density, ultimately composed of the ground state N-electron wavefunction $|\Psi_0\rangle$, and the energy E_0 is the ground state total electronic energy.

However, the HK theorems provide no guidance for constructing $F[n]$, but only prove its existence, and no exact form of $F[n]$ is known for any system of more than one electron. Was it not for the subsequent work of Kohn and Sham, who provided a framework for constructing approximate but useful density functionals for systems of many electrons, DFT would probably have remained a curiosity of condensed matter physics.

2.2.1 Kohn–Sham formalism

Kohn and Sham proposed in their seminal 1965 work [20] to replace the interacting many-body system of the Hohenberg–Kohn theorems with a fictitious (auxiliary) independent-particle system that is more easily solved. The Kohn–Sham (KS) formalism is based on the *ansatz* that the ground state electron density of the fully interacting system equals that of the auxiliary system. The HK approach can then be reformulated in terms of solvable independent-particle equations, with all the most difficult electron–electron interactions collected in a single exchange–correlation functional of the electron density. Any auxiliary Hamiltonian will do the job, as long as it has the correct density in the ground state!

The KS Hamiltonian is thus one for the fictitious system of non-interacting electrons,

$$\widehat{H}_{\text{KS}} = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 + v_{\text{s}}(\boldsymbol{r}), \tag{2.15}$$

where the effective KS singe-particle potential $v_{\text{s}}(\boldsymbol{r})$ is such that $n_0(\boldsymbol{r})$ of $\widehat{H}_{\text{KS}}$ equals that of the real (physical) Hamiltonian (2.2). Once we know $v_{\text{s}}(\boldsymbol{r})$, minimizing $\widehat{H}_{\text{KS}}$ is trivial: we simply solve the Schrödinger equation for each single-electron state $\phi_i(\boldsymbol{r})$

in the effective potential,

$$\left(-\frac{1}{2}\nabla^2 + v_{\mathrm{s}}(\boldsymbol{r})\right)\phi_i(\boldsymbol{r}) = \varepsilon_i\phi_i(\boldsymbol{r}), \tag{2.16}$$

where ε_i is the energy eigenvalue for electron i. The ground state of (2.15) is the Slater determinant composed of the N lowest eigenstates $\phi_i(\boldsymbol{r})$ computed from (2.16).

The KS effective potential consists of the Hartree potential, the external potential, and the XC potential $v_{\mathrm{xc}}(\boldsymbol{r})$,

$$v_{\mathrm{s}}(\boldsymbol{r}) = v_{\mathrm{H}}(\boldsymbol{r}) + v_{\mathrm{ext}}(\boldsymbol{r}) + v_{\mathrm{xc}}(\boldsymbol{r}). \tag{2.17}$$

This is obviously a mean-field approximation for the electrons, but a particularly clever one: While the majority of the KS Hamiltonian is exact and reasonably easy to treat computationally, simple but powerful approximations to the XC potential (and energy) enable electronic structure calculations for systems containing hundreds or thousands of electrons.[21]

2.2.2 Total energy and potentials

The KS total energy density functional is a reformulation of the HK one,

$$E_{\mathrm{KS}}[n] = T_{\mathrm{s}}[n] + U_{\mathrm{H}}[n] + \int v_{\mathrm{ext}}(\boldsymbol{r})n(\boldsymbol{r})d\boldsymbol{r} + E_{\mathrm{xc}}[n]. \tag{2.18}$$

The KS kinetic energy T_{s} is evaluated on the independent-particle KS orbitals,

$$T_{\mathrm{s}}[n] = T_{\mathrm{s}}[\{\phi_i\}] = \sum_{i=1}^{N}\langle\phi_i| - \frac{1}{2}\nabla^2|\phi_i\rangle, \tag{2.19}$$

and is therefore implicitly a functional of the electron density $n(\boldsymbol{r}) = \sum_{i=1}^{N}|\phi_i(\boldsymbol{r})|^2$. This is an important difference from the Hartree–Fock method, where the kinetic energy is evaluated on an approximation to the physical many-electron wavefunction.

Comparing the HK total energy expression (2.13) to the KS expression (2.18) we may write the exchange–correlation energy functional as

$$E_{\mathrm{xc}}[n] = F_{\mathrm{HK}}[n] - \left(T_{\mathrm{s}}[n] + U_{\mathrm{H}}[n]\right), \tag{2.20}$$

which can be rearranged to

$$E_{\mathrm{xc}}[n] = \left(\langle\widehat{T}\rangle - T_{\mathrm{s}}[n]\right) + \left(\langle\widehat{V}_{\mathrm{int}}\rangle - U_{\mathrm{H}}[n]\right), \tag{2.21}$$

where $\langle\widehat{T}\rangle$ and $\langle\widehat{V}_{\mathrm{int}}\rangle$ are evaluated on the true many-body wavefunction. Writing the XC energy like this explicitly shows how the important, but complicated, many-particle

electron–electron interactions in the KS formalism are grouped together in a single functional: $E_{\rm xc}$ is the difference between the kinetic energy plus internal interaction energies of the interacting many-electron system, and those of the fictitious KS system with electron–electron interactions replaced by the classical Hartree energy.

With the XC potential defined as the functional derivative of the XC energy with respect to the density,

$$v_{\rm xc}(\boldsymbol{r}) = \frac{\delta E_{\rm xc}[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})}, \tag{2.22}$$

the KS total energy for N electrons of density $n(\boldsymbol{r})$ may be written

$$\begin{aligned} E_{\rm KS}[n(\boldsymbol{r})] = &\sum_{i=1}^{N} \langle\phi_i| -\frac{1}{2}\nabla^2|\phi_i\rangle + \int v_{\rm ext}(\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r} \\ &+ \frac{1}{2}\int \frac{n(\boldsymbol{r})n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} d\boldsymbol{r} d\boldsymbol{r}' + E_{\rm xc}[n(\boldsymbol{r})]. \end{aligned} \tag{2.23}$$

It must be emphasized that this expression is exact if an exact exchange–correlation energy functional of the density is applied. Unfortunately, no exact form suitable for numerical representation and application exists, so in practice we are forced to resort to approximate representations of $E_{\rm xc}[n(\boldsymbol{r})]$. Thus, the XC energy and potential is formally the only approximation that must be made in actual KS-DFT simulations of electronic structure.

2.2.3 Force theorem

The total energy (2.23) is determined by the electron density and parametrically by the ion positions $\boldsymbol{R}_I$,

$$E[n(\boldsymbol{r})] = E(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_M), \tag{2.24}$$

If the ionic positions does not correspond to an energy minimum, there will be forces $\boldsymbol{F}_I$ acting on the ions,

$$\begin{aligned} \boldsymbol{F}_I = -\frac{\partial E[n]}{\partial \boldsymbol{R}_I} &= -\langle\Psi|\frac{\partial \widehat{H}}{\partial \boldsymbol{R}_I}|\Psi\rangle - \frac{\partial E_{II}}{\partial \boldsymbol{R}_I} \\ &= -\int n(\boldsymbol{r})\frac{\partial V_{\rm ext}(\boldsymbol{r})}{\partial \boldsymbol{R}_I} d\boldsymbol{r} - \frac{\partial E_{II}}{\partial \boldsymbol{R}_I}, \end{aligned} \tag{2.25}$$

where the Hellmann–Feynman force theorem is applied, $n(\boldsymbol{r})$ is the unperturbed electron density, and E_{II} is classical ion–ion interactions. [15] In practice, efficient algorithms are used for relaxing the real-space structure of a collection of atoms by minimizing the first principles forces (2.25) acting on them. This should lead to the global equilibrium struture $\boldsymbol{R}_I^0$. Moreover, zero-point energies may be obtained from the vibrational spectrum of all atoms around $\boldsymbol{R}_I^0$ in the harmonic approximation. This requires calculating (or approximating) second-order derivatives of the total energy, $\partial^2 E/\partial^2 \boldsymbol{R}_I|_{\boldsymbol{R}_I=\boldsymbol{R}_I^0}$. [22]

2.2.4 Adiabatic connection fluctuation-dissipation formalism

Though rarely used in its entire complexity for DFT calculations, the adiabatic connection fluctuation-dissipation (ACFD) formalism [23–25] provides a formally exact expression for the KS exchange–correlation energy. It does so by combining the principle of adiabatic connection, where the non-interacting Kohn–Sham system is adiabatically connected to the fully interacting many-electron one, and the fluctuation-dissipation theorem of DFT. [24,25] The latter relates the response of an electronic system after an external pertubation to fluctuation properties of the system in its ground state. The ACFD formalism is therefore closely connected to the field of time-dependent DFT, where electronic structure pertubations out of the ground state (excitations) are considered.

The starting point of ACFD theory is the Hamiltonian for N electrons interacting with each other via Coulomb interaction, while moving in a local external potential $v_{\text{ext}}(\boldsymbol{r})$,

$$\widehat{H} = \widehat{T} + v_{\text{eff}} = \widehat{T} + v_{\text{ext}} + v_{\text{int}}, \tag{2.26}$$

where $v_{\text{int}}(\boldsymbol{r}, \boldsymbol{r}') = 1/|\boldsymbol{r} - \boldsymbol{r}'|$ is the nonlocal Coulomb interaction between electrons. This fully interacting Hamiltonian has $n_0(\boldsymbol{r})$ as its ground state density. Now, the basic idea is to establish an adiabatic connection between this Hamiltonian and the corresponding Kohn–Sham one, $\widehat{H}_{\text{KS}}$. [26] To that end, we introduce a set of Hamiltonians $\widehat{H}^\lambda$ which describe electrons interacting through Coulomb interactions *scaled* with the scalar λ,

$$\widehat{H}^\lambda = \widehat{T} + v_{\text{eff}}^\lambda = \widehat{T} + \widehat{v}_{\text{ext}}^\lambda + \lambda v_{\text{int}}, \tag{2.27}$$

where $0 \leq \lambda \leq 1$, and the external potential v_{ext}^λ is such that $n_0(\boldsymbol{r})$ is unchanged for all λ. When $\lambda = 0$, these electrons are non-interacting and the fictitious system is nothing but the Kohn–Sham system, $v_{\text{eff}}^{\lambda=0} = v_{\text{ext}}^{\lambda=0} = v_{\text{s}} = v_{\text{H}} + v_{\text{ext}} + v_{\text{xc}}$. In the other limit, for $\lambda = 1$, the Hamiltonian is the physical one, $v_{\text{eff}}^{\lambda=1} = v_{\text{ext}} + v_{\text{int}}$.

If we denote by $|\Psi^\lambda\rangle$ the ground state of $\widehat{H}^\lambda$, the point is now that the total electron–electron interaction may be expressed as a *coupling-constant integral* over λ,

$$E_{\text{int}} = U_{\text{H}} + E_{\text{xc}} = \int_0^1 d\lambda \langle \Psi^\lambda | v_{\text{int}} | \Psi^\lambda \rangle. \tag{2.28}$$

Introducing the imaginary-frequency density–density response function $\chi_\lambda(\boldsymbol{r}, \boldsymbol{r}', iu)$, the ACFD expression for the XC energy may then be written [27]

$$E_{\text{xc}}^{\text{ACFD}} = -\frac{1}{2\pi} \int_0^1 d\lambda \int_0^\infty du \times \int v_{\text{int}}(\boldsymbol{r}, \boldsymbol{r}') \left[\chi_\lambda(\boldsymbol{r}, \boldsymbol{r}', iu) + n(\boldsymbol{r}) \delta(\boldsymbol{r} - \boldsymbol{r}') \right] d\boldsymbol{r} d\boldsymbol{r}'. \tag{2.29}$$

This may be separated into exact expressions for the exchange and correlation energies. With $\chi_0 = \chi_{\lambda=0}$ we find the exact exchange (EXX) energy,

$$E_x^{\text{ACFD}} = -\frac{1}{2\pi} \int_0^\infty du \int v_{\text{int}}(\boldsymbol{r}, \boldsymbol{r}') \left[\chi_0(\boldsymbol{r}, \boldsymbol{r}', iu) + n(\boldsymbol{r}) \delta(\boldsymbol{r} - \boldsymbol{r}') \right] d\boldsymbol{r} d\boldsymbol{r}', \tag{2.30}$$

the DFT analog of Hartree–Fock exchange, but evaluated on KS orbitals. The remaining part must by definition be a formally exact expression for the correlation energy,

$$E_c^{\mathrm{ACFD}} = -\frac{1}{2\pi}\int_0^1 d\lambda \int_0^\infty du \int v_{\mathrm{int}}(\boldsymbol{r},\boldsymbol{r}')\left[\chi_\lambda(\boldsymbol{r},\boldsymbol{r}',iu) - \chi_0(\boldsymbol{r},\boldsymbol{r}',iu)\right] d\boldsymbol{r}d\boldsymbol{r}'. \quad (2.31)$$

This is an incredibly complicated formulation of correlation effects. In particular, suitable forms of the response function χ_λ are in general unknown, and (2.31) is in any case extremely demanding from a computational point of view. However, a simple approximation for χ_λ yields the (direct) random phase approximation (RPA),[26,27] which has proven highly useful but is still computationally heavy. This EXX+RPA method is perhaps the most high-level density funtional method in practical use today. The total electronic energy is usually evaluated on KS orbitals obtained from calculations using simpler XC approximations, eliminating the need for constructing the XC potential for exact exhange and RPA correlation.[26,28–31]

2.3 Implementation of density functional theory

Assuming a good approximation to exchange–correlation effects is readily at hand, calculating the ground state density and total energy is done using the iterative self-consistent field procedure: Starting from some trial electron density $n(\boldsymbol{r})$ we compute the KS effective potential $v_{\mathrm{s}}(\boldsymbol{r})$ for all electrons and solve (2.16). But the new singe-particle solutions $\phi_i'(\boldsymbol{r})$ correspond to a different density $n'(\boldsymbol{r}) = \sum_{i=1}^N |\phi_i'(\boldsymbol{r})|^2$ than the input one. The output density is then used in a new and improved guess for an input density. The SCF loop continues until a self-consistent solution is reached, that is, until a well-converged ground state electronic structure is obtained.

Efficient computational *representations* of the physical quantities entering this algorithm (densities, potentials, wavefunctions, etc.) are in practice needed. Common representations include expansion in atom-centered orbitals or plane waves, and numerical representation on a real-space grid.[15] Moreover, physical and numerical approximations are introduced in order to limit the computational cost of the density functional calculations. This means that the KS equations are hardly ever represented in their full complexity. It is therefore fair to say that whereas the choice of XC approximation for a DFT calculation introduces a strichtly physical approximation, the computational method employed adds further approximations, physical as well as numerical.

Wavefunctions are the major problem. Since the atomic wavefunctions are eigenfunctions of the atomic Hamiltonian, they must be mutually orthogonal. The atomic core states are localized around the ion, while the valence states are more delocalized in space. However, in order to maintain their orthogonality to the core states, the valence states must oscillate rapidly in the core region. This makes them hard to represent computationally. The frozen-core approximation is often applied, meaning that core

orbitals are not allowed to relax in the SCF procedure. This approximation works well as long as the core states are not significantly affected by the chemical environment around the atom, which is the case for most purposes.

Using pseudopotentials is a common implementation of the frozen-core approximation. The pseudopotential is a smoothly varying potential constructed to mimic the effect of the ion and the core electrons on the valence electrons. Using pseudopotentials for all ions and core electrons, one solves the KS equations for the valence electrons only, decreasing the computational load significantly. On the other hand, the method also discards all information about the full KS single-particle states close to the ion.

All-electron DFT methods, such as the augmented plane wave (APW) method introduced by Slater,[32] do not suffer from this. The approach is to divide space up into two types of regions: atom-centered augmentation spheres in which wavefunctions are taken as atom-like partial waves in order to reproduce the afore mentioned rapid oscillations, and an interstitial bonding region between all atoms where wavefunctions are expanded in more smoothly varying envelope functions, e.g., plane waves. The partial waves and envelope functions are matched with value and derivative at the augmentation sphere boundaries.[15,33]

All of the DFT calculations presented here are performed using the GPAW software package.[34–36] It employs the projector-augmented wave (PAW) method of Blöchl,[33,37] primarily in a real-space grid implementation. Furthermore, the Atomic Simulation Environment[36,38] has provided a convenient interface to GPAW.

2.3.1 Grid-based projector-augmented wave method

The projector-augmented wave method is a very general approach to practical DFT, extending the augmented-wave methods as well as the pseudopotential approach.[39] The prime feature of PAW is a reformulation of the original all-electron Kohn–Sham problem with rapidly oscillating wavefunctions $|\psi_n\rangle$ near the atomic cores, into a problem posed in terms of smooth auxiliary wavefunctions $|\tilde{\psi}_n\rangle$ and cleverly chosen atomic corrections inside the augmentation spheres. This is illustrated in Fig. 2.1.

2.3.1.1 PAW transformation

The switch of representation is done by introducing a linear transformation $\mathcal{T} = \mathbf{1} + \sum_a \hat{\mathcal{T}}^a$ between the true KS single-particle wavefunction and the auxiliary wavefunction. It is defined from the requirement $|\psi_n\rangle = \mathcal{T}|\tilde{\psi}_n\rangle$, and accounts for the inadequate representation of $|\psi_n\rangle$ by $|\tilde{\psi}_n\rangle$ inside the atomic core regions a. This leads to the following expansion of the true wavefunction,

$$|\psi_n\rangle = |\tilde{\psi}_n\rangle + \sum_a (|\psi_n^a\rangle - |\tilde{\psi}_n^a\rangle) = |\tilde{\psi}_n\rangle + \sum_{i,a} (|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle)\langle \tilde{p}_i^a | \tilde{\psi}_n\rangle, \tag{2.32}$$

where the sums are over atomic corrections inside the augmentation sphere for each atom. These are formulated in terms of partial waves that exhibit the proper oscillatory behavior, denoted $|\phi_i^a\rangle$ and $|\tilde{\phi}_i^a\rangle$, and local projectors $|\tilde{p}_i^a\rangle$. The true KS wavefunction is therefore divided into an interstitial part for valence electrons outside the atomic cores and atom-centered contributions. This is computationally convenient because the two may then be treated in different numerical schemes. The electron density is similarly split up,

$$n(\boldsymbol{r}) = \tilde{n}(\boldsymbol{r}) + \sum_a \left[n^a(\boldsymbol{r}) - \tilde{n}^a(\boldsymbol{r})\right]. \tag{2.33}$$

Obviously, the full Kohn–Sham problem is affected by the transformation. The transformed KS Hamiltonian $\widehat{\tilde{H}} = \mathcal{T}^\dagger \widehat{H}_{\mathrm{KS}} \mathcal{T}$ contains an auxiliary effective potential $\tilde{v}_{\mathrm{eff}}(\boldsymbol{r}) = \tilde{v}_{\mathrm{H}}(\boldsymbol{r}) + \tilde{v}_{\mathrm{xc}}(\boldsymbol{r})$ evaluated on "smooth" auxiliary quantities such as $\tilde{n}(\boldsymbol{r})$, and atomic corrections. The PAW total energy functional is therefore also separated into a smooth part $\tilde{E}$ and atomic corrections ΔE^a,

$$E = \tilde{E} + \sum_a \Delta E^a = \tilde{E} + \sum_a \left(E^a - \tilde{E}^a\right), \tag{2.34}$$

where the atomic corrections can be pre-calculated and stored for each chemical element such that they are ready for use in the SCF cycle. The main computational effort can then be focused on the valence electrons.

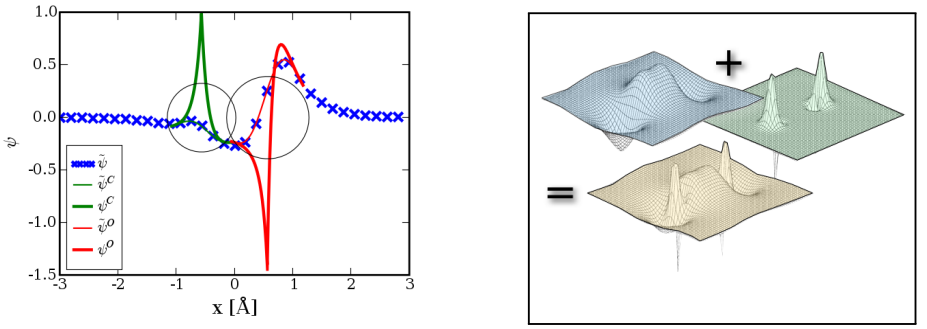

Figure 2.1: The PAW method. Left: Decomposition of the physical wavefunction $|\psi\rangle$ for the $2\sigma^*$ orbital of the CO molecule in terms of the auxiliary wavefunction $|\tilde{\psi}\rangle$ in all of space (blue crosses) and expansions of $|\psi\rangle$ (thick lines) and $|\tilde{\psi}\rangle$ (thin lines) in partial waves inside the augmentation spheres (black circles). Right: 3D schematic of the decomposition illustrating contributions from valence and core electrons separately.

2.3.1.2 The GPAW code

The PAW method is implemented in GPAW using a real-space multigrid discretization of the PAW quantities and transformed KS equations. Physical quantities living in 3D real space (wavefunctions, potentials, densities, etc.) are therefore represented numerically by their value in points on a grid. Integration of a quantity over space is approximated by a summation over grid-point values, while differentation with respect to spatial coordinates are performed within the finite-difference (FD) approximation. The Laplacian of a quantity in some grid point is therefore approximated from values of the quantity in the neighboring grid points. The accuracy of the FD approximation to local derivatives is in principle only determined by the grid-point spacing h and the number of neighboring grid-point values taken into account by the FD stencil; the smaller the h and the less "near-sighted" stencil, the better the FD approximation. A 3-point central stencil is default.

For isolated atomistic systems the domain in which the KS equations should be solved can be restricted to a finite region determined by the decay of the effective potential, and the appropriate supercell boundary conditions are simple closed (Dirichlet) boundary conditions. For infinite systems the domain is in principle unrestricted. However, if the system has translational symmetry, the Bloch theorem

$$\psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n,\boldsymbol{k}}(\boldsymbol{r}), \tag{2.35}$$

is applied to map all wavefunctions of the infinite lattice into the first Brillouin zone (BZ) described by the Bloch wave vector $\boldsymbol{k}$. Exploiting symmetry properties of the supercell we may even restrict calculations to the irreducible part of the BZ. Clever methods for sampling only a representative set of $\boldsymbol{k}$-points in reciprocal space have been devised. Monkhorst–Pack sampling[40] is the GPAW default.

2.4 Summary

The Kohn–Sham density functional approach to the many-body problem for mutually interacting electrons in the potential of static ions has been outlined. This would be an exact theory for the correlated electrons if the exact exchange–correlation density functional was known. However, it is not, and even if we did know it, this "divine" functional would probably be far too demanding for most applications in computational materials science. We must therefore resort to approximate XC functionals. Meanwhile, actual implementation of DFT in working computer codes that allow efficient and reliable solution of the KS equations introduce even more approximations. The projector-augmented wave method implemented in GPAW is a particularly clever approach to minimizing the casualties when electronic structure quantities are represented numerically.

Chapter 3

Exchange–Correlation Approximations in DFT

The remainder of this thesis is concerned with development and application of XC density functional approximations (DFAs). A veritable zoo of DFAs are currently available to the DFT user, many of which differ only little, but spanning a wide range of target applicabilities and computational costs. An overview of the DFA landscape is here given in terms of Perdew's metaphorical ladder leading to the exact density functional, with emphais on the types of condensed matter interactions and classes of matter that are well described by these approximations.

3.1 Strong and weak bonding

The main mechanisms behind inter-atomic bonding in matter are often divided into five groups;[15,41] Ionic, covalent, metallic, van der Waals (vdW) like, and hydrogen bonds. The first three result from strong interactions between nuclei and their electron densities, and are characterized by relatively small bond lengths and significant density overlaps.[42] They are responsible for the stability of most of the matter surrounding us, counteracting repulsion due to ion–ion electrostatics (Coulomb) and Pauli exclusion for short ion–ion distances.

Bonds in ionic crystals (e.g., solid NaCl) are formed from elements with a large differ-

ence in electronegativity. Charge transfer leads to approximately closed-shell ions and significant electrostatic attractions. Covalent bonds in molecules and solid crystals arise from a complete redistribution of the electronic structure due to hybridization between pairs of overlapping electron valence states. Completely new states (bonding and antibonding) result from this. The covalent bond is therefore often described as "charge sharing" between atoms. Lastly, metallic bonds are formed in the solid state when valence electrons are delocalized over the entire crystal. The positive ions are therefore "immersed" in a communal sea of electrons with which they interact, the Fermi gas of (nearly) free conduction electrons. In DFT, the homogeneous electron gas[41] is therefore a natural starting point for DFAs suitable for describing the electronic structure of many solids. However, while the chemical bond is predominantly covalent, strong bonding in the solid state is often mixed, and may simultaneously have metallic, covalent, and ionic character.

Hydrogen bonding and vdW interactions are most often considerably weaker than the interactions described above, but have important contributions to the stability of biological matter (e.g., proteins and DNA), as well as rare-gas chemistry and soft condensed matter in general. The hydrogen bond is an attractive interaction between two species that arises from a link of the form A$-$H$\cdots$B, where A and B are strongly electronegative atoms such as F, O, and N. It is thus a longer-ranged and directional interaction caused by electrostatic interactions between the polarized hydrogen atom (the electronegative host A has largely snatched the single electron of H) and the electronegative species B. Interaction energies are usually of the order 0.1 eV,[41] which may be compared to the dihydrogen covalent bond strength of 4.8 eV.

Finally, van der Waals forces are the weakest interactions considered here, but also ones that are virtually always present, even in the limit of large separation between the interacting fragments with no density overlap. These arise primarily from spontaneous charge density fluctuations, resulting in transient electromagnetic fields that induce dipole and multipole moments in distant charge densities. In quantum mechanical terms the fluctuations are related to zero-point motion of the electrons. The correlation of the temporary fields leads to net forces of attraction, which are long-ranged and decay algebraically with separation. We consider only the non-retarded regime where the finite speed c of light does not retard the correlation of separated fields, that is, the distance R between interacting fragments is small enough that the light transit time $\gamma = R/c \ll \tau$, where τ is the response time of charges.[30,42]

There are several ways to derive the physical contributions to vdW interactions. A particularly illustrative decomposition is based on a partubation theory (PT) treatment of the electrostatic potential between two fragments A and B with non-overlapping charge densities. To second order it yields the Casimir–Polder formula for the interaction between A and B,

$$E_{AB} = -\frac{1}{2\pi}\int_0^\infty du \int d\boldsymbol{r}_1 d\boldsymbol{r}_1' d\boldsymbol{r}_2 d\boldsymbol{r}_2' \times V(\boldsymbol{r}_2' - \boldsymbol{r}_1)\chi_A(\boldsymbol{r}_1, \boldsymbol{r}_1', iu)V(\boldsymbol{r}_1' - \boldsymbol{r}_2)\chi_B(\boldsymbol{r}_2, \boldsymbol{r}_2', iu), \quad (3.1)$$

where the first integral is over all imaginary frequencies, and χ_A is the density response

of A. In the particularly simple case of isotropic systems, a multipole expansion of the Coulomb potential V leads to lowest order to the well-known vdW dispersion energy,

$$E_{\text{disp}} = C_6 R^{-6}, \tag{3.2}$$

where $R = |\boldsymbol{r}_A - \boldsymbol{r}_B|$ and C_6 depends on the polarizabilities of the fragment densities. This is also known as the London dispersion energy and is the purely quantum mechanical effect of induced multipole–induced multipole interactions (a long-range correlation effect). Lower-order interaction energies ($C_n R^{-n}$ with $n < 6$) arising from electrostatatics between permanent poles and induction between permanent and induced poles may also be derived, and higher-order PT leads to even more contributions to the vdW energy ($C_n R^{-n}$ with $n > 6$).[27,42,43] However, dispersion is the first term that never completely vanishes, not even for strongly overlapped electron densities, and we identify the van der Waals energy with London dispersion throughout.

3.2 A density functional hierarchy

The five-rung "Jacob's ladder" of Perdew[44] represents a systematic classification of the exchange–correlation approximations available for DFT. Each rung adds new ingredients to the approximation, and so should enable better density functionals, but also adds to the computational cost. In order of increasing complexity, the ladder consists of the local (spin-)density approximation[20] (LDA), the generalized-gradient approximation (GGA), meta-GGA (MGGA), hyper-GGA, and finally the generalized random phase approximation. The LDA uses only the local density as input, while rungs 2 and 3 introduce semilocal dependence of the density (GGA) and the KS orbitals (MGGA). Hyper-GGAs introduce nonlocal dependence of the occupied KS orbitals in the exact exchange energy density, and fifth-rung approximations calculate correlation energies from the unoccupied KS orbitals. The latter is computationally heavy, but RPA-type methods are the only DFAs in this five-rung hierarchy that can possibly account for vdW dispersion between non-overlapped electron densities.[17,45] It is useful to consider the ladder a progression of increasingly more complex *models* of the exact density functional.

LDA The XC energy is in general written as a spatial integral over the XC energy density per particle $\varepsilon_{\text{xc}} = \varepsilon_x + \varepsilon_c$,

$$E_{\text{xc}} = \sum_{\sigma=\uparrow,\downarrow} \int \varepsilon_{\text{xc}} n(\boldsymbol{r}) d\boldsymbol{r}. \tag{3.3}$$

where ε_{xc} usually depends on the electron density $n(\boldsymbol{r}) = n_\uparrow(\boldsymbol{r}) + n_\downarrow(\boldsymbol{r})$ and other quantities. However, the local density approximation takes the XC energy density in the point $\boldsymbol{r}$ as that af the homogeneous electron gas (HEG) with that density,

$$E_{\text{xc}}^{\text{LDA}} = \int \varepsilon_{\text{xc}}^{\text{HEG}}(n(\boldsymbol{r})) n(\boldsymbol{r}) d\boldsymbol{r}. \tag{3.4}$$

A simple but exact analytic form for LDA exchange energy is known, and accurate parametrizations of LDA correlation exist.[15]

The LDA has had great success in the solid state physics community, where it has proven surprisingly reliable. One reason for this is that the valence electrons in many solid crystals are nearly a HEG, but favorable cancellation of erros between exchange and correlation is also part of the explanation. However, several atomistic systems of common interest, e.g., molecules, are far from homogeneous, and LDA is a poor model of the exact XC functional for such systems. The predicted thermochemistry is simply too incorrect; intra-molecular covalent bond energies are vastly overestimated.[46] No surprise, the LDA was never really adopted in the quantum chemistry community.

GGA The first obvious improvement on LDA is to include the gradient of the electron density as an ingredient in the DFA. This is most often done within the GGA method, where the magnitude of the local density gradient $|\nabla n(\boldsymbol{r})|$ is used. This renders GGAs semilocal approximations.

The GGA exchange energy is then formulated in terms of an exchange enhancement factor $F_x(s)$, depending on the reduced density gradient $s = |\nabla n|/2k_F n \in [0, \infty]$, where $k_F = (3\pi^2 n)^{1/3}$ is the Fermi wave vector of the HEG. The function $F_x(s)$ locally scales $\varepsilon_x^{\mathrm{HEG}}(n)$ depending on the density gradient,

$$\begin{aligned} \varepsilon_x^{\mathrm{GGA}}(n, \nabla n) &= \varepsilon_x^{\mathrm{HEG}}(n) F_x(s(n, \nabla n)), \\ E_x^{\mathrm{GGA}}[n, \nabla n] &= \int \varepsilon_x^{\mathrm{HEG}}(n) F_x(s) n(\boldsymbol{r}) d\boldsymbol{r}. \end{aligned} \tag{3.5}$$

The lowest-order terms in the expansion of $F_x(s)$ for slowly varying densities are[15]

$$F_x(s) = 1 + \frac{10}{81} s^2 + \frac{146}{2025} \left(\frac{|\nabla^2 n|}{(2k_F)^2 n} \right)^2 + \cdots . \tag{3.6}$$

The Laplacian of the density and higher-order derivatives are not included in GGA DFAs. This form obeys the LDA limit $F_x(s = 0) = 1$ for vanishing density gradients. Another known constraint is the Lieb–Oxford lower bound,[47] which implies $F_x(s) \leq 1.804$. A wealth of different GGA exchange approximations have been devised, that is, different functional forms of $F_x(s)$. The Perdew–Burke–Ernzerhof[48] (PBE) exchange functional is a standard example of semilocal exchange.

GGA correlation approximations are usually more complicated, but also relies on correction the LDA correlation in a density gradient dependent fashion. The PBE and the Lee–Yang–Parr[49] (LYP) correlation functionals are particularly popular.

In total, the GGA may be considered an extrapolation of LDA to the slowly-varying high-density (small-s) regime, and are often capable of improving over LDA on many accounts with little extra computational overhead. GGAs are now a standard tool in computational chemistry and materials science. It is, however, important to note the

limitations of such a model. The s-parameter is simply a measure of inhomogeneity, and the GGA attempts to derive information about the system from this parameter (does the density belong to a system which is nearly a HEG, or a system which is quite far from the HEG?). Though the electron density distribution, and thereby the distribution of s-values, in ionic and metallic solid crystals may often be different,[50,51] this is merely an indirect indication of the interactions at play. Given a density with a local reduced gradient $s = 2$, the GGA can not tell whether the atomistic system is a solid or a molecule.

The DFA zoo is particularly populated with GGAs. Some functional forms for $F_x(s)$ are found well suited for studies in solid state physics because they do not deviate too strongly from LDA in the small-s regime, while GGAs tailored for studies in theoretical chemistry may be substantially different. Popular variants include PBE and its PW91 predecessor,[52,53] as well as the PBEsol[54] and RPBE[55] revisions of PBE. While PBEsol was designed as a "GGA for solids" with the aim of accurately capturing bonding in the solid state, the RPBE exchange model was targeted at studies in chemistry.[55,56] Lastly, also the combination of the B88 exchange of Becke[57] and LYP correlation, commonly denoted BLYP, has had a great impact on bringing DFT to the quantum chemistry community.

MGGA The third rung on Jacob's ladder expands on GGA by adding to the list of ingredients the Laplacian of the electron density and/or the kinetic energy density (KED) $\tau(\boldsymbol{r})$ of the occupied KS orbitals,

$$\tau_\sigma(\boldsymbol{r}) = \frac{1}{2}\sum_i^{\text{occ.}} |\nabla\phi_{i\sigma}(\boldsymbol{r})|^2. \tag{3.7}$$

The MGGA exchange enhancement factor is therefore of the form $F_x(n, \nabla n, \nabla^2 n, \tau)$, and has the possibility of more variational freedom in the space of its arguments than the more limited GGA exchange. The KED may be expressed in a dimensionless form as $\alpha = \left(\tau - \tau^{\text{W}}\right)/\tau^{\text{HEG}} \in [0, \infty]$, where $\tau^{\text{W}} = |\nabla n|^2/8n$ is the von Weizsäcker KED, and $\tau^{\text{HEG}} = \frac{3}{10}(3\pi^2)^{2/3}n^{5/3}$ is the KED of the homogeneous electron gas. The limit $\alpha = 0$ corresponds to $\tau = \tau^{\text{W}}$, which is characteristic of electron densities with single-electron (iso-orbital) character,[58] that is, densities not belonging to a HEG, whereas $\alpha = 1$ means $\tau = \tau^{\text{HEG}}$. The local KED is thus able to discriminate between the "environment" of different densities even though they are characterized by the same s. Much unlike GGAs, the MGGA exchange model is therefore able present densities of identical s but different τ with different enhancement factors. Thus, the model space of MGGA exchange is considerably expanded as compared to that of GGA.

Since the KED is a semilocal functional of the occupied KS orbitals, which are readily available in most DFT calculations, MGGA approximations are usually cossidered semilocal DFAs[17] (as opposed to fully nonlocal ones) and the extra computational overhead of evaluating the MGGA total energy and orbital-dependent XC potential is modest if the method of Neumann *et al.* is used (Ref. 59). Significant MGGA functionals include the VSXC of Voorhis and Scuseria,[60] the TPSS[58] and its revTPSS[61] revision, and the M06-L of Zhao and Truhlar.[62]

Hyper-GGA and hybrids Fourth-rung density functionals introduce the exact exchange energy density, a fully nonlocal functional of the occupied KS orbitals. Hybrid functionals, which mix a fraction of EXX into otherwise lower-rung DFAs, are very popular in quantum chemistry because of their good description of molecular thermochemistry.[63–65] However, the long-ranged nature of the Coulomb potential renders hybrid DFAs computationally demanding for calculations on periodic crystals, especially metallic systems. Screening out the long-range part of the Coulomb potential for EXX is one possible solution (range separation).[66] The most well-known hybrid DFA is probably B3LYP,[67] but there are many others, including PBE0[68] and the HSE family of range-separated hybrids.[66,69]

3.3 Strategies for DFA development

Put in simple terms, two paradigms for developing new and improved density functionals are dominant: that of constraint satisfaction by reduction and that of fitting to empirical data.[45,70] Both have contributed greatly to the successful applications of DFT in many different fields. Reductionists impose constraints based on analytic properties of the exact density functional, and strive for *nonempirical* functionals that fulfill as many constraints as possible on each rung of Jacob's ladder. The LDA limit and LO bound are examples of such constraints, but other exact constraints are known.[17] Empirically oriented DFA developers use experimental or high-level theoretical training data to *optimize* the XC model description of one or more materials properties. Note that the term "materials properties" is here used in the widest possible meaning, covering all from molecular bond energetics to, e.g., band gaps of solid crystals and stabilities of surface facets against reconstruction.

Reduction is arguably the most systematic approach to density functional development, and has had a significant impact on KS-DFT. However, choices are often made as to what types of physics and chemistry the DFA should describe well. The several revisions of PBE are good examples of this: The PBEsol revision sacrificed accuracy for chemical bonds but improved the desciption of equilibrium crystal volumes by imposing the second-order gradient expansion for exchange ($F_x(s) \to 1 + 10/81s^2$ for $s \to 0$), while the revPBE modification of PBE exchange biased the XC model towards describing total energies of free atoms and chemical bonding in small molecules. The consequence was significantly worse predictions of equilibrium crystal volumes.

The empirical approach is fundamentally a matter of explicitly making these choices, and parametrize and train an XC model to suit personal preferences for performance in practice. This makes overfitting the training data and transferability of the optimized DFA to systems and materials properties not contained in the training data a central issue,[45] which we shall return to in Chapter 7.

3.4 Challenges to current density functionals

Two major issues with standard lower-rung density functionals are briefly discussed: The complete lack of the physics of vdW dispersion interactions in rung 1–3 DFAs, and the ever haunting issue of self-interaction. These are largely absent in RPA-like approaches (and in the exact density functional, obviously) but for large-scale materials design studies from first principles we are currently forced to resort to the lower rungs of Jacob's ladder for workhorse DFAs.[71]

Dispersion It was realized early on that the local LDA and semilocal GGAs are not able to successfully account for dispersion interactions.[72] This should be no surprise; the physics of dispersion is nowhere to be found in such XC models. Particularly GGAs, where the enhancement of exchange interactions over the LDA ones usually acts repulsively, often predict no attraction between fragments which are known to be dispersion bonded. However, since hydrogen bonds are fundamentally electrostatic effects over not-too-long distances, describing these are not completely outside the realm of GGA-DFT, and several semilocal DFAs are known to do so reasonably well. On the other hand, LDA exchange alone can actually produce significant (and often vastly exaggerated) dispersion-like attractions from density–density overlap effects. This is spurious and outright wrong since pure dispersion is solely a correlation effect.[72–74] Lastly, since MGGA approximations are essentially semilocal, there is not much hope for them to reliably produce dispersion forces either. It should be noted that a reasonable emulation of short-range dispersion has been reported for the M06-L MGGA,[62,75] but reliable description of medium- to long-range dispersion should not be expected.

Self-interaction Most current DFAs are not self-interaction free, that is, the spurious interaction of an electron with itself in the Hartree energy is not cancelled out by exchange functionals other than EXX. This can have significant implications, and is largely responsible for underestimated molecular reaction energy barriers, erroneous single-particle properties like ionization potentials and electron affinities, and underestimated band gaps of solid crystals. Self-interaction corrections attempting to remedy this issue is a highly active field of research, efficient application of the Perdew–Zunger approach dating back to 1981 being particularly popular.[15,76,77]

3.5 Summary

A truly general-purpose density functional for the energetics of condensed matter interactions faces a tough challenge: The important inter-atomic and inter-fragment forces range from strong interactions for large density–density overlaps between atoms in close vicinity, leading to ionic, covalent, and metallic types of bonding in molecules and the solid state, to weaker but long-ranged van der Waals type interactions in more

sparse systems. The London dispersion force is special in that it is the lowest-order vdW-type interaction that is ubiquitous. The XC density functional levels of theory have been introduced in the context of Jacob's ladder of Perdew, which forms the basis for the reductionist's approach to DFA development, where satisfaction of constraints on approximate XC functionals imposed by the exact functional is *modus operandi*. On the other hand, empirically oriented DFA developers seek to optimize DFAs by fitting to benchmark data. Two current challenges to DFT were furthermore highlighted; the complete lack of the physics of vdW dispersion interactions in lower-level DFAs, and the ever haunting issue of self-interaction. While the former is a central topic of this thesis, the latter is not considered in further detail.

Chapter 4

Representations of Condensed Matter Interactions

Reliable compilations of materials properties and chemical observables are very important for validation and development of density functionals. If accurate enough, such datasets are representations of the condensed matter interactions we meet in every-day applications of DFT. The exact density functional obviously captures these correctly [exactly if the data is exact and a strictly exact all-electron DFT code is used (such one probably does not exist)], and the question is often how well our XC approximations work in practice and how these may be improved. This chapter presents datasets to be used throughout the rest of this thesis, as well as the computational procedures employed to calculate the different quantities.

4.1 Datasets of materials properties

Benchmark data may be experimentally determined or calculated from high-level theory such as CCSD(T). The latter may not yield exact benchmarks in the true meaning of the word, but the most elaborate wavefunction methods are accurate enough to be considered essentially exact for molecular properties. The CCSD(T) "model chemistry" is a standard for benchmark data that may be directly compared to DFT results.

Experimental data are often not as accurate as theoretical benchmarks, but many atomistic systems are simply impossible to treat with expensive wavefunction theory, e.g., the extended lattices of solid crystals. Careful experiments, preferably at low temperature and possibly with extrapolation to the zero Kelvin limit and corrections for zero-point motion of atoms, must then be resorted to.

The benchmark datasets used in this work are either adapted from literature or compiled here from published works. They are presented in the following. Certain additional information, particularly extensive tables, are placed in Appendix A.1.

4.1.0.3 G3/99 and G2/97: Molecular formation energies

The molecular formation enthalpies of the G3/99 thermochemical test set of Curtiss and co-workers[78] represent intramolecular bond energetics. It has become very popular for benchmarking and calibrating electronic structure methods. The 223 molecules may be divided into three subsets denoted G3-1, G3-2, and G3-3 comprising 55, 93, and 75 molecules, respectively. The G3-1 and G3-2 subsets constitute G2/97, in which case the two subsets may be denoted G2-1 and G2-2, respectively.

The formation enthalpies are experimentally determined. In accordance with the procedure of Ref. 79 they are extrapolated to zero Kelvin by correcting for thermal and vibrational contributions. Thermal corrections and zero-point energies from Refs. 79 and 46 are used. The result is 233 electronic-only static-nuclei formation energies $\Delta_f E$, i.e., negatively signed atomization energies, which are directly comparable to predictions from ground state DFT. Contributions to $\Delta_f E$ from spin-orbit effects are not corrected for. This is expected to be of little overall consequence.[79]

Theoretical G3/99 formation energies are calculated from the difference between molecular and atomic total energies as

$$\Delta_f E = E_M - \sum_A E_A, \tag{4.1}$$

where A runs over all atoms in the molecule M, while E_M and E_A are ground state molecular and atomic total energies, respectively.

4.1.0.4 RE42: Molecular reaction energies

Molecular formation energies lend themselves well to compilation of gas phase reaction energies. The RE42 compilation contains 42 zero-Kelvin reaction energies involving 45 different molecules from G3/99. A complete list of the reactions and reaction energies $\Delta_r E$ is given in Table A.1 in the Appendix. Theoretical reaction energies are calculated from total electronic energies as

$$\Delta_r E = \sum_P E_P - \sum_R E_R, \tag{4.2}$$

where the sums run over reactant (R) and product (P) molecules.

4.1.0.5 DBH24/08: Molecular reaction barriers

As discussed in the Introduction to this thesis, chemical reactant and product states are often separated by an energy barrier, which must be surmounted if the reaction is to proceed. The DBH24/08 set of Zheng *et al.*[80] comprises 12 forward (V_f) and 12 backward (V_b) benchmark barriers for simple gas phase reactions. Ground- and transition-state molecular geometries, calculated using the quadratic configuration interaction with single and double excitations (QCISD) wavefunction method, are from Ref. 81. Density functional barrier heights are computed from the transition state total energy ($E_\ddagger$) and the initial (E_i) and final (E_f) state total energies as

$$\begin{aligned} V_f &= E_\ddagger - E_i, \\ V_b &= E_\ddagger - E_f. \end{aligned} \tag{4.3}$$

4.1.0.6 S22 and S22x5: Noncovalent bonding

The S22 dataset of Hobza and co-workers[82] represents van der Waals interactions and hydrogen bonding by considering noncovalent bonding between molecular dimers and complexes. It has been widely used for assessment[65,83–88] and parametrization[62,85,89–92] of density functional methods for vdW type interactions, though it may be somewhat superseded by the newer and larger S66 set.[93,94] The S22 set consists of CCSD(T) interaction energies between relatively small molecular complexes, but includes also noncovalent bonding between the somewhat larger DNA and RNA bases adenine, thymine, and uracil, as well as 2-pyridoxine and 2-aminopyridine. It is therefore biased towards biomolecular structures, but is very useful for benchmarking electronic structure methods in general. The 22 complexes are divided into three groups according to the type of interaction predominantly responsible for stabilizing the complex; hydrogen bonding, dispersion interactions, and a mixture of dispersion and electrostatic interactions. This categoratization was made on the basis of interaction energy decompositions using the symmetry-adapted pertubation theory method.

MP2 or CCSD(T) geometries at equilibrium intermolecular separations from the original work in Ref. 82 are used. Benchmark CCSD(T) interaction energies with extrapolation to the complete basis set (CBS) limit were reported in that same publication. However, most likely due to the computing resources available at the time, different basis sets were used for small and large complexes. Later works[95,96] have therefore revised the S22 interaction energies, employing larger and identical basis sets for all complexes at the original geometries. For the larger complexes the reported basis set effects are significant, so the CCSD(T)/CBS energies of Takatani *et al.*[95] are adopted here as the current best-estimate of the true S22 interaction energies.

Recently the S22x5[97] extension was proposed. In addition to the near-equilibrium

intermolecular distances, S22x5 contains for each complex four non-equilibrium binding energies. Thus, CCSD(T) potential-energy curves (PECs) for each complex are mapped out at relative interaction distances d of 0.9, 1.0, 1.2, 1.5, and 2.0 as compared to S22, totalling 110 interaction energies on 22 PECs. For convenience we here divide S22x5 into five subsets according to interaction distance, e.g., "S22x5-0.9".

The computational procedure used for S22x5 was identical to the S22 one,[97] so we expect the basis set deficiencies to persist in S22x5. The non-equilibrium data points on each PEC are therefore corrected according to the difference between original and revised S22x5-1.0 CCSD(T) energies. This correction scheme is described in Appendix A.1, where Table A.4 also lists the resulting energy corrections. These are very small on average but significant for certain larger complexes.

Each S22x5 density functional interaction energy E^d_{int} is computed as the difference between the total electronic energy of the interacting complex E^d_0 and those of its two isolated molecular constituents, E^d_1 and E^d_2,

$$E^d_{\text{int}} = E^d_0 - E^d_1 - E^d_2, \tag{4.4}$$

which obviously applies to S22 also ($d = 1$). Computational accuracy is enhanced by keeping all atoms in the monomers in the same positions in the supercell as those atoms have when evaluating the total energy of the complex. With the sign convention in (4.4) stable intermolecular bonding is here taken to mean negative interaction energy.

4.1.0.7 Crystalline solids

Energetic and structural properties of crystalline solids are represented by four different sets of experimental data. Three of these comprise lattice constants as well as cohesive energies, and one contains only cohesive energies. The four sets reflect a gradual progress from small experimental datasets without corrections for zero-point effects towards larger sets of low-temperature data with appropriate zero-point corrections.

Sol20 crystals Contains lattice constants and cohesive energies of 20 Period 3–6 pure crystals in the fcc, bcc, hcp, and diamond structures. The included elements are found in Fig. 5.4 (p. 45). Zero-point effects are not considered.

Sol34Ec Cohesive energies of 34 Period 2–6 pure crystals in fcc, bcc, diamond, and hcp lattices. Zero-point effects are not considered.

Sol27 It was recently shown[50] that removal of thermal and zero-point contributions to experimentally determined lattice constants and bulk moduli may be important

when benchmarking density functional methods. Experimental zero-Kelvin lattice constants and cohesive energies (E_c) contain zero-point vibrational contributions, leading to zero-point anharmonic expansion (ZPAE) of the lattice and zero-point vibrational energy (ZPVE) contributions to E_c. As discussed in Ref. 98, an estimate of the ZPVE may be obtained from the Debye temperature Θ_D of the solid according to

$$\mathrm{ZPVE} = -\frac{9}{8} k_\mathrm{B} \Theta_\mathrm{D}. \tag{4.5}$$

The vibrational contribution is subtracted from the cohesive energy, leading to increased stability of the crystal towards atomization. The same reference derived a semi-empirical estimate of the ZPAE contribution to the volume of cubic crystals. A recent study[69] calculating the ZPAE from first principles largely validates this approach.

The Sol27LC and Sol27Ec sets of zero Kelvin lattice constants and cohesive energies of 27 fcc, bcc, and diamond structured bulk solids are appropriately corrected for zero-point phonon effects. Details are given in Table A.3.

Sol53 Our use of solid state datasets peaks with the Sol53 lattice constants (Sol53LC) and Sol53 cohesive energies (Sol3Ec), which are essentially extensions of the Sol27 sets to include also mixed-element compounds in the rock-salt, cesium chloride, and zincblende cubic crystal structures. The low-temperature zero-point exclusive data are from Refs. 69 and 99 (Sol53LC) and Ref. 69 (Sol53Ec), respectively.

Equilibrium quantities The crystal cohesive energy for a given lattice constant a is calculated from

$$E_c = E_A - E_B, \tag{4.6}$$

where E_A is the total energy of the free atom and E_B the bulk total energy per atom. The equilibrium (maximum) cohesive energy of a stable solid is thus a positive quantity. Equilibrium lattice constants a_0 are determined from fitting the SJEOS[98] equation of state to cohesive energies sampled in five points in a small interval around the maximum of the $E_c(a)$ curve. For hcp-structured crystals the c/a lattice constant ratio is fixed at the experimental one.

4.1.0.8 CE17 and CE27: Chemisorption on solid surfaces

The CE17 and CE27 datasets contains chemisorption energies of simple molecules on late transition-metal surfaces. They are derived from temperature programmed desorption experiments or from microcalorimetry, most often at low coverage. The 27 chemisorption energies in CE27 have been critically chosen from literature with emphasis on reliability as well as covering a reasonably wide range of substrates and adsorbates. CE17 is a subset of CE27. All data is listed in Table A.2 along with details regarding adsorption mode, adsorption site, and references.

Most of the surface reactions are associative adsorption processes at 0.25 ML coverage. In that case the chemisorption energy ΔE is computed according to

$$\Delta E = E_{AM} - E_M - xE_A, \tag{4.7}$$

where E_{AM} is the total electronic energy of the adsorbate A on metal surface M, and E_A and E_M total energies of the isolated adsorbate and metal surface, respectively. The constant x equals 1 for associative adsorption and N_2 dissociation on Fe(100), while $x = \frac{1}{2}$ for dissociative H_2 chemisorption. In the case of NO dissociation on Ni(100) at 0.25 ML coverage the chemisorption energy is

$$\Delta E = E_{AM} + E_{BM} - 2E_M - E_{AB}, \tag{4.8}$$

where AB is the NO molecule.

Computational setup With these definitions of chemisorption energies we consider extended surface slab models with 2×2 atoms in each layer and 5 layers in total. The slabs are periodic in the surface plane and a vacuum width of 20 Å separates periodically repeated slabs perpendicularly to the surface planes. Well-converged chemisorption energies are obtained using a 10×10×1 **k**-point mesh and a real-space grid spacing around 0.16 Å. Self-consistently determined lattice constants are used for the slabs. During structure relaxations the bottom two layers of the 2×2×5 slab models are fixed in the bulk structure.

4.2 General computational procedures

Self-consistent density functional calculations using GPAW are in general performed using grid-point spacings of 0.16 Å for high-quality computations of simple properties such as molecular binding energies. Properties of solid crystals are calculated using somewhat denser grids with a spacing around 0.13 Å and dense Monkhorst–Pack $\boldsymbol{k}$-point sampling (at least 12 × 12 × 12 reducible points in the first BZ) and 0.1 eV Fermi smearing of electron occupations. Real-space structure relaxation is applied to the G3/99 molecules (including G2/97) and CE17/CE27 chemisorption systems with 0.05 eV/Å as the criterion of maximum Hellmann–Feynman force on each relaxing atom. Molecular and single-atomic systems are centered in supercells with at least 7 Å vacuum to the box boundaries. Vacuum widths of 10 Å are applied for the S22x5 complexes. All calculations are spin-polarized when appropriate.

Chapter 5

van der Waals Density Functionals

The semilocal nature of GGA XC models is unable to correctly capture long-ranged van der Waals dispersion. Including dispersion forces in DFT is therefore a vibrant field of research these days, and several approaches have been proposed. These range from computationally free-of-charge force-field methods to extremely heavy methods such as the random phase approximation with exact exchange and others.

Meanwhile, the so-called van der Waals density functionals (vdW-DFs) attempt to unite the best of both worlds; computational efficiency for relevant system sizes, and accurate description of dispersion forces. Interest is increasing in the materials science community to apply such computational methods, and we here review the characteristics of the vdW-DF approach and evaluate its general applicability in surface science studies.

5.1 Rutgers–Chalmers correlation approximation

The vdW-DF [100–102] family of density functionals aims at capturing short- to medium-range van der Waals interactions in an approximate but computationally feasible fashion using only the electron density and its gradients as input. At the heart of the original vdW-DF approach [100] is the Rutgers–Chalmers (RC) nonlocal correlation approximation. This has proven useful in several studies of condensed matter systems

with significant vdW-type interactions, and it is generally acknowledged that the RC approximation captures the leading term, $E \sim C_6 R^{-6}$, of the nonretarded dispersion energy. Moreover, the introduction of efficient numerical implementations [83,103–105] for evalutating the RC correlation energy has rendered vdW-DF calculations near-standard in the present day electronic structure toolbox. In fact, several vdW-DF variants using the RC approach now exist, and we shall consider most of them in later chapters.

A vdW-DF type XC approximation is in general written

$$E_{\rm xc}^{\rm vdW\text{-}DF}[n, \nabla n] = E_x^{\rm GGA} + E_c^0 + E_c^{\rm nl}, \tag{5.1}$$

where $E_x^{\rm GGA}$ is a suitably chosen GGA exchange functional, $E_c^0 := E_c^{\rm LDA}$ is usually chosen to be LDA correlation, and $E_c^{\rm nl}[n, \nabla n]$ is the RC nonlocal correlation functional, which allows a mutual interaction between all densities in space. The approximation $E_c = E_c^{\rm LDA} + E_c^{\rm nl}$ is obviously the crucial difference between vdW-DFs and GGAs.

Regardless of the exchange functional used in vdW-DF calculations, the method provides an approximate correction to the LDA correlation energy of a non-uniform electron density. Starting out from the ACFD expression for the exact ground state correlation energy, Eq. (2.31), which is repeated here,

$$E_c^{\rm ACFD} = -\frac{1}{2\pi}\int_0^1 d\lambda \int_0^\infty du \int v_{\rm int}(\boldsymbol{r}, \boldsymbol{r}') \left[\chi_\lambda(\boldsymbol{r}, \boldsymbol{r}', iu) - \chi_0(\boldsymbol{r}, \boldsymbol{r}', iu)\right] d\boldsymbol{r} d\boldsymbol{r}', \tag{5.2}$$

a few approximations and assumptions lead to a nonlocal but compact approximation to the non-LDA part of the ACFD correlation energy, [27,31,100]

$$E_c^{\rm nl} = \frac{1}{2}\iint n(\boldsymbol{r})\phi(\boldsymbol{r}, n, \nabla n; \boldsymbol{r}', n', \nabla n') n(\boldsymbol{r}') d\boldsymbol{r} d\boldsymbol{r}', \tag{5.3}$$

where $n' = n(\boldsymbol{r}')$. This is a 6-dimensional integral over the interaction kernel ϕ, which depends on the densities and density gradients in all pairs of spatial points $\boldsymbol{r}$ and $\boldsymbol{r}'$.

The main steps taken in going from the ACFD expression (5.2) to the approximation (5.3) are introduction of

1. an approximate subtraction of LDA correlation from the ACFD correlation expression,
2. an approximate coupling-constant integration over λ of what remains, resulting in
$$E_c^{\rm nl} = \frac{1}{2\pi}\int_0^\infty du\, {\rm Tr}\ln\left[\epsilon^{-1}\left(1 - v_{\rm int}\tilde{\chi}\right)\right], \tag{5.4}$$
where $v_{\rm int}$ is the electron–electron Coulomb interaction, $\tilde{\chi}$ the density response, ϵ a dielectric function relating the polarization due to a pertubation of the potential to an induced charge density, and u the imaginary frequency,
3. a second-order expansion of (5.4) in powers of $S = 1 - \epsilon^{-1}$,

4. a model dielectric function ϵ based on a plasmon-pole type approximation.

The resulting interaction kernel, ϕ, has several appealing features.[100] Since the asymptotic form is

$$\phi \sim |\boldsymbol{r} - \boldsymbol{r}'|^{-6} \text{ for } |\boldsymbol{r} - \boldsymbol{r}'| \to \infty, \tag{5.5}$$

the nonlocal correlation energy has the desired $-C_6 R^{-6}$ form for well-separated densities which is missing in local approximations. Symmetry properties of ϕ also leads to the nice feature that $E_c^{\text{nl}} = 0$ when $\nabla n(\boldsymbol{r}) = 0$, that is, the nonlocal correlation energy vanishes for HEG-like densities, such that the total XC energy in (5.1) reduces to the LDA one,

$$E_{\text{xc}}^{\text{vdW-DF}}[n, \nabla n = 0] = E_{\text{xc}}^{\text{LDA}}[n]. \tag{5.6}$$

The density gradient dependence of the interaction kernel enters through a local wave vector denoted $q_0(\boldsymbol{r})$. It is defined as the Fermi wave vector modulated by an energy ratio,

$$q_0(\boldsymbol{r}) = \frac{\varepsilon_{\text{xc}}^0}{\varepsilon_x^{\text{LDA}}} k_{\text{F}}(\boldsymbol{r}), \tag{5.7}$$

where $\varepsilon_x^{\text{LDA}} = \varepsilon_x^{\text{LDA}}[n]$, and $\varepsilon_{\text{xc}}^0 = \varepsilon_{\text{xc}}^0[n, \nabla n]$ contains a gradient correction,

$$\varepsilon_{\text{xc}}^0 = \varepsilon_{\text{xc}}^{\text{LDA}} - \varepsilon_x^{\text{LDA}} \left[\frac{Z}{9} \left(\frac{\nabla n}{k_{\text{F}} n} \right)^2 \right], \tag{5.8}$$

with $Z = -0.8491$. In fact, the kernel may be written in terms of the density–density separation $r_{12} = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$ and the local wave vectors $q_1 = q_0(\boldsymbol{r}_1)$ and $q_2 = q_0(\boldsymbol{r}_2)$, which may even be reduced to the two dimensionless variables $d_1 = r_{12} \cdot q_1[n_1, \nabla n_1]$ and $d_2 = r_{12} \cdot q_2[n_2, \nabla n_2]$, where $n_i = n(\boldsymbol{r}_i)$.

Even though E_c^{nl} uses only semilocal density information as input, it is still computed from a 6D spatial integral over all possible density pairs (the factor of $1/2$ in (5.3) accounts for double-counting). Moreover, the corresponding KS correlation potential[101] $v_c^{\text{nl}} = \delta E_c^{\text{nl}} / \delta n$ is needed for self-consistent calculations. This is a significant computational task. However, the fast Fourier transformation procedure of Román-Pérez and Soler (Ref. 104) has proven particularly suited for efficient evaluation of E_c^{nl} and v_c^{nl} in periodic DFT calculations, and is adopted in GPAW. In essence, the interaction kernel is approximately factorized and the 6D real-space integral is transformed to a 3D integral over convolutions in reciprocal space.

5.2 vdW-DF variants

The choice of GGA exchange functional is the final approximation of vdW-DF type calculations. As outline above the vdW-DF method was designed primarily to yield the

correct $-C_6R^{-6}$ asymptotics for well-separated fragments, where inter-fragment exchange interactions are supposed to be repulsive or neglible, as shown by HF potential-energy curves (PECs) for weakly interacting molecular and noble-gas dimers.[106,107] However, it was realized early on that HF (or EXX) itself is not necessarily a good match to the vdW-DF correlation.[108,109] Far too strong interaction energies are found when using the DFA $E_{\rm xc} = E_x^{\rm HF} + E_c^{\rm LDA} + E_c^{\rm nl}$, as compared to MP2 or CCSD(T) wavefunction calculations. Even though the vdW-DF correlation model is derived from the exact ACFD expression, inadequate cancellation of errors between exchange and correlation may be the reason for this.[102] Furthermore, such a procedure would be computationally intractable for most extended systems anyway due to the cost of evaluating HF/EXX.

Instead, a suitably chosen or designed GGA exchange functional is used, $E_x^{\rm GGA}$ in (5.1). The revPBE exchange was initially chosen due to its close resemblence to HF in simple test cases.[100,102] This defines the original vdW-DF as

$$E_{\rm xc}^{\rm vdW\text{-}DF} = E_x^{\rm revPBE} + E_c^{\rm LDA} + E_c^{\rm nl}, \tag{5.9}$$

which has been successfully applied in a range of sparse matter studies, but also has its flaws. Most notably, this particular combination of exchange and correlation models predicts in general too long equilibrium distances between interacting fragments, and underestimate the strength of hydrogen bonds.[90,102]

Improvements of vdW-DF have mainly taken three different approaches: (i) optimization of the GGA exchange approximation employed, (ii) modification of the nonlocal kernel and a new choice of GGA exchange, and (iii) fundamental modification of the nonlocal correlation approximation.

Exchange-optimized vdW-DF variants Several GGA exchange functionals have been constructed especially for use in vdW-DF type calculations. The optPBE and optB88 exchanges of Ref. 89 are optimized versions of the PBE[48] and B88[57] exchange functionals, respectively, and the corresponding vdW-DF variants are denoted optPBE-vdW and optB88-vdW. They were generated by fitting the parameters in the exchange enhancement factors such as to minimize the optPBE-vdW and optB88-vdW errors on the S22 benchmark dataset of noncovalent interactions. Very recently also the optB86b-vdW was proposed.[51]

Another vdW-DF variant is C09-vdW.[84] In order to reduce the short-range exchange repulsion partly responsible for the poor vdW-DF equilibrium distances, a less steep form of $F_x(s)$ was adopted for small reduced-density gradients (similar in spirit to the PBEsol GGA exchange), and an asymptotic approach to the revPBE upper bound, $F_x(s \to \infty) = 2.245$, was enforced for large s. A smooth function was fitted to interpolate between the two regimes. C09-vdW improves over vdW-DF on the S22 set, and provides a more accurate prediction of the interlayer distance in graphite, albeit at the cost of a somewhat overestimated interlayer binding energy.[84,110]

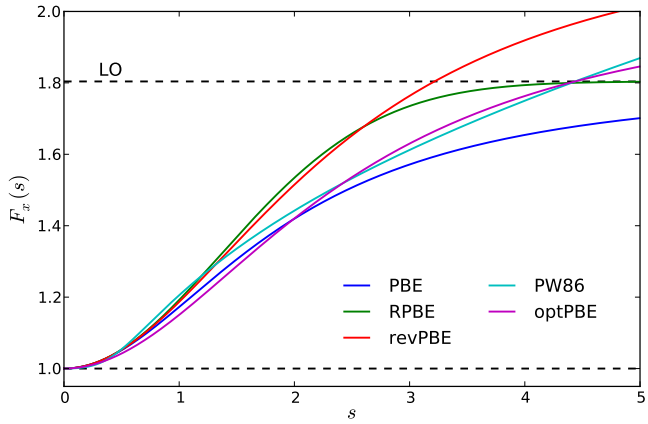

Figure 5.1: Various literature GGA exchange enhancement factors. The LDA is indicated by the bottom dashed line, and the Lieb–Oxford bound at the top.

vdW-DF2 revision The vdW-DF2 functional is a revision of the original vdW-DF.[90] It employs not only a different GGA exchange model but also a slightly modified nonlocal correlation kernel. As discussed around (5.7) and (5.8), a semilocal energy fraction determines the local wave vector $q_0(\boldsymbol{r})$. In vdW-DF2 the correlation kernel is modified to employ a stronger gradient dependence for determining q_0 by choosing $Z = -1.887$ in (5.8). Furthermore, following studies showing that PW86 exchange[111] particularly closely emulates HF for weakly bonded systems,[73,107] this GGA exchange functional was chosen for vdW-DF2.[112]

From now on the vdW-DF family of DFAs shall be understood as consisting of vdW-DF, vdW-DF2, optPBE-vdW, optB88-vdW, and C09-vdW. A few of the relevant exchange enhancement factors are shown in Fig. 5.1.

VV09 and VV10 approximations A rather different route towards improving van der Waals density functional theory has been taken by Vydrov and Van Voorhis, who have proposed several modifications of the original vdW-DF framework.[91,113–115] The vdW-DF-09[113] introduced a different model for the dielectric function and a few other simplifications, leading to a nonlocal correlation kernel with an empirical parameter. Together with a density dependent gradient correction the E_c^{nl} of vdW-DF-09 was then applied as a correction to a range-separated hybrid functional. Later on, the VV09 correlation approximation was proposed.[114,115] It is based on a different analytic form for the interaction kernel, and employs a rather elaborate damping function such as to match the spin-dependent $E_c^{\text{nl-VV09}}$ with LDA correlation. More recently also the VV10 revision has appeared.[91] It builds on many of the ideas from the VV09 development, but differs by employing two empirical parameters that allow high accuracy and saturation of the nonlocal correlation model for short-range density–density

interactions, i.e., in the covalent regime. It is therefore intended as a correction to existing GGA correlation functionals, and yields impressive performance for noncovalently as well as covalently bonded molecular systems.[91,116] However, an efficient scheme for self-consistent evaluation of the VV10 nonlocal correlation for extended systems (crystalline solids, metallic surfaces, etc.) has yet to appear. This DFA has therefore not been applied in any of the studies reported here.

5.3 Explicit dispersion corrections

Another category of approaches to accouting for dispersion in DFT is addition of more or less empirical corrections to the KS total energy expression. Some of the simplest of these so-called DFT-D methods are those of Grimme,[117,118] where the correction is based on the assumption that the total dispersion interaction can be described as a sum of contributions from all pairs of atoms,

$$\begin{aligned} E_{\text{DFT-D}} &= E_{\text{KS}} + E_{\text{disp}}, \\ E_{\text{disp}} &= -s_6 \sum_{\text{pairs}} f_{\text{damp}}(R) C_6 R^{-6}, \end{aligned} \tag{5.10}$$

where R is the distance between the atoms in a pair and C_6 a dispersion coefficient. The Fermi-type damping function $f_{\text{damp}}(R)$ is fitted such as to avoid the singularity for $R \to 0$ and to match E_{disp} to the XC functional chosen for calculating E_{KS}. The scaling factor s_6 is also fitted. Such a force-field type energy is very easily calculated, even for systems containing thousands of atoms, but is obviously outside the realm of KS-DFT. It is not electron density dependent and does not provide a potential for the KS equations. In particular, the dispersion coefficients should be electronic structure dependent rather than constant and empirically determined.

Later developments within DFT-D does, however, offer improvements by using system-dependent C_6 coefficients. The starting point is often the Casimir–Polder formula (3.1) in a simplified form, which then relates the C_6^{AB} coefficient for dispersion interactions between fragments A and B to their polarizabilities, which may be calculated from first principles in several ways. These methods include the exchange-hole dipole moment approach of Becke and Johnson,[119,120] the TS09 method of Tkatchenko and Scheffler,[121] Grimme's DFT-D3 method,[92] and the approach of Sato and Nakai for their local-response dispersion model,[122] amongst others. Higher-order multipole terms may also be included, as well as nonadditive three-body terms.[92]

None of these methods were applied in the present studies.

5.4 Assessment of vdW-DF variants

With a growing number of vdW-DF type methods appearing in literature it is important to be able to discriminate between them in terms of computational accuracy and general applicability. This section summarizes two studies (Paper I and Paper II) of the reliability of the vdW-DF family for predicting various condensed matter quantities, as compared to the standard PBE and RPBE GGA functionals. The question of whether or not to include semilocal correlation contributions in vdW-DF type calculations is also discussed, that is, other choices than $E_c^0 = E_c^{\mathrm{LDA}}$ for (5.1).

5.4.1 Standard materials properties

Three datasets are chosen to represent three vastly different materials properties of common interest in computational physics and chemistry: the G2-1 molecular formation energies, the S22 noncovalently bonded complexes, and the Sol20 lattice constants and cohesive energies of crystalline solids. These were all introduced in Section 4.1.

Most density functional computations presented in this section were done using self-consistent RPBE electron densities. Total energies for all other DFAs were thus evaluated non-self-consistently. Earlier studies have largely validated this approach to vdW-DF type calculations.[89,101]

5.4.2 Explicit semilocal correlation for vdW-DF

It is interesting to investigate whether purely local LDA correlation is necessarily the optimum choice for E_c^0. It is the obvious choice if correlation double-counting is to be completely avoided; the nonlocal correlation kernel vanishes in the local limit, $\phi(\boldsymbol{r}, \boldsymbol{r}', |\boldsymbol{r}-\boldsymbol{r}'|) \to 0$ for $|\boldsymbol{r}-\boldsymbol{r}'| \to 0$, indicating that E_c^{nl}, by design, has no strictly local contributions. However, semilocal approximations often prove reasonably reliable in interatomic bonding situations dominated by non-vdW type interactions. To conserve this property in vdW-DF variants one must therefore rely on E_c^{nl} to contain (at least) the corrections to LDA correlation offered by successful GGA functionals for strongly interacting systems. Does E_c^{nl} accomplish this? And does explicit inclusion of GGA correlation contributions to the vdW-DF method offer a generally improved description of interacting matter anyhow?

To investigate this, the RPBE+nl functional is tentatively introduced here,

$$\begin{aligned} E_{\mathrm{xc}}^{\mathrm{RPBE+nl}} &= E_x^{\mathrm{RPBE}} + E_c^{\mathrm{PBE}} + E_c^{\mathrm{nl}} \\ &= E_{\mathrm{xc}}^{\mathrm{RPBE}} + E_c^{\mathrm{nl}}, \end{aligned} \tag{5.11}$$

which replaces LDA correlation with semilocal PBE correlation in the vdW-DF method. This is obviously a vdW-extension of RPBE, and poses a risk of correlation double-

counting. Indeed, some overlap between GGA and nonlocal correlation interactions is quite probable, but RPBE+nl is here tested for illustrative purposes only. To facilitate comparison, a vdW-DF variant using RPBE exchange is also used. This approximation was called RPBE-vdW in Ref. 123. Finally, an XC model mixing local and semilocal correlation approximations for RPBE-vdW is introduced and denoted RPBEc2/3+nl. Details of this model are given later on, in Section 5.4.3.4.

5.4.3 Comparative assessment

All tested density functionals are evaluated and compared on the basis of the mean-signed and mean absolute deviation from benchmark quantities (MSD and MAD, respectively), as well as their relative counterparts (MSRD and MARD). The sign convention is

$$\text{deviation} = \text{DFT} - \text{benchmark}. \tag{5.12}$$

5.4.3.1 S22 complexes

Figure 5.2 shows signed deviations from CCSD(T) benchmarks of S22 interaction energies, E_{int}. Deviation statistics for each DFA are listed in Table 5.1. It is seen that both PBE and RPBE in particular underestimate E_{int} on average (positive MSD and MSRD), while the functionals including nonlocal correlation contributions improve on GGA results, as previously found in Ref. 89 and other studies using the S22 set. The improvement when applying RPBE-vdW instead of RPBE is particularly illustrative since only the treatment of correlation is changed, that is, the substitution $E_c^{\text{PBE}} \to E_c^{\text{LDA}} + E_c^{\text{nl}}$ is made. As seen in Table 5.1 the effect of this substitution is massive, decreasing the MSRD by more than a factor of 10 (101% → 8%). This illustrates that the vdW-DF approach is indeed able to capture noncovalent interactions otherwise out of the realm of ordinary GGAs. Also note from Fig. 5.2 that vdW-DF2 improves

Table 5.1: Statistics of the S22 interaction energy deviations shown in Fig. 5.2. All energies in meV.

E_{xc}	MSD	MAD	MSRD	MARD
PBE	120	120	59%	59%
RPBE	225	225	101%	101%
RPBE-vdW	57	61	8%	21%
vdW-DF	71	72	20%	25%
vdW-DF2	48	48	15%	17%
optPBE-vdW	3	20	−8%	12%
RPBE+nl	−37	39	−29%	29%

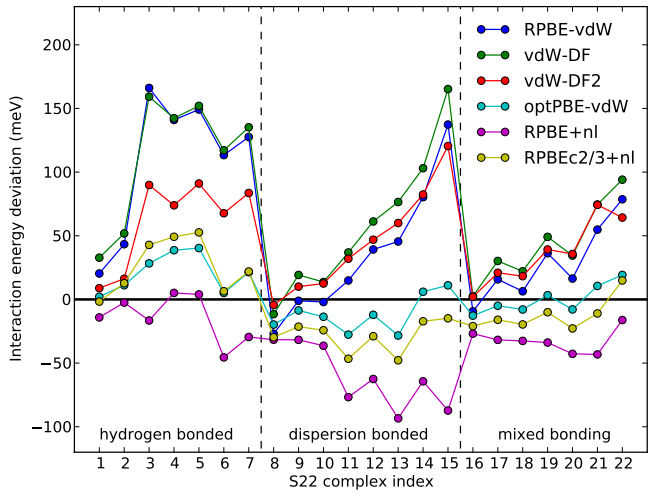

Figure 5.2: Deviations of calculated S22 interaction energies from the CCSD(T) benchmarks of Takatani *et al.*[95] The complexes are divided into three groups according to the predominant type of bonding: hydrogen, dispersion, and mixed bonding. Positive energy deviation means interaction energy underestimate as compared to the benchmark.

on vdW-DF for the energetics of hydrogen bonds in particular, as also reported in Ref. 90.

Using an exchange functional that is explicitly fitted to the S22 benchmark data the optPBE-vdW functional performs very well with a MAD of 20 meV. The corresponding relative deviations are also quite small, and the optPBE-vdW data points in Fig. 5.2 are significantly closer to the zero-line than those of its vdW-DF predecessor, particularly for dispersion dominated complexes.

Interestingly, exchanging LDA for PBE correlation, RPBE+nl yields a smaller MAD than that of the RPBE-vdW, and proves highly useful for hydrogen bonding. This remarkable result is obtained without any modification of the exchange functional but comes about solely from the use of semilocal rather than strictly local correlation in conjunction with E_c^{nl}. However, a significant tendency to overestimate the strength of dispersion dominated bonding is apparent from the middle part of Fig. 5.2, where some RPBE+nl deviations reach almost -90 meV. Also the negative RPBE+nl MSD and MSRD in Table 5.1 indicate a mean overbinding of this XC model, which is most likely due to some degree of correlation double-counting.

Self-consistent calculations were also performed with a few of the DFAs, and was found to change the statistics in Table 5.1 by only a few meV, so non-self-consistent calculations are adequate for the present purpose of DFA cross-comparison.

Table 5.2: Stastistics of the G2-1 formation energy deviations shown in Fig. 5.3. Negative deviation signifies intramolecular overbinding. All energies in eV.

E_{xc}	MSD	MAD	MSRD	MARD
PBE	−0.26	0.34	−4.1%	5.7%
RPBE	0.10	0.23	0.1%	3.6%
RPBE-vdW	−0.12	0.17	−2.8%	3.4%
vdW-DF	−0.16	0.18	−3.0%	3.5%
vdW-DF2	−0.18	0.20	−3.1%	3.8%
optPBE-vdW	−0.44	0.45	−6.4%	6.7%
RPBE+nl	−0.11	0.27	−2.7%	5.1%

5.4.3.2 G2-1 molecules

Figure 5.3 shows deviations of computed molecular formation energies from the G2-1 benchmarks. Statistics are listed in Table 5.2. It is easily seen that RPBE offers on average a quantitatively more accurate prediction of intramolecular bond strengths as compared to PBE: The RPBE yields a smaller G2-1 MAD and a modest average underbinding of the molecules (MSD > 0). This should not surprise since improved predictions of molecular bond strengths was a main argument for the design of the RPBE exchange approximation and its revPBE predecessor.[55,56]

Conversely, the five van der Waals functionals all tend to overbind on average. Even so, the MAD for all except optPBE-vdW are smaller than or comparable to that of RPBE, and the MARD of RPBE-vdW, vdW-DF, and vdW-DF2 appear very reasonable. On the other hand, the optPBE-vdW model, with an exchange component somewhat less repulsive in the small-s regime than the other exchange functionals considered here, is clearly overbinding the G2-1 dataset. Only a few optPBE-vdW points in Fig. 5.3 are above the zero-line and so almost all of the 0.45 eV MAD stems from overbinding. This illustrates one possible drawback of fitting density functionals to a single electronic structure property only. Finally, it is seen from the RPBE and RPBE+nl points in Fig. 5.3 that the effect of adding the nonlocal correlation term is system dependent: For nearly all 55 molecules it introduces either essentially no change to the formation energy, or a negative contribution to $\Delta_f E$, which may be rather large (−0.6 eV for SO_2).

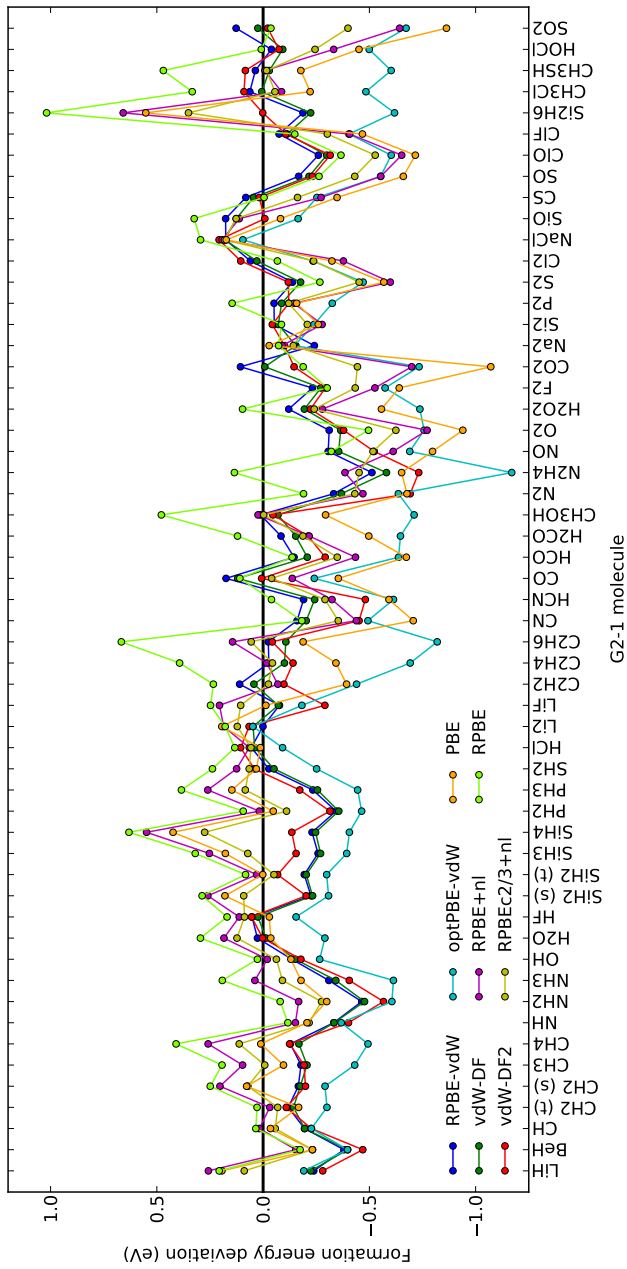

Figure 5.3: Deviations of electronic-only molecular formation energies from semi-experimental benchmarks derived from the G2-1 set of formation enthalpies. Negative energy deviation signify an overestimate of the bond strength.

5.4.3.3 Sol20 crystals

Figure 5.4 reports deviations from experimental values of calculated equilibrium lattice parameters a_0 and cohesive energies E_c for the Sol20 crystals. Results are summarized in Table 5.3.

The PBE approximation appears very well-balanced for these types of materials properties. The experimental lattice constants are rather accurately reproduced with a small average overstimate, while cohesive energies are slightly underestimated. The RPBE predicts softer crystal lattices, resulting in a larger a_0 MSD and more negative E_c MSD.

All of the three functionals RPBE-vdW, vdW-DF, and vdW-DF2 essentially double the RPBE a_0 mean error, vdW-DF2 in this respect performing worst of all considered DFAs. The optPBE exchange functional appears more attractive for solids than both the RPBE and revPBE ones. Crystal volumes are smaller on average than with vdW-DF, and cohesive energies are correspondingly larger. Again, this may be attributed to the functional form for small reduced density gradients. Similar conclusions were drawn from a recent study [51] of the performance of vdW-DF variants for solid state

Table 5.3: Statistics of the deviations from experimental values of calculated Sol20 lattice constants and cohesive energies shown in Fig. 5.4. For details see text and the caption for Fig. 5.4.

E_{xc}	MSD	MAD	MSRD	MARD
	Lattice constants (Å)			
PBE	0.036	0.040	0.9%	1.1%
RPBE	0.063	0.065	1.7%	1.7%
RPBE-vdW	0.123	0.123	3.4%	3.4%
vdW-DF	0.113	0.113	3.1%	3.1%
vdW-DF2	0.141	0.141	3.8%	3.8%
optPBE-vdW	0.069	0.071	1.9%	1.9%
RPBE+nl	0.024	0.033	0.6%	0.9%
	Cohesive energies (eV/atom)			
PBE	−0.11	0.33	−4%	9%
RPBE	−0.69	0.71	−19%	19%
RPBE-vdW	−0.86	0.86	−22%	22%
vdW-DF	−0.78	0.78	−20%	20%
vdW-DF2	−0.80	0.80	−21%	21%
optPBE-vdW	−0.24	0.33	−7%	9%
RPBE+nl	0.39	0.42	9%	10%

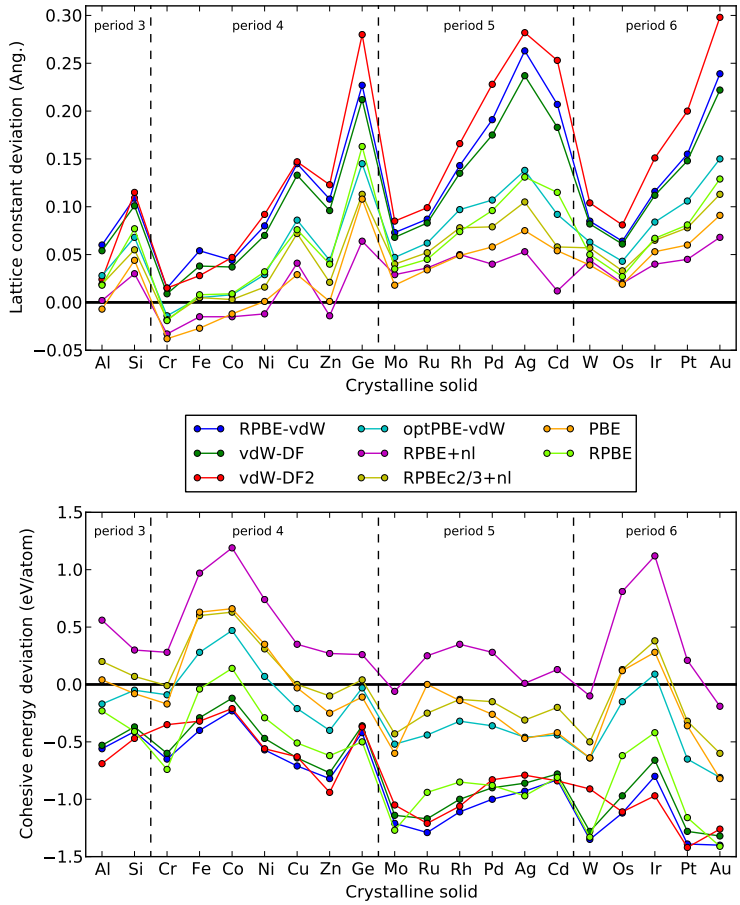

Figure 5.4: Deviations of the Sol20 materials properties from experimental values. The 20 bulk solids are ordered ascendingly according to their Periodic Table period and atomic number. Top: Lattice constant deviations $a_0^{\mathrm{DFT}} - a_0^{\mathrm{exp}}$. Experimental values without zero-point corrections are from Ref. 124. Bottom: Cohesive energy deviations $E_c^{\mathrm{DFT}} - E_c^{\mathrm{exp}}$ with $E_c^{\mathrm{exp}} > 0$. Experimental data without zero-point corrections from Ref. 41.

properties. It is by now widely recognized that the Rutgers–Chalmers functionals vdW-DF and vdW-DF2 predict in general much too soft crystal lattices, and that this may be remedied by optimizing the GGA exchange model.

It is seen from Fig. 5.4 and Table 5.3 that explicit inclusion of semilocal correlation contributions also has the potential to offer such improvements. The RPBE+nl model appears highly accurate for lattice constants but also vastly overstimate the cohesive energies. Correlation interaction double-countings may be at play again. However, in Fig. 5.4 we also see that the so-called RPBEc2/3+nl DFA offers cohesive energies quite similar to the optPBE-vdW ones, while actually predicting lattice constants even closer to experimental values.

5.4.3.4 RPBEc2/3+nl

It is apparent from the analysis above that the amount of semilocal correlations introduced in the vdW-DF method (if any) must be balanced against those already accounted for by $E_c^{\rm nl}$. The RPBE+nl mean overbinding of the S22 systems and mean overestimate of bulk cohesive energies are probably due to such issues. From mere inspection of the data in Tables 5.1–5.3 the RPBEc2/3+nl functional is therefore constructed. It mixes LDA end GGA correlation to average out this problem,

$$E_{\rm xc}^{\rm RPBEc2/3+nl} = E_x^{\rm RPBE} + \tfrac{1}{3}E_c^{\rm LDA} + \tfrac{2}{3}E_c^{\rm PBE} + E_c^{\rm nl}, \tag{5.13}$$

where the $E_c^{\rm nl}$ of vdW-DF is used. As shown in Table 5.4 this approximation performs remarkably well when applied non-self-consistently to the three previously considered groups of benchmark systems. The S22 interaction energies are very accurately captured with only a slight average overbinding, while the G2-1 formation energies turn out similar to those of the vdW-DF variants (i.e., better than optPBE-vdW). Moreover, RPBEc2/3+nl performs comparably to optPBE-vdW for predicting the bulk solid properties and avoids the significant overbinding of crystals arising when fully substituting LDA correlation for PBE correlation. In total, this looks like a very promising density functional approximation. In the following section it is applied in an example of every-day surface science studies.

Table 5.4: Mean deviations from experiments or high-level theory of calculated quantities performed with the RPBEc2/3+nl density functional defined in (5.13).

Dataset	unit	MSD	MAD	MSRD	MARD
S22 complexes	meV	−6	29	−16%	20%
G2-1 molecules	eV	−0.12	0.20	−2.8%	4.1%
Lattice constants	Å	0.052	0.054	1.4%	1.4%
Cohesive energies	eV/atom	−0.03	0.27	−1%	6%

5.4.4 Benzene adsorption on the Cu(111), Ag(111), and Au(111) surfaces

Benzene adsorption on the noble (111) surfaces of Cu, Ag, and Au is mediated primarily by dispersive van der Waals forces. These surfaces are the most close-packed facets of the first three Group 11 transition metals, sometimes termed the coinage metals. Such physisorption processes are therefore prime examples of the inability of ordinary rung 1–3 density functionals in correctly describing dispersion, and accurate description of physisorption is one main target for vdW-DFT. Several DFAs are here tested for their ability to capture the benzene/M(111) interactions (M = Cu, Ag, Au) responsible for physisorption. Apart from the DFAs applied in the preceeding section, these include the LDA, TPSS, and revTPSS functionals, the latter two of which are meta-GGA type approximations.

5.4.4.1 Experiments

Experimental evidence suggests that benzene (C_6H_6) interacts only weakly with the Cu(111), Ag(111), and Au(111) surfaces at low temperatures, resulting in relatively small adsorption energies and neglible distortion of the molecular geometry upon adsorption.[125–132] This is a signature of the decisive role of vdW dispersion in these three adsorption systems. At low coverage, the weak adsorbate–surface interactions cause C_6H_6 to adsorb with its aromatic ring parallel to the noble surfaces in a physisorbed state from which it can desorb reversibly. Low-energy electron diffraction (LEED) experiments at low coverage and temperature indicate benzene desorption from primarily three-fold hollow adsorption sites.

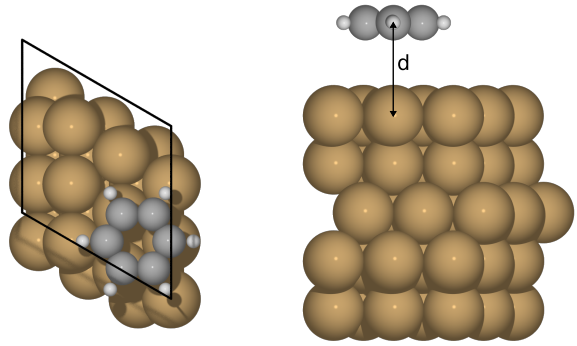

Figure 5.5: Side-view of the atomic geometry for benzene adsorption on the noble Cu(111) surface. Five metal layers are used in the Cu(111) repeated slab model and the C_6H_6 molecule adsorbs with its aromatic ring parallel to the surface at the equilibrium adsorbate–metal distance d.

Actual measurements of the low-coverage benzene adsorption energy $E_{\rm ads}$ on Cu(111), Ag(111), and Au(111) are scarce, but some numbers may be found in the surface science literature. In the case of C_6H_6/Au(111), an adsorption energy of −0.64 eV was found in temperature-programmed desorption (TPD) experiments by Koel and co-workers (Ref. 130). Another experiment reports $E_{\rm ads} = -0.6$ eV (Ref. 133), though not properly documented in that reference. Literature experimental adsorption energies for benzene on Ag(111) are somewhat more scattered. Using Redhead analysis,* Wöll and co-workers (Ref. 133) converted the 160 K benzene desorption temperature from Ref. 134 to $E_{\rm ads} = -0.42$ eV on Ag(111). However, later TPD experiments have found somewhat stronger benzene–Ag(111) binding energies: $E_{\rm ads} = -0.58$ eV was inferred in Ref. 130, while Rockey and co-workers (Ref. 132) found adsorption energies of −0.57 eV and −0.52 eV for physisorption of benzene in two different rotations in three-fold hollow sites on Ag(111). Since there is no reason to doubt the reliability of these later experiments, the experimental value $E_{\rm ads} = -0.57$ eV is here adopted for C_6H_6/Ag(111). Lastly, benzene adsorption on the Cu(111) surface was investigated by Stevens and co-workers in Ref. 127. It was concluded that the C_6H_6/Cu(111) binding is weaker than −0.62 eV. Later, Wöll and co-workers (Ref. 131) used TPD spectra of Ref. 135 to determine the adsorption energy more accurately, obtaining $E_{\rm ads} = -0.59$ eV. In summary, the experimental low-temperature and low-coverage adsorption energies of benzene on Cu(111), Ag(111), and Au(111) are reasonably taken as −0.59 eV, −0.57 eV, and −0.64 eV, respectively. Unfortunately, there are no experimental data available about the equilibrium benzene–M(111) distance.

5.4.4.2 Computational details

Potential-energy curves (PECs) for C_6H_6/M(111) physisorption are calculated using 3×3×5 metal slabs, which are repeated periodically in the surface planes to model the M(111) surfaces. For each slab, the top-three metal layers are relaxed with PBE. The fully relaxed benzene molecule is placed with its molecular plane parallel to the relaxed surface in a hcp three-fold hollow adsorption site at the benzene–metal distance $d =$ 4.0 Å (see Fig. 5.5) and then relaxed again. With the resulting intramolecular geometry fixed, the adsorption energy is then computed for varying interaction distances, $d =$ 2.4–6.0 Å, using different DFAs. The model system geometries are therefore those found self-consistently with PBE. We expect this to have only a minor impact on the PECs for other DFAs.

5.4.4.3 Benzene–M(111) potential-energy curves

Figures 5.6–5.8 illustrate the calculated PECs for benzene physisorption on Cu(111), Ag(111), and Au(111), respectively. In all three figures, the experimental adsorption energy is indicated by a horizontal dashed line. The minima of PECs obtained with the

* Redhead analysis: A simple relation between the activation energy for desorption of the adsorbate and the measured temperature of maximum desorption rate.

computationally more expensive MP2 wavefunction method are shown for comparison. These data are from Ref. 133, where a cluster model approach was employed for C_6H_6 adsorption in the ontop adsorption site.

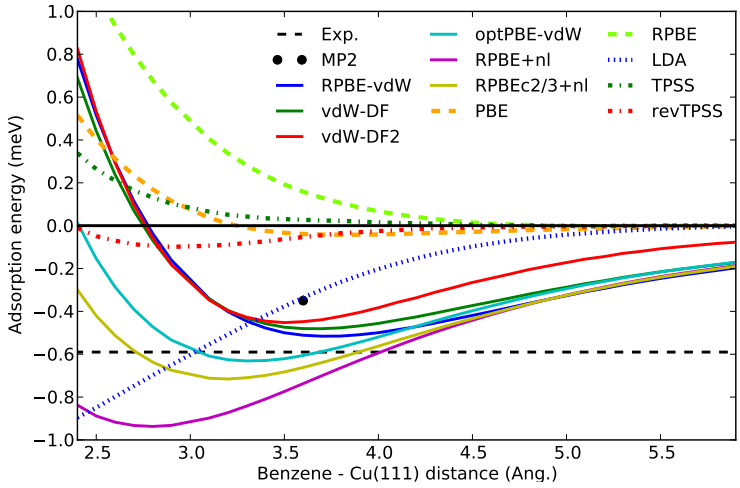

Figure 5.6: Potential-energy curves for benzene adsorption on the Cu(111) surface obtained using different density functionals. The black dot indicates the PEC minimum found in Ref. 133 using the MP2 wavefunction method. Horizontal dashed line indicates the experimental adsorption energy.

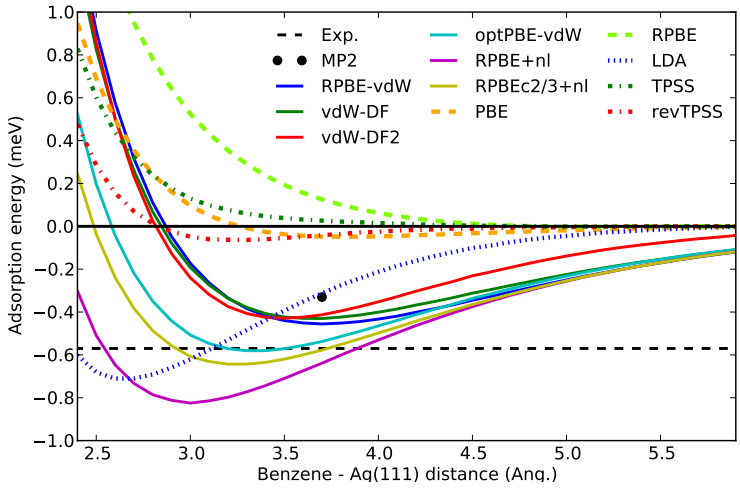

Figure 5.7: Benzene/Ag(111) PECs calculated using different XC models.

As the RPBE PECs are everywhere repulsive (positive E_{ads}) it is again demonstrated that ordinary RPBE calculations does not even attempt to account for dispersion-mediated binding. The same is true for TPSS, while PBE and revTPSS do predict some, but insignificant, stabilization of the benzene molecule on the coinage metals. Decomposition of the PBE and LDA PECs into exchange and correlation contributions (not shown in the figures) indicate that the shallow PEC minima for PBE stems from correlation only, ruling out spurious exchange-mediated attraction. The opposite is the case for LDA, where exchange is solely responsible for the minima. Such spurious effects of LDA exchange are well known (e.g., Refs. 72–74), and especially the LDA PEC for C_6H_6/Cu(111) adsorption in Fig. 5.6 is strongly exaggerated as compared to the vdW-DF type PECs, and is overly deep when compared to experiment.

Turning attention to the XC models containing a nonlocal correlation term, significant adsorption energy minima are introduced in Figs. 5.6–5.8. Calculated equilibrium C_6H_6–M(111) distances and adsorption energies are listed in Table 5.5 along with experimental and MP2 data.

The general trends across Table 5.5 appear somewhat different for the case of Cu(111) than for Ag(111) and Au(111). Note that experimental measurements of E_{ads} for C_6H_6/Cu(111) is particularly scarce (there is essentially only one). Furthermore, it was found in Ref. 133 that adsorbate-to-substrate charge donation is by far strongest in the case of Cu(111). Therefore, disregarding Cu(111) for a while, the following general trends are observed for benzene on Ag(111) and Au(111). First of all, the RPBE-vdW, vdW-DF, and vdW-DF2 adsorption energies are all roughly 20–25% too small as compared to experiments. Second, it is well known that vdW-DF tends to overestimate noncovalent binding separations,[90] so we should probably expect the C_6H_6–M(111) equilibrium distance to be somewhat smaller than the 3.6–3.7 Å consistently found with vdW-DF. Note that vdW-DF2 accomplishes this but without any extra binding.

Table 5.5: Equilibrium benzene–M(111) distances d (Å) and adsorption energies E_{ads} (eV).

	Bz/Cu(111)		Bz/Ag(111)		Bz/Au(111)	
XC	d	E_{ads}	d	E_{ads}	d	E_{ads}
MP2[a]	3.6	−0.35	3.7	−0.33	3.8	−0.31
RPBE-vdW	3.7	−0.52	3.7	−0.46	3.6	−0.50
vdW-DF	3.7	−0.48	3.7	−0.43	3.6	−0.48
vdW-DF2	3.5	−0.45	3.5	−0.43	3.5	−0.48
optPBE-vdW	3.3	−0.63	3.3	−0.58	3.3	−0.64
RPBE+nl	2.8	−0.94	3.0	−0.82	3.0	−0.90
RPBEc2/3+nl	3.2	−0.72	3.2	−0.64	3.2	−0.70
Exp.[b]	–	−0.59	–	−0.57	–	−0.64

[a] Ref. 133.
[b] Refs. 130–132.

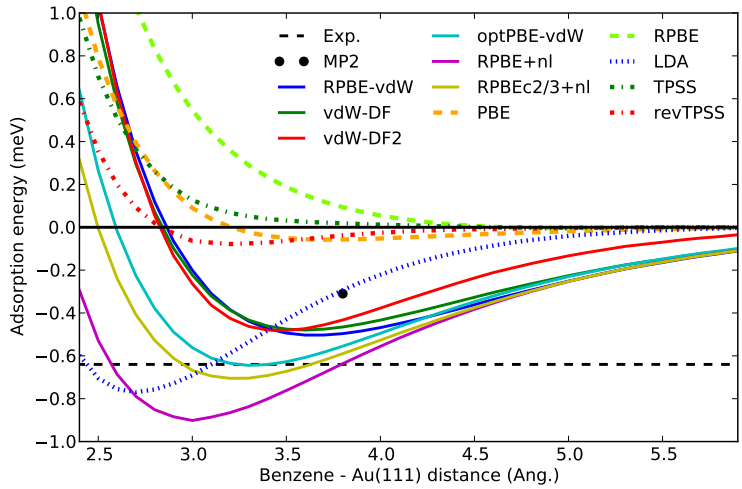

Figure 5.8: Potential-energy curves for benzene physisorption on the Au(111) surface calculated using different rung 1–3 density functionals as well as XC models including the Rutgers–Chalmers nonlocal correlation approximation. As for C_6H_6 on Cu(111) and Ag(111), the PEC minimum predicted by the MP2 wavefunction method appears to fall very close to the LDA PEC, though far from the LDA minimum.

Of the van der Waals XC functionals considered here, only the optPBE-vdW and the RPBEc2/3+nl appear to perform well for C_6H_6 on Ag(111) and Ag(111). Both predict smaller separations than vdW-DF, and optPBE-vdW in particular proves highly accurate for the physisorption energetics. The PECs for both DFAs furthermore show an asymptotic approach to the non-interacting limit ($d \to \infty$) that is very similar to those of RPBE-vdW and vdW-DF, which are expected to be reasonably accurate in this limit of vanishing density–density overlap between adsorbate and surface.

5.5 Discussion and summary

The vdW-DF approach to inclusion of vdW dispersion interactions in DFT is promising. It employs the Rutgers–Chalmers density dependent nonlocal correlation, an approximate but computationally tractable representation of the beyond-LDA part of the exact ACFD correlation energy. In contrast to the many DFT-D methods, which add a more or less empirical dispersion correction to the KS total energy, the RC correlation approximation provides a KS potential and thus also Hellmann–Feynman forces. In standard vdW-DF variants it is combined with a GGA exchange approximation and strictly local LDA correlation.

It is clear from this study and many others that vdW-DF introduces stabilizing interactions between noncovalently bonded fragments, as it was designed to do. Such interactions are not at all accounted for by, e.g., the RPBE density functional, which has become popular for studies in surface chemistry and catalysis. However, it is also clear that the three vdW-DF variants RPBE-vdW, vdW-DF, and vdW-DF2 do not offer quantitative agreement with high-level theory for noncovalent bonds when available. Those three XC models furthermore predict much too soft crystal lattices. Even though molecular formation energies are described at the RPBE level, this arguably hampers truly general applicability of these three DFAs in theoretical surface science.

The optPBE-vdW approximation amends some of these issues by explicitly fitting an exchange functional for use in the vdW-DF method. This approach works very well for noncovalent bonds and crystalline solids, but also leads to severe overestimates of intramolecular binding energies of the G2-1 molecules. Offhand, this questions the applicability of such exchange-optimized vdW-DF variants in studies involving chemical transformations, a dominant field in materials science. This is investigated more closely in Chapter 7.

Including semilocal correlation contributions in the vdW-DF approach was here explored as an alternative approach towards a generally applicable, high-accuracy van der Waals density functional. Not surprising, fully substituting LDA for GGA correlation produces significant correlation double-countings, so local and explicitly semilocal correlations must be balanced. The RPBEc2/3+nl construction accomplishes this very well when assessed in terms of quantitative performance on the benchmark datasets considered here. Moreover, together with optPBE-vdW, this approach convincingly outperforms RPBE-vdW, vdW-DF, and vdW-DF2 for studies of benzene physisorption on the noble Ag(111) and Au(111) surfaces. Note that these are rather generic examples of target applications of vdW density functionals. For C_6H_6/Cu(111) the picture is, however, less clear-cut.

Even so, the rest of this thesis takes the point of view that high-performance and general-purpose van der Waals density functionals should be obtained from optimizing the exchange approximation *and* the fraction of GGA correlation used in E_c^0.

Chapter 6

Statistical Background and Machine Learning Methods

Having in the previous chapter investigated the performance of vdW-DF variants with different exchange approximations and the possible influence of semilocal correlation contributions, the rest of this thesis reports on methods to construct completely new XC models through semi-empirical fitting. Before diving into this it is useful to first consider some of the statistical methods underlying these works.

This chapter is therefore concerned with the basics of the approach to statistical modelling later used for developing density functionals with built-in error estimation, which involves a combination of methods inspired by Bayesian statistics and tools developed in the field of machine learning. Fear not, we shall not undertake a complete review of these topics. Instead, methodologies are sketched with emphasis on the underlying philosophy. This chapter is based on the thorough exposition in Ref. 136.

6.1 Bayesian statistics

Statistical modelling is essentially all about maximizing our confidence that some mathematical model can be used for predicting future events or results of measurements, that is, the value of data drawn from some distribution. However, often we only have a finite (and possibly small) number of data samples available from this distribu-

tion, and the data drawn from it may even be noisy and subject to uncertainty. This is a common situation when modelling experimentally determined data. Moreover, the distribution underlying the data may be essentially unknown to us, making it hard to assess wether the final model really captures the essential characteristics of the distribution, or if it merely fits the random noise. The latter would be an *overfitted* model with limited or no ability to reliably predict the outcome of future "measurements" of data. This is a very dangerous situation for the statistical modeller, but there are ways around it. We start off with some basic probability theory before discussing aspects of the present approach to statistical modelling.

6.1.1 Basic probability theory

The two basic rules for working with probabilities are the product rule and the sum rule. If we consider two random variables X and Y, the product rule states that the joint probability $P(X, Y)$ of both variables taking on some variables (say, x and y) is given by the probability $P(Y)$ that $Y = y$, times the conditional probability $P(X|Y)$ that $X = x$, given that Y has already taken on the value y,

$$P(X \cap Y) = P(X, Y) = P(X|Y)P(Y). \tag{6.1}$$

The sum rule states that the probability for a random variable X may be found from summing over joint probabilities,

$$P(X) = \sum_{Y} P(X, Y), \tag{6.2}$$

where the sum is over mutually exclusive possibilities for Y.

Let us now consider the situation where we have a dataset D of target data which we will attempt to describe using a model M with the parameters $\boldsymbol{a}$ (a vector). The present discussion implicitly assumes the existence of some "true" model $\mathcal{M}$ which generated the target data, and it is our aim to represent $\mathcal{M}$ with the parametrized model $M(\boldsymbol{a})$. Even though the target data were generated by the true model, the "measurement" process may have introduced noise on the data. In that case we can never be absolutely certain of what the model parameters should be. However, using the product rule (6.1) and the relation $P(X, Y) = P(Y, X)$, we can define a *probability distribution* $P(M\boldsymbol{a}|D)$ for the parameters given the known data,

$$P(M\boldsymbol{a}|D) = \frac{1}{P(D)} P(D|M\boldsymbol{a})P(M\boldsymbol{a}) \propto P(D|M\boldsymbol{a})P(M\boldsymbol{a}). \tag{6.3}$$

This is the well-known Bayes' theorem, which plays as central role in the field of machine learning and computational pattern recognition (e.g., Refs. 137–140). The Bayesian understanding of probabilities is different from the better-known frequentist interpretation. While the frequentist view is that the probability of an event is the limit of its relative frequency in a large number of trials, in a Bayesian interpretation it is rather a measure of plausability. Thus, the relation (6.3) may be written

$$\text{posterior} \propto \text{likelihood} \times \text{prior}, \tag{6.4}$$

which relates the *posterior* probability $P(M\boldsymbol{a}|D)$ to the *likelihood* $P(D|M\boldsymbol{a})$ and the *prior* probability $P(M\boldsymbol{a})$. In a Bayesian setting, these probabilities have the following interpretation: The prior is the probability that model $M(\boldsymbol{a})$ is a good model before the data D are known, the likelihood is the conditional probability that model $M(\boldsymbol{a})$ actually generated data D, given that $M(\boldsymbol{a})$ is a good model, and the posterior is the probability that $M(\boldsymbol{a})$ is in fact a good model after all evidence (data) has been considered.

6.1.2 Maximum-likelihood solutions

Let us consider now the situation where we have sampled N_d data points with Gaussian noise of width σ_0 from the true model $\mathcal{M}$. We will fit these data with a third-order polynomial z,

$$z(x, \boldsymbol{v}) = v_0 + v_1 x + v_2 x^2 + v_3 x^3 = \sum_{j=0}^{3} v_j x^j, \tag{6.5}$$

where $\boldsymbol{v}$ is the coefficient vector model solution to be determined using the maximum posterior method, that is, by maximizing the posterior in (6.3) and (6.4).[137]

Assuming unit prior probability, maximizing the posterior amounts to maximizing the likelihood. We are then looking for the maximum-likelihood (ML) model solution. Now, since the data points are normal distributed with a width of σ_0, the relation between posterior and likelihood may be explicitly written in terms of a Gaussian distribution with variance σ_0^2,

$$P(M\boldsymbol{v}|D) \propto P(D|M\boldsymbol{v}) \propto \exp\left(-\sum_n \frac{\{z(x_n, \boldsymbol{v}) - y_n\}^2}{2\sigma_0^2}\right). \tag{6.6}$$

Maximizing this probability amounts to minimizing the sum in the exponential, which is identical to minimizing the conventional least-squares loss function $L(\boldsymbol{v})$,

$$L(\boldsymbol{v}) = \frac{1}{2}\sum_n^{N_d} \{z(x_n, \boldsymbol{v}) - y_n\}^2. \tag{6.7}$$

6.1.2.1 Overfitting

As already alluded to, a well-known problem with fitting models is overfitting. If the model is too complex (contains too many parameters), and/or the set of target data is too small, the least-squares (or ML) best-fit models may be highly unrealistic and thus useless for reliable predictions of new data not contained in the dataset. Examples of this are shown in Fig. 6.1, where ML solutions when fitting a sine function using polynomial degrees of 3, 7 and 11 are plotted. The models, denoted M3 (–), M7 (–), and M11 (–), respectively, become increasingly oscillatory as the polynomial order

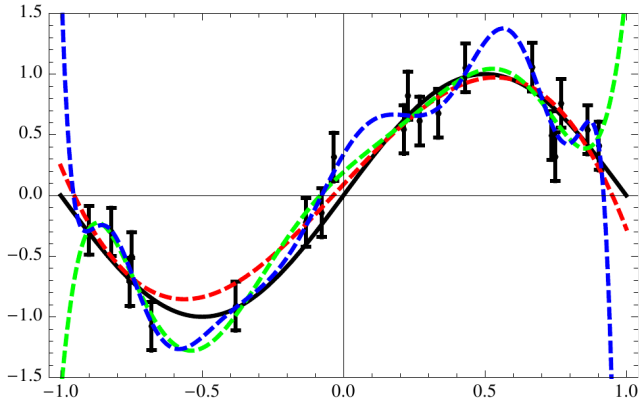

Figure 6.1: Maximum-likelihood solutions when fitting a noisy sine function (black curve) with polynomials of degrees 3 (red), 7 (green) and 11 (blue). The error bars on the N_d black target data points indicate the width σ_0 of the Gaussian noise. Adapted from Ref. 136.

increases, and for M11 it is clear that the flexible model space is used to fit the noise rather than the underlying sine function. That model is clearly more complex than necessary, indicating overfitting and thus a loss of model *generalization.*[139] On the other hand, any first-order polynomial approximation would be a poor representation of a noise-free sine function as well. Such model would be *underfitted.*

6.1.3 Prior expectation and regularization

Model space complexity is obviously the real issue with models M7 and M11 in Fig. 6.1; the M11 solution is clearly allowed to adjust to the noise and scarcely looks like a sine function. However, our *a priori* expectations when optimizing linearly expanded models is often well-behaved solutions with reasonable expansion coefficients and modest-valued derivatives. The prior distribution of Bayes' theorem allows us to express such expectations before even considering the target data (before evaluating the likelihood). This can be cleverly used to tune the model complexity by making certain parts of a highly flexible model space unlikely and effectively inaccessible.

A simple choice for the prior probability takes the form of the squares of all coefficients,

$$P(M\boldsymbol{v}) = \exp\left(-\frac{\boldsymbol{v}^T\boldsymbol{v}}{2\sigma_p^2}\right) = \exp\left(-\omega^2\boldsymbol{v}^T\boldsymbol{v}\right), \tag{6.8}$$

where $\boldsymbol{v}^T\boldsymbol{v} = \|\boldsymbol{v}\|^2 = \sum_j v_i^2$ is the length of $\boldsymbol{v}$, and σ_p is the width of the Gaussian prior, here expressed in terms of the prior strength $\omega^2 = \left(2\sigma_p^2\right)^{-1} > 0$. Minimizing the

posterior (6.4) with this prior amounts to minimizing a loss function C of the form

$$C(\boldsymbol{v},\omega)=\sum_n \frac{\{z(x_n,\boldsymbol{v})-y_n\}^2}{2\sigma_0^2}+\omega^2\boldsymbol{v}^T\boldsymbol{v}, \tag{6.9}$$

such that the *regularization* term $R=\omega^2\boldsymbol{v}^T\boldsymbol{v}$ penalizes the occurence of large coefficients in the minimizing solution $\boldsymbol{v}_0$. This particular choice of a quadratic regularizer is well documented in statistical literature under such different names as ridge regression, parameter shrinkage, and weight decay.[137] For $\omega^2 \gg 1$ the regularizer will quench the large coefficients in models M7 and M11 in Fig. 6.1, leading to significantly less overfitted models at the expence of sligthly worse fitness of the models in the least-squares sense.

However, we can only make this statement because we already know the true model in Fig. 6.1 (a sine). The next section deals with the more general case where we do not know $\mathcal{M}$, and where we can not even be sure that the employed model space can reasonably represent $\mathcal{M}$, whatever it is.

6.2 Machine learning procedures for linear model parametrization

Until now we have implicitly assumed models linearly expanded in a set of expansion functions, e.g., polynomials, and we shall continue to do so, but a more explicit description of the model space is now needed. Furthermore, error estimators developed for machine learning are used for determining the regularization strength.

6.2.1 Regularized cost function

We aim at approximating the true model $\mathcal{M}(x)$, which generated the N_d target data points in dataset D, by the objective function $f(x)$, which is linearly expanded in N_p basis functions $f_n(x)$,

$$f(x)=\sum_n^{N_p} f_n(x)a_n=\boldsymbol{f}^T(x)\boldsymbol{a}, \tag{6.10}$$

where $\boldsymbol{a}$ is a column vector of expansion coefficients and $\boldsymbol{f}^T(x)$ a row vector of basis functions. The model prediction $z_i^{\boldsymbol{a}}$ of the target data point y_i at x_i is therefore $z_i^{\boldsymbol{a}}=\boldsymbol{f}^T(x_i)\boldsymbol{a}$, which in matrix notation for all data points is written $\boldsymbol{z}=\boldsymbol{A}\boldsymbol{a}$, where $\boldsymbol{A}$ is a $N_d\times N_p$ matrix $A_{in}=f_n(x_i)$. Using these definitions we now write a regularized *cost function* of the form (6.9) as

$$\begin{aligned} C(\boldsymbol{a},\omega) &= \widetilde{C}(\boldsymbol{a})+R(\boldsymbol{a},\omega) \\ &= (\boldsymbol{A}\boldsymbol{a}-\boldsymbol{y})^T(\boldsymbol{A}\boldsymbol{a}-\boldsymbol{y})+R(\boldsymbol{a},\omega), \end{aligned} \tag{6.11}$$

where $\widetilde{C}(\boldsymbol{a})$ is the maximum-likelihood part and $R(\boldsymbol{a},\omega)$ the regularizer. The model solution $\boldsymbol{\alpha}$ that minimizes the cost function for some regularization strength ω_0^2, that is, $C(\boldsymbol{\alpha}) = \min_{\boldsymbol{a}} C(\boldsymbol{a},\omega_0) \equiv C_0$, is found from requiring the first derivative of the cost be zero with respect to all a_n.

However, $C(\boldsymbol{a},\omega)$ contains much more information than simply the model solution. The Hessian matrix $\boldsymbol{H} = C''$, more formally defined by

$$H_{kl}(\boldsymbol{a}) = \left.\frac{\partial^2 C(\boldsymbol{a},\omega)}{\partial a_k \partial a_l}\right|_{\boldsymbol{a}=\boldsymbol{\alpha},\omega=\omega_0}, \tag{6.12}$$

describes the cost function curvature around $\boldsymbol{\alpha}$. Since $C(\boldsymbol{a},\omega)$ is quadratic in the model parameters it may in the model space vicinity of $\boldsymbol{\alpha}$ be expressed in terms of the Hessian as

$$C(\boldsymbol{a},\omega) = C_0 + \frac{1}{2}\left(\boldsymbol{a}-\boldsymbol{\alpha}\right)^T \boldsymbol{H}\left(\boldsymbol{a}-\boldsymbol{\alpha}\right). \tag{6.13}$$

Importantly, an eigenvector decomposition of the Hessian for the ML part of the cost function alone, $\widetilde{C}(\boldsymbol{a})$, allows identification of strong (large eigenvalue) and weak (small eigenvalue) eigenmodes of $\widetilde{C}(\boldsymbol{a})$. These correspond directly to strong and weak directions in model space, and gives rise to the concept of *sloppy models*,[141] that is, models containing one or more poorly determined parameters. A sloppy mode is thus a direction in model space which may dramatically change its weight in the model solution without significantly changing the apparent quality of the fit (the ML cost). Such a parameter is not sufficiently constrained by the target data, and thus poorly determined. This may give rise to overfitting issues when the trained model is used to predict data outside the training data. Similar conclusions may be drawn from considering a singular value decomposition of $\widetilde{C}(\boldsymbol{a})$, see Refs. 136 and 142.

The regularizer has a direct impact on the eigenvalue spectrum of the ML cost function; it adds a curvature scaled with ω^2 to all modes in the Hessian. Thereby, small eigenvalues of $\widetilde{C}(\boldsymbol{a})$ are lifted while large eigenvalues are essentially unaffected. For intermediate regularization strength this means that sloppy cost function modes (small eigenvalues) are effectively frozen out of the regularized cost function $C(\boldsymbol{a},\omega)$, leaving behind only the most important modes. For very strong regularization ($\omega^2 \to \infty$) the simple ridge regression regularizer forces the length of $\boldsymbol{\alpha}$ to zero, most often leading to a far too simplistic model where also the important modes have been suppressed.

6.2.1.1 Effective number of parameters

Since regularization limits the effective degrees of freedom in model space contributing to the model solution, it is natural to speak of the *effective number of parameters* left in the model after regularization, N_{eff}. For the ridge regression regularizer the effective number of model parameters may be calculated from the eigenvalues w_n of the ML cost function Hessian,[139]

$$N_{\text{eff}}(\omega^2) = \sum_n \frac{w_n}{w_n + \omega^2}, \tag{6.14}$$

with the limits

$$\lim_{\omega^2 \to 0} N_{\text{eff}} = \sum_n 1 = N_p,$$
$$\lim_{\omega^2 \to \infty} N_{\text{eff}} = \sum_n 0 = 0. \tag{6.15}$$

Thus, the effective number of parameters can be intuitively understood as the number of eigenmodes in the cost function that are not (significantly) affected by regularization.

6.2.2 Estimators of the generalization error

Given some regularization method, all we need to do is determine a suitable regularization strength, that is, to choose the desired model complexity. Now, what characterizes the optimum regularization strength? Since we anticipate to apply this machiney for development of semilocal exchange-correlation approximations for density functional theory, where we know *a priori* that the exact XC functional will not be available in the model space, avoiding overfitting such simple XC models is very important. We therefore wish to suppress the most sloppy model eigenmodes (which are harmful to transferability of the fit) without freezing out the strongest ones (which are essential to quantitative performance of the XC model). Determining the scale of ω^2's that satisfy these requirements for *model selection* is not trivial, and definite or analytic solutions to this problem does not exist. However, resampling methods developed for machine learning applications are useful tools for this problem. [137]

If the true model is known we may define the expected prediction error PE when the fit $\mu(x)$ attempts to estimate the data point $y_i = \mathcal{M}(x_i)$ by $\mu(x_i)$ in terms of the bias and variance of $\mu(x)$, [139]

$$\text{PE} = \sigma_\varepsilon^2 + \text{Bias}^2 + \text{Var} = \langle\langle\{\mu_D(x) - \mathcal{M}(x)\}^2\rangle_D\rangle_x, \tag{6.16}$$

where σ_ε^2 is an irreducible error due to noise on the data, and the right-hand expression is the deviation between the best-fit model for dataset D and the true model, averaged over datasets and data points. The squared bias quantitifies the deviation between the average best-fit model and the true model, while the variance is the expected squared deviation of the prediction from its mean. The variance is in general high for flexible models with many parameters, since these are able to adapt to different datasets representing the true model differently. On the other hand, the bias expresses the inability of an approximate model to represent the true model even for large datasets with low noise levels, and is in geneeral high for overly simple models. However, since we do not in general know $\mathcal{M}(x)$, we can not directly evaluate and minimize (6.16).

Assessment of the overall quality of a model is key in many machine learning problems, e.g., regression, pattern recognition, and classification. Efficient methods have therefore been developed for *estimating* the model PE on real-life datasets. This is a

well-documented scientific field in itself (e.g., Refs. 137,139,143,144), but only a small subset of the methods are of interest here.

The most simple estimator of the prediction error is the apparent error rate $\overline{\text{err}}$,

$$\overline{\text{err}} = \frac{1}{N_d} \sum_{i}^{N_d} \{\mu(x_i) - y_i\}^2, \tag{6.17}$$

which is the data-point average of the squared error for model μ, and similar to the least-squares loss function. It is well-known and easily understood that $\overline{\text{err}}$ tends to underestimate the true generalization error (the PE), since the data used for training the model is here also used for testing it. The apparent error is thus an overly optimistic estimate of the generalization error.

Bootstrapping is a prominent class of resampling methods that attempt to alleviate this problem. In particular, the bootstrap ".632 estimator" proposed by Efron in Ref. 145 is designed to provide a balanced representation of the bias–variance tradeoff and reliable estimates of the PE. The essential feature is random generation of new datasets from the original target data by drawing N_s samples of N_d data points from D while allowing for repetition. The bootstrap out-of-bag estimator $\overline{\text{Err}}$ is then defined as

$$\overline{\text{Err}} = \frac{1}{N_d} \sum_{i \in D} \frac{1}{N_i} \sum_{\widetilde{D} | i \notin \widetilde{D}} \{\mu_{\widetilde{D}}(x_i) - y_i\}^2, \tag{6.18}$$

where N_i is the number of bootstrap samples $\widetilde{D}$ that do not contain data point i. The parenthesis thus evaluates the squared error on model predictions only for models $\mu_{\widetilde{D}}$ which were not trained on target data i. However, $\overline{\text{Err}}$ is itself an overly pessimistic estimate of the generalization error, so the total .632 estimator is defined as a weighted average of $\overline{\text{err}}$ and $\overline{\text{Err}}$,

$$\text{EPE}^{\text{BS.632}} = \left(0.368 \cdot \overline{\text{err}} + 0.632 \cdot \overline{\text{Err}}\right)^{1/2}, \tag{6.19}$$

where the weights stem from considering the asymptotic ($N_d \to \infty$) probability $P \approx 1 - e^{-1} = 0.632$ that a data point will be included in a particular bootstrap sample.[139] The bootstrap .632 estimator (6.19) is used here to determine the optimum model complexity, that is, optimum model solutions with a minimum of overfitting, by minimizing $\text{EPE}^{\text{BS.632}}$ over the regularization strength.

6.3 Error estimation

The maximum-likelihood solutions considered here maximize the probabiliaty distribution for the model parameters in Bayes' theorem, which peaks at the best-fit solution $\boldsymbol{\alpha}$. The concept of an ensemble of model solutions around the optimum one is therefore intrinsic to the present approach to model parametrization. As the number of

target data points N_d increases, the ensemble contracts around $\boldsymbol{\alpha}$, and for $N_d \to \infty$ the ensemble spread vanishes, even though $M(\boldsymbol{\alpha})$ may be a poor model. The spread of the maximized posterior distribution can clearly not be used for inferring the actual errors made by the model when applied to predict new data.

Nevertheless, a model ensemble with error estimation capabilities may be rather useful, so we simply design one. Considering a model $\widetilde{M}$ with optimized parameters $\tilde{\boldsymbol{a}}$, we seek to define a probability distribution $\widetilde{P}(\boldsymbol{a})$ to be used to estimate errors for model $\widetilde{M}(\tilde{\boldsymbol{a}})$. If we consider a data point $y(x)$, this point has a deviation from the model given by $\Delta y(x) = \tilde{y}(x) - y(x)$. We would therefore like the distribution $\widetilde{P}(\boldsymbol{a})$ to exhibit a width of about this size for the prediction of $y(x)$. In other words, if we define $\delta y^{\boldsymbol{a}}(x) = y^{\boldsymbol{a}}(x) - \tilde{y}(x)$ for the deviation between the model given by $\boldsymbol{a}$ and $\tilde{\boldsymbol{a}}$, we would prefer if the distribution obeyed $\langle \delta y^{\boldsymbol{a}}(x)^2 \rangle = \Delta y(x)^2$, that is, the spread of the ensemble predictions should reproduce the actual error. This can not be obtained at every data point, but if we make this a requirement on average for all data points,

$$\sum_i \langle (\delta y_i^{\boldsymbol{a}})^2 \rangle = \sum_i \Delta y_i^2, \tag{6.20}$$

a probability distribution $\widetilde{P}_0$ for the model parameters in terms of the cost function may be derived,[136]

$$\widetilde{P}_0 \propto \exp\left(-C/T\right), \tag{6.21}$$

where the "temperature" T scales the ensemble fluctiations so that errors between model and target data are on average reproduced. The temperature may be expressed in terms of the minimized cost function and the effective number of parameters in the model,

$$T = 2 \sum_i \Delta y_i^2 / N_p = 2C_0/N_p, \tag{6.22}$$

which implies $C_0 = N_p T/2$. The latter form shows that each of the N_p harmonic degrees of freedom in the cost function contributes $T/2$ to the cost, as well-known from the equipartition theorem of statistical physics. Several more details may be found in Ref. 142.

6.4 Summary and a standard procedure

The basics of the approach to statistical modelling used in the following chapters for developing density functionals with error estimation ensembles have been introduced. Apart from the design of the ensemble, the whole methodology is aimed at tackling overfitting issues when parametrizing highly flexible models in manifestly incomplete model spaces. It is here appropriate to summarize the use of a regularized cost function, an estimator of the model generalization error, and scaling of the fluctuations in the Bayesian ensemble such as to enable error estimation, in a "standard procedure" for XC model fitting.

Given is the simple situation of a single dataset of electronic structure materials properties. With the aim of generating an DFA suitable for accurate but transferable DFT based predictions of the materials properties represented by the target data, as well as an error estimation ensemble, a highly flexible linear-expansion XC model space is employed. To minimize the risk of overfitting, which will most certainly be the result of parameter adjustment without regularization, an optimum model complexity must therefore be determined. The bootstrap .632 estimator is used for this in a procedure where the regularized cost function is minimized over a wide range (several orders of magnitude) of k different regularization strengths in the set $\{\omega_k^2\}$, yielding the set of model solutions $\{\boldsymbol{\alpha}_k\}$. The solution corresponding to the ω_k^2 that minimizes the estimator is identified as the "globally" optimum model $\boldsymbol{\alpha}_{\mathrm{opt}}$ in this multi-objective optimization problem of simultaneously minimizing model bias and variance. The cost function at $\boldsymbol{\alpha}_{\mathrm{opt}}$ and the effective number of parameters in the solution form the basis for generation of a Bayesian error estimation ensemble around $\boldsymbol{\alpha}_{\mathrm{opt}}$.

Chapter 7

A Semi-empirical Approach to Density Functional Development

The machine learning methods introduced in the previous chapter are here applied in a novel approach to semi-empirical density functional development. The goal is a general-purpose density functional for surface science and catalysis studies, which should accurately describe bond breaking and formation in chemistry, solid state physics, and surface chemistry, and should also account for van der Waals dispersion interactions. In this approach we consider XC functionals more or less accurate models of the exact density functional, and strive for computationally efficient but reliable XC models. From this point of view, the basis function expansion of XC model space, model selection procedure, and high-quality target dataset(s) representing the materials properties of main interest, are essential.

An error estimation ensemble of model solutions for the final XC model, named BEEF-vdW, is also generated. A normal DFT calculation does not provide any information about the uncertainty of the result from using an approximate DFA. One method to obtain an estimate of the uncertainty is performing several calculations using different functionals, and observe the variations in the prediction of the quantity of interest. An ensemble of functionals designed to provide this error estimate is a more systematic approach.

7.1 Pitfalls of fitting density functionals

The semi-empirical approach to DFA development was introduced in broad terms in Section 3.3. It has become a popular path to specialized XC functionals of high accuracy for the intended applications, and a wealth of semi-empirically adjusted functionals on nearly all levels of Jacob's ladder exist today.[70]

A few different methods of DFA parametrization are in general use. On the simplest level the parameters in existing DFAs, which may originally have been "nonempirically" chosen from physical arguments, are reoptimized towards specific applications. The exchange-modified vdW-DF variants in Section 5, as well as the revPBE GGA and many others, are examples of this. Linear combinations of existing DFAs, as in the B3LYP and PBE0 hybrids, and smooth switching between exchange functionals as in C09-vdW and the AM05 and HTBS[146] GGAs, are other simple approaches. These are often taken far further in more or less full expansions of the functional. The expansion coefficients in a suitably chosen basis function expansion of, e.g., the GGA exchange enhancement factor, are then optimized by model training on one or more benchmark datasets representing the target materials properties for the resulting XC model. Polynomial basis functions expressed in dimensionless electronic structure dependent variables are often used. Spin-dependent scaling of the LDA correlation energy density may be similarly expanded, and the total XC models may end up containing more than 40 parameters.[64,147] Some well-known DFAs generated this way include VSXC, the HCTH family, and the Minnesota functionals of the Truhlar group, which include the M06-L meta-GGA.[62,65] However, as the number of parameters increases, and the polynomial order in particular, this procedure entails an increasing risk of overfitting.

This was realized early on by Becke and others.[64,148] Using polynomial expansions of GGA exchange and correlation in least-squares fitting procedures, polynomial orders above 4 were found to yield increasingly oscillatory and unphysical XC functionals. Becke described this in Ref 64 as "a transition to mathematical nonsense". Nevertheless, semi-empirical DFAs containing many parameters continue to be constructed with only limited attention to the overfitting issue. Figure 7.1 illustrates a heavily overfitted GGA exchange enhancement factor. Though the XC model performs very well on the dataset it was fitted to (G3/99), we should expect the generalization error to be large when considering quantities very different from those included in the training dataset.

Transferability of a DFA parametrization depends not only on the degree of overfitting to a single set of molecular or condensed matter properties, but also on how many physically different properties the approximate model was trained on. A key issue is here model tradeoffs arising from a "competition" between vastly different datasets in determining the XC model solution. Such a competition is unavoidable when employing a rather incomplete XC model space, such as GGA, and the result is necessarily an XC *model compromise*. Implicitly acknowledging this, many semi-empirically oriented DFA developers assign more or less arbitrary weights to each dataset.[62,147,149] Here we take a more elaborate but less biased approach to choosing the optimum compromise.

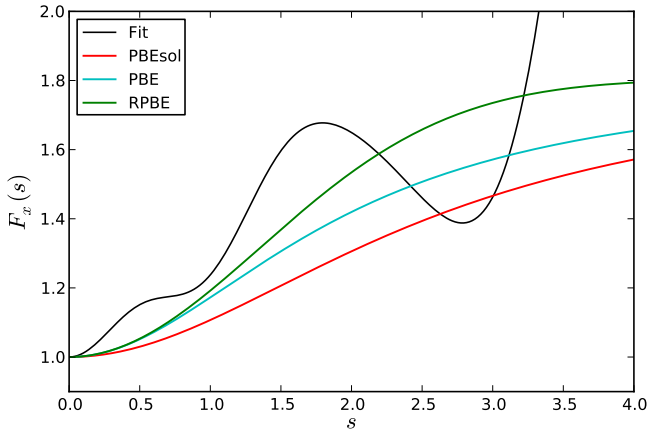

Figure 7.1: An overfitted GGA exchange enhancement factor (black). The model was least-squares optimized on the G3/99 molecular formation energies using a 10^{th}-order polynomial expansion forced to respect the HEG limit. Combined with PBE correlation this GGA is apparently very good for thermochemistry, yielding a mean absolute deviation of 0.16 eV on the G3/99 molecules, significantly lower than the 0.21 eV MAD for B3LYP,[46] a standard high-accuracy hybrid XC functional in quantum chemistry. But alas, the DFA is overfitted in the sense that we should not expect good transferability of this performance when considering quantities very different from molecular formation energies, or even formation energies of molecules not included in the G3/99 set.

7.2 GGA+vdW model space

Following the results of Section 5.4, where introduction of semilocal correlation contributions to the vdW-DF method was found useful, the XC model space to be employed here is GGA exchange and a linear combination of LDA and PBE correlation in conjunction with nonlocal correlation of the vdW-DF2 type.

As already reviewed in Section 3.2, the GGA exchange energy density $\varepsilon_x^{\text{GGA}}(n, \nabla n)$ is expressed in terms of the exchange energy density of the homogenous electron gas, $\varepsilon_x^{\text{HEG}}(n)$, and the exchange enhancement factor $F_x(s)$. The latter is (usually) a reasonably well-behaved function of the reduced density gradient s,

$$s = \frac{|\nabla n|}{2k_{\text{F}}n} \propto \frac{|\nabla n|}{n^{4/3}} \ , \ 0 \leq s \leq \infty, \tag{7.1}$$

and the total GGA exchange energy is written $E_x^{\text{GGA}}[n, \nabla n] = \int \varepsilon_x^{\text{HEG}}(n) F_x(s) n(\boldsymbol{r}) d\boldsymbol{r}$. A highly flexible exchange model space is obtained by expanding $F_x(s)$ in a basis of M_x Legendre polynomials $B_m(t(s))$ of orders 0 to $M_x - 1$ in a *transformed* reduced

density gradient $t(s)$,

$$\begin{aligned} t(s) &= \frac{2s^2}{q+s^2} - 1 \ , \ -1 \leq t \leq 1, \\ F_x(s) &= \sum_m a_m B_m(t(s)), \end{aligned} \tag{7.2}$$

where a_m are expansion coefficients and $q = 4$. The first few basis functions are illustrated in Fig. 7.2. Note that $t(s)$ is essentially a Padé approximant to the PBE $F_x(s)$. In fact, choosing $F_x(s) := 1.4 + 0.4 \cdot t(s)$ with $q := \kappa/\mu = 0.804/0.21951 \approx 3.663$ would quite accurately approximate the PBE $F_x(s)$, However, for simplicity, we choose $q = 4$ which does not make much difference, and note that due to the mapping $t(s)$, PBE exchange is nearly expandable in the two lowest-order polynomial basis functions $B_0(t(s))$ and $B_1(t(s))$.

The total GGA exchange energy is therefore expanded as

$$\begin{aligned} E_x^{\mathrm{GGA}}[n, \nabla n] &= \sum_m a_m \int \varepsilon_x^{\mathrm{UEG}}(n) B_m(t) n(\boldsymbol{r}) d\boldsymbol{r}, \\ &= \sum_m a_m \mathcal{E}_m[n, \nabla n], \end{aligned} \tag{7.3}$$

where $\mathcal{E}_m$ is the exchange energy corresponding to B_m.

Semilocal approximations to electron correlation effects beyond GGA exchange are not easily cast in terms of a single variable, such as s. In fact, the popular PBE GGA correlation approximation is far from simple, and even depends on the relative spin polarization. Meanwhile, correlation energy contributions to most materials properties

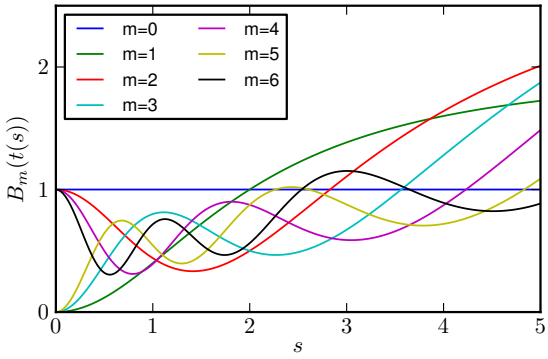

Figure 7.2: Legendre polynomial exchange basis functions $B_m(t(s))$ illustrated for polynomial orders 0 to 6. Though not explicitly illustrated, the PBE exchange enhancement factor is almost expanded as $F_x(s) := 1.4 \cdot B_0 + 0.4 \cdot B_1 = 1.4 + 0.4 \cdot t(s)$.

of interest here are significantly smaller than the exchange ones.[15] This motivates the use of the rather simple correlation model space given by

$$E_c[n, \nabla n] = \alpha_c E_c^{\mathrm{LDA}} + (1 - \alpha_c) E_c^{\mathrm{PBE}} + E_c^{\mathrm{nl}}, \tag{7.4}$$

where purely local Perdew–Wang[150] correlation is mixed with the PBE semilocal correlation through the parameter α_c, and the relative amount of contributions from E_c^{nl} is kept constant and equal to one.

In total, the XC model space consequently consists of GGA exchange expanded in Legendre polynomials in addition to local, semilocal, and nonlocal correlation,

$$E_{\mathrm{xc}} = \sum_{m=0}^{M_x-1} a_m \mathcal{E}_m + \alpha_c E_c^{\mathrm{LDA}} + (1 - \alpha_c) E_c^{\mathrm{PBE}} + E_c^{\mathrm{nl}}, \tag{7.5}$$

where $M_x = 30$, and the total number of parameters is $M = M_x + 1 = 31$.

Note that none of the commonly imposed constraints on GGA exchange are invoked, e.g., the HEG limit $F_x(0) = 1$ and recovery of the correct gradient expansion for slowly varying densities, nor the Lieb–Oxford bound $F_x(s \to \infty) = 1.804$ for large electron density gradients and/or small densities. However, the sum of LDA and PBE correlation is constrained to unity, and should preferably fulfill $0 \leq \alpha_c \leq 1$.

7.3 XC model selection and model compromises

Choices are obviously made when developing semi-empirical density functionals. These are both explicit and implicit choices pertaining to what the functional is to be designed for, that is, for the selection of an optimum exchange–correlation model that captures the materials properties of main interest when applying the approximation. We aim here to explicate some of these choices, and develop a set of principles for the model selection process. The main point is that an XC model compromise is inevitably faced when attempting to describe significantly different quantities in chemistry and condensed matter physics using an incomplete model space. We acknowldge this situation and attempt to deal with it. The alternative is overfitting the XC model, or generation of individual-purpose density functionals rather than general-purpose ones.

DFA development is in this approach divided into two steps. First an individual model selection for a number of datasets is carried out, and subsequently a simultaneous model selection is made, compromising between the individual fits.

7.3.1 Individual materials properties

Model training for individual datasets is conducted using the "standard procedure" of Section 6.4. That is, a regularized cost function in the model parameters is minimized

for given regularization strength ω^2. The model solution that minimizes the bootstrap .632 estimator of the generalization error is identified as the optimum solution, which should represent an appropriate bias–variance tradeoff and avoid overfitting the XC model.

7.3.1.1 Tikhonov regularization

The simple redge regression regularizer was introduced in Section 6.1.3. It is quadratic in the model parameters and simply punishes large solution vector coefficients, which is often sufficient to limit overfitting. However, experience tells us that fitting exchange approximations in the quite incomplete GGA model space requires special attention to the smoothness of the obtained solutions. Indeed, significantly non-smooth exchange solutions have been shown to degrade transferability of fitted exchange functionals to systems outside the training data, even though an EPE minimization was performed.[136] Such effects should clearly be avoided, and physically it is also very reasonable to require $F_x(s)$ to be a smooth and preferably injective function of s, after all, why should it *not* be so?

The Tikhonov regularization method[151] is therefore applied. This regularizer is also quadratic in the coefficient vector $\boldsymbol{a}$, but instead of penalizing according to $\|\boldsymbol{a}\|^2$, it penalizes non-smooth variations away from a suitably chosen prior vector $\boldsymbol{a}_p$. The Tikhonov regularizer we write as

$$R_{\mathrm{T}}(\boldsymbol{a}, \omega^2) = \omega^2 \boldsymbol{\Gamma}^2 \left(\boldsymbol{a} - \boldsymbol{a}_p\right)^2 , \tag{7.6}$$

where $\boldsymbol{\Gamma}$ is the Tikhonov matrix, and the parenthesis measures the difference between coefficients in $\boldsymbol{a}$ and the prior vector. The Tikhonov matrix is primarliy defined from a smoothness criterion on the exchange basis functions, even though it also contains entries for the correlation parameter α_c.

Consider first the exchange part of $\boldsymbol{\Gamma}$, which is defined from the overlap of the second derivative of the exchange basis functions with respect to $t(s)$,

$$\boldsymbol{\Gamma}^2_{ij} = \int_{-1}^{1} \frac{d^2 B_i(t)}{dt^2} \frac{d^2 B_j(t)}{dt^2} dt. \tag{7.7}$$

Note that this overlap is zero for the order 0 and 1 basis functions, which are therefore considered "infinitely smooth" and not affected by the Tikhonov regularizer. Instead, the integrated squared second derivative of the exchange fit is directly penalized for any finite regularization strength. This can be understood as penalizing a measure of non-smoothness of the fitted exchange enhancement factor. In effect, the $\boldsymbol{\Gamma}$ matrix scales the ω^2 acting on each basis function in (7.6). One subtle detail about $R_{\mathrm{T}}(\boldsymbol{a}, \omega^2)$ is particularly important to note: Since the exchange part of $\boldsymbol{\Gamma}$ acts in the transformed $t(s)$-space, the transformation (7.2) causes the regularization penalty on exchange to be strongest in the large-s regime, where information from the target data about the optimum behavior of $F_x(s)$ is expected to be scarce.[50,55] The correlation part of $\boldsymbol{\Gamma}$ has one in the diagonal and zeros in the off-diagonal elements.

7.3.1.2 Cost function and bootstrapping

The cost function to be minimized for the i'th dataset consists of the maximum-likelihood cost and the Tikhonov regularizer,

$$\begin{aligned} C_i(\boldsymbol{a}) &= \widetilde{C}_i(\boldsymbol{a}) + R_{\mathrm{T}}(\boldsymbol{a}, \omega^2) \\ &= (\boldsymbol{X}_i \boldsymbol{a} - \boldsymbol{y}_i)^2 + \omega^2 \boldsymbol{\Gamma}^2 (\boldsymbol{a} - \boldsymbol{a}_p)^2 , \end{aligned} \tag{7.8}$$

where $C_i(\boldsymbol{a}) = C_i(\boldsymbol{a}, \omega^2)$ for brevity. In accordance with (7.5), the data matrix $\boldsymbol{X}_i$ consists of XC contributions to a materials property for each system in the dataset from the M basis functions. These are evaluated non-self-consistently on RPBE electron densities. The target vector $\boldsymbol{y}_i$ contains the target XC contribution to each quantity in the set.

The role of the Tikhonov regularizer in this cost function is emphasized: It suppresses solution vector coefficients for higher-order exchange basis functions unless they are essential for obtaining a satisfactory fit. This leads to a model selection preference for solutions that are smooth variations away from the the prior vector $\boldsymbol{a}_p$, which is the XC model solution for $\omega^2 \to \infty$. The prior we therefore also call the origo for regularization. The exchange part of $\boldsymbol{a}_p$ is chosen as $1.4 \cdot B_0(t) + 0.4 \cdot B_1(t)$, while the prior for correlation is $\alpha_c = 0.75$.

In order to minimize the cost function (7.8) it is transformed by $\boldsymbol{\Gamma}^{-1}$. Ones are therefore inserted in the first two diagonal elements of $\boldsymbol{\Gamma}$ to avoid numerical issues. The solution vector $\boldsymbol{a}_i$ that minimizes C_i is found from

$$\boldsymbol{a}_i = \boldsymbol{\Gamma}^{-1} \left(\boldsymbol{X}_i'^T \boldsymbol{X}_i' + \boldsymbol{L}^2 \omega^2 \right)^{-1} \left(\boldsymbol{X}_i'^T \boldsymbol{y}_i + \omega^2 \boldsymbol{L}^2 \boldsymbol{a}_p' \right), \tag{7.9}$$

where $\boldsymbol{X}_i' = \boldsymbol{X}_i \boldsymbol{\Gamma}^{-1}$, $\boldsymbol{a}_p' = \boldsymbol{\Gamma} \boldsymbol{a}_p$, and $\boldsymbol{L}^2$ is the identity matrix with zeros in the first two diagonal elements. Well-behaved correlation solutions are enforced by requiring $0 \leq \alpha_c \leq 1$.

The bootstrap .632 procedure introdued in Section 6.2.2 is conducted using 500 randomly generated target data samples for each ω^2. The regularization strength that minimizes the estimator is found by a smooth fitting of the slightly scattered estimator plot near its minimum. To properly regularize the S22x5 subsets with long interaction distances, a condition $F_x(s = \infty) \geq 1$ is enforced.

7.3.1.3 Individually trained XC models

Table 7.1 and Fig. 7.3 show details and statistics for XC models obtained when the model selection procedure outlined above is applied to molecular, solid state, surface chemical, and vdW dominated energetics as represented by the CE17, RE42,

Table 7.1: Model selection results of individually training the XC model in (7.5) to 10 different datasets. M_{eff} is the effective number of parameters in a model, see (7.23). The $s = 0$ and $s \to \infty$ limits of the obtained exchange enhancement factors are also shown. MSD, MAD and STD are mean signed, mean absolute, and standard deviation, respectively, all in meV. Note that these are non-selfconsistent results.

	α_c	M_{eff}	$F_x(0)$	$F_x(\infty)$	MSD	MAD	STD
CE17	0.90	4.7	0.97	2.15	−10	96	116
RE42	1.00	4.2	1.06	1.21	19	168	207
DBH24/08	0.00	3.7	1.14	3.14	1	116	142
G2/97	0.27	7.2	1.10	2.53	−13	109	149
Sol34Ec	0.00	7.7	0.97	1.25	−4	168	208
S22x5-0.9	0.81	3.2	0.96	1.68	0	9	11
S22x5-1.0	0.82	3.1	0.98	1.87	0	8	10
S22x5-1.2	0.40	5.7	1.04	2.38	0	4	6
S22x5-1.5	0.85	4.0	1.02	1.91	−1	3	4
S22x5-2.0	1.00	3.3	0.95	1.37	2	3	3

DBH24/08, G2/97, and Sol34Ec datasets, as well as the five S22x5 subsets (all introduced in Section 4.1). Each model is therefore trained on a single materials property only, and their features differ significantly.

The DBH24/08 set appears to favor GGA exchange that substantially violates the LDA limit ($F_x(0) = 1.14$) along with inclusion of full PBE correlation ($\alpha_c = 0$). The model furthermore overshoots the LO bound significantly ($F_x(\infty) = 3.14$). The XC

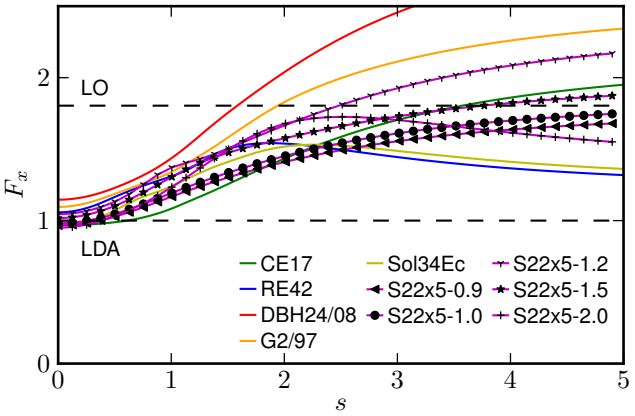

Figure 7.3: Exchange enhancement factors of the individually trained XC models listed in Table 7.1.

model optimized to the G2/97 set shows similar trends with respect to GGA exchange and PBE correlation, but is less extreme.

In the other end of the spectrum is the model optimized to the Sol34Ec cohesive energies. These favor GGA exchange starting out slightly below $F_x = 1$, then reaching a maximum at $s \approx 2$, and finally declining slowly towards $F_x = 1.25$. Best agreement with experimental cohesive energies is found with full PBE correlation in addition to nonlocal correlation. The occurrence of a maximum in the exchange enhancement factor should, however, not be overemphasized. It has been shown [50,55] that only small GGA exchange contributions to chemical and solid state binding energetics can be attributed to reduced density gradients above 2.5. In the region of large s, where the smoothness criterion on exchange is strongly enforced, the regularization term in the cost function (7.8) will therefore be dominant in determining the solution for such systems. The regularizer may therefore well determine the behavior of $F_x(s)$ in the large-s regime.

For the remaining datasets in Table 7.1, the optimized XC models appear reasonable, with all exchange enhancement factors starting out near the LDA limit and intermediate correlation mixing parameters.

It is illustrative to investigate how the XC models perform for datasets on which they were not trained. The standard deviation, here denoted STD, is a natural measure

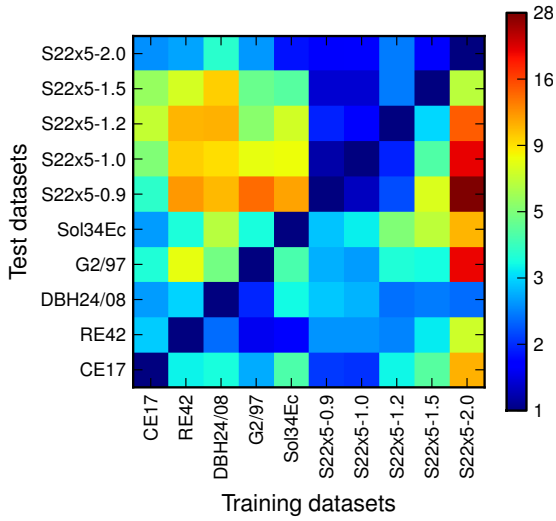

Figure 7.4: Color map of the relative standard deviations obtained when non-self-consistently applying the XC models found individually for each training dataset, listed on the abscissa, to all 10 training datasets along the ordinate.

of performance, and should be understood as the root-mean-square deviation defined in (7.25) on p. 83. We also define the relative standard deviation rSTD on some dataset with some XC model as the STD obtained by that model relative to the STD of the model that was fitted to that dataset. The rSTD is thus a measure of model transferability. Figure 7.4 shows a color map of the rSTD for all 10 XC models evaluated on all 10 training datasets. The diagonal from bottom left to top right is, by definition, ones. In a background of blue and yellow-green squares, the map features two distinct areas of mostly reddish squares. To the far right, the S22x5-2.0 model yields rSTD > 5 for all other sets than DBH24/08, and rSTD ≈ 28 for S22x5-0.9. Furthermore, a 5×4 square in the top left corner illustrates that XC models trained on chemical or solid state datasets perform significantly worse on vdW type energetics, than models fitted to the latter. It is also interesting to see that the S22x5-2.0 rSTDs are largely unaffected by changing XC models (top horizontal row). With little density–density overlap between many of the S22x5-2.0 complexes, the constant nonlocal correlation in all 10 models is likely the main XC contribution to intermolecular binding.

In summary, the deviation statistics in Table 7.1 illustrate that the XC model space considered here most certainly spans the model degrees of freedom necessary to obtain well-performing density functionals with smooth exchange enhancement factors and sound correlation components. However, Fig. 7.3 clearly illustrates that different materials properties are best captured by significantly different XC models in this incomplete model space, and a high degree of transferability between the datasets should not be expected for several of the models.

7.3.2 Several materials properties

Fundamentally, a compromise has to be made between how well different materials properties are reproduced by the same semi-empirical density functional on the lower rungs of Jacob's ladder. This is expressed as a compromise between how well the functional quantitatively performs on different training datasets. What the compromise should be can only be determined by the target applications of the functional, and one challenge is to make this choice as explicit as possible. This section presents one route towards a methodology for optimizing an XC model to simultaneously describe several different materials properties.

7.3.2.1 Characteristics of the model compromise

We first establish aspects of the XC model compromise by considering the problem of simultaneously fitting two datasets. The obvious approach is to let the compromise be described by a total cost function $\Lambda(\boldsymbol{a})$, given as the sum of the two individual cost functions,

$$\Lambda(\boldsymbol{a}) = \mathcal{W}_1 C_1(\boldsymbol{a}) + \mathcal{W}_2 C_2(\boldsymbol{a}), \tag{7.10}$$

where $\mathcal{W}_i$ is the weight on dataset i.

The model solution $\boldsymbol{b}$ that minimizes $\Lambda(\boldsymbol{a})$ is found by setting the derivative to zero: Since the summed cost function is quadratic in $\boldsymbol{a}$, as the individual cost functions C_i are, we may express it in terms of the individual solutions $\boldsymbol{a}_i$ as

$$\Lambda(\boldsymbol{a}) = \sum_{i=1,2} \mathcal{W}_i \left(C_i^0 + \frac{1}{2}(\boldsymbol{a} - \boldsymbol{a}_i)^T \boldsymbol{H}_i (\boldsymbol{a} - \boldsymbol{a}_i) \right), \tag{7.11}$$

where $C_i^0 = C_i(\boldsymbol{a}_i)$ is the minimized cost of dataset i, and $\boldsymbol{H}_i$ is the Hessian of $C_i(\boldsymbol{a})$, see Section 6.2.1. The minimizing solution $\boldsymbol{b}$ may then be found from the individual solutions $\boldsymbol{a}_i$ as

$$\boldsymbol{b} = \left(\sum_{i=1,2} \mathcal{W}_i \boldsymbol{H}_i \right)^{-1} \left(\sum_{i=1,2} \mathcal{W}_i \boldsymbol{H}_i \boldsymbol{a}_i \right). \tag{7.12}$$

Now, the regularized cost functions for each training dataset, $C_i(\boldsymbol{a})$, contain information about the costs associated with deviating from $\boldsymbol{a}_i$ along all directions in model space. The individual costs all increase when moving away from $\boldsymbol{a}_i$ due to deterioration of the fits, increased overfitting, or a combination of both. Define now the relative cost for each dataset, denoted rCost$[\,i\,]$, as the individual cost for set i evaluated at the compromising solution $\boldsymbol{b}$ relative to the individual cost at $\boldsymbol{a}_i$, hence

$$\mathrm{rCost}[\,i\,] = \frac{C_i(\boldsymbol{b})}{C_i(\boldsymbol{a}_i)} = \frac{C_i(\boldsymbol{b})}{C_i^0} \geq 1. \tag{7.13}$$

Thus defined, the relative cost for each training dataset is a simple measure of how unfavorable it is for the set to be fitted by the compromising solution $\boldsymbol{b}$ instead of $\boldsymbol{a}_i$.

The main panel of Fig. 7.5 illustrates XC model compromises between the G2/97 and S22x5-1.0 datasets. The curve maps out the relative costs on both materials properties obtained from model solutions $\boldsymbol{b}$ when systematically varying the weights in $\Lambda(\boldsymbol{a})$. The weight fraction f is introduced, see the figure caption. Note also the different scaling of the figure axes. A wide range of poorly compromising models can obviously be produced, sacrificing a lot of relative cost on one set while gaining little on the other. However, if both materials properties represented by the datasets are considered important, the optimum (or least inferior) compromise is somewhere midway between the asymptotic extrema. The unfortunate conclusion must be that the present GGA+vdW model space does not offer the possibility of very high accuracy for the G2/97 formation energies and the S22x5-1.0 interactions energies at the same time, when we demand smooth exchange enhancement factors and sound correlation components (not even with exchange models that deviate from the LDA limit).

7.3.2.2 Least inferior compromise

We are thus faced with the challenge of choosing the least inferior compromise possible given a number of different training datasets. Extending the approach of Fig. 7.5 to

several datasets and look for candidate optima in a multi-dimensional space of relative costs is clearly not feasible in general, nor a particularly unbiased methododology. Switching our representation of XC model compromises to a space of relative standard deviations offer no improvements; it is still a multi-objective optimization problem.

The insert in Fig. 7.5 directs us to a viable and efficient approach. The product of the relative costs varies with f in a particular, but intuitively reasonable, manner: To the right along the abscissa, where the fraction increasingly favors G2/97, the rCost product increases rapidly. These model solutions are seriously disfavored by the S22x5-1.0 set, which is also seen in the color map, Fig. 7.4. To the left along the abscissa in the Fig. 7.5 inset, the increase of the rCost product is much smaller, but a minimum is located in between. At least one intermediate minimum is always present, since the slopes in the two asymptotic regions in the main panel are $-\infty$ and 0, respectively. This property is induced by the variational property around the two original minima of the individual cost functions. Similar conclusions apply to any combination of two or more datasets that do not favor the same directions in the incomplete model space.

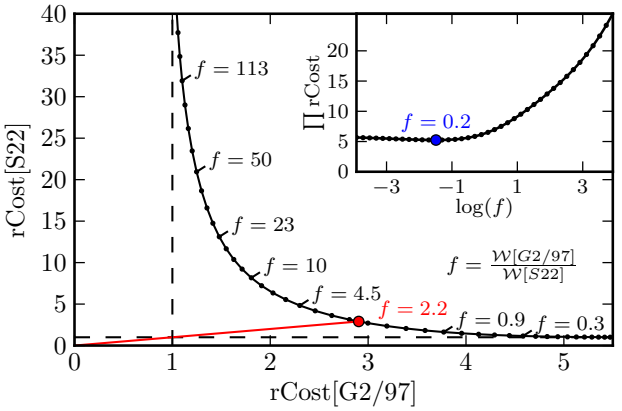

Figure 7.5: Main panel: XC model compromises between the G2/97 and S22x5-1.0 datasets illustrated in terms of relative costs (rCost) for both sets when the weight fraction $f = \mathcal{W}[\text{G2/97}]/\mathcal{W}[\text{S22x5-1.0}]$ is varied and the summed cost function (7.10) minimized. A range a different solutions are obtained, many of which are essentially fitting one dataset only (rCost ≈ 1) while sacrificing model performance on the other (rCost $\gg 1$). A red dot marks the point of equal rCost. The fact that an XC model with rCost[G2/97] = rCost[S22x5-1.0] = 1 is not obtainable illustrates the neccessity of a model compromise. Insert: The product of relative costs display a minimum (blue dot) for a certain weight fraction.

7.3.2.3 Minimizing the product of relative costs

We find in general that the condition of minimizing the product of relative costs is well-suited for choosing cost function weights for arbitrary numbers of training sets, if the aim is a general-purpose XC model. This condition, which is identical to minimizing the product of costs, is applied henceforth by minimizing a *product cost function* $\Phi(\boldsymbol{a})$,

$$\Phi(\boldsymbol{a}) = \prod_i C_i(\boldsymbol{a})^{w_i}, \tag{7.14}$$

where w_i is a constant weight, and C_i is again an individual cost function. The minimizing solution $\boldsymbol{c}$ is the desired model solution for arbitrary numbers of training datasets. The constant weight w_i is an important feature of $\Phi(\boldsymbol{a})$, since it allows inclusion of training sets which are perceived significantly less important than others. It is thus chosen from personal preferences given the purpose of the functional, and we shall see that $\boldsymbol{c}$ minimizes the product of costs *given* this choice.

For the case of two datasets, the stationary point between the two individual solutions in model space is found by differentiating the logarithm of $\Phi(\boldsymbol{a})$ with respect to $\boldsymbol{a}$, and solving

$$\sum_i \frac{w_i}{C_i}\frac{dC_i}{d\boldsymbol{a}} = 0. \tag{7.15}$$

Using the method outlined above for $\Lambda(\boldsymbol{a})$, the solution vector that minimizes $\Phi(\boldsymbol{a})$ is found in terms of the individual solutions as

$$\boldsymbol{c} = \left(\sum_i \frac{w_i}{C_i}\boldsymbol{H}_i\right)^{-1}\left(\sum_i \frac{w_i}{C_i}\boldsymbol{H}_i\boldsymbol{a}_i\right), \tag{7.16}$$

where $C_i = C_i(\boldsymbol{c})$, and w_i simply scales the individual costs. We see that this solution corresponds to $\mathcal{W}_i := w_i/C_i$ in (7.11) and (7.12). Thus, minimizing the product of costs has introduced a natural weight C_i^{-1}, while w_i still leave room for deliberately biasing the model solution.

From here on, the product solution is therefore used to find the desired XC model solution: Since C_i is evaluated at $\boldsymbol{c}$, the latter is found iteratively, using C_i^{-1} as an iterator while searching for a converged minimum of the product cost function, given the constant weights.[152]

7.3.2.4 BEEF-vdW density functional

The BEEF-vdW exchange–correlation functional was designed using the set of weights w listed in Table 7.2. In principle these should all equal one, however, correlations between some of the datasets have led us to lower the constant weight on some of them: Since the RE42 set is based on G2/97 molecules, the data in RE42 is correlated

Table 7.2: The BEEF-vdW model compromise. The effective weight in determining the XC model solution is w/C for each dataset, as iteratively found from minimizing the product cost function (7.14). The relative standard deviation (rSTD) is the ratio of the STD at the BEEF-vdW compromise to the STD at the regularized individual solution in Table 7.1. The relative cost (rCost) are defined similarly, but includes regularization, see Eq. (7.13).

	w	w/C	rCost	rSTD
CE17	1.0	1.80	1.7	1.3
RE42	0.5	0.62	2.5	1.8
DBH24/08	1.0	0.65	4.9	2.3
G2/97	0.5	0.62	2.6	1.6
Sol34Ec	1.0	0.43	7.5	2.8
S22x5-0.9	0.1	0.01	28.6	5.4
S22x5-1.0	0.1	0.04	9.1	2.9
S22x5-1.2	0.1	0.09	3.5	2.1
S22x5-1.5	0.1	0.08	4.1	2.1
S22x5-2.0	0.1	0.18	1.8	1.5

with some of the data in G2/97. Both weights were therefore lowered to 0.5. The same reasoning applies to the S22x5 subsets, where the same complexes are found in all five subsets, albeit at different interaction distances. A weight of $1/5 = 0.2$ on each S22x5 subset would therefore be natural, but for reasons of performance of the final functional, constant weights of 0.1 were chosen.

The resulting model compromise is also tabulated in Table 7.2, showing the effective

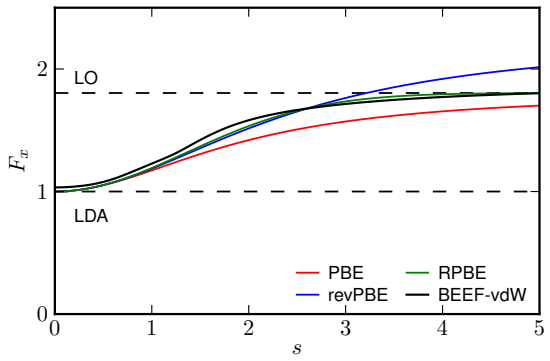

Figure 7.6: The BEEF-vdW exchange enhancement factor compared to those of a few standard GGA exchange functionals. The corresponding BEEF-vdW correlation functional is composed of 0.6 LDA, 0.4 PBE, and 1.0 nonlocal correlation.

Table 7.3: Expansion coefficients a_m for the BEEF-vdW Legendre exchange basis functions B_m. The correlation mixing parameter, α_c in (7.5), is 0.6001664769.

m	a_m	m	a_m
0	1.516501714	15	$-8.018718848\times10^{-4}$
1	4.413532099×10^{-1}	16	$-6.688078723\times10^{-4}$
2	$-9.182135241\times10^{-2}$	17	1.030936331×10^{-3}
3	$-2.352754331\times10^{-2}$	18	$-3.673838660\times10^{-4}$
4	3.418828455×10^{-2}	19	$-4.213635394\times10^{-4}$
5	2.411870076×10^{-3}	20	5.761607992×10^{-4}
6	$-1.416381352\times10^{-2}$	21	$-8.346503735\times10^{-5}$
7	6.975895581×10^{-4}	22	$-4.458447585\times10^{-4}$
8	9.859205137×10^{-3}	23	4.601290092×10^{-4}
9	$-6.737855051\times10^{-3}$	24	$-5.231775398\times10^{-6}$
10	$-1.573330824\times10^{-3}$	25	$-4.239570471\times10^{-4}$
11	5.036146253×10^{-3}	26	3.750190679×10^{-4}
12	$-2.569472453\times10^{-3}$	27	2.114938125×10^{-5}
13	$-9.874953976\times10^{-4}$	28	$-1.904911565\times10^{-4}$
14	2.033722895×10^{-3}	29	7.384362421×10^{-5}

dataset weight w/C, rCost, and rSTD for all datasets used in model training. It is clearly seen that especially the S22x5-0.9 interaction energies are hard to fit simultaneously with the other sets within the XC model space employed here: The relative cost for the set is high, allowing the model to adapt mostly to the other datasets by lowering w/C for this set. This is furthermore reflected in the rSTD of 5.4, indicating that the BEEF-vdW performance on this dataset is significantly worse than obtained in the individual fit to the S22x5-0.9 complexes reported in Table 7.1. Even so, the remaining S22x5 subsets appear to share XC model space with the datasets representing formation and rupture of interatomic bonds to a significantly greater extent. Therefore, even a reasonably accurate description of the balance of strong and weak interactions in the S22x5-0.9 complexes is nearly incompatible with at least one of the other sets of materials properties, when demanding well-behaved DFAs in the present model space.

Table 7.3 lists the BEEF-vdW expansion coefficients. The suppression of high-order solution coefficients by the Tikhonov regularizer is clearly seen. The correlation functional consists of 0.6 LDA, 0.4 PBE, and 1.0 nonlocal correlation. The qualitative shape of the BEEF-vdW exchange enhancement factor is shown in Fig. 7.6, with $s = 0$ and $s \to \infty$ limits of 1.034 and 1.870, respectively. Thus, BEEF-vdW exchange does not exactly comply with the LDA limit for $s = 0$, but is 3.4% higher. The enhancement factor is above most GGA exchange functionals up to $s \approx 2.5$, from where it approaches the LO bound with a small overshoot in the infinite limit.

The lack of exact fulfillment of the LDA limit for exchange indicates a conflict between this limit, the training data, and the employed preference for smooth exchange models.

The G2/97 and DBH24/08 chemical datasets are found to give particular preference to exchange models with $F_x(0) \approx 1.1$, and enforcing $F_x(0) = 1.0$ for these sets leads to severely non-smooth exchange solutions for $s \to 0$. Similar behavior was found in Ref. 136. Note that MGGA approximations are able to achieve exchange models with $F_x(0) \neq 1.0$ for densities different from the HEG, while stille obeying the LDA limit for HEG-like densities. The BEEF-vdW $F_x(s)$ also has small "bump" at $s \approx 1.3$. This is not essential to the quality of the model and is not expected to harm its transferability. However, completely removing such features require overly strong regularization.

7.4 Bayesian error estimation enesemble

As outlined in Section 6.3, the concept of an error estimation ensemble of models is fundamental to the present approach to model optimization. The construction of such an ensemble for the BEEF-vdW model is here presented. Note that a very similar approach was employed in Ref. 153 for the purpose of generating ensembles for rather simple, but not necessarily optimum, GGA exchange–correlation approximations.

In summary, the object of primary interest is the probability distribution P for the model parameters $\boldsymbol{a}$ given the model θ and training data D,

$$P(\boldsymbol{a}|\theta D) \sim \exp(-\Phi(\boldsymbol{a})/\tau), \tag{7.17}$$

where $\Phi(\boldsymbol{a})$ is the product cost function of (7.14), and τ is the ensemble temperature. Given D, a model perturbation $\delta\boldsymbol{a}$ has a certain probability associated with it, and this defines an ensemble of different XC functionals around BEEF-vdW. The temperature is to be chosen such that the spread of the ensemble model predictions of the training data reproduces the errors observed when using BEEF-vdW self-consistently.

Ensemble matrix

As in Section 6.3, the ensemble is defined through the cost function Hessian scaled with the temperature, yielding the ensemble matrix $\boldsymbol{\Omega}^{-1}$,

$$\boldsymbol{\Omega}^{-1} = \tau\, \boldsymbol{H}^{-1}, \tag{7.18}$$

with eigenvalues $\boldsymbol{w}^2_{\Omega^{-1}}$ and eigenvectors $\boldsymbol{V}_{\Omega^{-1}}$. Using an ensemble of k vectors $\boldsymbol{v}_k$, each of length M with elements randomly drawn from a normal distribution with zero mean and variance one, the BEEF-vdW ensemble coefficient vectors $\boldsymbol{a}_k$ are then computed according to

$$\boldsymbol{a}_k = \boldsymbol{V}_{\Omega^{-1}} \cdot \mathbf{1}\boldsymbol{w}_{\Omega^{-1}} \cdot \boldsymbol{v}_k. \tag{7.19}$$

However, the temperature and the Hessian are needed.

Hessian The BEEF-vdW Hessian is calculated directly from

$$\boldsymbol{H} = 2\sum_{i}^{N} \frac{w_i}{C_i(\boldsymbol{a}_p)} \boldsymbol{\Gamma}^{-1} \left(\boldsymbol{X}_i'^{T} \boldsymbol{X}_i' + \omega_i^2 \boldsymbol{L}^2 \right) \boldsymbol{\Gamma}^{-1^T}, \tag{7.20}$$

where the sum is over training datasets and ω_i^2 is the regularization strength that minimized the individual cost C_i.

Temperature and effective number of parameters As in (6.22) the temperature is related to the effective number of parameters in the model, here calculated from the effective regularization ω_{eff}^2,

$$\omega_{\text{eff}}^2 = \sum_{i}^{N} \frac{w_i}{C_i(\boldsymbol{c})} \omega_i^2. \tag{7.21}$$

Additionally, diagonalization of the combined square of the transformed data matrix

$$\boldsymbol{\Sigma}' = \boldsymbol{V}^T \left(\sum_{i}^{N} \frac{w_i}{C_i(\boldsymbol{c})} \boldsymbol{X}_i'^{T} \boldsymbol{X}_i' \right) \boldsymbol{V}, \tag{7.22}$$

where $\boldsymbol{\Sigma}'$ contains the eigenvalues along the diagonal and $\boldsymbol{V}$ the eigenvectors, allows the effective number of parameters left in the model after regularization, M_{eff}, to be computed as

$$M_{\text{eff}} = \sum_{m}^{M} \frac{{\Sigma_m'}^2}{{\Sigma_m'}^2 + \omega_{\text{eff}}^2 L_m^2}. \tag{7.23}$$

Since $M_{\text{eff}} = 7.11$ in the BEEF-vdW model compromise, more than 75% of the initially 31 model degrees of freedom have been suppressed by regularization.

The temperature calculation itself is slightly modified from that in (6.22) and Ref. 153 in order to construct an unbiased error estimation. This reflects that a larger error is expected when BEEF-vdW is applied to systems not included in the training datasets. The temperature is therefore calculated as

$$\tau = 2\frac{C(\boldsymbol{c})}{M_{\text{eff}}} \cdot \frac{N}{N - M_{\text{eff}}}, \tag{7.24}$$

where N is the total number of systems in the training sets. The second term is close to unity since $N \gg M_{\text{eff}}$.

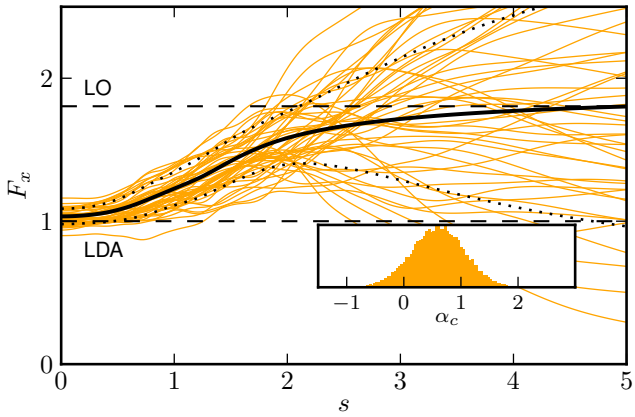

Figure 7.7: Bayesian ensemble of XC functionals around BEEF-vdW. Main panel: Black solid line is the BEEF-vdW exchange enhancement factor, while the orange lines depict $F_x(s)$ for 50 samples of the randomly generated ensemble. Dashed black lines mark the exchange model pertubations that yield DFT results ± 1 standard deviation away from BEEF-vdW results. Insert: Histogram of the distribution of correlation parameters in an ensemble containing 20,000 samples. The distribution is centered around $\alpha_c = 0.6$.

Application

An illustration of the resulting BEEF-vdW ensemble is shown in Fig. 7.7. For each data point in each dataset this ensemble may be applied non-self-consistently to BEEF-vdW electron densities. The spread of the ensemble predictions of a quantity is then the ensemble estimate of the BEEF-vdW standard deviation on that quantity. Naturally, the exchange enhancement ensemble expands after $s \approx 2$, where most of the chemistry and solid state physics have already happened.[50,55]

The predictive performance of the Bayesian ensemble has been evaluated using 20 000 ensemble functionals. In practice, however, far fewer draws from the ensemble distribution suffice for well-converged error estimates at a neglible computational overhead. Estimated standard deviations on the training datasets are compared to those from self-consistent calculations in Table 7.4. The ensemble performance on the dataset level should be assessed in combination with observing the error predictions on a system-to-system basis. Figure 7.8 illustrates the BEEF-vdW ensemble error estimates for the RE42 reaction energies, and compares BEEF-vdW results to those of other functionals. Similar figures for more datasets are found in the Supplemental Material (SM) for Paper II (the SM is not included here).

On the dataset level, the overall predictive performance of the Bayesian ensemble is satisfactory. The ensemble errors in Table 7.4 are slightly overestimated for the CE17,

Table 7.4: Comparison of self-consistent BEEF-vdW standard deviations to those predicted by the ensemble of functionals around BEEF-vdW. All energies in meV.

	BEEF-vdW	Ensemble estimate
CE17	143	209
RE42	372	253
DBH24	331	144
G2/97	242	312
SolEc34	576	436
s22x5-0.9	171	197
s22x5-1.0	94	181
s22x5-1.2	36	137
s22x5-1.5	8	67
s22x5-2.0	5	18

G2/97, and S22x5-0.9 datasets, while the ensemble underestimates the errors for RE42, DBH24/08, and Sol34Ec. For the remaining S22x5 subsets the error estimates are too large.

Importantly, Fig. 7.8 illustrates strengths and weaknesses of the present approach to error estimation. Many of the reaction energies are accurately reproduced by BEEF-vdW and the ensemble estimates a relatively small error on those data. However, some of the reactions for which BEEF-vdW yields larger errors are assigned too small error bars. The water-gas shift reaction $CO+H_2O \rightarrow CO_2+H_2$ is one of these. The reason for this is indicated by the fact that all tested GGA, MGGA and vdW-DF type functionals yield nearly identical reaction energies for this reaction. One simply has to move rather far in XC model space to find a functional that predicts a reaction energy significantly different from the BEEF-vdW result. This causes the ensemble to underestimate the actual error for that reaction. Since the hybrid functionals appear to break the overall trends observed for the lower-rung functionals in Fig. 7.8, inclusion of exact exchange in the model space might remedy such limitations of the BEEF-vdW functional and its Bayesian ensemble.

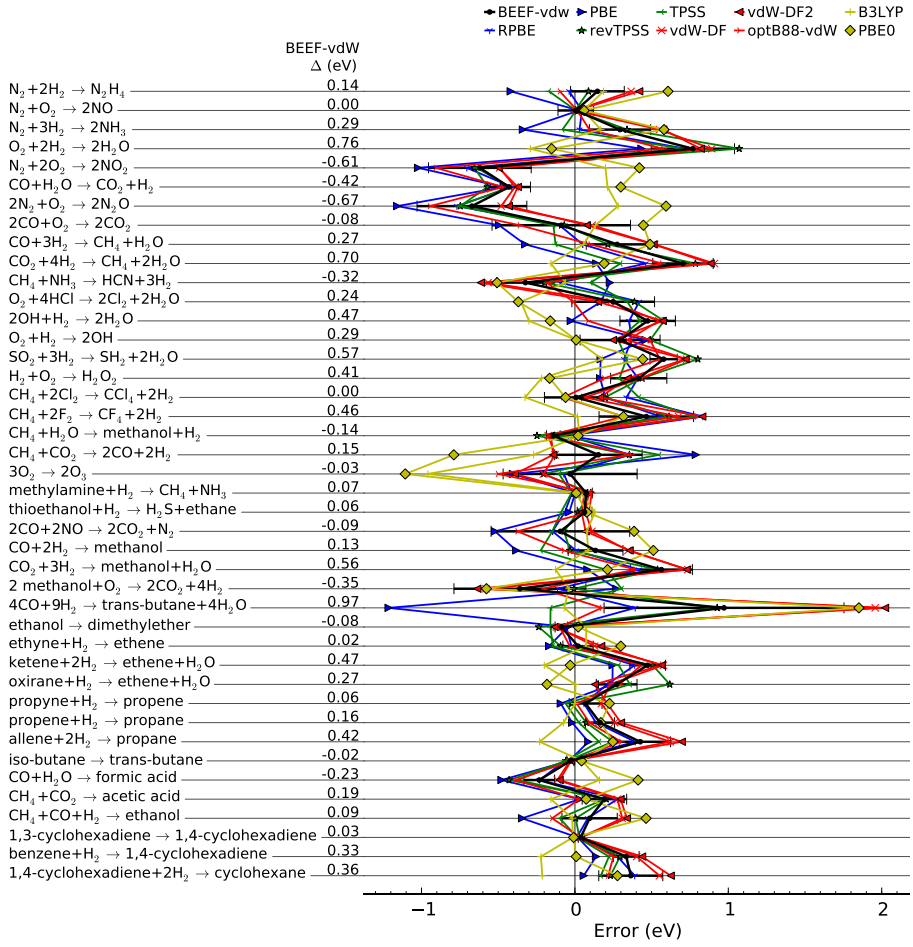

Figure 7.8: Deviations $\Delta = \Delta_r E^{\mathrm{DFT}} - \Delta_r E^{\mathrm{exp}}$ between the RE42 molecular reaction energies calculated using representative XC functionals and experiment. Color codes are BEEF-vdW: black, GGA: blue, MGGA: green, vdW-DF type: red, and hybrid: yellow. BEEF-vdW ensemble error estimates are indicated by horizontal error bars. The numbers are the self-consistent BEEF-vdW deviations from experiment.

7.5 Benchmark study

The following is a comparative assessment of BEEF-vdW and a selection of literature XC functionals of the LDA, GGA, MGGA, vdW-DF, and hybrid types. These are listed in Table 7.5 along with a simple characterization. The vdW-DF variants are here assigned to rung 3.5 on Jacob's ladder. In terms of computational complexity, the vdW-DF method contains fully nonlocal density–density information without explicit use of the Kohn–Sham orbitals. From this point of view it should fit between rungs 3 and 4. Note that nonlocal exchange approximations, designed to partially mimic exact exchange at a reduced computational cost, have recently been proposed in Refs. 154 and 155 as belonging to a rung 3.5.

The benchmark datasets used are partly the sets on which BEEF-vdW was trained, and partly other high-quality data. This totals to the G3/99, RE42, CE27, DBH24/08, S22x5, Sol27Ec, and Sol27LC datasets. The G3/99 extension of G2/97 and both Sol27 sets of solid state materials properties were introduced in Section 4.1. Note that hybrid functionals have not been applied to the extended systems in CE27, Sol27Ec, and Sol27LC.

Statistics on deviations of computed quantities from experimental or high-level theoretical references, Δx, are evaluated for each density functional in terms of the mean signed (MSD), mean absolute (MAD), and root-mean-square deviation (RMSD, standard deviation):

$$\begin{aligned} \Delta x &= \text{DFT} - \text{reference}, \\ \text{MSD} &= \frac{1}{N}\sum_i^N \Delta x_i, \\ \text{MAD} &= \frac{1}{N}\sum_i^N |\Delta x_i|, \\ \text{RMSD} &= \left(\frac{1}{N}\sum_i^N (\Delta x_i)^2\right)^{1/2}, \end{aligned} \tag{7.25}$$

which may have relative counterparts (MSRD, MARD, RMSRD) defined by the relative deviation $\Delta_\text{r} x$,

$$\Delta_\text{r} x = \frac{\Delta x}{|\text{reference}|}, \tag{7.26}$$

which conserves the sign on the error. Most of the statistics are tabulated in Tables A.5–A.8 in Appendix A.2 (pages 113–116).

All data are futhermore available online in the Computational Materials Repository (CMR).[158] The repository contains all information about the individual DFT calculations which form the basis for the results presented here, including atomic configurations and GPAW specific parameters. Access to search, browse, and download these data is provided through the CMR web-interface.[159]

Table 7.5: A selection of density functionals at the LDA (1), GGA (2), MGGA (3), vdW-DF (3.5), and hybrid (4) rungs of Jacob's ladder.

	Type	Target[a]	Ref.
LDA	1	–	150
PBE	2	general	48
RPBE	2	chemistry	55
BLYP	2	chemistry	49,57
HCTH407	2	chemistry	149
PBEsol	2	solid state	54
WC	2	solid state	156
AM05	2	solid state	157
TPSS	3	general	58
revTPSS	3	general	61
vdW-DF	3.5	vdW	100
vdW-DF2	3.5	vdW	90
optPBE-vdW	3.5	vdW	89
optB88-vdW	3.5	vdW	89
C09-vdW	3.5	vdW	84
B3LYP	4	chemistry	67
PBE0	4	chemistry	68

[a] Should be understood as a very general characterization of the main target of a functional, and does not consider underlying principles of design.

Molecular formation energies

The G3/99 thermochemical test set has become standard for validating density functional methods, and the present calculations are well in line with published benchmark data for this set.[46] Deviation statistics are illustrated in Fig. 7.9, where RMSD, MSD, and MAD are indicated by gray, white, and red bars, respectively. The data are tabulated in Table A.6, which also includes statistics for G2/97 and the G3-3 extension (G3/99 = G2/97 + G3-3).

With a MSD of −5.25 eV, the local spin-density approximation (not shown in Fig. 7.9) grossly overestimates the G3/99 molecular formation energies. The interatomic bonds in a molecule are therefore much too strong within LDA. Significant improvements are found for the GGAs in Fig. 7.9, where DFAs mainly designed to capture molecular energetics (RPBE, BLYP, HCTH407) yield MADs below 0.6 eV, while those targeted at solid-state properties (AM05, PBEsol, WC) perform significantly worse: their MSEs are large and negative, indicating severe overbinding. The TPSS and revTPSS MGGA approximations perform very well for predicting the G3/99 thermochemistry, and sig-

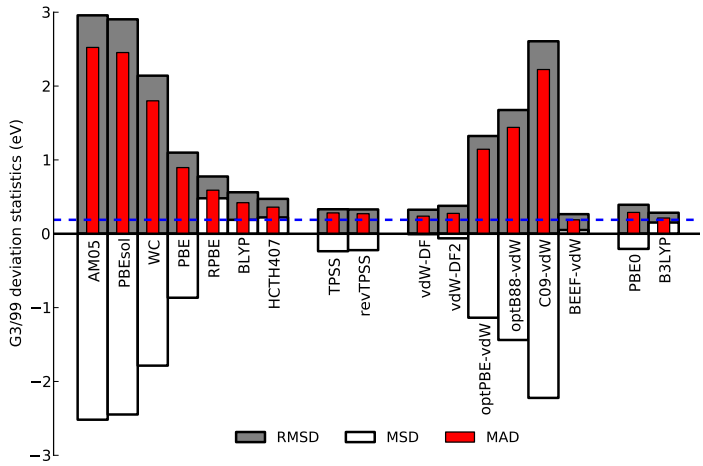

Figure 7.9: Bar plot comparison of the performance of popular GGA, MGGA, vdW-DF, and hybrid density functionals when predicting the formation energies of the G3/99 thermochemical test set. The dashed blue line indicates the BEEF-vdW MAD. LDA statistics are not shown, as they are far out of the scale.

nificantly improve over their PBE ancestor and the "GGAs for chemistry".

Turning to the vdW-DF variants, good description of the G3/99 formation energies is also found for vdW-DF and vdW-DF2. This, however, is not the case for the optPBE-vdW, optB88-vdW, and C09-vdW functionals, for which the GGA exchange components were optimized with vdW-dominated energetics in mind. This approach apparently leads to intramolecular overbinding, as previously noted in Sec. 5.4.3.

For comparison, Fig. 7.9 also shows deviation statistics for the B3LYP and PBE0 hybrids. As the wide application of hybrid XC functionals in the quantum chemistry community suggests, B3LYP and PBE0 accurately describe molecular bond energetics, the B3LYP parametrization performing best with a MAD of 0.21 eV. It is furthermore seen that also the BEEF-vdW XC model performs very well in predicting molecular thermochemistry. With a MAD of 0.19 eV and MSE nearly zero, it is the most accurate DFA tested on G3/99, closely followed by B3LYP. Note that G3/99 contains 75 extra molecules as compared to G2/97, the latter of which BEEF-vdW was trained on.

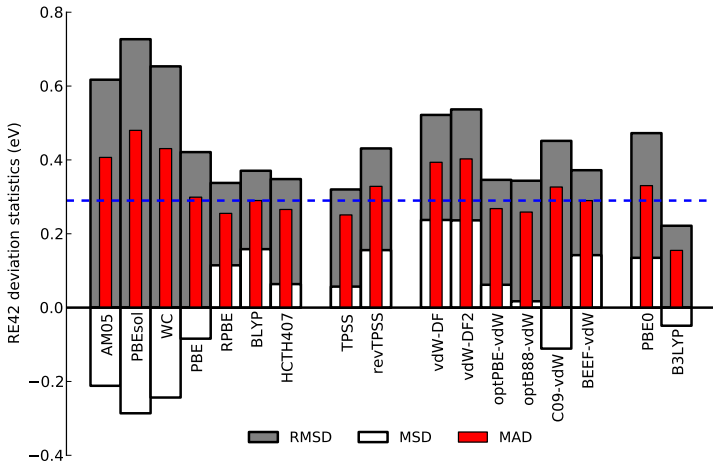

Figure 7.10: Bar plot comparison of density functional performance for calculating the RE42 molecular reaction energies. The dashed blue line indicates the BEEF-vdW MAD. The corresponding LDA statistics (not shown) are 1.62, −0.55, and 1.06 eV for RMSD, MSD, and MAD, respectively.

Molecular reaction energies

Figure 7.10 summarizes deviation statistics for the RE42 dataset, which are also found in Table A.6. Even though the reaction energies are derived from G3/99 formation energies, the former appear difficult to capture accurately with LDA, GGA, MGGA, and vdW-DF type functionals. None of them yield a RMSD less than 0.3 eV. The B3LYP hybrid proves significantly more accurate in this respect. Interestingly, the optPBE-vdW and optB88-vdW functionals, which both severely overestimate the G3/99 formation energies, prove as reliable for calculating gas-phase reaction energies as the best GGA (RPBE), and compare very well to TPSS. Apparently, cancellation of erros between reactants and products, which are both too strongly bound intramolecularly, leads in his case to very reasonably determined reaction energies. The BEEF-vdW yields a marginally higher MAD than optB88-vdW (0.29 eV vs. 0.26 eV), but a non-zero MSE.

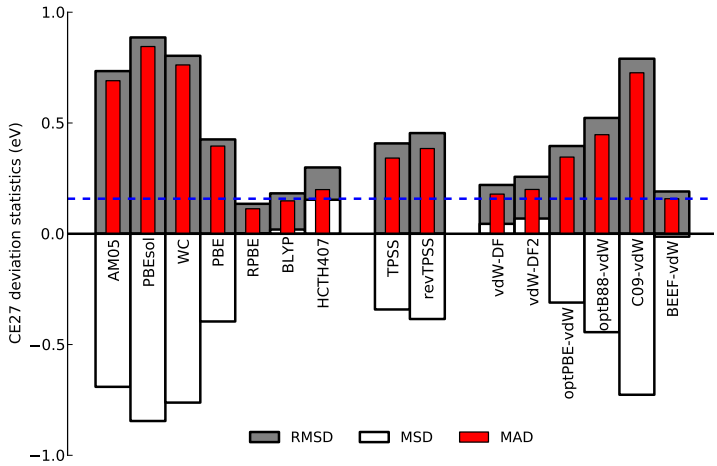

Figure 7.11: Bar plot comparison of density functional deviation statistics on reproducing the CE27 experimental chemisorption energies. The dashed blue line indicates the BEEF-vdW MAD. The corresponding LDA statistics (not shown) are 1.42, −1.33, and 1.33 eV for RMSD, MSD, and MAD, respectively.

Chemisorption on solid surfaces

The deviation statistics for the CE27 dataset are very important in this study since accurate surface chemistry is one main target for a truly general-purpose density functional. The computed statistics are illustrated in Fig. 7.11, and tabulated in Table A.7 along with those of CE17. Note that BEEF-vdW was trained on CE17, while the CE27 set contains 10 extra target data points, mostly covering dissociative H_2 adsorption.

With MADs ≥ 0.7 eV, the GGAs designed for application in solid state physics problems are clearly overbinding simple adsorbates to solid surfaces (negative MSEs). As expected, the chemically biased DFAs are significantly more reliable for calculation of chemisorption energies, RPBE performing best with a MAD of 0.11 eV for both CE17 and CE27. Also vdW-DF and vdW-DF2 yield MADs of 0.20 eV or less and small MSEs on CE27, while the two MGGAs overbind on average. Again, a significant overbinding is found for the three exchange-modified vdW-DF flavors, though optPBE-vdW compares well to TPSS for predicting this materials property.

Lastly, it is clearly seen from Fig. 7.11 that BEEF-vdW is among the DFAs offering most accurate predictions of chemisorption energies of simple adsorbates on solid surfaces. The standard deviation is down to 0.19 eV, and the MSE is nearly zero, very similarly to BLYP.

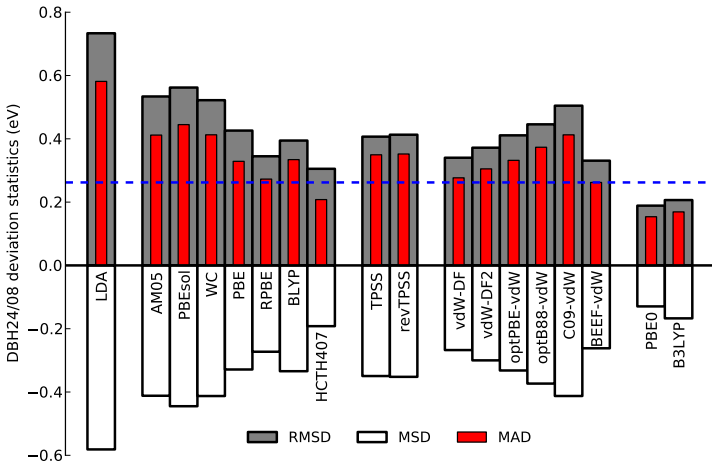

Figure 7.12: Bar plot comparison of density functional deviation statistics on predicting the DBH24/08 benchmark molecular reaction barrier heights. The dashed blue line indicates the BEEF-vdW MAD.

Molecular reaction barriers

The DBH24/08 reaction barrier heights belong to a class of systems for which a fraction of exact exchange is known to increase computational accuracy significantly over GGAs.[77,160] This is supported by the DBH24/08 data in Fig. 7.12, where the two hybrids clearly outperform the lower-rung DFAs. Considering the corresponding statistics for BEEF-vdW as well as for the individual DBH24/08 XC model reported in Table 7.1, where a MAD of 0.12 eV was obtained, it is clear that the BEEF-vdW model has moved significantly away from the part of model space favored by these barrier heights. Nevertheless, BEEF-vdW is among the best non-hybrid functionals for such quantities, but only marginally better than many others. Exact exchange or self-interaction corrections would probably enable significantly better performance on this set, but at least BEEF-vdW is not overfitted in an attempt to make up for the missing directions in model space, even though the XC model was trained on this materials property.

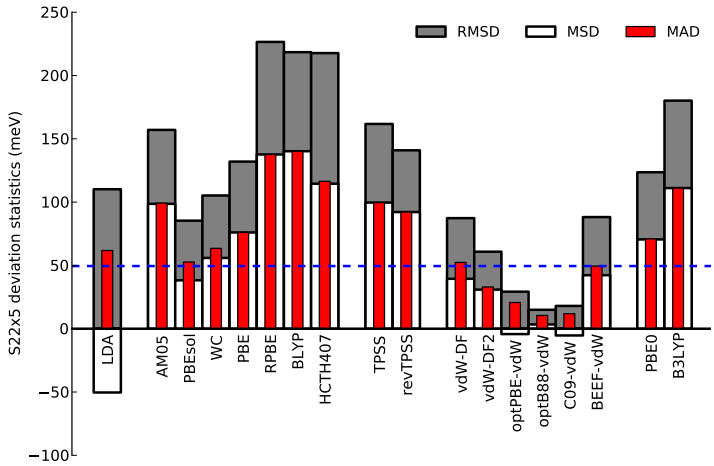

Figure 7.13: Bar plot comparison of density functional deviation statistics on reproducing the S22x5 CCSD(T) interaction energies. The dashed blue line indicates the BEEF-vdW MAD.

Noncovalent interactions

Next in line are computed deviations from the S22x5 noncovalent interaction energies. Statistics for these are illustrated in Fig. 7.13, and tabulated in Table A.7. Note that the figure considers the total S22x5 set.

As previously found in Section 5.4.3 and several literature studies of the original S22 dataset (e.g., Refs. 83,89–91), vdW dominated interactions are well described by vdW-DF type density functionals, especially those with an optimized exchange component. With MADs of 21 meV or less over all 110 points on the 22 potential-energy curves, the optPBE-vdW, optB88-vdW, and C09-vdW functionals prove highly accurate in this respect. The vdW-DF2 also captures vdW interactions well, but the positive MSE signifies that most of the deviations from the CCSD(T) benchmark energies stem from underbinding. For vdW-DF and BEEF-vdW this is somewhat more pronounced.

None of the tested MGGA or hybrid DFAs are supposed to truly capture vdW dispersion interactions, which constitute a major part of the bonding in S22x5. However, two of the most weakly gradient enhancing GGAs, PBEsol and WC, appear in Fig. 7.13 to yield deviation statistics comparable to those of vdW-DF and BEEF-vdW. Does these GGAs then account for long-range dispersion forces? The answer is no, they are merely exaggerating the energetics of interactions between somewhat overlapping densities of gas-phase species. This is easily realized when considering the appreciable LDA overbinding of the S22x5 complexes (MSD < 0). Note that this spurious LDA

behavior was also observed for benzene physisorption on (111) coinage-metal surfaces in Figs. 5.6–5.8 (pages 49–51).

These issues are more closely investigated in Fig. A.1 in Appendix A.2, where *relative* deviation statistics are shown for the S22x5 subsets at $d \in [0.9, 1.0, 1.5]$ (top, middle, and bottom panel, respectively). Data is also listed in Table A.8 for vdW-DF variants and BEEF-vdW. One main conclusion to be drawn from that figure is that though performing reasonably well on S22x5 as a whole, the vdW-DF, vdW-DF2, and BEEF-vdW functionals severely underestimate the interaction energies of the S22x5-0.9 complexes (MARDs of 198%, 143%, and 214%, respectivaly), whereas the exchange-optimized vdW-DF variants appear rather accurate. Note that this subset also ended up with the smallest weight in the BEEF-vdW model compromise after applying the model selection procedure. The choice of GGA exchange approximation for use in the vdW-DF method therefore has a massive impact on the description of the $d = 0.9$ complexes. However, we have also seen that these "locally" optimum choices may generalize poorly to prediction of chemical observables dominated by stronger covalent interactions.

Properties of solid crystals

We consider finally the performance of the tested DFAs for predicting the Sol27 lattice constants and cohesive energies of pure cubic crystal lattices. Figures 7.14 and 7.15 summarize the deviation statistics. The data are tabulated in Table A.5 as well.

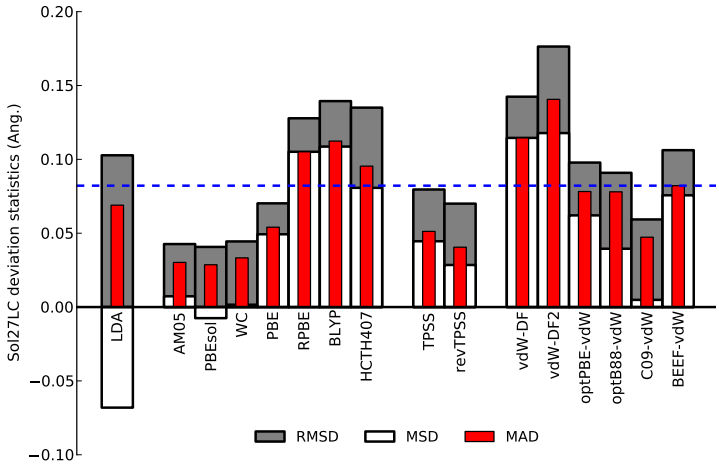

Figure 7.14: Bar plot comparison of density functional deviation statistics on reproducing the Sol27LC semi-experimental lattice constants of mono-atomic solids in cubic crystal structures. The dashed blue line indicates the BEEF-vdW MAD.

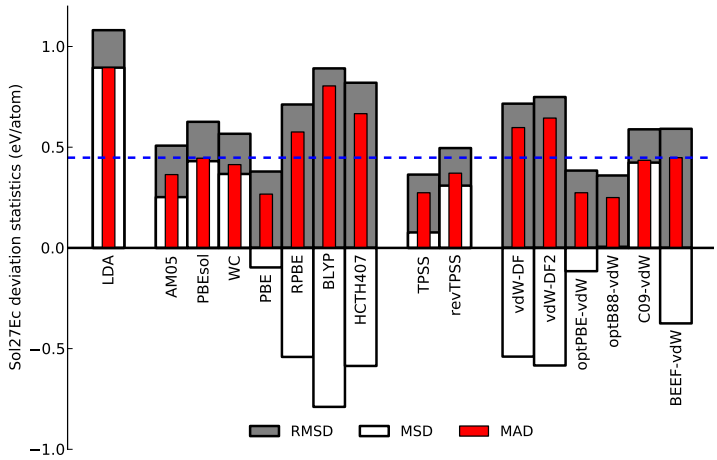

Figure 7.15: Bar plot comparison of density functional deviation statistics on predicting the Sol27Ec semi-experimental cohesive energies of mono-atomic solids in cubic crystal structures. The dashed blue line indicates the BEEF-vdW MAD.

The lattice constant statistics in Fig. 7.14 are in clear favor of the AM05, PBEsol, and WC GGA functionals, but also C09-vdW and to some extend the revTPSS approximation. Their standard deviations are small and the MSEs are close to zero Å. However, Fig. 7.15 shows that these remarkably accurate predictions of equilibrium crystal volumes come at the price of overall overestimated cohesive energies.

The picture is opposite for vdW-DF and vdW-DF2. Lattice constants are overestimated, and more so than with any other XC functional tested, vdW-DF2 yielding a RMSD of 0.18 Å. Furthermore, those two DFAs notably underestimate the cohesive energies, so the vdW-DF and vdW-DF2 density functionals severely underestimate average interatomic bond strength in the rather broad range of bulk solids constituting the Sol27 set. The "opt"-vdW functionals appear to offer a more balanced description of solid state energetics than that of C09-vdW. As PBE and TPSS, these two vdW-DF variants yield the most accurate cohesive energies without (very) large MSDs on the corresponding lattice constants. These findings closely match those reported in recent studies assessing the performance of GGA, MGGA, and vdW-DF type XC functionals for solid state properties (see Refs. 50,51,99,161,162).

Benchmarking finally BEEF-vdW, we find in Figs 7.14 and 7.15 that it performs reasonably well for cohesive energies and lattice constants, though still predicting softer crystal lattices than the optimized vdW-DF variants. With BEEF-vdW these two bulk materials properties are, however, significantly closer to agreement with experiments than predictions by vdW-DF, vdW-DF2, and most of the GGAs designed mainly for chemistry.

7.6 Examples of BEEF-vdW application

Following the extensive benchmark study of the previous section, it is suitable to consider also cases of actual application, especially when the overall goal is versatile density functionals. Two applications of BEEF-vdW to problems of current interest in the surface science community are here discussed: graphene adsorption on the close-packed Ni(111) surface, and the trends observed when applying lower-rung density functionals in calculations of the binding energy of CO to Pt(111) and Rh(111) substrates as well as the surface energy of those substrates.

7.6.1 Graphene adsorption on Ni(111)

The remarkable electronic properties of monolayer graphene[165–167] and its potential application in electronics technology[166,168] motivate investigation of the interactions between graphene sheets and metallic surfaces. The nature of graphene adsorption on metals is highly metal-dependent, some surfaces binding graphene only weakly and others forming strong covalent bonds to the carbon sheet.[169,170] The Ni(111) surface belongs to the latter group, graphene forming a (1×1) overlayer at a graphene–

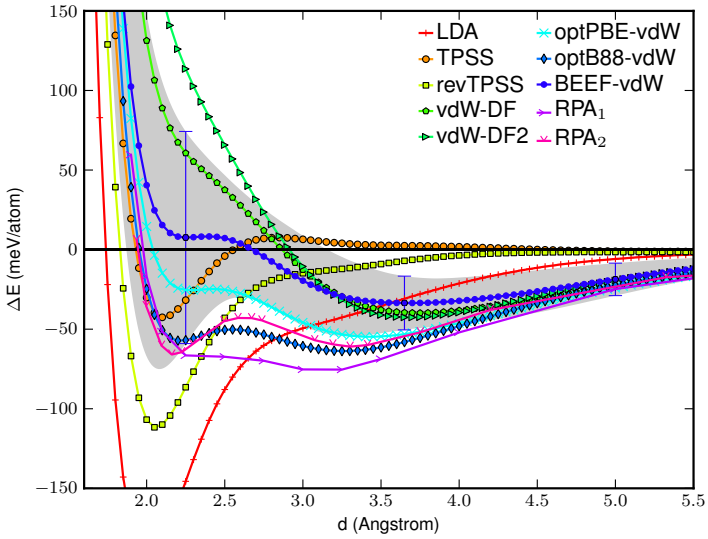

Figure 7.16: Potential-energy curves for graphene adsorption on the Ni(111) surface. Random phase approximation data are from Refs. 163 (RPA_1) and 164 (RPA_2). The gray area indicates the region spanned by the estimated standard deviations along the BEEF-vdW PEC.

metal distance $d = 2.1$ Å.[171] Furthermore, a bandgap is induced in graphene upon adsorption, underlining the strong hybridization responsible for changing the electronic structure of the carbon sheet.[172,173]

Several theoretical studies have investigated the graphene/Ni(111) potential-energy curve, with mixed results.[110,174–179] However, based on RPA calculations, it is by now established that this particular adsorption process is a delicate competition between strong interactions close to the surface and vdW forces further from the surface.[163,164]

Figure 7.16 shows calculated PECs for graphene adsorption on Ni(111) using LDA, MGGA, and vdW-DF type density functionals, as well as BEEF-vdW. Computational details are given in Appendix A.3 Additionally, two sets of RPA data are shown for comparison, indicating that graphene adsorption on Ni(111) is characterized by a physisorption minimum at $d = 3.0$–3.5 Å and a chemisorbed state at $d \approx 2.2$ Å, the latter in good agreement with experiments.[171] However, as previously found in literature for GGAs and vdW-DF variants (e.g., Refs. 110,163,164,178), rung 1–2 density functionals, as well as vdW-DF and vdW-DF2, fail to simultaneously describe both qualitative features. This is here shown to be the case for the rung 3 TPSS and revTPSS approximations also. Conversely, the optPBE-vdW and optB88-vdW PECs in Fig. 7.16 are increasingly closer to RPA data. The BEEF-vdW PEC shows qualitatively similar features, but the local minimum at $d = 2.25$ Å is very shallow and yields a positive adsorption energy.

Figure 7.16 also shows ensemble error estimates along the BEEF-vdW PEC. Especially two aspects of these are of interest. First of all, the error bars do not straddle the zero line for large graphene–metal distances, indicating that confidence in the presence of a physisorption minimum is high. Secondly, the error bars, indicated by the shaded area, enlarge notably at smaller distances from $d = 2.6$ Å and inwards, reflecting that these BEEF-vdW data points are associated with a significantly larger uncertainty. Recalling how the ensemble error estimate is designed (Sec. 7.4), the error estimates indicate that the graphene/Ni(111) PEC is very sensitive to the choice of XC functional in the chemically interesting range. Put differently, the ensemble suggests that we should not trust the BEEF-vdW prediction of a positive PEC for $d < 2.7$ Å as a definite result. This would be a very nice piece of information if the characteristics of the graphene/Ni(111) system were more or less unknown to us on beforehand.

7.6.2 Surface chemistry and stability

As outlined in the introduction to this thesis, chemisorption energies of molecules on surfaces are vital quantities in heterogeneous catalysis and surface science. However, accurate computation of surface energies, E_γ, can be critical as well, since minimization of surface energy is a driving force determining the morphology and composition of surfaces, interfaces, and nanoparticles.[13] GGA density functionals, however, often underestimate E_γ, and the GGAs yielding most accurate surface energies also vastly overbind molecules to surfaces.[180] It thus appears that accurate computation of

chemisorption energies on a surface as well as the stability of that surface is not possible with the same GGA approximation, underscoring a fundamental incompleteness of the GGA XC model space.

The issue is here investigated for vdW-DF variants and BEEF-vdW. Figure 7.17 shows ontop chemisorption energies of CO on Pt(111) and Rh(111) against surface energies of those substrates, calculated using GGA, MGGA and vdW-DF type DFAs, and BEEF-vdW with error estimation. These are compared to RPA results and experimental data.

As also observed elsewhere,[180,184] the GGA data points fall along an approximately straight line, which is significantly offset from the experimental data, thus illustrating the issue discussed above. This is here shown to be the case for vdW-DF variants also. The dashed vdW-DF lines are parallel to the solid GGA lines, and are only slightly offset from the latter, especially for Rh(111). The vdW-DF and vdW-DF2 data points are quite close to RPBE. Larger surface energies are found with the exchange-modified vdW-DF variants, albeit at the expense of overestimated chemisorption energies. Note that such a correlation should be expected from Figs. 7.11 and 7.15 and a linear relation between E_γ and the solid cohesive energy.[13]

Though BEEF-vdW contains the vdW-DF2 nonlocal correlation functional as an es-

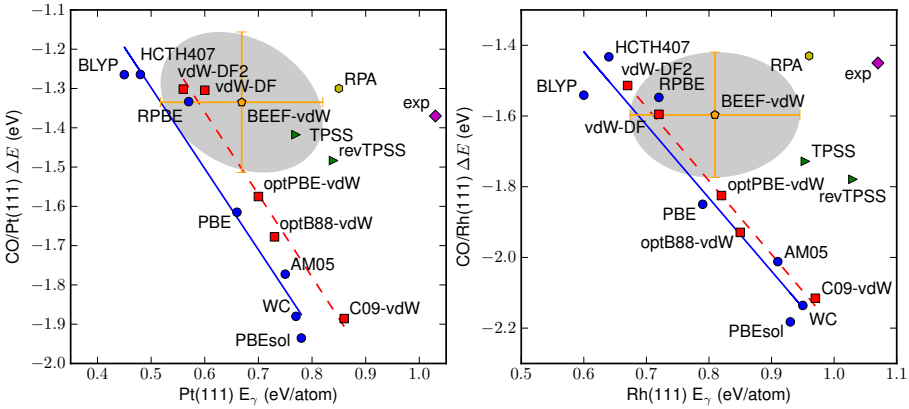

Figure 7.17: Ontop CO chemisorption energies ΔE versus surface energies E_γ for Pt(111) and Ru(111). Red and blue lines are linear fits to GGA and vdW-DF type data points, respectively. MGGA data in green. RPA data from Ref. 180 in yellow. Estimated standard deviations are indicated by error bars around the orange BEEF-vdW data points. All points $(E_\gamma, \Delta E)$ inside the gray areas are within one standard deviation from the BEEF-vdW point for both quantities. Experimental surface energies from liquid-metal data (Refs. 181 and 182), and experimental CO chemisorption energies from Ref. 183.

sential component, the former predicts larger surface energies than the latter without sacrificing accuracy of the CO–metal binding energy. We expect that this ability of BEEF-vdW to "break" the vdW-DF line is primarily due to the larger GGA model space. Furthermore, the BEEF-vdW error estimates appear very reasonable (shaded areas and orange error bars in Fig. 7.17). The experimental CO chemisorption energies are straddled for both Pt(111) and Rh(111), and the error estimates along E_γ almost fill out the gap between the GGA lines to the left and the RPA and C09-vdW surface energies to the right.

Lastly, it is seen from the green TPSS and revTPSS data points in Fig. 7.17, as also noted in Ref. 184, that the third rung of Jacob's ladder may offer the possibility of quite accurate surface energies with only moderately overbound surface adsorbates.

7.7 Discussion and summary

A machine learning inspired approach to semi-empirical density functional development has been presented and evaluated. Focus has been on general applicability of the resulting DFA to both strong and weak interactions in chemistry and condensed matter physics, including surface chemistry. Transferability and avoiding overfitting are thus key issues, and the present methodology therefore relies primarily on 1) a variety of datasets chosen to represent vastly different interactions and bonding situations, 2) a very flexible XC model space expansion at a computationally feasible GGA+vdW level of approximation, and 3) XC model selection procedures designed to "tame" the flexible model space and yield XC approximations which properly compromise between describing different types of physics and chemistry.

The combination of Tikhonov regularization and bootstrap cross-validation is a powerful route to smooth and high-performance GGA exchange models. The cost function for a single dataset has both weak (sloppy) and strong (important) eigenmodes in a sufficiently flexible model space, and regularization is used to suppress the weak modes in order to facilitate a physically sensible model and maximize transferability. The regularized model ensemble thus contracts around the strong modes, and the optimum model can, to some extent, be regarded an average of the ensemble solutions. For single datasets, this leads to accurate and well-behaved XC models, where effectively 3–8 of the initially 31 parameters are allowed to survive.

However, it is also clear that computationally efficient general-purpose XC functionals must compromise between describing physically and chemically different materials properties in an incomplete model space. Determining the optimum model tradeoff is thus a complex multi-objective problem, but a simple principle for finding the position in model space of *one* properly compromising DFA was formulated an applied. The three most important aspects of semi-empirical DFA design were thus considered in detail: datasets, model space, and model selection. This resulted in the BEEF-vdW approximation.

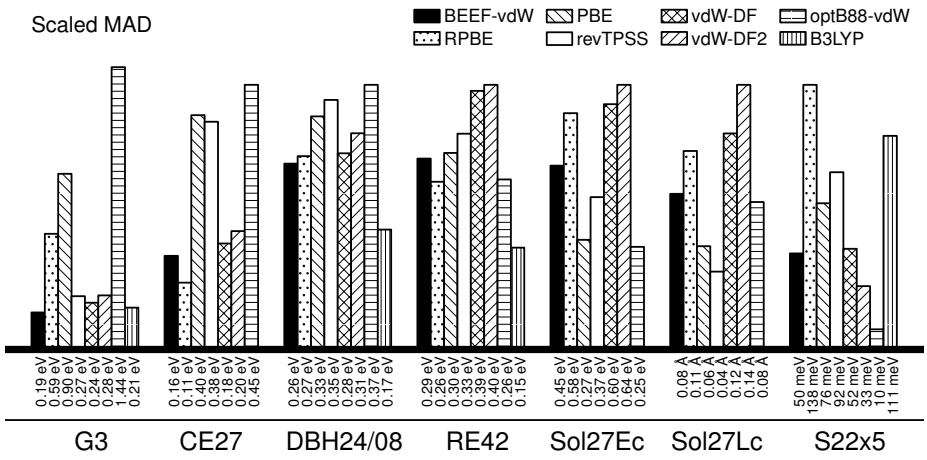

Figure 7.18: Bar plot comparison of the accuracy of different density functionals in predicting various materials properties. For each dataset, the bars illustrate proportionally scaled mean absolute deviations. B3LYP calculations were not performed for extended systems.

A benchmark of BEEF-vdW against popular GGA, MGGA, vdW-DF type, and hybrid XC functionals for energetics in chemistry and condensed matter physics was conducted. The results are summarized in Fig. 7.18, and show that the BEEF-vdW model compromise is indeed a very agreeable one. For none of the datasets is the average BEEF-vdW error among the largest, while several other functionals are highly biased towards certain types of materials properties. This is especially true for vdW-DF2 and optB88-vdW, displaying severely erroneous description of binding energetics for bulk solids and molecules, respectively. However, the BEEF-vdW description of vdW-dominated energetics certainly leaves room for improvement when compared to exchange-optimized vdW-DF variants.

A BEEF-vdW error estimation ensemble was also generated. It is designed to provide an easily obtainable (non-self-consistent) estimate of the computational error. It is based on a probability distribution for the model parameters, and was applied in the benchmark as well as in qualitative assessments for molecular surface adsorption, surface energies, and graphene adsorption on Ni(111). Since the ensemble is confined to live in an incomplete model space, but was designed to reproduce training set errors by fluctuations within this space, one may think of the error estimates as somehow midway between a systematic sensitivity analysis and actual evaluation of the errors on predictions by DFT.

Chapter 8

meta-BEEF Density Functionals

The approach to semi-empirical density functional development presented in the previous chapter is here applied to the more complex case of a meta-GGA exchange model space. This naturally involves an extension of the regularization methodology used for generating BEEF-vdW, but also allows for significant simplifications of the search for an optimum model compromise.

The contents of this chapter constitute the forefront of an ongoing effort to improve over the BEEF-vdW XC model, and all presented methods and results from applying them are preliminary. As these lines are written, thousands of CPU cores are converting electrons to zeros and ones to advance the project, which is run by Keld T. Lundgaard and the author in collaboration with Karsten W. Jacobsen and Thomas Bligaard.

8.1 Expanded exchange model space

The semi-local exchange–correlation energy is written in the usual way as an integral over the XC energy density per particle, $E_{xc} = \sum_\sigma \int \epsilon_{xc}(n_\sigma, \nabla n_\sigma, \tau_\sigma) n_\sigma(\boldsymbol{r}) d\boldsymbol{r}$, where n_σ is the local electron spin density, ∇n_σ its gradient, and $\tau_\sigma = \frac{1}{2}\sum_i |\nabla \phi_{i\sigma}|^2$ the local KS spin-orbital kinetic energy density. The MGGA exchange energy is then expressed in terms of an exchange enhancement factor $F_x(n, \nabla n, \tau) = F_x(s, \alpha)$, where we have

dropped the spin arguments, s is the dimensionless reduced density gradient, and α a dimensionless form of the KED, see Section 3.2.

The semilocal exchange model space is therefore exptended from that in the previous chapter by including the electronic-structure dependent parameter α, and we aim now for a 2D basis expansion of $F_x(s,\alpha)$. This is easily done by introducing an expansion in *products* of one-dimensional Legendre polynomials $B_m(t_s(s))$ and $B_n(t_\alpha(\alpha))$,

$$F_x(s,\alpha) = \sum_m \sum_n a_{mn} B_m(t_s(s)) B_n(t_\alpha(\alpha)), \tag{8.1}$$

where the polynomials depend on the transformed quantities $t_s(s)$ and $t_\alpha(\alpha)$,

$$\begin{aligned} t_s(s) &= \frac{2s^2}{q+s^2} - 1, \\ t_\alpha(\alpha) &= \frac{(1-\alpha^2)^3}{1+\alpha^3+\alpha^6}, \end{aligned} \tag{8.2}$$

which are both confined to the range $[-1,1]$, and $q = 0.804/(10/81) = 6.5124$. These transformations have been chosen to easily represent the exchange gradient expansion of the slowly varying HEG. The first was used in the PBEsol exchange functional,[54] while the latter has recently been introduced by Sun and co-workers[185] to capture the gradient expansion up to 4th order. Note that this property is only strictly conserved in (8.1) for $a_{m,n=0} = 1$ and $a_{m,n>0} = 0$, which is in general not the case here.

The correlation model space of BEEF-vdW is largely retained, that is, a linear combination of LDA, GGA, and nonlocal correlation of the vdW-DF2 type,

$$E_{xc} = \sum_m \sum_n E_x^{mn} + \theta_c E_c^{\mathrm{LDA}} + (1-\theta_c) E_c^{\mathrm{GGA}} + E_c^{\mathrm{nl}}, \tag{8.3}$$

where $E_x^{mn}[n,\nabla n,\tau] = \int \varepsilon_x^{\mathrm{HEG}}(n) F_x^{mn}(s,\alpha) n(\boldsymbol{r}) d\boldsymbol{r}$ is a basis-function exchange energy contribution, and the PBEsol correlation is chosen for E_c^{GGA}.

8.2 Model selection

Regularization Several parts of the BEEF-vdW model selection procedure are reused here, but certain modifications are made. Thus, minimization of the cost function (7.8) with Tikhonov regularization is employed for training of individual XC models, and the bootstrap .632 estimator directs us to the optimum regularization strength. However, the regularizer should now penalize non-smooth exchange model deviations away from the prior vector in the 2D exchange model space given by s and α. The exchange part of the Tikhonov matrix is therefore given by the overlap of the Laplacian in both directions,

$$\boldsymbol{\Gamma}^2_{nmkl} = \int_{-1}^{1}\int_{-1}^{1} dt_s\, dt_\alpha \left(\left[\frac{d^2}{dt_s^2} + \gamma \frac{d^2}{dt_\alpha^2} \right] B_m B_n \right) \left(\left[\frac{d^2}{dt_s^2} + \gamma \frac{d^2}{dt_\alpha^2} \right] B_l B_k \right), \tag{8.4}$$

where the scaling factor $\gamma \sim 10^2$ is used to penalize non-smoothness in the t_α-space a few orders of magnitude more than in the t_s-space. The correlation mixing parameter θ_c should in principle not be regularized at all. However, this will not do numerically, so the diagonal entry γ' for θ_c in the Tikhonov matrix is nonzero but scaled as $\gamma' \sim 10^{-5}\gamma$, so that regularization of θ_c has essentially no weight as compared to regularization of exchange. In the bootstrapping procedure, a scan over regularization strengths from $\omega^2 = 10^{10}$ and downwards is performed, and the model solution associated with the first EPE minimum encountered is chosen as the optimum one.

Summed cost function The model selection procedure for more than one dataset is pivotal, and a neatly simplified procedure is used here. The BEEF-vdW XC model compromise was found by iteratively minimizing a weighted product cost function, which was composed of individually regularized models in their quadratic form. Here we choose to directly regularize the compromising XC models using a summed cost function,*

$$C = \sum_i \frac{w_i}{k_i Q_i^2} \left(\boldsymbol{X}_i \boldsymbol{a} - \boldsymbol{y}_i\right)^2 + \omega^2 \boldsymbol{\Gamma}^2 (\boldsymbol{a} - \boldsymbol{a}_p)^2, \tag{8.5}$$

where k_i is the number of systems in the i'th dataset, Q_i is the RMS deviation for an individual fit to dataset i, and w_i is the dataset weight. The first cost function term is thus a normalized and weighted error. Following the discussion of optimum XC model compromises in the previous chapter, the weights w_i must then be determined. Since the cost function (8.5) is not explicitly formulated in terms of individually regularized cost functions, choosing to minimize the product of dataset RMSDs rather than individual costs is natural. We therefore consider the product $P = \Pi_i S_i(\{w_i\})$, where S_i is the dataset RMSD for the jointly regularized and compromising XC model, which obviously depends on the set of cost function weights $\{w_i\}$. The optimum set is therefore found iteratively by minimizing P over the weights.

A few more details about the procedure are appropriate: Legendre polynomial expansions of exchange to fourth order in t_s and t_α have been employed, yielding $5 \times 5 = 25$ exchange basis-function contributions to total exchange energies. The order 0 and 1 polynomials in the t_s and t_α model space directions have been fixed in order to fulfill the gradient expansion for exchange to second order in s and α. Similarly to the BEEF-vdW study, the prior vector $\boldsymbol{a}_p$ is chosen such that infinite regularization strength yields an exchange enhancement factor $F_x(s \to \infty, \alpha = 1) = 1.804$. The origo for regularization of the correlation model space is $\theta_c = 0.5$.

* We tried doing this already in the BEEF-vdW work, but found the GGA exchange model space too limited for this to work properly. The bootstrap .632 is not a good estimator of the generalization error in situations of heavy overfitting,[143] and the resulting EPE-minimized XC model solutions appeared to reflect this: The exchange solutions were overly oscillatory in order to accommodate the vastly different datasets, leading us to conclude that the bootstrap EPE was not able to tell that those solutions were really overfits. Only individually regularized model solutions ended up properly regularized. Introducing the MGGA exchange model space and choosing ω^2 as the first EPE minimum in a scan from very strong regularization and downwards remedies this issue.

Datasets The collection of benchmark datasets used in model training reflects the overall goal of the XC models to be generated; versatile approximations applicable to a wide range of condensed matter. The datasets chosen are the G3/99 molecular formation energies, the RE42 molecular reaction energies, the S22x5 noncovalent interaction energies, the CE27 reactions energies for chemisorption, and finally the Sol53LC and Sol53Ec lattice constants and cohesive energies of solid crystals. Note that, as compared to the solid-state sets used in BEEF-vdW training, the Sol53 sets comprise data for pure solids as well as mixed-element compounds, as discussed in Section 4.1.0.7.

Procedure Development of two new density functionals with built-in error estimation is the prime object in this work — a meta-GGA with MGGA exchange and GGA correlation (mBEEF), and the natural extension to include the Rutgers–Chalmers nonlocal correlation (mBEEF-vdW). An important issue when optimizing the XC models to a dataset such as CE27 is self-consistency in geometries as well as electron densities. This was not discussed in connection with the BEEF-vdW development, but is important in this work. The reason is simple: The one-shot BEEF-vdW fit turned out to predict crystal lattice constants and geometries in general not significantly different from those of RPBE. Since RPBE geometries and densities was used for the BEEF-vdW fit, the non-self-consistent predictions of performance were not far from those found self-consistently afterwards. This is not the case in the current study, where the expanded model space offers the possibility of lattice constants significantly closer to experiment while largely preserving good predictions of thermochemistry. Non-self-consistent RPBE-density estimates of the performance of the optimized MGGA functionals on especially CE27 are therefore not adequate.

An iterative approach towards near-self-consistent model optimization is therefore taken. This procedure starts off with optimizing the mBEEF using PBE densities for non-self-consistent XC contributions to the data matrix in (8.5). The model resulting from applying the model selection method outlined above is then used to generate new self-consistent densities including full geometry-optimization of mainly CE27. This loop should in principle be continued intil self-consistency, and the "globally optimized" mBEEF can then serve as base for optimizing the mBEEF-vdW in a simple manner. This is, however, a lengthy procedure requiring an immense number of DFT calculations, so the results presented in the next section are for preliminary versions of the mBEEF and mBEEF-vdW functionals.

8.3 Results

The mBEEF presented here is a "version 2", that is, it is obtained from optimization on first GGA electron densitites, and then once on the resulting MGGA-type densitites. When it became clear just how long time it might take to reach the desired near-self-consistency, it was decided to start off the cycle for mBEEF-vdW also, taking the current mBEEF as starting point. The mBEEF-vdW is in that sense a one-shot fit

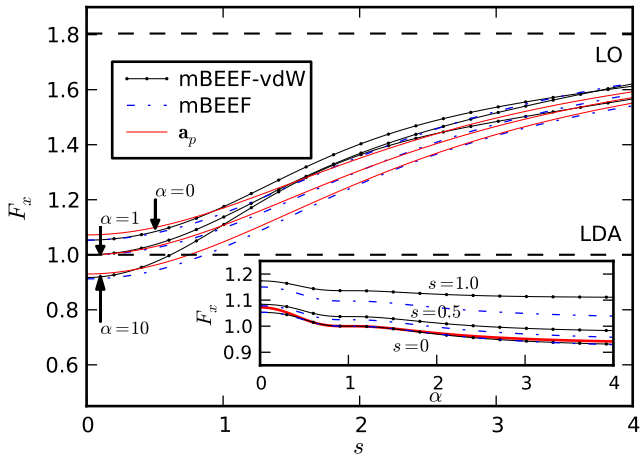

Figure 8.1: The mBEEF (blue) and mBEEF-vdW (black) exchange enhancement factors represented in the 2D (s, α)-space. The main panel illustrates $F_x(s, \alpha)$ along s for different constant α-values, as well as the prior vector $\boldsymbol{a}_p$ (red), which corresponds to the PBEsol $F_x(s)$. The inset shows the exchange enhancement along the α-direction for different values of constant s, and the prior for KED dependence at $s = 0$ in red. Note that $F_x(s = 0, \alpha = 1) = 1$ is the only condition for complying with the HEG limit, which is imposed on both XC models.

and highly preliminary, but promising results have been obtained using both DFAs.

Preliminary functional forms Figure 8.1 shows the exchange enhancement factors of the two obtained DFAs, mBEEF in blue and mBEEF-vdW in black. Both models are exceedingly smooth variations away from the prior solution (red), and $F_x(s, \alpha)$ for mBEEF-vdW in particular is notably different from $\boldsymbol{a}_p$. Note that the HEG limit $F_x(s = 0, \alpha = 1) = 1$ is imposed on and smoothly satisfied by the exchange models. The enhancement factors for iso-orbital like densities ($\tau \approx 0$) appear rather intent on $F_x(s = 0, \alpha = 0) \approx 1.05$, that is, slowly varying/high density regions different from the HEG are not treated as a HEG. While the mBEEF obviously employs full PBEsol correlation, the mBEEF-vdW correlation mixing parameter is $\theta_c = 0.38$, i.e., 38% LDA and 62% PBEsol correlation in conjunction with full E_c^{nl}. The effective numbers of parameters left in the models after regularization are 2.4 (mBEEF) and 5.5 (mBEEF-vdW), where the constant GGA correlation in mBEEF is not included.

Performance on training datasets The performance of mBEEF and mBEEF-vdW on the training datasets is illustrated in the bar-plot statistics in Fig. 8.2. Note that mBEEF was not trained on CE27, while mBEEF-vdW was. Since only half of the required calculations for S22x5 with mBEEF-vdW have finished at thesis deadline, the illustrated statistics for this DFA are not absolutely complete, but very near so.

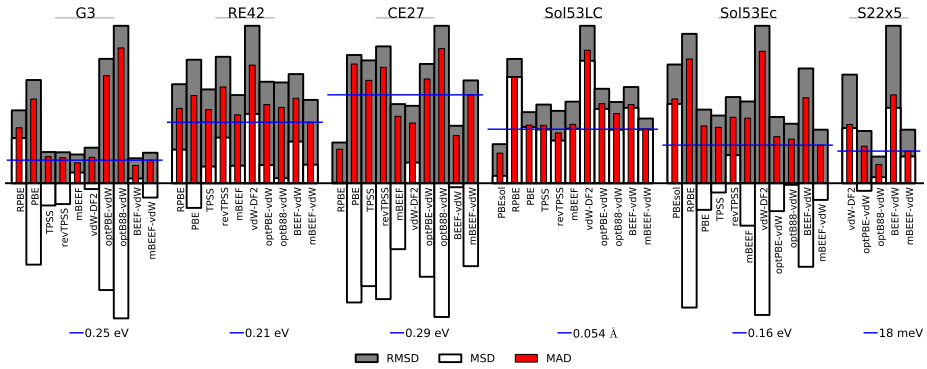

Figure 8.2: Bar plot comparison of the mean accuracy of different density functionals for predicting the materials properties used for training the mBEEF and mBEEF-vdW XC approximations. The statistics for each dataset is normalized to the largest RMSD illustrated for each set. Horizontal blue lines indicate the mBEEF-vdW MAD. Note that the S22x5 statistics for mBEEF-vdW are based on only half of the 110 target data points, see text for details.

The missing data are rather randomly distributed over the whole set, and if these are omitted when computing the statistics for, e.g., optB88-vdW, the results change only by a few meV.

Some important conclusions can be drawn from Fig. 8.2. The mBEEF is clearly a serious competitor to the established TPSS and revTPSS meta-GGA functionals when it comes to broad applicability: On all five relevant sets of materials properties does it appear to offer equal or better computational accuracy, and predicts significantly less overestimated CE27 chemisorption energies. This indicates a clear step forward from the GGA model space in terms of the model compromises available. Many GGAs have been generated (and discarded) in our group, but never before has such a general-purpose semilocal DFA emerged from our efforts. Furthermore, the mBEEF-vdW extension of mBEEF improves significantly on the Sol53Ec cohesive energies, and appears highly accurate for capturing the vdW interactions in S22x5, but is worse on the CE27 set. This model compromise is therefore a different one than for mBEEF, which should not surprise, as two extra datasets (CE27 and S22x5) have been introduced in the multi-objective minimization problem, as well as extra degrees of freedom in the correlation model space. If desired, more favorable CE27 statistics might be obtainable by taking a few more steps in the optimization cycle towards near-self-consistency, but probably at the expence of worse performance on other datasets and/or a less smooth model solution with more effective parameters. Whatever the personal prefence, the statistics in Fig. 8.2 lend promise to general-purpose density functionals well beyond what the more simple GGA+vdW model space can reasonably offer.

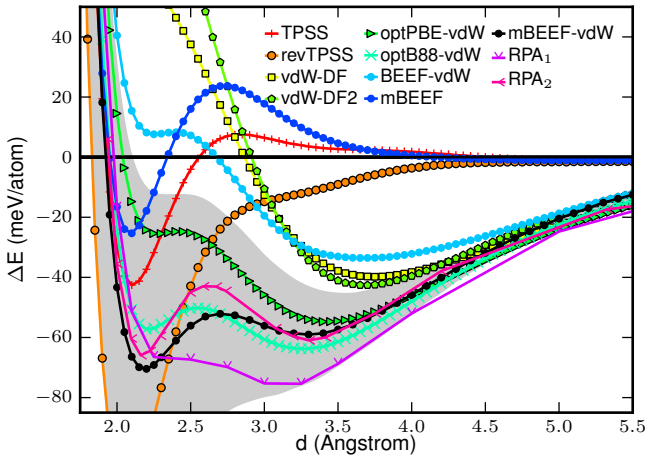

Figure 8.3: Potential-energy curves for graphene adsorption on the Ni(111) surface. Most data are from Fig. 7.16, but here the mBEEF (dark blue) and mBEEF-vdW (black) PECs are added. The gray area indicates the region around the mBEEF-vdW PEC spanned by the corresponding ensemble error estimates.

We consider in the next few paragraphs how mBEEF and mBEEF-vdW perfom in the familiar cases of graphene adsorption on Ni(111) and the correlation between the computed CO chemisorption energy on Pt(111) and Rh(111) and surface energy of the substrates.

Graphene adsorption on Ni(111) The mBEEF and mBEEF-vdW PECs for benzene adsorption on the Ni(111) surface are shown in Fig. 8.3 along with many of the PECs previously presented in Fig. 7.16. The blue mBEEF PEC captures a chemisorption minimum small for graphene–metal distances, but as expected not the stable physisorption state predicted by RPA. On the other hand, similarly to optB88-vdW, the mBEEF-vdW PEC is very close to RPA data, and clearly predicts the presence of two PEC minima. The mBEEF-vdW error estimation ensemble furthermore supports the presence of both bonding states and narrows significantly for $d \to \infty$.

Surface chemistry and stability Predictions by mBEEF and mBEEF-vdW of the CO adsorption energy against surface energy for the Pt(111) and Rh(111) substrates are illustrated in Fig. 8.4 along with previously considered data. The meta-GGA functionals, including mBEEF, are able to produce significantly larger (and more accurate) estimates of surface energies than GGAs of similar accuracy for chemisorption energies. Again, the terms of a model compromise have changed from the GGA case. The change when moving to the mBEEF-vdW model space is not overwhelming.

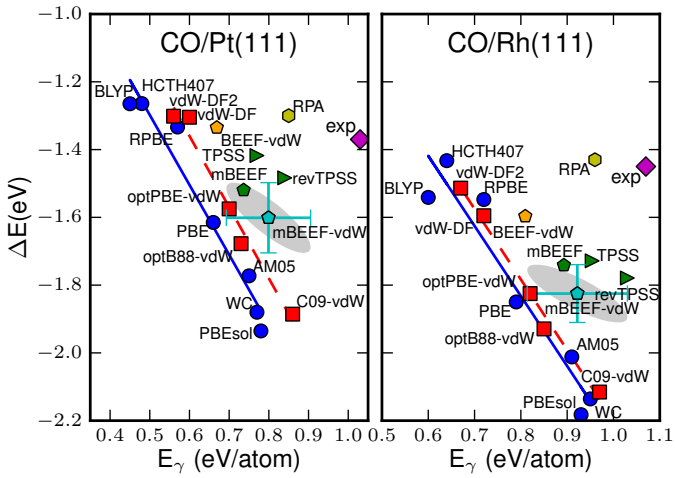

Figure 8.4: Chemisorption energy vs. surface energy for CO on Pt(111) and Rh(111). These data were also shown in Fig. 7.17, but here data points for mBEEF (green diamond) and mBEEF-vdW (cyan diamond) are added. Estimated standard deviations are indicated by error bars around the the mBEEF-vdW data points, and all points $(E_\gamma, \Delta E)$ inside the gray areas are within one standard deviation from the mBEEF-vdW point for both quantities.

8.4 Discussion and summary

It is apparent from the individually trained XC models in Chapter 7 that significantly different materials properties favor different directions in the GGA+vdW model space, and the individual GGA exchange enhancement factors obtained were very different, also in the $s < 4$ regime determining most bonding energetics in the solid state and chemistry. This makes XC model compromising at that level of theory largely a game of choosing to describe either chemistry or solids well, which is also reflected in the BEEF-vdW model compromise.

Recent literature suggests that the MGGA model space may improve over the GGA one at a modest extra computational cost. By applying our machine learning approaches to the MGGA and MGGA+vdW model spaces, both containing MGGA exchange and the latter also linear-combination correlation with nonlocal contributions, it was here shown that the MGGA step on Jacob's ladder offers a new ballpark for compromising XC model solutions, and one in which the HEG limit for exchange is easily fulfilled with smooth general-purpose density functionals. Furthermore, neatly following the several demonstrations of the ambiguity of a suitable functional form of GGA exchange for vdW-DF calculations, meta-GGA exchange was here found a very promising partner for our linear-combination correlation models. We expect this conclusion to hold in general, and using properly optimized MGGA exchange for the "hot seat" in the more restricted Rutgers–Chalmers method should enable significant improvements as well.

Chapter 9

Summary and Outlook

This thesis is based on three published papers and preliminary results from an onging extension of particularly the last of these studies. The overall theme is computationally tractable approximations for the exchange–correlation density functional of KS-DFT, and the first few steps towards development of highly versatile density functionals targeted at general application in computational surface science and heterogeneous catalysis have been taken.

Such density functional approximations should accurately describe bond breaking and formation in chemistry, solid state physics, and the surface chemistry interface between them, but also capture the vdW dispersion forces responsible for somewhat more exotic phenomena in materials science. Dispersion is not accounted for by conventional XC approximations.

Broad assessments of the vdW-DF method for inclusion of vdW interactions in DFT were here conducted, and they validated its applicability for describing the energetics of chemical bonds. However, optimization of the exchange functional employed is needed for high-accuracy prediction of noncovalent bond energies. Such optimization leads to improved description of bonding in the solid state as well, but at the expense of significant intramolecular overbinding. Inclusion of explicitly semilocal correlation contributions to the vdW-DF method were here proposed as an alternative path to improvements, yielding strikingly uniform refinements over the significantly different materials properties considered. Application to studies of benzene physisorption on the noble (111) surfaces of Cu, Ag, and Au supported this notion.

A machine learning inspired approach to semi-empirical density functional development was developed and applied. With the aim of a general-purpose van der Waals density functional with built-in error estimation on computed quantities, a highly flexible GGA+vdW model space expansion was explored. Issues of preserving transferability of the final DFA and avoiding overfitting were tackled by a combination of Tikhonov regularization and bootstrap cross-validation, leading to a powerful procedure yielding smooth and high-performance XC models that retain only the most important degrees of freedom. However, it was also shown that such computationally efficient XC models must necessarily compromise between describing physically and chemically different materials properties in an incomplete model space. A simple principle for determining the position in this space of a properly compromising model was formulated an applied. Benchmarking the resulting BEEF-vdW functional largely validated it as a general-purpose approximation, but also underscored the model compromises made. A corresponding error estimation ensemble of XC models was designed for computationally cheap estimates of the DFT error on predicted quantities, and was applied in studies of graphene adsorption on Ni(111) and the correlation between DFT predictions of molecular CO adsorption and surface energies.

The exchange–correlation model compromises necessary at the GGA level of theory are often grave. The DFA designer must choose between high computational accuracy for bonding in chemistry, the solid state, or none of them. By applying the developed methodology to the MGGA and MGGA+vdW model spaces it was shown that MGGA offers a highly favorable ballpark for compromising models. This is therefore a promising path towards truly versatile van der Waals density functionals with built-in error estimation for surface science applications. Introduction of self-interaction corrections of some sort to the model space may prove advantageous in the long run as well.

Appendix A

Appendix

A.1 Datasets

Table A.1: Gas-phase molecular reactions and reaction energies (in eV) constituting the RE42 dataset. The experimental reaction energies are compiled from the G3/99 static-nuclei formation energies. $\Delta_r E < 0$ means exothermic.

Reaction	$\Delta_r E$
$N_2 + 2H_2 \rightarrow N_2H_4$	0.41
$N_2 + O_2 \rightarrow 2NO$	1.88
$N_2 + 3H_2 \rightarrow 2NH3$	−1.68
$O_2 + 2H_2 \rightarrow 2H_2O$	−5.45
$N_2 + 2O_2 \rightarrow 2NO_2$	0.62
$CO + H_2O \rightarrow CO_2 + H_2$	−0.31
$2N_2 + O_2 \rightarrow 2N_2O$	1.57
$2CO + O_2 \rightarrow 2CO_2$	−6.06
$CO + 3H_2 \rightarrow CH_4 + H_2O$	−2.80
$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$	−2.50
$CH_4 + NH_3 \rightarrow HCN + 3H_2$	3.32
$O_2 + 4HCl \rightarrow 2Cl_2 + 2H_2O$	−1.51
$2OH + H_2 \rightarrow 2H_2O$	−6.19
$O_2 + H_2 \rightarrow 2OH$	0.74
$SO_2 + 3H_2 \rightarrow SH_2 + 2H_2O$	−2.62

Reaction	$\Delta_r E$
$H_2 + O_2 \rightarrow H_2O_2$	−1.68
$CH_4 + 2Cl_2 \rightarrow CCl_4 + 2H_2$	0.19
$CH_4 + 2F_2 \rightarrow CF_4 + 2H_2$	−8.60
$CH_4 + H_2O \rightarrow$ methanol $+ H_2$	1.33
$CH_4 + CO_2 \rightarrow 2CO + 2H_2$	3.11
$3O_2 \rightarrow 2O_3$	2.92
methylamine $+ H_2 \rightarrow CH_4 + NH_3$	−1.15
thioethanol $+ H_2 \rightarrow H_2S +$ ethane	−0.71
$2CO + 2NO \rightarrow 2CO_2 + N_2$	−7.94
$CO + 2H_2 \rightarrow$ methanol	−1.48
$CO_2 + 3H_2 \rightarrow$ methanol $+ H_2O$	−1.17
2 methanol $+ O_2 \rightarrow 2CO_2 + 4H_2$	−3.11
$4CO + 9H_2 \rightarrow$ trans-butane $+ 4H_2O$	−9.00
ethanol $\rightarrow$ dimethylether	0.53
ethyne $+ H_2 \rightarrow$ ethene	−2.10
ketene $+ 2H_2 \rightarrow$ ethene $+ H_2O$	−1.92
oxirane $+ H_2 \rightarrow$ ethene $+ H_2O$	−1.56
propyne $+ H_2 \rightarrow$ propene	−2.00
propene $+ H_2 \rightarrow$ propane	−1.58
allene $+ 2H_2 \rightarrow$ propane	−3.64
iso-butane $\rightarrow$ trans-butane	0.08
$CO + H_2O \rightarrow$ formic acid	−0.39
$CH_4 + CO_2 \rightarrow$ acetic acid	0.28
$CH_4 + CO + H_2 \rightarrow$ ethanol	−0.91
1,3-cyclohexadiene $\rightarrow$ 1,4-cyclohexadiene	−0.01
benzene $+ H_2 \rightarrow$ 1,4-cyclohexadiene	−0.01
1,4-cyclohexadiene $+ 2H_2 \rightarrow$ cyclohexane	−2.94

Table A.2: The 27 experimental reaction energies ΔE for chemisorption on late transition metal surfaces constituting the CE27 dataset. The somewhat smaller CE17 dataset is a subset of CE27. Reactions in CE17 are marked with a "⋆". All chemisorption energies are in eV per adsorbate at a surface coverage of 0.25 ML, except where otherwise noted. The adsorption mode is indicated by "m" (molecular) or "d" (dissociative), along with the adsorption site. Chemisorption energies for O have been evaluated as $\frac{1}{2}\{\Delta E(O_2) - E_b(O_2)\}$ with $E_b(O_2) = 118$ kcal/mol[195] for the dioxygen bond energy.

		mode	site	ΔE	reference(s)
CO/Ni(111)	⋆	m	fcc	−1.28	183
CO/Ni(100)		m	hollow	−1.26	186
CO/Rh(111)	⋆	m	top	−1.45	183
CO/Pd(111)	⋆	m	fcc	−1.48	183
CO/Pd(100)[a]	⋆	m	bridge	−1.60	186–189
CO/Pt(111)	⋆	m	top	−1.37	183
CO/Ir(111)	⋆	m	top	−1.58	183
CO/Cu(111)	⋆	m	top	−0.50	183
CO/Co(0001)	⋆	m	top	−1.20	183
CO/Ru(0001)	⋆	m	top	−1.49	183
O/Ni(111)	⋆	m	fcc	−4.95	189
O/Ni(100)	⋆	m	hollow	−5.23	189
O/Rh(100)	⋆	m	hollow	−4.41	189
O/Pt(111)	⋆	m	fcc	−3.67	190
NO/Ni(100)	⋆	d	hollow	−3.99	186
NO/Pd(111)	⋆	m	fcc	−1.86	191
NO/Pd(100)	⋆	m	hollow	−1.61	192
NO/Pt(111)		m	fcc	−1.45	190
N_2/Fe(100)[b]		d	hollow	−2.3	193
H_2/Pt(111)	⋆	d	fcc	−0.41	194
H_2/Ni(111)		d	fcc	−0.98	194
H_2/Ni(100)		d	hollow	−0.93	194
H_2/Rh(111)		d	fcc	−0.81	194
H_2/Pd(111)		d	fcc	−0.91	194
H_2/Ir(111)		d	fcc	−0.55	194
H_2/Co(0001)		d	fcc	−0.69	194
H_2/Ru(0001)[c]		d	fcc	−1.04	194

[a] ΔE is the average of −1.58, −1.67, −1.69, and −1.45 eV.
[b] The coverage of atomic nitrogen is 0.5 ML.
[c] ΔE is the average of −0.83 and −1.24 eV, both from Ref. 194.

Table A.3: Experimental solid state properties of 27 cubic bulk solids. The ZPAE exclusive Sol27LC zero-Kelvin lattice constants a_0 (Å) are from Ref. 99, while the zero-Kelvin Sol27Ec cohesive energies E_c (eV/atom) are from Ref. 41 and corrected for ZPVE contributions. Strukturberichtsymbols are indicated in parenthesis for each solid. A1: fcc, A2: bcc, A3: hcp, A4: diamond.

	Sol27LC	Sol27Ec	
Solid	a_0	E_c	ZPVE[a]
Li (A2)	3.451	1.66	−0.033
Na (A2)	4.209	1.13	−0.015
K (A2)	5.212	0.94	−0.009
Rb (A2)	5.577	0.86	−0.005
Ca (A1)	5.556	1.86	−0.022
Sr (A1)	6.040	1.73	−0.014
Ba (A2)	5.002	1.91	−0.011
V (A2)	3.024	5.35	−0.037
Nb (A2)	3.294	7.60	−0.027
Ta (A2)	3.299	8.12	−0.023
Mo (A2)	3.141	6.86	−0.044
W (A2)	3.160	8.94	−0.039
Fe (A2)	2.853	4.33	−0.046
Rh (A1)	3.793	5.80	−0.047
Ir (A1)	3.831	6.98	−0.041
Ni (A1)	3.508	4.48	−0.044
Pd (A1)	3.876	3.92	−0.027
Pt (A1)	3.913	5.86	−0.023
Cu (A1)	3.596	3.52	−0.033
Ag (A1)	4.062	2.97	−0.022
Au (A1)	4.062	3.83	−0.016
Pb (A1)	4.912	2.04	−0.010
Al (A1)	4.019	3.43	−0.041
C (A4)	3.544	7.59	−0.216
Si (A4)	5.415	4.69	−0.063
Ge (A4)	5.639	3.89	−0.036
Sn (A4)	6.474	3.16	−0.019

[a] ZPVE corrections are calculated according to Eq. (4.5) using Debye temperatures from Ref. 41.

S22x5 interaction energies

The revised CCSD(T) interaction energies of Ref. 95 are used for the S22 dataset. However, all points on the 22 PECs in the S22x5 extension were originally calculated using a somewhat less accurate CCSD(T) procedure. In order to shift the equilibrium point on each S22x5 PEC to the revised S22 energies, and approximately correct the remaining data points, a modification of the (possibly) slightly inaccurate S22x5 interaction energies is here introduced as

$$E_{\mathrm{int}}^{d} := \varepsilon_{\mathrm{int}}^{d} \times \frac{E_{\mathrm{int}}^{1.0}}{\varepsilon_{\mathrm{int}}^{1.0}}, \tag{A.1}$$

where E_{int}^{d} and $\varepsilon_{\mathrm{int}}^{d}$ denote modified and original S22x5 energies at the relative intermolecular distance d, respectively. For $E_{\mathrm{int}}^{1.0} = \varepsilon_{\mathrm{int}}^{1.0}$ Eq. (A.1) obviously reduces to $E_{\mathrm{int}}^{d} = \varepsilon_{\mathrm{int}}^{d}$ for all distances. The obtained corrections to $\varepsilon_{\mathrm{int}}^{d}$ are listed in Table A.4. The maximum correction of 11.4% amounts to 25.6 meV for the indole–benzene complex in a stacked geometry, while the total mean signed correction to all the 110 interaction energies is 0.1 meV.

Table A.4: Corrections to the original S22x5 interaction energies, $E^d_{\text{int}} - \varepsilon^d_{\text{int}}$, computed from Eq. (A.1). Reported statistics are most negative (min), most positive (max), mean signed (msc), and mean absolute (mac) interaction energy correction at each distance. Furthermore, the total mean signed (MSC) and total mean absolute (MAC) energy correction over all 110 energies is reported in the bottom rows. All energies in meV.

		Relative interaction distance, d				
Complex	$E^{1.0}_{\text{int}}/\varepsilon^{1.0}_{\text{int}}$	0.9	1.0	1.2	1.5	2.0
1	−1.0%	−1.0	−1.3	−1.0	−0.5	−0.1
2	−1.0%	−1.9	−2.2	−1.8	−1.0	−0.4
3	−1.1%	−8.0	−9.1	−7.6	−4.5	−1.8
4	−1.1%	−6.5	−7.3	−6.1	−3.7	−1.6
5	−1.1%	−9.2	−10.0	−8.4	−5.1	−2.2
6	−1.8%	−11.8	−13.0	−10.8	−6.4	−2.5
7	−2.3%	−14.7	−16.0	−13.0	−7.3	−2.5
8	0.0%	0.0	0.0	0.0	0.0	0.0
9	−1.2%	−0.4	−0.8	−0.4	−0.1	0.0
10	3.2%	1.5	2.1	1.6	0.7	0.2
11	6.8%	0.4	8.3	5.7	1.6	0.2
12	6.9%	5.1	13.5	9.0	2.9	0.6
13	1.3%	3.8	5.6	3.6	1.4	0.4
14	11.4%	10.5	25.6	17.8	5.3	0.5
15	4.6%	15.9	24.3	16.4	6.5	1.8
16	−1.4%	−0.7	−0.9	−0.7	−0.3	−0.1
17	−0.6%	−0.8	−0.9	−0.7	−0.4	−0.1
18	1.3%	1.1	1.3	1.0	0.5	0.2
19	−0.7%	−1.2	−1.3	−1.1	−0.6	−0.2
20	3.2%	3.1	3.9	3.1	1.6	0.5
21	2.1%	4.5	5.2	4.4	2.5	1.0
22	−0.6%	−1.6	−1.8	−1.5	−0.9	−0.4
min	−2.3%	−14.7	−16.0	−13.0	−7.3	−2.5
max	11.4%	15.9	25.6	17.8	6.5	1.8
msc	1.2%	−0.5	1.1	0.4	−0.4	−0.3
mac	2.5%	4.7	7.0	5.3	2.4	0.8
MSC	0.1					
MAC	4.0					

A.2 Benchmark statistics

Table A.5: Deviation statistics for the Sol27Ec cohesive energies (eV/atom) and Sol27LC lattice constants (Å). Zero-point vibrational effects have been removed from both experimental datasets.

	Sol27Ec (27)			Sol27LC (27)		
Method	MSD	MAD	STD	MSD	MAD	STD
LDA	0.89	0.89	1.08	−0.07	0.07	0.10
PBE	−0.10	0.27	0.38	0.05	0.06	0.07
RPBE	−0.54	0.58	0.71	0.11	0.11	0.13
PBEsol	0.43	0.45	0.63	−0.01	0.03	0.04
BLYP	−0.79	0.80	0.89	0.11	0.11	0.14
AM05	0.25	0.36	0.51	0.01	0.03	0.04
WC	0.37	0.41	0.57	0.00	0.03	0.04
HCTH407	−0.59	0.67	0.82	0.08	0.10	0.14
TPSS	0.08	0.27	0.36	0.05	0.05	0.08
revTPSS	0.31	0.37	0.50	0.03	0.04	0.07
vdW-DF	−0.54	0.60	0.72	0.12	0.12	0.14
vdW-DF2	−0.58	0.64	0.75	0.12	0.14	0.18
optPBE-vdW	−0.12	0.27	0.38	0.06	0.08	0.10
optB88-vdW	0.01	0.25	0.36	0.04	0.08	0.09
C09-vdW	0.42	0.43	0.59	0.01	0.05	0.06
BEEF-vdW	−0.37	0.45	0.59	0.08	0.08	0.11

Table A.6: Deviation statistics on the G2/97, G3-3, and G3/99 thermochemical datasets, as well as the RE42 set of molecular reaction energies. All energies in eV.

	G2/97 (148)			G3-3 (75)			G3/99 (223)			RE42 (42)		
Method	MSD	MAD	STD	MSD	MAD	STD	MSD	MAD	STD	MSD	MAD	STD
LDA	−3.69	3.69	4.27	−8.35	8.35	8.78	−5.25	5.25	6.16	−0.55	1.06	1.62
PBE	−0.64	0.68	0.84	−1.32	1.32	1.48	−0.87	0.90	1.10	−0.08	0.30	0.42
RPBE	0.25	0.40	0.51	0.94	0.96	1.13	0.48	0.59	0.78	0.11	0.26	0.34
PBEsol	−1.69	1.70	2.00	−3.94	3.94	4.14	−2.45	2.45	2.90	−0.29	0.48	0.73
BLYP	0.00	0.32	0.43	0.57	0.62	0.76	0.19	0.42	0.56	0.16	0.29	0.37
AM05	−1.77	1.78	2.07	−4.00	4.00	4.19	−2.52	2.52	2.96	−0.21	0.41	0.62
WC	−1.24	1.26	1.51	−2.86	2.86	3.03	−1.79	1.80	2.14	−0.24	0.43	0.65
HCTH407	0.09	0.26	0.35	0.48	0.55	0.65	0.22	0.36	0.47	0.06	0.27	0.35
TPSS	−0.22	0.28	0.33	−0.26	0.29	0.33	−0.24	0.28	0.33	0.06	0.25	0.32
revTPSS	−0.21	0.28	0.34	−0.24	0.26	0.31	−0.22	0.27	0.33	0.16	0.33	0.43
vdW-DF	−0.10	0.24	0.33	0.18	0.24	0.32	−0.01	0.24	0.33	0.24	0.39	0.52
vdW-DF2	−0.15	0.28	0.39	0.11	0.26	0.36	−0.06	0.28	0.38	0.24	0.40	0.54
optPBE-vdW	−0.84	0.85	0.98	−1.72	1.72	1.82	−1.14	1.14	1.32	0.06	0.27	0.35
optB88-vdW	−1.04	1.04	1.20	−2.22	2.22	2.34	−1.44	1.44	1.68	0.02	0.26	0.34
C09-vdW	−1.55	1.55	1.80	−3.55	3.55	3.72	−2.22	2.22	2.61	−0.11	0.33	0.45
B3LYP[a]	0.05	0.14	0.19	0.36	0.37	0.41	0.15	0.21	0.28	−0.05	0.15	0.22
PBE0[a]	−0.10	0.21	0.28	−0.40	0.44	0.55	−0.20	0.29	0.39	0.13	0.33	0.47
BEEF-vdW	−0.02	0.16	0.24	0.19	0.25	0.31	0.05	0.19	0.27	0.14	0.29	0.37

[a] B3LYP and PBE0 data adapted from Ref. 46.

Table A.7: Deviation statistics on the CE17 and CE27 chemisorption energies, DBH24/08 reaction barriers, and the S22x5 interaction energies of noncovalently bonded complexes. All energies in eV, except S22x5 which is in meV.

	CE17 (17)			CE27 (27)			DBH24/08 (24)			S22x5 (110)		
Method	MSD	MAD	STD	MSD	MAD	STD	MSD	MAD	STD	MSD	MAD	STD
LDA	−1.34	1.34	1.39	−1.33	1.33	1.42	−0.58	0.58	0.73	−50	62	110
PBE	−0.42	0.42	0.44	−0.40	0.40	0.43	−0.33	0.33	0.43	76	76	132
RPBE	−0.02	0.11	0.13	0.00	0.11	0.14	−0.27	0.27	0.34	138	138	227
PBEsol	−0.85	0.85	0.87	−0.85	0.85	0.89	−0.44	0.44	0.56	38	53	85
BLYP	−0.04	0.13	0.16	0.02	0.15	0.18	−0.33	0.33	0.39	140	140	218
AM05	−0.70	0.70	0.73	−0.69	0.69	0.73	−0.41	0.41	0.53	99	99	157
WC	−0.76	0.76	0.78	−0.76	0.76	0.80	−0.41	0.41	0.52	56	63	105
HCTH407	0.11	0.17	0.22	0.15	0.20	0.30	−0.19	0.21	0.31	115	116	218
TPSS	−0.32	0.32	0.37	−0.34	0.34	0.41	−0.35	0.35	0.41	100	100	162
revTPSS	−0.38	0.38	0.43	−0.38	0.38	0.45	−0.35	0.35	0.41	92	92	141
vdW-DF	−0.05	0.12	0.14	0.04	0.18	0.22	−0.27	0.28	0.34	39	52	87
vdW-DF2	−0.04	0.13	0.18	0.07	0.20	0.26	−0.30	0.31	0.37	31	33	61
optPBE-vdW	−0.39	0.39	0.42	−0.31	0.35	0.40	−0.33	0.33	0.41	−4	21	29
optB88-vdW	−0.52	0.52	0.56	−0.44	0.45	0.52	−0.37	0.37	0.45	3	10	15
C09-vdW	−0.78	0.78	0.81	−0.73	0.73	0.79	−0.41	0.41	0.50	−5	12	18
B3LYP	–	–	–	–	–	–	−0.17	0.17	0.21	111	111	180
PBE0	–	–	–	–	–	–	−0.13	0.15	0.19	71	71	124
BEEF-vdW	−0.08	0.12	0.14	−0.01	0.16	0.19	−0.26	0.26	0.33	42	50	88

Table A.8: Detailed statistics on the deviations of calculated S22x5 interaction energies from CCSD(T) benchmarks using van der Waals density functionals in all five points along the intermolecular potential-energy curve. Mean signed and mean absoulte deviations are in meV. Mean signed relative (MSRD) and mean absolute relative (MARD) deviations are also listed.

Method	MSD	MAD	MSRD	MARD
$\boldsymbol{d = 0.9}$				
vdW-DF	140	140	198%	198%
vdW-DF2	99	99	143%	143%
optPBE-vdW	29	31	28%	35%
optB88-vdW	17	19	26%	26%
C09-vdW	−13	21	−13%	35%
BEEF-vdW	136	137	214%	214%
$\boldsymbol{d = 1.0}$				
vdW-DF	70	71	20%	25%
vdW-DF2	43	44	13%	15%
optPBE-vdW	−1	20	−9%	13%
optB88-vdW	5	13	3%	6%
C09-vdW	−3	13	1%	6%
BEEF-vdW	72	74	20%	28%
$\boldsymbol{d = 1.2}$				
vdW-DF	4	32	−16%	23%
vdW-DF2	5	13	−2%	7%
optPBE-vdW	−25	28	−29%	30%
optB88-vdW	−4	13	−6%	9%
C09-vdW	−3	13	−8%	11%
BEEF-vdW	6	27	−12%	18%
$\boldsymbol{d = 1.5}$				
vdW-DF	−13	15	−39%	40%
vdW-DF2	2	4	4%	6%
optPBE-vdW	−20	20	−44%	44%
optB88-vdW	−3	6	−12%	13%
C09-vdW	−6	11	−26%	28%
BEEF-vdW	−5	6	−13%	14%
$\boldsymbol{d = 2.0}$				
vdW-DF	−4	4	−20%	20%
vdW-DF2	5	5	34%	34%
optPBE-vdW	−5	5	−20%	21%
optB88-vdW	1	2	3%	8%
C09-vdW	−2	2	−13%	15%
BEEF-vdW	2	3	27%	28%

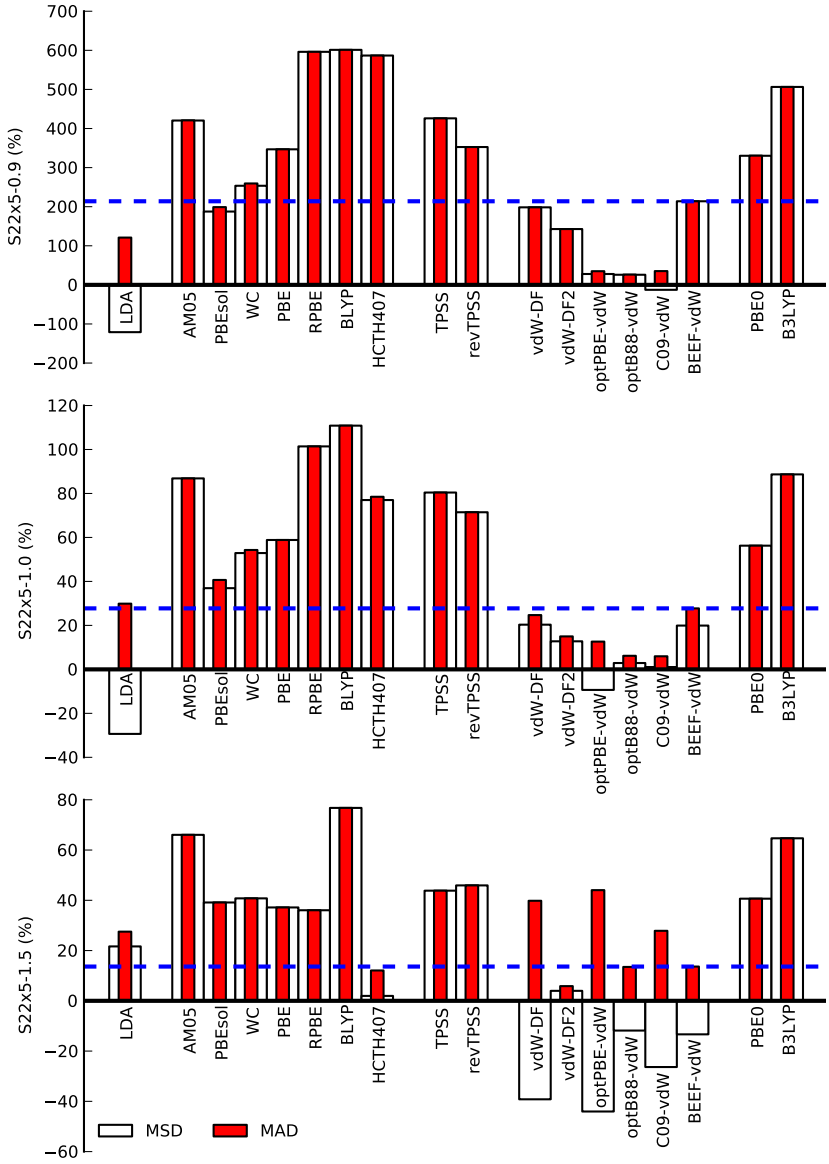

Figure A.1: Bar plot comparison of relative deviations in density functional calculations of the S22x5 CCSD(T) interaction energies. The dashed blue line indicates the BEEF-vdW MARD.

A.3 Graphene adsorption on Ni(111): Computational details

Adsorption of graphene on Ni(111) was modelled using a $1 \times 1 \times 5$ surface slab, a Ni(fcc) lattice constant of 3.524 Å as determined with the PBE density functional, and 20 Å vacuum width. The top three atomic layers were fully relaxed with PBE using a grid spacing of 0.15 Å and a $(20 \times 20 \times 1)$ k-point mesh. Carbon atoms were placed in atop and fcc adsorption sites, respectively.

Bibliography

[1] J. Bernholc. Computational materials science: The era of applied quantum mechanics. *Phys. Today*, 52(9):30, 1999.

[2] J. Hafner, C. Wolverton, and G. Ceder. Toward computational materials design: The impact of density functional theory on materials research. *MRS Bull.*, 31 (9):659, 2006.

[3] B. Hammer and J. K. Nørskov. Theoretical surface science and catalysis – calculations and concepts. In B. C. Gates and H. Knözinger, editors, *Impact of Surface Science on Catalysis*, volume 45 of *Advances in Catalysis*, page 71. Academic Press, 2000.

[4] M. J. Field. Simulating enzyme reactions: Challenges and perspectives. *J. Comput. Chem.*, 23(1):48, 2002.

[5] K. M. Neyman and F. Illas. Theoretical aspects of heterogeneous catalysis: Applications of density functional methods. *Catal. Today*, 105:2, 2005.

[6] R. A. Friesner and V. Guallar. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.*, 56:389, 2005.

[7] H. M. Senn and W. Thiel. QM/MM Methods for biomolecular systems. *Angew. Chem. Int. Ed.*, 48(7):1198, 2009.

[8] M. Neurock. Perspectives on the first principles elucidation and the design of active sites. *J. Catal.*, 216:73, 2003.

[9] J. Greeley and M. Mavrikakis. Alloy catalysts designed from first principles. *Nat. Mater.*, 3(11):810, 2004.

[10] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen. Towards the computational design of solid catalysts. *Nat. Chem.*, 1:37, 2009.

[11] J. K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard. Density functional theory in surface chemistry and catalysis. *PNAS*, 108:937, 2011.

[12] G. Ertl. Reactions at surfaces: From atoms to complexity (Nobel lecture). *Angew. Chem. Int. Ed.*, 47(19):3524, 2008.

[13] I. Chorkendorff and J. W. Niemantsverdriet. *Concepts of Modern Catalysis and Kinetics*. Wiley-VCH, second edition, 2007.

[14] L. E. Ballentine. *Quantum Mechanics: A Modern Development*. World Scientific Publishing Co. Pte. Ltd., Singapore, 1998.

[15] R. M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004.

[16] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 84(20):457, 1927.

[17] J. P. Perdew and S. Kurth. Density functionals for non-relativistic Coulomb systems in the new century. In M. Marques, C. Fiolhais, and M. Marques, editors, *A Primer in Density Functional Theory*, page 1. Springer, 2003.

[18] T. Helgaker, P. Jørgensen and J. Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons Ltd., Chichester, 2000.

[19] P. Hohenberg and W. Kohn. Inhomogenous electron gas. *Phys. Rev. B*, 136(3B): B864, 1964.

[20] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133, 1965.

[21] W. Kohn. Nobel lecture: Electronic structure of matter — wave functions and density functionals. *Rev. Mod. Phys.*, 71:1253, 1999.

[22] D. S. Sholl and J. A. Steckel. *Density Functional Theory: A Practical Introduction*. Wiley-Interscience, first edition, 2009.

[23] O. Gunnarsson and B. I. Lundqvist. Exchange and correlation in atoms, molecules, and solids by spin-density functional formalism. *Phys. Rev. B*, 13 (10):4274, 1976.

[24] D. C. Langreth and J. P. Perdew. Exchange-correlation energy of a metallic surface. *Solid State Commun.*, 17(11):1425, 1975.

[25] D. C. Langreth and J. P. Perdew. Exchange-correlation energy of a metallic surface — wave-vector analysis. *Phys. Rev. B*, 15(6):2884, 1977.

[26] Y. M. Niquet, M. Fuchs, and X. Gonze. Exchange-correlation potentials in the adiabatic connection fluctuation-dissipation framework. *Phys. Rev. A*, 68: 032507, 2003.

[27] J. F. Dobson and T. Gould. Calculation of dispersion energies. *J. Phys.: Condens. Matter*, 24(7):073201, 2012.

[28] J. Harl and G. Kresse. Accurate bulk properties from approximate many-body techniques. *Phys. Rev. Lett.*, 103:056401, 2009.

[29] B. G. Janesko, T. M. Henderson, and G. E. Scuseria. Long-range-corrected hybrids including random phase approximation correlation. *J. Chem. Phys.*, 130(8):081105, 2009.

[30] S. Lebègue, J. Harl, Tim Gould, J. G. Ángyán, G. Kresse, and J. F. Dobson. Cohesive properties and asymptotics of the dispersion interaction in graphite by the random phase approximation. *Phys. Rev. Lett.*, 105:196401, 2010.

[31] D. Lu, H.-V. Nguyen, and G. Galli. Power series expansion of the random phase approximation correlation energy: The role of the third- and higher-order contributions. *J. Chem. Phys.*, 133(15):154110, 2010.

[32] J. C. Slater. Wavefunction in a periodic potential. *Phys. Rev*, 51:846, 1937.

[33] P. E. Blöchl, C. J. Först, and J. Schimpl. Projector augmented wave method: ab initio molecular dynamics with full wave funcsions. *Bull. Mater. Sci.*, 26(1): 33, 2003.

[34] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B*, 71(3):035109, 2005.

[35] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys.: Condens. Matter*, 22:253202, 2010.

[36] The GPAW and ASE codes are available as parts of the CAMPOS software: `http://www.camd.dtu.dk/Software`.

[37] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953, 1994.

[38] S. R. Bahn and K. W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.*, 4:56, 2002.

[39] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758, 1999.

[40] H. J. Monkhorst and J. D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13:5188, 1976.

[41] C. Kittel. *Indtroduction to Solid State Physics*. John Wiley & Sons, Inc., 8th edition, 2005.

[42] A. Salam. *Molecular Quantum Electrodynamics: Long-Range Intermolecular Interactions*. Wiley, first edition, 2009.

[43] B. Jeziorski, R. Moszynski, and K. Szalewicz. Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.*, 94(7):1887, 1994.

[44] J. P. Perdew and K. Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. In V. Van Doren, C. Van Alsenoy, and P. Geerlings, editors, *Density Functional Theory and its Application to Materials*, volume 577 of *AIP Conference Proceedings*, page 1, 2001.

[45] J. P. Perdew, A. Ruzsinszky, J. M. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *J. Chem. Phys.*, 123(6):062201, 2005.

[46] V. N. Staroverov, G. E. Scuseria, J. M. Tao, and J. P. Perdew. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.*, 119(23):12129, 2003.

[47] E. H. Lieb and S. Oxford. Improved Lower Bound on the Indirect Coulomb Energy. *Int. J. Quantum Chem.*, 19:427, 1981.

[48] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.

[49] C. Lee, W. Yang, and R. G. Parr. Development of the Colic-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785, 1988.

[50] G. I. Csonka, J. P. Perdew, A. Ruzsinszky, P. H. T. Philipsen, S. Lebegue, J. Paier, O. A. Vydrov, and J. G. Angyan. Assessing the performance of recent density functionals for bulk solids. *Phys. Rev. B*, 79:155107, 2009.

[51] J. Klimes, D. R. Bowler, and A. Michaelides. Van der Waals density functionals applied to solids. *Phys. Rev. B*, 83:195131, 2011.

[52] J. P. Perdew. Unified theory of exchange and correlation beyond the local density approximation. In P. Ziesche and H. Eschrig, editors, *Electronic Structure of Solids '91*, volume 17, page 11, Berlin, 1991. Akademie Verlag.

[53] J. P. Perdew, K. Burke, and Y. Wang. Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Phys. Rev. B*, 54:16533, 1996.

[54] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100:136406, 2008.

[55] B. Hammer, L. B. Hansen, and J. K. Nørskov. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B*, 59:7413, 1999.

[56] Y. K. Zhang and W. T. Yang. Comment on "Generalized gradient approximation made simple". *Phys. Rev. Lett.*, 80(4):890, 1998.

[57] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098, 1988.

[58] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, 91:146401, 2003.

[59] R. Neumann, R. H. Nobes, and N. C. Handy. Exchange functionals and potentials. *Mol. Phys.*, 87(1):1, 1996.

[60] T. Van Voorhis and G. E. Scuseria. A novel form for the exchange-correlation energy functional. *J. Chem. Phys.*, 109(2):400, 1998.

[61] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun. Workhorse semilocal density functional for condensed matter physics and quantum chemistry. *Phys. Rev. Lett.*, 103:026403, 2009.

[62] Y. Zhao and D. G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.*, 125:194101, 2006.

[63] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.

[64] A. D. Becke. Density-functional thermochemistry .5. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.*, 107(20):8554, 1997.

[65] Y. Zhao and D. G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Account.*, 120:215, 2008.

[66] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.*, 118(18):8207, 2003.

[67] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98:11623, 1994.

[68] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110:6158, 1999.

[69] L. Schimka, J. Harl, and G. Kresse. Improved hybrid functional for solids: The HSEsol functional. *J. Chem. Phys.*, 134:024116, 2011.

[70] G. E. Scuseria and V. N. Staroverov. Recent progress in the development of exchange-correlation functionals. In C. Dykstra *et al.*, editor, *Theory and Applications of Computational Chemistry: The First Forty Years*, page 669. Elsevier B.V., 2005.

[71] E. A. Carter. Challenges in modeling materials properties without experimental input. *Science*, 321(5890):800, 2008.

[72] S. Kristyán and P. Pulay. Can (semi)local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.*, 229(3):175, 1994.

[73] F. O. Kannemann and A. D. Becke. Van der Waals interactions in density-functional theory: Rare-gas diatomics. *J. Chem. Theory Comput.*, 5(4):719, 2009.

[74] D. J. Lacks and R. G. Gordon. Pair interactions of rare-gas atoms as a test of exchange-energy-density functionals in regions of large density gradients. *Phys. Rev. A*, 47:4681, 1993.

[75] L. Ferrighi, G. K. H. Madsen, and B. Hammer. Self-consistent meta-generalized gradient approximation study of adsorption of aromatic molecules on noble metal surfaces. *J. Chem. Phys.*, 135(8):084704, 2011.

[76] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048, 1981.

[77] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Challenges for density functional theory. *Chem. Rev.*, 112(1):289, 2012.

[78] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople. Assessment of Gaussian-3 and density functional theories for a larger experimental test set. *J. Chem. Phys.*, 112:7374, 2000.

[79] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople. Assessment of Gaussian-2 and density functional methods for the computation of enthalpies of formation. *J. Chem. Phys.*, 106:1063, 1997.

[80] J. Zheng, Y. Zhao, and D. G. Truhlar. The DBH24/08 Database and Its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights. *J. Chem. Theory Comput.*, 5(4):808, 2009.

[81] J. Zheng, Y. Zhao, and D. G. Truhlar. Representative benchmark suites for barrier heights of diverse reaction types and assessment of electronic structure methods for thermochemical kinetics. *J. Chem. Theory Comput.*, 3:569, 2007.

[82] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.*, 8:1985, 2006.

[83] A. Gulans, M. J. Puska, and R. M. Nieminen. Linear-scaling self-consistent implementation of the van der Waals density functional. *Phys. Rev. B*, 79: 201105(R), 2009.

[84] V. R. Cooper. Van der Waals density functional: An appropriate exchange functional. *Phys. Rev. B*, 81:161104(R), 2010.

[85] F. O. Kannemann and A. D. Becke. van der Waals interactions in density-functional theory: Intermolecular complexes. *J. Chem. Theory Comput.*, 6:1081, 2010.

[86] L. A. Burns, A. Vazquez-Mayagoitia, B. G. Sumpter, and C. D. Sherrill. Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (DFT-D), exchange-hole dipole moment (XDM) theory, and specialized functionals. *J. Chem. Phys.*, 134:084107, 2011.

[87] F. Hanke. Sensitivity analysis and uncertainty calculation for dispersion corrected density functional theory. *J. Comput. Chem.*, 32:1424, 2011.

[88] L. Goerigk and S. Grimme. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.*, 13:6670, 2011.

[89] J. Klimes, D. R. Bowler, and A. Michaelides. Chemical accuracy for the van der Waals density functional. *J. Phys.: Condens. Matter*, 22:022201, 2010.

[90] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth. Higher-accuracy van der Waals density functional. *Phys. Rev. B*, 82:081101, 2010.

[91] O. A. Vydrov and T. Van Voorhis. Nonlocal van der Waals density functional: The simpler the better. *J. Chem. Phys.*, 133:244103, 2010.

[92] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132:154104, 2010.

[93] J. Rezac, K. E. Riley, and P. Hobza. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.*, 7(8):2427, 2011.

[94] L. Goerigk, H. Kruse, and S. Grimme. Benchmarking density functional methods against the S66 and S66x8 datasets for non-covalent interactions. *Chem. Phys. Chem.*, 12(17):3421, 2011.

[95] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill. Basis set consistent revision of the S22 test set of noncovalent interaction energies. *J. Chem. Phys.*, 132:144104, 2010.

[96] R. Podeszwa, K. Patkowski, and K. Szalewicz. Improved interaction energy benchmarks for dimers of biological relevance. *Phys. Chem. Chem. Phys.*, 12: 5974, 2010.

[97] L. Grafova, M. Pitonak, J. Rezac, and P. Hobza. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. *J. Chem. Theory Comput.*, 6:2365, 2010.

[98] A. B. Alchagirov, J. P. Perdew, J. C. Boettger, R. C. Albers, and C. Fiolhais. Energy and pressure versus volume: Equations of state motivated by the stabilized jellium model. *Phys. Rev. B*, 63:224115, 2001.

[99] P. Haas, F. Tran, and P. Blaha. Calculation of the lattice constant of solids with semilocal functionals. *Phys. Rev. B*, 79:085104, 2009.

[100] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der Waals density functional for general geometries. *Phys. Rev. Lett.*, 92:246401, 2004.

[101] T. Thonhauser, V. R. Cooper, S. Li, A. Puzder, P. Hyldgaard, and D. C. Langreth. Van der Waals density functional: Self-consistent potential and the nature of the van der Waals bond. *Phys. Rev. B*, 76:125112, 2007.

[102] D. C. Langreth, B. I. Lundqvist, S. D. Chakarova-Kack, V. R. Cooper, M. Dion, P. Hyldgaard, A. Kelkkanen, J. Kleis, L. Kong, S. Li, P. G. Moses, E. Murray, A. Puzder, H. Rydberg, E. Schröder, and T. Thonhauser. A density functional for sparse matter. *J. Phys.: Condens. Matter*, 21(8):084203, 2009.

[103] O. A. Vydrov, Q. Wu, and T. Van Voorhis. Self-consistent implementation of a nonlocal van der Waals density functional with a Gaussian basis set. *J. Chem. Phys.*, 129(014106), 2008.

[104] G. Román-Pérez and J. M. Soler. Efficient implementation of a van der Waals density functional: Application to double-wall carbon nanotubes. *Phys. Rev. Lett.*, 103:096102, 2009.

[105] JuNoLo – Jülich nonlocal code for parallel post-processing evaluation of vdW-DF correlation energy. *Comput. Phys. Commun.*, 181(2):71, 2010.

[106] X. Wu, M. C. Vargas, S. Nayak, V. Lotrich, and G. Scoles. Towards extending the applicability of density functional theory to weakly bound systems. *J. Chem. Phys.*, 115(19):8748, 2001.

[107] E. D. Murray and D. C. Lee, K. Langreth. Investigation of exchange energy density functional accuracy for interacting molecules. *J. Chem. Theory Comput.*, 5(10):2754, 2009.

[108] A. Puzder, M. Dion, and D. C. Langreth. Binding energies in benzene dimers: Nonlocal density functional calculations. *J Chem. Phys.*, 124(16):164105, 2006.

[109] T. Thonhauser, A. Puzder, and D. C. Langreth. Interaction energies of monosubstituted benzene dimers via nonlocal density functional theory. *J. Chem. Phys.*, 124(16):164106, 2006.

[110] I. Hamada and M. Otani. Comparative van der Waals density-functional study of graphene on metal surfaces. *Phys. Rev. B*, 82:153412, 2010.

[111] J. P. Perdew and W. Yue. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B*, 33: 8800, 1986.

[112] In fact, a refitted version [107] of PW86 exchange was shown to yield slightly better vdW-DF2 results.

[113] O. A. Vydrov and T. Van Voorhis. Improving the accuracy of the nonlocal van der Waals density functional with minimal empiricism. *J. Chem. Phys.*, 130(10): 104105, 2009.

[114] O. A. Vydrov and T. Van Voorhis. Nonlocal van der Waals density functional made simple. *Phys. Rev. Lett.*, 103:063004, 2009.

[115] O. A. Vydrov and T. Van Voorhis. Implementation and assessment of a simple nonlocal van der Waals density functional. *J. Chem. Phys.*, 132:164113, 2010.

[116] W. Hujo and S. Grimme. Performance of the van der Waals density functional VV10 and (hybrid)GGA variants for thermochemistry and noncovalent interactions. *J. Chem. Theory Comput.*, 7(12):3866, 2011.

[117] S. Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.*, 25(12):1463, 2004.

[118] S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787, 2006.

[119] A. D. Becke and E. R. Johnson. A density-functional model of the dispersion interaction. *J. Chem. Phys.*, 123(15), 2005.

[120] A. D. Becke and E. R. Johnson. Exchange-hole dipole moment and the dispersion interaction: High-order dispersion coefficients. *J. Chem. Phys.*, 124, 2006.

[121] A. Tkatchenko and M. Scheffler. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.*, 102:073005, 2009.

[122] T. Sato and H. Nakai. Density functional method including weak interactions: Dispersion coefficients based on the local response approximation. *J. Chem. Phys.*, 131(22):224104, 2009.

[123] P. G. Moses, J. J. Mortensen, B. I. Lundqvist, and J. K. Nørskov. Density functional study of the adsorption and van der Waals binding of aromatic and conjugated compounds on the basal plane of MoS2. *J. Chem. Phys.*, 130(10): 104709, 2009.

[124] D. R. Lide, editor. *CRC Handbook of Chemistry and Physics, 90th Edition (Internet Version 2010)*. CRC Press/Taylor and Francis, Boca Ranton, FL.

[125] R. Dudde, K. H. Frank, and E. E. Koch. The electronic structure of benzene adsorbed on Ag(111) studied by angle resolved photoemission. *Surf. Sci.*, 225 (3):267, 1990.

[126] F. P. Netzer. Surface structure and reactivity of aromatic molecules: Can we track down trends in the Periodic Table? *Langmuir*, 7(11):2544, 1991.

[127] M. Xi, M. X. Yang, S. K. Jo, B. E. Bent, and P. Stevens. Benzene adsorption on Cu(111): Formation of a stable layer. *J. Chem. Phys.*, 101(10):9122, 1994.

[128] K. Weiss, S. Gebert, M. Wühn, H. Wadepohl, and C. Wöll. Near edge X-ray absorption fine structure study of benzene adsorbed on metal surfaces: Comparison to benzene cluster complexes. *J. Vac. Sci. Technol. A*, 16(3):1017, 1998.

[129] C. Wöll. Small hydrocarbons on metal surfaces: Adsorption-induced changes in electronic and geometric structure as seen by X-ray absorption spectroscopy. *J. Synchrotron Radiat.*, 8:129, 2001.

[130] D. Syomin, J. Kim, G. B. Ellison, and B. E. Koel. Identification of adsorbed phenyl (C6H5) groups on metal surfaces: Electron-induced dissociation of benzene on Au(111). *J. Phys. Chem. B*, 105(35):8387, 2001.

[131] P. S. Bagus, K. Hermann, and C. Woll. The interaction of C_6H_6 and C_6H_{12} with noble metal surfaces: Electronic level alignment and the origin of the interface dipole. *J. Chem. Phys.*, 123(18):184109, 2005.

[132] T. J. Rockey, M. Yang, and H. L. Dai. Adsorption energies, inter-adsorbate interactions, and the two binding sites within monolayer benzene on Ag(111). *J. Phys. Chem. B*, 110(40):19973, 2006.

[133] R. Caputo, B. P. Prascher, V. Staemmler, P. S. Bagus, and C. Wöll. Adsorption of benzene on coinage metals: A theoretical analysis using wavefunction-based methods. *J. Phys. Chem. A*, 111(49):12778, 2007.

[134] X.-L. Zhou, M. E. Castro, and J. M. White. Interactions of UV photons and low energy electrons with chemisorbed benzene on Ag(111). *Surf. Sci.*, 238:215, 1990.

[135] H. Koschel, G. Held, and H. P. Steinruck. The orientation of benzene on bimetallic surfaces. *Surf. Rev. Lett.*, 6(5):893, 1999.

[136] V. Petzold, T. Bligaard, and K. W. Jacobsen. Construction of new electronic density functionals with error estimation through fitting. *Top. Catal.* (Accepted).

[137] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1 edition, 2006.

[138] M. E. Tipping. Bayesian inference: An introduction to principles and practice in machine learning. In Bousquet, O. and VonLuxburg, U. and Ratsch, G., editor, *Advanced Lectures On Machine Learning*, volume 3176 of *Lecture Notes In Artificial Intelligence*, page 41, 2004.

[139] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2 edition, 2009.

[140] C. M. Bishop. A new framework for machine learning. In Zurada, J. M. and Yen, G. G. and Wang, J., editor, *Computational Intelligence: Research Frontiers*, volume 5050 of *Lecture Notes In Computer Science*, page 1, 2008.

[141] K. S. Brown and J. P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, 68:021904, 2003.

[142] V. Petzold. *New density functionals with error estimation applied to atomic-scale systems.* Phd thesis, Department of Physics, Technical University of Denmark, 2010.

[143] S. Borra and A. Di Ciaccio. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Comput. Stat. Data Anal.*, 54(12):2976, 2010.

[144] J. A. Khan, S. Van Aelst, and R. H. Zamar. Fast robust estimation of prediction error based on resampling. *Comput. Stat. Data Anal.*, 54(12):3121, 2010.

[145] B. Efron. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Amer. Statistical Assoc.*, 78:316, 1983.

[146] P. Haas, F. Tran, P. Blaha, and K. Schwarz. Construction of an optimal GGA functional for molecules and solids. *Phys. Rev. B*, 83:205117, 2011.

[147] R. Peverati and D. G. Truhlar. M11-L: A local density functional that provides improved accuracy for electronic structure calculations in chemistry and physics. *J. Phys. Chem. Lett.*, 3(1):117, 2012.

[148] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.*, 109(15): 6264, 1998.

[149] A. D. Boese and N. C. Handy. A new parametrization of exchange–correlation generalized gradient approximation functionals. *J. Chem. Phys.*, 114:5497, 2001.

[150] J. P. Perdew and Y. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45:13244, 1992.

[151] A. N. Tikhonov, A. Goncharsky, V. V. Stepanov, and A. G. Yagola. *Numerical Methods for the Solution of Ill-Posed Problems.* Mathematics and Its Applications. Kluwer Academic Publishers, 2010.

[152] If one or more of the individual cost functions have very strong modes, several stationary points of the product cost function can exist. In this case one must carefully determine all stationary points, and select the one which represents the best compromise. For the subsequent fitting of the BEEF-vdW, this was a minor issue.

[153] J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen, J. K. Nørskov, J. P. Sethna, and K. W. Jacobsen. Bayesian error estimation in density-functional theory. *Phys. Rev. Lett.*, 95:216401, 2005.

[154] B. G. Janesko. Rung 3.5 density functionals. *J. Chem. Phys.*, 133(10), 2010.

[155] B. G. Janesko and A. Aguero. Nonspherical model density matrices for Rung 3.5 density functionals. *J. Chem. Phys.*, 136(2), 2012.

[156] Z. Wu and R. E. Cohen. More accurate generalized gradient approximation for solids. *Phys. Rev. B*, 73:235116, 2006.

[157] R. Armiento and A. E. Mattsson. Functional designed to include surface effects in self-consistent density functional theory. *Phys. Rev. B*, 72:085108, 2005.

[158] D. D. Landis, J. S. Hummelshøj, S. Nestorov, J. Greeley, M. Dułak, T. Bligaard, J. K. Nørskov, and K. Jacobsen. The Computational Materials Repository. *Comput. Sci. Eng.*, 99, 2012. DOI:10.1109/MCSE.2012.16.

[159] The CMR web-interface is found at `http://cmr.fysik.dtu.dk`.

[160] A. D. Boese and J. M. L. Martin. Development of density functionals for thermochemical kinetics. *J. Chem. Phys.*, 121(8):3405, 2004.

[161] J. Sun, M. Marsman, G. I. Csonka, A. Ruzsinszky, P. Hao, Y.-S. Kim, G. Kresse, and J. P. Perdew. Self-consistent meta-generalized gradient approximation within the projector-augmented-wave method. *Phys. Rev. B*, 84:035117, 2011.

[162] P. Hao, Y. Fang, J. Sun, G. I. Csonka, P. H. T. Philipsen, and J. P. Perdew. Lattice constants from semilocal density functionals with zero-point phonon correction. *Phys. Rev. B*, 85:014111, 2012.

[163] T. Olsen, J. Yan, J. J. Mortensen, and K. S. Thygesen. Dispersive and covalent interactions between graphene and metal surfaces from the random phase approximation. *Phys. Rev. Lett.*, 107:156401, 2011.

[164] F. Mittendorfer, A. Garhofer, J. Redinger, J. Klimes, J. Harl, and G. Kresse. Graphene on Ni(111): Strong interaction and weak adsorption. *Phys. Rev. B*, 84:201401(R), 2011.

[165] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov. Electric field effect in atomically thin carbon films. *Science*, 306(5696):666, 2004.

[166] A. K. Geim and K. S. Novoselov. The rise of graphene. *Nat. Mater.*, 6(3):183, 2007.

[167] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim. The electronic properties of graphene. *Rev. Mod. Phys.*, 81:109, 2009.

[168] F. Schwierz. Graphene transistors. *Nat. Nanotech.*, 5(7):487, 2010.

[169] J. Wintterlin and M.-L. Bocquet. Graphene on metal surfaces. *Surf. Sci.*, 603: 1841, 2009.

[170] Y. S. Dedkov and M. Fonin. Electronic and magnetic properties of the graphene-ferromagnet interface. *New J. Phys.*, 12:125004, 2010.

[171] Y. Gamo, A. Nagashima, M. Wakabayashi, M. Teria, and C. Oshima. Atomic structure of monolayer graphite formed on Ni(111). *Surf. Sci.*, 374:61, 1997.

[172] A. Grüneis and D. V. Vyalikh. Tunable hybridization between electronic states of graphene and a metal surface. *Phys. Rev. B*, 77:193401, 2008.

[173] A. Varykhalov, J. Sánchez-Barriga, A. M. Shikin, C. Biswas, E. Vescovo, A. Rybkin, D. Marchenko, and O. Rader. Electronic and magnetic properties of quasifreestanding graphene on Ni. *Phys. Rev. Lett.*, 101:157601, 2008.

[174] G. Bertoni, L. Calmels, A. Altibelli, and V. Serin. First-principles calculation of the electronic structure and EELS spectra at the graphene/Ni(111) interface. *Phys. Rev. B*, 71:075402, 2005.

[175] G. Kalibaeva, R. Vuilleumier, S. Meloni, A. Alavi, G. Ciccotti, and R. Rosei. Ab initio simulation of carbon clustering on an Ni(111) surface: A model of the poisoning of nickel-based catalysts. *J. Phys. Chem. B*, 110(8):3638, 2006.

[176] G. Giovannetti, P. A. Khomyakov, G. Brocks, V. M. Karpan, J. van den Brink, and P. J. Kelly. Doping graphene with metal contacts. *Phys. Rev. Lett.*, 101: 026803, 2008.

[177] M. Fuentes-Cabrera, M. I. Baskes, A. V. Melechko, and M. L. Simpson. Bridge structure for the graphene/Ni(111) system: A first principles study. *Phys. Rev. B*, 77:035405, 2008.

[178] M. Vanin, J. J. Mortensen, A. K. Kelkkanen, J. M. Garcia-Lastra, K. S. Thygesen, and K. W. Jacobsen. Graphene on metals: A van der Waals density functional study. *Phys. Rev. B*, 81:081408, 2010.

[179] C. Gong, G. Lee, B. Shan, E. M. Vogel, R. M. Wallace, and K. Cho. First-principles study of metal–graphene interfaces. *J. Appl. Phys.*, 108:123711, 2010.

[180] L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse. Accurate surface and adsorption energies from many-body perturbation theory. *Nat. Mater.*, 9:741, 2010.

[181] W. R. Tyson and W. A. Miller. Surface free energies of solid metals: Estimation from liquid surface tension measurements. *Surf. Sci.*, 62:267, 1977.

[182] L. Vitos, A. V. Ruban, H. L. Skriver, and J. Kollar. The surface energy of metals. *Surf. Sci.*, 411:186, 1998.

[183] F. Abild-Pedersen and M. P. Andersson. CO adsorption energies on metals with correction for high coordination adsorption sites – A density functional study. *Surf. Sci.*, 601:1747, 2007.

[184] J. Sun, M. Marsman, A. Ruzsinszky, G. Kresse, and J. P. Perdew. Improved lattice constants, surface energies, and CO desorption energies from a semilocal density functional. *Phys. Rev. B*, 83:121410(R), 2011.

[185] J. Sun, B. Xiao, and A. Ruzsinszky. Effect of orbital-overlap dependence in density functionals. *ArXiv e-prints*, 2012.

[186] W. A. Brown, R. Kose, and D. A. King. Femtomole adsorption calorimetry on single-crystal surfaces. *Chem. Rev.*, 98:797, 1998.

[187] H. Conrad, G. Ertl, J. Koch, and E. E. Latta. Adsorption on co on pd single crystal surfaces. *Surf. Sci.*, 43:462, 1974.

[188] R. J. Behm, K. Christmann, and G. Ertl. Adsorption of co on pd(100). *J. Chem. Phys.*, 73:2984, 1980.

[189] Q. Ge, R. Kose, and D. A. King. Adsorption energetics and bonding from femtomole calorimetry and from first principles theory. *Adv. Catal.*, 45:207, 2000.

[190] V. Fiorin, D. Borthwick, and D. A. King. Microcalorimetry of o2 and no on flat and stepped platinum surfaces. *Surf. Sci.*, 603:1360, 2009.

[191] R. D. Ramsier, Q. Gao, H. N. Neergaard Waltenburg, K.-W. Lee, O. W. Nooij, L. Lefferts, and J. T. Yates Jr. No adsorption and thermal behavior on pd surfaces. a detailed comparative study. *Surf. Sci.*, 320:209, 1994.

[192] Y. Y. Yeo, L. Vattuone, and D. A. King. *J. Chem. Phys.*, 106:1990, 1997.

[193] F. Bozso, G. Ertl, M. Grunze, and M. Weiss. Interaction of nitrogen with iron surfaces .1. fe(100) and fe(111). *J. Catal.*, 49:18, 1977.

[194] K. Christmann. Interaction of hydrogen with solid surfaces. *Surf. Sci. Rep.*, 9: 1, 1988.

[195] H. B. Gray. *Chemical Bonds: An Introduction to Atomic and Molecular Structure*. University Science Books, Mill Valley, California, 1994.

Paper I

RPBE-vdW Description of Benzene Adsorption on Au(111)

J. Wellendorff, A. Kelkkanen, J. J. Mortensen, B. I. Lundqvist, and T. Bligaard

Topics in Catalysis **53**, 378 (2010)

Top Catal (2010) 53:378–383
DOI 10.1007/s11244-010-9443-6

ORIGINAL PAPER

RPBE-vdW Description of Benzene Adsorption on Au(111)

Jess Wellendorff · André Kelkkanen · Jens Jørgen Mortensen · Bengt I. Lundqvist · Thomas Bligaard

Published online: 4 February 2010

Abstract Density functional theory has become a popular methodology for the analysis of molecular adsorption on surfaces. Despite this popularity, there exist adsorption systems for which commonly used exchange–correlation functionals fail miserably. Particularly those systems where binding is due to van der Waals interactions. The adsorption of benzene on Au(111) is an often mentioned such system where standard density functionals predict a very weak adsorption or even a repulsion, whereas a significant adsorption is observed experimentally. We show that a considerable improvement in the description of the adsorption of benzene on Au(111) is obtained when using the so-called RPBE-vdW functional.

Keywords Benzene · Au(111) · Van der Waals · RPBE-vdW · Density functional theory

1 Introduction

The adsorption of aromatic molecules on transition metal surfaces is a topic of great interest in surface science. Understanding the interactions of an aromatic adsorbate with a metallic surface also has important industrial applications. Catalytic conversion of aromatic compounds is a key reaction in many petrochemical processes [1, 2] and has been studied experimentally for decades [3].

Benzene is the smallest aromatic molecule, and the close-packed (111) surface of gold is among the least reactive surfaces known. For coverages up to one monolayer (ML) of benzene on Au(111) there appears to be a general consensus that benzene adsorbs with its molecular plane parallel to the surface in a physisorbed state from which it can desorb reversibly [4, 5]. The physisorption is believed to be mediated by dispersive van der Waals (vdW) interactions. Contrary to the case of adsorption of benzene on the much more reactive Pt(111) surface, Wöll and co-workers [4, 6] observed no non-planar benzene distortion (bending of C–H bonds) in near edge X-ray absorption fine structure (NEXAFS) studies of benzene monolayers physisorbed on Au(111). From temperature programmed desorption (TPD) experiments, Koel and co-workers [5] used a desorption peak temperature of 239 K to estimate the desorption energy of physisorbed benzene on Au(111) to 0.64 eV (14.7 kcal/mol) at a sub-monolayer coverage.

Theoretical first principles calculations have also been applied to the benzene/Au(111) interaction. Indeed, density functional theory (DFT) using exchange and correlation (xc) functionals within the generalized gradient approximation (GGA) has well proven its usefulness for describing many strongly interacting systems [7] also with respect to adsorption on transition and noble metal surfaces [8, 9] where in particular the RPBE xc-functional [10] has proven useful. However, when the adsorbate-substrate interaction is dominated by van der Waals interactions, the GGA-level of theory proves insufficient for calculating adsorption energies [11, 12]. This is clearly exemplified in a recent DFT study [13] of benzene adsorbed on (111) surfaces of Cu, Ag, and Au. At all three surfaces the binding energy was dramatically underestimated and often fell below the

J. Wellendorff · A. Kelkkanen · J. J. Mortensen · B. I. Lundqvist · T. Bligaard (✉)
Department of Physics, Center for Atomic-Scale Materials Design (CAMD), Technical University of Denmark, Building 307, 2800 Kgs. Lyngby, Denmark
e-mail: bligaard@fysik.dtu.dk

J. Wellendorff · T. Bligaard
Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

anticipated accuracy of the methodology itself. Even though other qualitative features of experimental results were reproduced by the calculations, the largest calculated adsorption energy of benzene on Au(111) was −0.08 eV (−1.92 kcal/mol). Insufficient description of the correlation effects responsible for vdW interactions was the reason for this. Similar conclusions were drawn from another recent GGA-DFT study of benzene/Au(111) interactions [14]. These authors furthermore concluded that GGA-DFT is well-suited for describing the ground state electronic structure of closed-shell molecules such as benzene on metallic surfaces, but that the adsorption structure (e.g., equilibrium adsorbate-metal distance) not necessarily is correctly captured due to the absence of van der Waals interactions in the calculations.

Ab initio MP2 calculations have been applied with some success to describe the vdW-interactions between benzene and Au(111). In recent MP2 calculations benzene adsorption on the (111) surfaces of Cu, Ag, and Au was considered [15]. The calculated adsorption energy for the Au(111) surface of −0.31 eV at an equilibrium benzene–metal distance of 3.6 Å was somewhat weaker than the experimental value of −0.6 eV reported in the same study.

A density functional that includes the nonlocal correlation effects responsible for van der Waals interactions has recently been developed [16]. This van der Waals density functional (vdW-DF) describes the exchange and correlation energy as a sum of GGA exchange and a correlation term consisting of both local correlation (evaluated in the local density approximation) and nonlocal correlation (depending nonlocally on the electron density) [17]:

$$E_{xc}[n] = E_x^{GGA}[n] + E_c^0[n] + E_c^{nl}[n] \quad (1)$$

The nonlocal correlation term is expressed as

$$E_c^{nl}[n] = \frac{1}{2}\int d\vec{r}_1 d\vec{r}_2 n(\vec{r}_1)\phi(q_1, q_2, r_{12})n(\vec{r}_2) \quad (2)$$

where ϕ is the vdW-DF kernel, $r_{12} = |\vec{r}_1 - \vec{r}_2|$ and q_1 and q_2 are the values of a universal function $q_0\left(n(\vec{r}), \left|\vec{\nabla} n(\vec{r})\right|\right)$ evaluated at the two points $\vec{r}_1$ and $\vec{r}_2$. The vdW-DF has recently been used to treat a number of diverse systems such as, e.g., layered structures, dimers, adsorbed molecules, carbon nanotubes, metal organic frameworks, and DNA with promising results [17]. In this study we use RPBE [10] exchange for the GGA exchange term in the vdW-DF. This approximation was called RPBE-vdW in Ref. [18], where it was also shown to give adsorption-energy values that deviate only about 0.04 eV from the vdW-DF values for a number of different cyclic molecules adsorbed on the basal plane of MoS_2.

In the present study we use first principles DFT calculations with the RPBE-vdW functional to calculate the adsorption energy of benzene on the Au(111) surface.

2 Computational Methods

We used the grid-based real-space code GPAW [19] for all the reported calculations. It is an implementation of the projector augmented wave (PAW) method of Blöchl [20], which is an all-electron full-potential method within the frozen core approximation. The Atomic Simulation Environment (ASE) [21] provided an interface to GPAW. Wave functions, electron densities and potentials are represented on grids in real space and a grid spacing of 0.20 Å was used in all directions of the supercell. Brillouin-zone integrations were performed using a 4 × 4 × 1 Monkhorst–Pack grid [22] and a Fermi smearing of 0.1 eV, and Pulay density mixing [23] was used.

All RPBE-vdW calculations were done non self-consistently; the RPBE xc-functional was used to obtain self-consistent GGA electron densities where after the vdW-DF scheme of Eq. 1 was applied to those densities as a post-GGA correction using RPBE exchange. The inclusion of vdW interactions in DFT calculations in other studies have shown to modify the total energy significantly without appreciably affecting the self-consistent electronic structure itself [13, 14, 24, 25]. Therefore, we assume that performing geometry optimizations within RPBE and *a posteriori* applying the RPBE-vdW to the density to be a reasonably reliable procedure for obtaining total energies. Evaluation of $E_c^{nl}[n]$ (Eq. 2) by direct summation in real space has an operation count that scales as N^2 for system size N, which turned out to require computation times of about 50% of those required for self-consistent RPBE calculations. Instead, we have implemented the fast Fourier transformation (FFT) evaluation technique introduced by Román-Pérez and Soler [26] where the scaling is $N \log N$. We have used 20 interpolation points, q_α, distributed between zero and $q_c = 5.0$ according to the formula $(q_{\alpha+1} - q_\alpha) = \lambda(q_\alpha - q_{\alpha-1})$ with $\lambda = 1.2$. The logarithmic divergence of the vdW-DF kernel ϕ is cut off smoothly at $r_{12}\left(q_1^2 + q_2^2\right)^{1/2} < 1.0$ with $\phi(q_1, q_2, 0)$ set to 0.5. For more details, see Ref. [26]. With the non self-consistent FFT evaluation of the vdW-DF parallelized over the 20 interpolation points, we typically find that the required computational time amounts to 0.5% of that required for a self-consistent RPBE calculation, thus becoming insignificant.

The three model systems treated in this study are (1) benzene adsorbed on the Au(111) surface [C_6H_6/Au(111)], (2) the Au(111) surface before adsorption [Au(111)] and (3) gas phase benzene [C_6H_6]. The extended Au(111) surface was modeled by an orthorhombic supercell with 16 atoms per metal layer. The surface lattice cell was repeated periodically in the surface plane to create an infinite slab, while non-periodic boundary conditions were employed in the direction perpendicular to the slab. The benzene molecule was added parallel to the surface with the center of

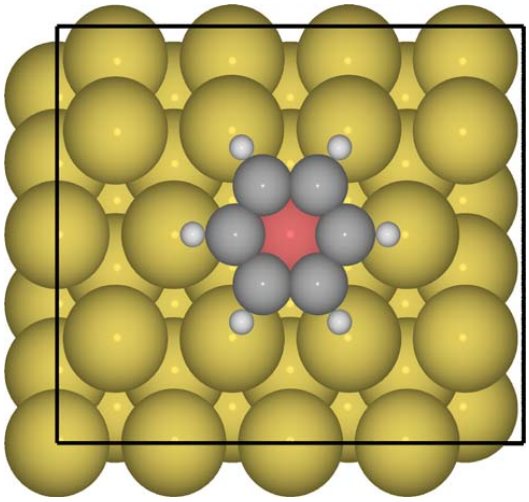

Fig. 1 Top-view of benzene adsorption over a central gold atom (in red) in the 16-metal atom rectangular surface cell

the C_6 ring at an on-top site directly above a central Au atom. A schematic top-view of this adsorption geometry is depicted in Fig. 1.

All geometry relaxation was done by minimizing the Hellmann–Feynman forces using the RPBE xc-functional. The two top layers of the Au(111) slab were allowed to relax freely while the bottom layers further from the surface were fixed at the bulk Au lattice constant of 4.21 Å, likewise determined within RPBE. With the parallel-oriented benzene placed 3.65 Å above the (111) surface the adsorption system was relaxed while constraining the benzene atoms to only relax in the plane parallel to the surface. Experimental [4, 6] and theoretical [14] studies have shown that the non-planar distortions of the benzene molecule upon adsorption on Au(111) are negligible. The van der Waals interaction responsible for physisorption is not expected to depend significantly on the adsorption geometry and theoretical studies of benzene adsorption on coinage metals often consider this completely planar benzene geometry in the on-top adsorption site [15, 27, 28]. Defining the benzene–Au(111) separation as the distance between the center of the aromatic ring and the top-layer Au atom residing exactly beneath the center of the C_6 ring unambiguously defines the molecule-surface distance. Gas phase benzene was fully relaxed in a non-repeated cell with more than 6 Å vacuum to the cell boundaries. This determined the equilibrium geometry of gas phase benzene to be completely planar with bond lengths of 1.404 Å and 1.092 Å for C–C and C–H bonds, respectively.

The finite grid spacing in real-space DFT calculations leads to the so-called egg-box effect where the total energy of a system depends on the positions of all atoms in the system relative to the grid points [29]. The effect can be reduced by reducing the grid spacing. However, since each atom in the system may introduce an egg-box error upon translation of all atoms in the supercell, the sum of egg-box errors may be significant when large systems are considered. Care was taken to avoid sources of numerical errors from egg-box effects in all calculations. As adsorption energies are calculated from comparison of the total energies of the C_6H_6/Au(111), Au(111) and C_6H_6 model systems, egg-box effects were minimized within each adsorption energy by, as far as possible, positioning the atoms in corresponding model systems in identical positions relative to the 3D real-space mesh.

3 Results and Discussion

In order to enable high accuracy and reliability in calculating the benzene adsorption energy on Au(111) care was taken to ensure the reliability of the employed model systems.

In DFT slab calculations with non-periodic boundary conditions perpendicular to the surface plane the electronic structure is forced to extinction at the supercell boundaries above and below the slab. It is therefore important to use vacuum layers thick enough to eliminate the risk of artificial quenching of the self-consistent electron density in both directions. The benzene adsorption energy, E_{ads}, we define as

$$E_{ads} = E_{tot}^{C_6H_6/Au(111)} - E_{tot}^{Au(111)} - E_{tot}^{C_6H_6} \quad (3)$$

such that $E_{ads} < 0$ corresponds to bonding. In Fig. 2 is shown the convergence of the calculated adsorption energy of benzene on Au(111) as the vacuum layer above and below the adsorption and the slab system is increased. The computations were done by first relaxing the C_6H_6/Au(111) model system with benzene initially placed in the on-top site 3.65 Å above the Au(111) surface. Vacuum layers of 6.33 Å were used above the C_6H_6 adsorbate and below the Au slab. Minor geometry changes were observed for the two top metal layers while the benzene did not relax its bond lengths at all. The central gold atom (colored red in Fig. 1) moved 0.04 Å towards the surface upon relaxation, causing the equilibrium benzene–metal distance to be 3.69 Å. The equilibrium geometry of the Au(111) surface before adsorption was determined by removing the benzene molecule from the supercell of the relaxed C_6H_6/Au(111) system and relaxing the bare surface. Self-consistent RPBE adsorption energies for these geometries were then obtained using different vacuum layer thicknesses and the corresponding RPBE-vdW adsorption energies were

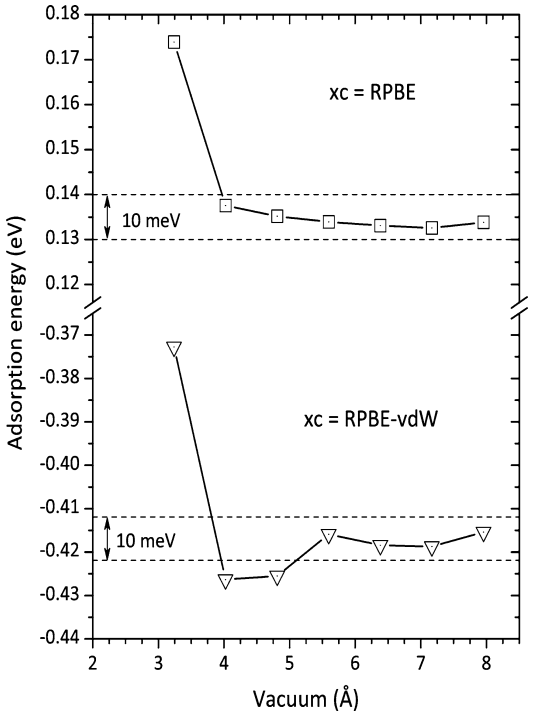

Fig. 2 Convergence of the RPBE and RPBE-vdW adsorption energy of benzene in the on-top site of Au(111) as the vacuum layers employed in the supercell are enlarged. For any vacuum layer >6 Å above and below the adsorption system the adsorption energy is converged to within 10 meV

computed. We see from Fig. 2 that both the RPBE and the RPBE-vdW adsorption energy are converged to within 10 meV for vacuum layers exceeding 6 Å above and below the system. RPBE repulsion of 0.13 eV was derived for the three-layer Au(111) slab while E_{ads} turned out exothermic by −0.42 eV within the RPBE-vdW.

To investigate the dependence of E_{ads} on the number of metal layers in the Au(111) slab we calculated the benzene adsorption energy for slabs containing up to ten Au layers (160 Au atoms in total). The geometries of the top-two metal layers in the C_6H_6/Au(111) and Au(111) model systems were taken from those previously found for three metal layers while the remaining layers were kept in their bulk geometry. The vacuum layers above and below the systems were at least 6.33 Å (with small variations introduced by the constant grid spacing).

The results are shown in Fig. 3 where the calculated benzene adsorption energies on Au(111) are plotted against the number of metal layers in the slab. Over the range of 3–10 metal layers both the RPBE and the RPBE-vdW adsorption energy fall within a 20 meV energy window.

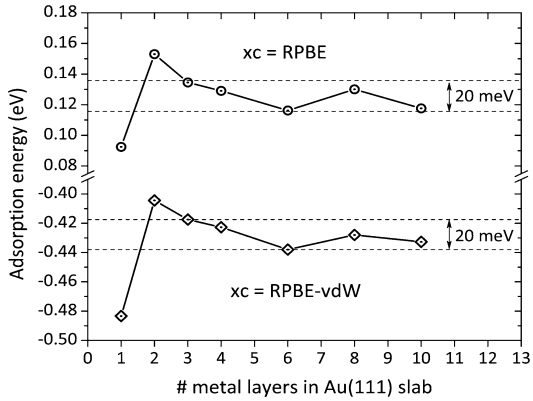

Fig. 3 Convergence of the RPBE and RPBE-vdW adsorption energy of benzene as the number of metal layers in the slab is increased from one through ten. For slabs comprised of three or more Au layers the calculated adsorption energies fall within a 20 meV energy window

Calculations using the three-layer Au(111) slab therefore capture the RPBE-vdW benzene/Au(111) interactions well.

Having established that closed-boundary DFT calculations with vacuum layers exceeding 6 Å above and below the three-metal-layer benzene/Au(111) adsorption system yields converged adsorption energies we calculated the benzene adsorption energy on the (111) surface plane of Au for a range of benzene–metal distances. Adsorption distances, d, in the range 0.75–6.05 Å were considered for the three-layer Au(111) substrate. While only varying the benzene–metal distance we kept the model system and supercell geometries fixed for calculation of E_{ads} for each d. To ensure computational accuracy over this range of separations a slightly enlarged supercell in the direction perpendicular to the surface plane was used. This implies that the vacuum layers above and below the adsorption system exceeded 6 Å for the case of the largest benzene–metal separation, $d = 6.05$ Å. With this enlarged supercell the previously determined equilibrium geometry of the C_6H_6/Au(111) model system with $d = 3.69$ Å yielded no relaxation. The geometries of the Au(111) and C_6H_6 reference model systems were determined by relaxation after removing the adsorbate and substrate from the enlarged supercell, respectively, as described above. This led to no relaxation of the bond lengths in gas phase benzene. The only atoms moving in the supercell when changing d were those of the C_6H_6 adsorbate.

The calculated RPBE-vdW adsorption energies are plotted against the benzene–Au(111) distance in Fig. 4. The bottom of a van der Waals type potential energy curve is mapped out. The minimum of the binding energy curve is found at $d = 3.7$ Å, making this the calculated equilibrium adsorption distance of benzene on the Au(111) surface when van der Waals interactions are taken into

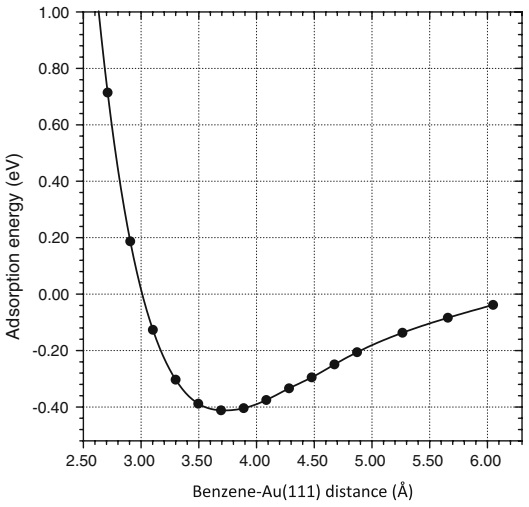

Fig. 4 RPBE-vdW binding energy curve for benzene adsorption on Au(111). The largest adsorption energy of −0.42 eV is obtained at the equilibrium benzene–Au(111) distance of 3.7 Å

account. The corresponding adsorption energy for the parallel-oriented C_6H_6 on Au(111) is −0.42 eV. To our knowledge, this is the first electronic structure study of benzene on Au(111) reporting an adsorption energy this close to the experimental data available, appearing to agree on C_6H_6/Au(111) adsorption energies of ∼0.6 eV. Also, the equilibrium binding energy obtained in the present study of benzene adsorption in the on-top site on Au(111) is an order of magnitude larger than that obtained for the same adsorption site in a recent GGA-DFT study of C_6H_6/Au(111) interface interactions [13]. The inclusion of dispersive interactions by using the RPBE-vdW is clearly seen from Figs. 2 and 3 to be the reason for this significant improvement; with RPBE-vdW corrections the adsorption energies are shifted down into the region of bonding. Furthermore, the van der Waals type binding energy curve in Fig. 4 shows that dispersive interactions are indeed important in benzene adsorption on Au(111). The calculated adsorption distance of 3.7 Å is similar to those found in MP2 calculations for benzene on the (111) surfaces of Au, Ag and Cu [15] and to DFT calculations using the vdW-DF of benzene adsorption on graphene [24] and on the basal plane of MoS_2 [18].

4 Conclusions

The experimentally observed temperature programmed desorption spectra for benzene on Au(111), combined with Redhead analysis, indicate a desorption energy of 0.64 eV for benzene on Au(111) [5]. Such TPD/Redhead experiments have some inherent uncertainty. It is certain from these experiments and standard density functional calculations, however, that GGA density functional theory significantly underestimates the adsorption energy of benzene on Au(111). Our fast implementation of the vdW-DF nonlocal correlation functional enables us to obtain a significantly improved adsorption energy for the purely van der Waals bonded benzene on Au(111) system. The improvement is comparable to the improvements observed using the computationally more expensive correlated electron method MP2, but is achievable at essentially no additional cost. It can also be applied to much larger systems than MP2, which we illustrate by applying the method to systems with up to 160 gold atoms.

Acknowledgments We gratefully thank Professor J.K. Nørskov for his inspiring enthusiasm and support over the years, and congratulate him with the Gabor A. Somorjai Award for Creative Research in Catalysis. We thank Professors G.A. Somorjai and C.T. Campbell for enlightening discussions. J.W. and T.B. thank the Somorjai group for kind hospitality during parts of this work. The Center for Atomic-scale Materials Design is funded by the Lundbeck Foundation, and this work was supported by the Danish Center for Scientific Computing.

References

1. Somorjai GA (1981) Chemistry in two dimensions: surfaces. Cornell University Press, Ithaca, NY
2. Stanislaus A, Cooper BH (1994) Catal Rev Sci Eng 36:75
3. Somorjai GA (1994) Introduction to surface chemistry and catalysis. Wiley–Interscience, New York
4. Wöll C (2001) J Synchrotron Radiat 8:129
5. Syomin D, Kim J, Ellison G, Koel E (2001) J Phys Chem B 105:8387
6. Weiss K, Gebert S, Wün M, Wadepohl H, Wöll C (1998) J Vac Sci Technol A 16:1017
7. Kohn W, Becke AD, Parr RG (1996) J Phys Chem 100:12974
8. Neurock M (2003) J Catal 216:73
9. Nørskov JK, Bligaard T, Rossmeisl J, Christensen CH (2009) Nature Chem 1:37
10. Hammer B, Hansen LB, Nørskov JK (1999) Phys Rev B 59:7413
11. Kristyán S, Pulay P (1994) Chem Phys Lett 229:175
12. Wu X, Vargas MC, Nayak S, Lotrich V, Scoles G (2001) J Chem Phys 115:8748
13. Bilić A, Reimers JR, Hush NS, Hoft RC, Ford MJ (2006) J Chem Theory Comput 2:1093
14. Schuster C, Schwingenschlögl U (2009) Chem Phys Lett 468:75
15. Caputo R, Prascher BP, Staemmler V, Bagus PS, Wöll C (2007) J Phys Chem A 111:12778
16. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Phys Rev Lett 92:246401
17. Langreth DC, Lundqvist BI, Chakarova-Käck SD, Cooper VR, Dion M, Hyldegaard P, Kelkkanen A, Kleis J, Kong L, Li S, Moses PG, Murray E, Puzder A, Rydberg H, Schröder E, Thonhauser T (2009) J Phys: Condens Matter 21:084203
18. Moses PG, Mortensen JJ, Lundqvist BI, Nørskov JK (2009) J Chem Phys 130:104709
19. Mortensen JJ, Hansen LB, Jacobsen KW (2005) Phys Rev B 71:035109

20. Blöchl PE (1994) Phys Rev B 50:17953
21. Bahn S, Jacobsen KW (2002) Comput Sci Eng 4:56
22. Monkhorst HJ, Pack JD (1976) Phys Rev B 13:5188
23. Pulay P (1980) Chem Phys Lett 73:393
24. Chakarova-Käck SD, Schröder E, Lundqvist BI, Langreth DC (2006) Phys Rev Lett 96:146107
25. Johnston K, Kleis J, Lundqvist BI, Nieminen RM (2008) Phys. Rev B 77:121404(R)
26. Román-Pérez G, Soler JM (2009) Phys Rev Lett 103:096102
27. Witte G, Lukas S, Bagus PS, Wöll C (2005) Appl Phys Lett 87:263502
28. Bagus PS, Hermann L, Wöll C (2005) J Chem Phys 123:184109
29. Nogueira F, Castro A, Marques MAL (2003) In: Fiolhais C, Nogueira F, Marques MAL (eds) A primer in density functional theory. Springer, Heidelberg

Paper II

On the Importance of Gradient-Corrected Correlation for van der Waals Density Functionals

J. Wellendorff and T. Bligaard

Topics in Catalysis **54**, 1143 (2011)

Top Catal (2011) 54:1143–1150
DOI 10.1007/s11244-011-9736-4

ORIGINAL PAPER

On the Importance of Gradient-Corrected Correlation for van der Waals Density Functionals

Jess Wellendorff · Thomas Bligaard

Published online: 13 September 2011

Abstract The van der Waals density functional (vdW-DF) family of exchange–correlation functionals is a promising step towards accounting for van der Waals interactions in density functional theory. This approach consists of a nonlocal correlation term in addition to semilocal generalized gradient approximation exchange and local density approximation correlation. It has proven useful for describing vdW bonded complexes but unfortunately deteriorates the prediction of solid-state properties such as bulk lattice parameters and cohesive energies, as compared to the underlying GGA functional. By considering a broad range of different condensed matter systems including weakly interacting complexes as well as strongly bonded molecules and bulk solids, we show that inclusion of gradient-corrected correlations in vdW-DF-type calculations may not only improve the accuracy for vdW bonded systems, but also amend vdW-DF deficiencies in predicting structural properties of solids. Based on this insight we construct a prototype vdW-DF which demonstrates high accuracy in describing the dispersive interactions responsible for benzene adsorption on the noble Au(111) surface.

Keywords Density functional theory · van der Waals · vdW-DF · RPBE + nl · RPBEc2/3 + nl · Au(111) · Benzene

J. Wellendorff · T. Bligaard
Department of Physics, Center for Atomic-scale Materials Design (CAMD), Technical University of Denmark, Building 307, 2800 Kgs., Lyngby, Denmark

T. Bligaard (✉)
SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
e-mail: bligaard@stanford.edu

1 Introduction

Density functional theory (DFT) [1] is a prominent electronic structure method within condensed matter physics and computational chemistry, where DFT offers a high ratio of accuracy to computational cost. The great success of modern generalized gradient approximations (GGAs) to exchange and correlation among electrons in strongly interacting systems has contributed significantly to this [2]. However, description of van der Waals (vdW) interactions within DFT appears more challenging and much research aims at designing exchange–correlation (xc) functionals that capture the often weak and long-ranged dispersion forces with reasonable accuracy [3].

A promising step towards including dispersion in DFT was the introduction of the Langreth–Lundqvist (LL) van der Waals density functional (vdW-DF) xc-functional which depends nonlocally on the electronic density [4, 5]. The vdW contribution to the DFT total energy is calculated as a double-integral over all electronic densities in space, mutually interacting through a correlation kernel. In the LL vdW-DF scheme this nonlocal correlation energy, E_c^{nl}, is added to a GGA exchange and local density approximation (LDA) correlation:

$$E_{xc}^{vdW-DF-type} = E_x^{GGA} + E_c^{LDA} + E_c^{nl}. \quad (1)$$

Several studies indicate the usefulness of the vdW-DF for weakly interacting systems [6]. Nevertheless, there is still room for improvement in accuracy and a range of different GGA exchange functionals have been proposed for E_x^{GGA} in Eq. 1 [7–9]. Meanwhile, it appears that the possible role of applying GGA semilocal correlation in the LL vdW-DF functional rather than the strictly local LDA correlation may have been overlooked. Indeed, recent work of Vydrov and Van Voorhis [10, 11] have successfully applied semilocal

correlation along with a modified expression for the nonlocal correlation.

In this study we first investigate the use of semilocal correlation in the Langreth–Lundqvist vdW-DF method when applying it to three very different groups of condensed matter systems; the S22 benchmark dataset of weakly interacting complexes [12], the G2-1 database of molecular formation energies [13], and a broad range of 20 bulk solids for which we calculate equilibrium lattice parameters and cohesive energies. In short, we find that exchanging LDA correlation for GGA correlation introduces an S22 overbinding that is comparable to vdW-DF-1 and vdW-DF-2 underbinding, offers comparable performance for the G2-1 molecules, and greatly amends vdW-DF-1 and vdW-DF-2 deficiencies in predicting structural properties of solids. This approach does, however, introduce notable overbinding in some cases. We therefore construct a simple vdW-DF using a suitable amount of semilocal correlations. This functional offers not only a balanced description of the three above mentioned groups of systems but also demonstrates high accuracy in describing benzene physisorption on the noble Au(111) surface.

The six different exchange–correlation functionals considered in the first part of this study are listed in Table 1. The only pure GGA functional is the RPBE. The following four xc functionals are all of the vdW-DF-type defined in Eq. 1 but with different exchange components. Furthermore, we tentatively introduce here an exchange–correlation functional pairing PBE GGA [15] correlation with nonlocal correlation:

$$E_{\mathrm{xc}}^{\mathrm{RPBE+nl}} \equiv E_{\mathrm{x}}^{\mathrm{RPBE}} + E_{\mathrm{c}}^{\mathrm{PBE}} + E_{\mathrm{c}}^{\mathrm{nl}} = E_{\mathrm{xc}}^{\mathrm{RPBE}} + E_{\mathrm{c}}^{\mathrm{nl}}, \qquad (2)$$

where $E_{\mathrm{c}}^{\mathrm{nl}}$ is that of the original vdW-DF-1. This is obviously a vdW-extension of the RPBE functional and we shall here term it RPBE + nl.

In acknowledging the risk of some degree of correlation double-counting, i.e. some possible overlap between GGA and nonlocal correlation interactions, we emphasize the tentative nature of parts of this study.

2 Computational Details

All calculations are done using a locally modified version of the GPAW density functional package [18] and the Atomic Simulation Environment [19]. GPAW is a real-space grid implementation of the PAW formalism [20] for all-electron DFT calculations within the frozen-core approximation. For calculations presented in Sect. 3 we use selfconsistent RPBE electronic densities in non-selfconsistent calculations for all other xc functionals. The GPAW implementation of a fast Fourier transformation method [21] for evaluation of $E_{\mathrm{c}}^{\mathrm{nl}}$ in Eqs. 1 and 2 is described elsewhere [22]. A nonlocal interaction cutoff of 10 Å is used along with real-space grid-spacings of 0.12–0.16 Å. Non-periodic supercell boundary conditions are applied in calculations for S22 and G2-1 systems as well as for single atoms in vacuum. Extended bulk lattices are treated periodically in all directions through Brillouin-zone integrations on a 12 × 12 × 12 Monkhorst-Pack [23] **k**-space grid and 0.1 eV Fermi smearing. Calculations for the S22 dataset employ supercells with minimum 10 Å vacuum in all directions, implying interaction energies converged at the meV level.

For evaluating and comparing the performance of different xc-functionals the statistical quantities we compute throughout this work are the mean deviation (MD) and mean absolute deviation (MAD) from reference quantities as well as their relative counterparts (MRD and MARD).

3 Results

3.1 S22 Weakly Bonded Dimers and Complexes

The S22 benchmark dataset of Hobza and co-workers contains high-level coupled-cluster and second-order Møller-Plesset (CCSD(T)/MP2) geometries and CCSD(T) interaction energies for 22 van der Waals bonded dimers and complexes. Geometries from the original S22 work in Ref. [12] are used while for reference interaction

Table 1 Overview of the composition of the exchange–correlation functionals, E_{xc}, considered in the first parts of this study, in terms of their exchange, E_{x}, and correlation, E_{c}, parts

E_{xc}	$E_{\mathrm{x}}^{\mathrm{GGA}}$	$E_{\mathrm{c}}^{\mathrm{LDA}}$	$E_{\mathrm{c}}^{\mathrm{GGA}}$	$E_{\mathrm{c}}^{\mathrm{nl}}$
RPBE	RPBE [14]	×	PBE [15]	×
RPBE-vdW	RPBE [14]	LDA	×	√
vdW-DF-1	revPBE [16]	LDA	×	√
vdW-DF-2	PW86 [17]	LDA	×	√[a]
optPBE-vdW	optPBE [7]	LDA	×	√
RPBE + nl	RPBE [14]	×	PBE [15]	√

[a] The vdW-DF-2 nonlocal correlation kernel is slightly modified [9]

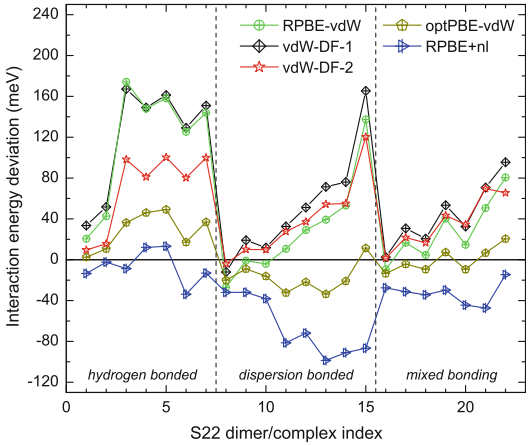

Fig. 1 Deviations of calculated interaction energies from CCSD(T) reference interaction energies in the S22 benchmark database. For details of the reference energies and geometries of the indexed dimers and complexes as well as their grouping into hydrogen bonded, dispersion bonded and mixed bonding, respectively, see Refs. [12] and [24]. Positive energy deviation means interaction energy underestimate as compared to the benchmark

Table 2 Statistics of the S22 interaction energy deviations shown in Fig. 1

xc-functional	MD	MAD	MRD (%)	MARD (%)
RPBE	225	225	101	101
RPBE-vdW	57	61	8	21
vdW-DF-1	71	72	20	25
vdW-DF-2	48	48	15	17
optPBE-vdW	3	20	−8	12
RPBE + nl	−37	39	−29	29

Positive energy deviation means interaction energy underestimate as compared to the benchmark. All energies in meV

energies we have chosen the recently revised CCSD(T) data of Sherrill and co-workers from Ref. [24]. Shown in Fig. 1 are deviations of S22 interaction energies (E_{int}, negative quantity) computed using all introduced vdW functionals from the CCSD(T) references. The averaged deviations are listed in Table 2, which also includes RPBE entries. It is seen that RPBE greatly underestimates E_{int}, while the vdW functionals improve on GGA results, as previously noted in e.g. Ref. [7]. We also note that the vdW-DF-2 improves on vdW-DF-1 interaction energies, especially for hydrogen bonds, as earlier reported in Ref. [9]. Using an exchange functional that is explicitly fitted to the S22 benchmark data the optPBE-vdW functional performs best with a MAD of 20 meV. Interestingly, exchanging LDA for PBE correlation, the RPBE + nl functional yields a smaller MAD than that of the RPBE-vdW, and proves highly useful for hydrogen bonding. We stress that this remarkable result is obtained without any fitting of the exchange functional but comes about solely from the use of semilocal rather than strictly local correlation. It is, however, apparent from the middle part of Fig. 1 and the signed errors (MD, MRD) in Table 2 that RPBE + nl tends to overestimate dispersion bonding.

Since the van der Waals calculations in this study are in general done non-selfconsistently we have also performed selfconsistent calculations for the S22 dataset and find that this would cause a difference of at most a few meV to the statistics in Table 2. We may also compare the vdW-DF-2 S22 interaction energies published in Ref. [11] to the coupled-cluster reference energies used in this study, and find a vdW-DF-2 MAD of 42 meV as compared to the 48 meV obtained for vdW-DF-2 in this study using GPAW. Furthermore, if we use as refence the original S22 interaction energies of Ref. [12] we find an optPBE-vdW MAD of 15 meV, which is in excellent agreement with the 15 meV reported for optPBE-vdW in Ref. [7], where those somewhat older CCSD(T) data were used as benchmark. We may therefore safely conclude that the non-selfconsistent GPAW calculations presented here are accurate enough for the present purpose of cross-comparing xc-functionals.

3.2 G2-1 Molecular Atomization Energies

We furthermore consider the G2-1 compilation of standard enthalpies of formation for 55 simple molecules in MP2 optimized geometries adapted from Ref. [13]. Using experimental formation enthalpies at 298 K, thermal corrections and zero-point energies from Refs. [13] and [25] we derive G2-1 reference electronic-only atomization energies. The obtained deviations from this G2-1 reference dataset using all introduced xc functionals are depicted in Fig. 2 and summarized in Table 2. As might be expected, the RPBE functional does rather well on average. The five van der Waals functionals all tend to overbind on average but the MAD for all except the optPBE-vdW are comparable to that for RPBE, thus hampering recommendation of one over all the others for DFT calculations on molecular systems. However, the optPBE-vdW functional, with an exchange component somewhat more attractive than the other exchange functionals considered here, is clearly overbinding the G2-1 dataset. Only a few optPBE-vdW points in Fig. 2 are above the zero-line and so almost all of the 0.45 eV MAD stems from overbinding. This illustrates one possible drawback of fitting density exchange–correlation functionals to a single electronic-structure property only.

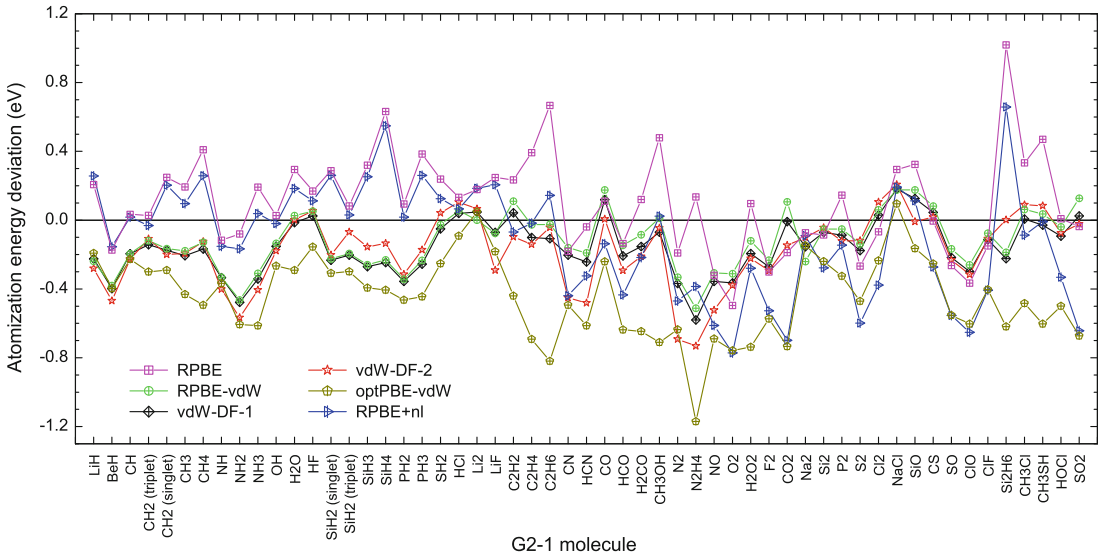

Fig. 2 Deviations of electronic-only molecular atomization energies from experimental reference values derived from the G2-1 database of formation enthalpies of 55 simple molecules. Negative energy deviation signify intra-molecular overbinding

3.3 Bulk Solid-State Properties

We show in Fig. 3 and summarize in Table 4 the deviations from experimental values of calculated equilibrium lattice parameters, a_0, and cohesive energies, E_{coh}, for 20 fcc, bcc, hcp, or diamond structured bulk solids. For each solid a_0 was obtained from fitting cohesive energies for a range of unit cell volumes to a third-order polynomium. For hcp-structured crystals the *c*/*a* lattice parameter ratio was fixed at the experimental one. Zero-point phonon effects have not been corrected for. The RPBE functional is found to overestimate lattice parameters with 1.7% on average, with a corresponding 19% underestimate of cohesive energies. All of the three functionals RPBE-vdW, vdW-DF-1, and vdW-DF-2 essentially double this equilibrium lattice parameter overestimate, vdW-DF-2 in this respect performing worst of all considered xc functionals. The optPBE exchange functional appears more attractive for solids than both the RPBE and revPBE ones. The a_0 deviation for optPBE-vdW is at the RPBE level, while the performance for cohesive energies is significantly better than RPBE. Similar conclusions were drawn from a recent study [28] of the performance of Langreth–Lundqvist type vdW-DF functionals for solid-state properties. While optPBE-vdW was found to perform at the PBE level with an MAE of 0.064 Å for the lattice constants in a somewhat different test set than the one used here, the vdW-DF-1 and vdW-DF-2 functionals did significantly worse with MADs of 0.105 and 0.116 Å, respectively. Those figures are well in line with the ones reported here.

It is furthermore seen from Table 4 that the RPBE + nl functional proves highly accurate for lattice parameters and provides a cohesive energy overestimate slightly larger than the optPBE-vdW underestimate. The RPBE + nl and optPBE-vdW functionals are therefore found to be significantly more accurate for structural properties of solid crystals than the GGA-type RPBE as well as the remaining van der Waals functionals. The RPBE + nl mean overestimate of cohesive energies does, however, point at possible correlation interaction double-countings.

3.4 Discussion

From the results presented above it is first of all apparent that the Langreth–Lundqvist vdW-DF scheme introduces attractive van der Waals interactions in the weakly bonded S22 systems. These interactions are not accounted for by GGA exchange–correlation functionals. It is also clear, however, that with mean S22 interaction energy underestimates around 50–70 meV the three vdW-DF-type functionals RPBE-vdW, vdW-DF-1, and vdW-DF-2 do not offer quantitative agreement with high-level theory for weak bonds. Furthermore, those three functionals significantly overestimate equilibrium lattice constants of bulk solids in essentially doubling the RPBE mean deviation for this property. The somewhat underestimated RPBE-vdW, vdW-DF-1, and vdW-DF-2 cohesive energies follow this trend of bulk solid underbinding, which arguably hampers truly general applicability of these three functionals within theoretical surface science.

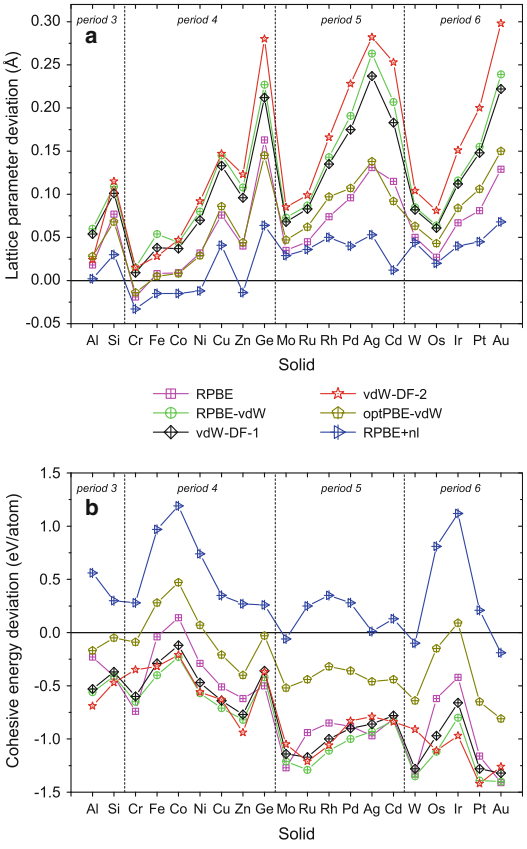

Fig. 3 Deviations of calculated solid-state properties from experimental values for 20 solids, ordered ascendingly according to their periodic table period and atomic number. **a** Lattice parameter deviations $a_0^{\mathrm{DFT}} - a_0^{\mathrm{exp}}$. Experimental values without zero-point corrections adapted from Ref. [26]. **b** Cohesive energy deviations $E_{\mathrm{coh}}^{\mathrm{DFT}} - E_{\mathrm{coh}}^{\mathrm{exp}}$ with $E_{\mathrm{coh}}^{\mathrm{exp}} > 0$. Experimental values without zero-point corrections adapted from Ref. [27]

The optPBE-vdW functional amends these issues by explicitly fitting an exchange functional for use in the vdW-DF method. This approach works very well for weak bonds as well as bulk solid structural properties but also leads to severe overestimates of intra-molecular binding energies of the G2-1 molecules. This probably renders the optPBE-vdW inappropriate for studies involving chemical transformations, a dominant field within condensed matter physics and chemistry.

The RPBE + nl functional offers an alternative approach towards a generally applicable, high-accuracy vdW-DF through the use of semilocal correlation contributions rather than exchange fitting. Using PBE correlation instead of LDA correlation in the LL vdW-DF scheme, the RPBE + nl overestimates the S22 interaction energies on average and completely changes the systematic trend with RPBE-vdW, vdW-DF-1, and vdW-DF-2 S22 underbinding in Fig. 1. The semilocal correlations introduce binding in bulk solids as well, leading to highly accurate lattice parameters at the expense of somewhat overestimated cohesive energies. Regarding molecular systems the RPBE + nl G2-1 atomization energies are comparable to the non-fitted vdW-DF-type functionals, which again compare well to RPBE GGA calculations.

3.5 A Convenient High-Performance vdW-DF

It is apparent from the above analysis that the amount of semilocal correlations introduced in the vdW-DF method must be balanced against those already accounted for in the fully nonlocal correlation term of Eq. 1 such that correlation double-counting is avoided. The RPBE + nl mean overbinding of the S22 systems and mean overestimate of bulk cohesive energies are due to such issues. From mere inspection of the data in Tables 2, 3 and 4 we therefore construct a vdW-DF mixing LDA end GGA correlation to average out this problem. In this work we shall term it the RPBEc2/3 + nl functional:

$$E_{\mathrm{xc}}^{\mathrm{RPBEc2/3+nl}} \equiv E_{\mathrm{x}}^{\mathrm{RPBE}} + \frac{1}{3}E_{\mathrm{c}}^{\mathrm{LDA}} + \frac{2}{3}E_{\mathrm{c}}^{\mathrm{PBE}} + E_{\mathrm{c}}^{\mathrm{nl}} \qquad (3)$$

As shown in Table 5 this functional form performs remarkably well when applied non-selfconsistently to the three previously considered groups of benchmark systems. The S22 interaction energies are very accurately captured with only a slight average overbinding, while the G2-1 molecular formation energies turn out similar to those of the vdW-DF-type functionals (and better than the optPBE-vdW). Moreover, the RPBEc2/3 + nl functional performs comparably to optPBE-vdW for predicting bulk solid properties and avoids the significant overbinding of solids arising when fully substituting LDA correlation for PBE correlation. As discussed above we expect insignificant changes to these conclusions had the RPBEc2/3 + nl density functional calculations been done selfconsistently.

4 Benzene Physisorption on Au(111)

Benzene adsorption on the noble (111) surfaces of Cu, Ag, and Au is mediated primarily by dispersive van der Waals forces and so offers a prime example of the inability of ordinary GGA-DFT in describing vdW interactions. We here assess the ability of the RPBE, RPBE-vdW, vdW-DF-1, vdW-DF-2, optPBE-vdW, and RPBEc2/3 + nl density functionals in describing benzene/Au(111) interactions.

Experimentally benzene is found to interact only weakly with the Cu(111), Ag(111), and Au(111) surfaces, resulting

Table 3 Statistics of the G2-1 molecular atomization energy deviations shown in Fig. 2

xc-functional	MD	MAD	MRD (%)	MARD (%)
RPBE	0.10	0.23	0.1	3.6
RPBE-vdW	−0.12	0.17	−2.8	3.4
vdW-DF-1	−0.16	0.18	−3.0	3.5
vdW-DF-2	−0.18	0.20	−3.1	3.8
optPBE-vdW	−0.44	0.45	−6.4	6.7
RPBE + nl	−0.11	0.27	−2.7	5.1

Negative energy deviation means intra-molecular overbinding. All energies in eV

Table 4 Statistics of the deviations from experimental values of calculated lattice parameters and cohesive energies for bulk solids shown in Fig. 3

xc-functional	MD	MAD	MRD (%)	MARD (%)
Lattice parameters (Å)				
RPBE	0.063	0.065	1.7	1.7
RPBE-vdW	0.123	0.123	3.4	3.4
vdW-DF-1	0.113	0.113	3.1	3.1
vdW-DF-2	0.141	0.141	3.8	3.8
optPBE-vdW	0.069	0.071	1.9	1.9
RPBE + nl	0.024	0.033	0.6	0.9
Cohesive energies (eV/atom)				
RPBE	−0.69	0.71	−19	19
RPBE-vdW	−0.86	0.86	−22	22
vdW-DF-1	−0.78	0.78	−20	20
vdW-DF-2	−0.80	0.80	−21	21
optPBE-vdW	−0.24	0.33	−7	9
RPBE + nl	0.39	0.42	9	10

For details see text and the caption for Fig. 3

Table 5 Mean deviations from experiments or high-level theory of calculated quantities performed with the RPBEc2/3 + nl density functional defined in Eq. 3

Dataset	Unit	MD	MAD	MRD (%)	MARD (%)
S22 complexes	meV	−6	29	−16	20
G2-1 molecules	eV	−0.12	0.20	−2.8	4.1
Lattice constants	Å	0.052	0.054	1.4	1.4
Cohesive energies	eV/atom	−0.03	0.27	−1	6

in relatively small adsorption energies and negligible distortion of the benzene (C_6H_6) molecular geometry upon adsorption [29–35]. This is a signature of the decisive role of dispersion interactions in these three adsorption systems. For the case of benzene on Au(111) at low coverage the weak adsorbate–surface interactions cause the C_6H_6 to adsorb with its molecular plane parallel to the surface in a physisorbed state from which it can desorb reversibly [29–31]. The adsorption energy of benzene on Au(111) was found to be −0.64 eV (−14.7 kcal/mol) at a sub-monolayer coverage in temperature programmed desorption (TPD) experiments by Koel and co-workers [30]. Another experiment reports $E_{ads} = -0.6$ eV [36].

In a collaboration with others the authors recently conducted a thorough study of C_6H_6/Au(111) adsorption using the RPBE-vdW functional non-selfconsistently on RPBE electronic densities [22]. The GPAW code was used in that study as well, and it was found that while the RPBE xc functional predicts essentially no bonding of benzene to the (111) gold surface, the RPBE-vdW introduces a significantly exothermic adsorption energy of −0.42 eV at an equilibrium C_6H_6–Au(111) distance of $d = 3.7$ Å. This latter result compares well to recent results obtained with the computationally more expensive MP2 wavefunction-based method, where the adsorption energy $E_{ads} = -0.31$ eV at $d = 3.6$ Å was found for the benzene adsorption on Au(111) in Ref. [36]. From these studies it may be concluded that benzene does indeed physisorb weakly onto the Au(111) surface without formation of actual chemical bonds. However, as encouraging the DFT RPBE-vdW and wavefunction MP2 methods may appear for describing the C_6H_6/Au(111) interactions, none of them offer quantitative agreement with experiments in terms of adsorption energy.

We here present selfconsistent potential-energy curves (PECs) for C_6H_6/Au(111) physisorption obtained using the RPBE xc functional as well as the RPBE-vdW, vdW-DF-1, vdW-DF-2, optPBE-vdW, and RPBEc2/3 + nl van der Waals functionals. An orthorhombic 4 × 4 × 4 metal slab with 16 Au atoms per metal layer and 4 layers in total is repeated periodically in the surface plane to model a Au(111) surface. The two top metal layers are relaxed by minimizing the Hellmann-Feynman forces. The fully relaxed benzene molecule is placed with its molecular plane parallel to the Au(111) surface (see Fig. 4), and the adsorption energy is computed for varying adsorbate–metal distances, d = 2–6 Å. Vacuum layers of 10 Å are used above and below the metal slab and computations are done at a real-space grid spacing of 0.16 Å. Brillouin-zone integrations are performed using a 4 × 4 × 1 Monkhorst-Pack **k**-space grid and a Fermi smearing of 0.1 eV. The most important differences of the computational method used here from the method used in our previous work on the benzene/Au(111) system are that (1) we use here more dense real-space grids for representing wavefunctions, densities, and potentials, (2) we use here four metal layers in the surface model rather than three, and (3) all calculations are now performed selfconsistently. We expect the present calculations to be somewhat more accurate than the previously reported ones.

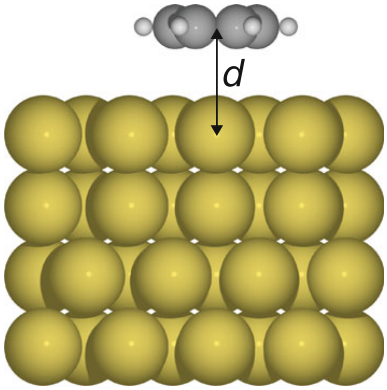

Fig. 4 Side-view of the atomic geometry for benzene adsorption on the noble Au(111) surface. Four metal layers are used in the Au(111) repeated slab model and the C_6H_6 molecule adsorbs with its molecular plane parallel to the surface at the equilibrium adsorbate–metal distance d

We show in Fig. 5 the obtained potential-energy curves for benzene on Au(111). As the RPBE PEC is almost everywhere repulsive (positive E_{ads}) we once again see that ordinary RPBE calculations does not account for vdW-mediated binding. As previously noted the RPBE-vdW changes the GGA picture by introducing an adsorption energy minimum. We find $E_{ads} = -0.47$ eV at $d = 3.7$ Å within RPBE-vdW. The vdW-DF-1 functional behaves very similarly and predicts an adsorption energy of -0.46 eV at $d = 3.6$ Å. Both adsorption energies are roughly 25% too small when compared to the $E_{ads} = -0.64$ eV obtained from TPD experiments in Ref. [30]. Furthermore, it is well known that the vdW-DF-1 tends to overestimate equilibrium binding separations [9 and references therein] so we should probably expect the C_6H_6–Au(111) equilibrium distance to be slightly smaller than 3.6 Å. It is seen from Fig. 5 that the vdW-DF-2 functional indeed accomplishes this as it predicts $d = 3.5$ Å. This makes sense as the vdW-DF-2 was designed in part to improve on the vdW-DF-1 equilibrium separations [9]. The vdW-DF-2 functional does, however, not predict a sufficiently deep potential-energy curve for the benzene/Au(111) system and the obtained adsorption energy of -0.44 eV is no better than that of its vdW-DF-1 predecessor.

Of the van der Waals exchange–correlation functionals considered here only the optPBE-vdW and the RPBEc2/3 + nl appears to perform well for the dispersive C_6H_6/Au(111) interactions. Both predict an equilibrium separation of 3.4 Å with adsorption energies straddling the experimental adsorption energy window, optPBE yielding $E_{ads} = -0.60$ eV and RPBEc2/3 + nl predicting $E_{ads} = -0.64$ eV. The PECs for both functionals furthermore show an asymptotic approach to the non-interacting limit ($d \to \infty$) that is very similar to those of the RPBE-vdW and vdW-DF-1. For a quantitatively correct

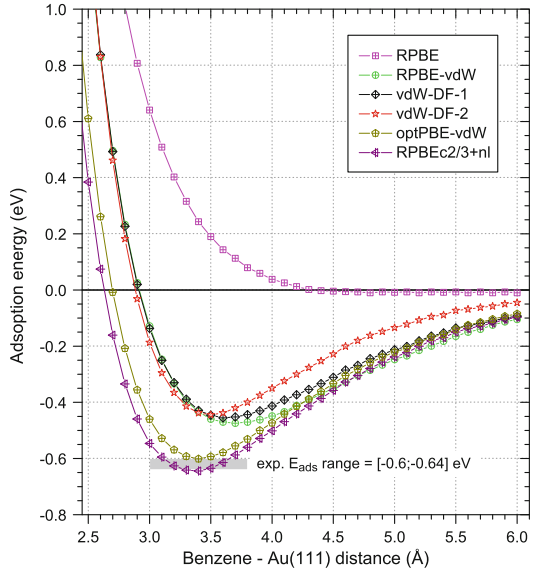

Fig. 5 Selfconsistently calculated potential-energy curves for benzene adsorption on the noble Au(111) surface obtained from the density functionals listed. The RPBE GGA predicts essentially no binding while the van der Waals functionals introduce significantly exothermic adsorption energy maxima below the zero-line. However, only the optPBE-vdW and RPBEc2/3 + nl density functionals are in quantitative agreement with the experimental adsorption energies falling in the range -0.6 to -0.64 eV

GGA-DFT description of the benzene–Au(111) interactions within the electron density dependent vdW-DF scheme it thus appears necessary to employ either an exchange functional that is explicitly fitted to a database of van der Wals interactions or a conveniently chosen mixture of local and semilocal correlation. The optPBE-vdW is an example of the former, while the RPBEc2/3 + nl developed in this work is one possible example of the latter.

5 Conclusions

Accounting for van der Waals interactions in DFT is currently a vibrant field of research. Several approaches to the field exist, one of which is the Langreth–Lundqvist vdW-DF density-dependent inclusion of nonlocal correlation interactions. We find that even though the vdW-DF method does introduce important attractive interactions between van der Waals bonded complexes, which are completely missing in the commonly used RPBE GGA-type approximation to effects of exchange and correlation among electrons, the vdW-DF scheme is not inherently accurate at a quantitative level when compared to advanced wave-function theory. Moreover, structural properties of bulk

solids are poorly described within this method resulting in significant overestimates of bulk lattice parameters and far too small bulk cohesive energies. These issues may be amended by explicitly fitting the exchange functional used for vdW-DF calculations. This, however, comes at the expense of severe overestimates of intra-molecular bond energies. By introducing the RPBEc2/3 + nl functional, which introduces 2/3 of the PBE semilocal correlation functional to the vdW-DF method, we show that a suitable amount of gradient-corrected correlations in vdW-DF-type calculations may provide high accuracy for weak van der Waals bonds as well as structural properties of strongly bonded solids, all while preserving the GGA-level description of intra-molecular bonding.

We furthermore consider benzene physisorption on the noble Au(111) surface, for which the adsorption is mediated by dispersion interactions. While the non-fitted vdW-DF-type functionals RPBE-vdW, vdW-DF-1, and vdW-DF-2 yield adsorption energies that are $\sim$25% too small, the exchange-fitted optPBE-vdW as well as the RPBEc2/3 + nl, not using a fitted exchange part but explicitly containing some semilocal correlations, both agree with experimental adsorption energies for benzene on Au(111). This emphasizes our view that explicit introduction of at least some semilocal correlation to the vdW-DF scheme constitutes a vital step forward towards a generally applicable method for van der Waals DFT, i.e. an exchange–correlation density functional that performs well for vdW interactions while preserving or even improving on the performance of GGA functionals for more strongly interacting systems. The apparent success of the RPBEc2/3 + nl construction illustrates this.

Acknowledgments The Center for Atomic-scale Materials Design is funded by the Lundbeck Foundation. We acknowledge support from the Danish Center for Scientific Computing. Support from the U.S. Department of Energy, Office of Basic Energy Sciences is gratefully acknowledged.

References

1. Kohn W (1999) Rev Mod Phys 71(5):1253
2. Kohn W, Becke AD, Parr RG (1996) J Phys Chem 100(31):12974
3. Sherrill CD (2010) J Chem Phys 132(11):110902
4. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Phys Rev Lett 92(24):246401
5. Thonhauser T, Cooper VR, Li S, Puzder A, Hyldgaard P, Langreth DC (2007) Phys Rev B 76(12):125112
6. Langreth DC et al (2009) J Phys Condens Matter 21(8):084203
7. Klimes J, Bowler DR, Michaelides A (2010) J Phys Condens Matter 22(2):022201
8. Cooper VR (2010) Phys Rev B 81(16):161104(R)
9. Lee K, Murray ED, Kong L, Lundqvist BI, Langreth DC (2010) Phys Rev B 82(8):081101(R)
10. Vydrov OA, Van Voorhis T (2009) J Chem Phys 130:104105
11. Vydrov OA, Van Voorhis T (2010) J Chem Phys 133:244103
12. Jurecka P, Sponer J, Cerny J, Hobza P (2006) Phys Chem Chem Phys 8(17):1985
13. Curtiss LA, Raghavachari K, Redfern PC, Pople JA (1997) J Chem Phys 106(3):1063
14. Hammer B, Hansen LB, Nørskov JK (1999) Phys Rev B 59(11):7413
15. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77(18):3865
16. Zhang YK, Yang WT (1998) Phys Rev Lett 80(4):890
17. Perdew JP, Yue W (1986) Phys Rev B 33(12):8800
18. Mortensen JJ, Hansen LB, Jacobsen KW (2005) Phys Rev B 71(3):035109
19. Bahn SR, Jacobsen KW (2002) Comput Sci Eng 4(3):56
20. Blöchl PE (1994) Phys Rev B 50(24):17953
21. Roman-Perez G, Soler JM (2009) Phys Rev Lett 103(9):096102
22. Wellendorff J, Kelkkanen A, Mortensen JJ, Lundqvist BI, Bligaard T (2010) Top Catal 53(5–6):378
23. Monkhorst HJ, Pack JD (1976) Phys Rev B 13(12):5188
24. Takatani T, Hohenstein EG, Malagoli M, Marshall MS, Sherrill CD (2010) J Chem Phys 132:144104
25. Staroverov VN, Scuseria GE, Tao JM, Perdew JP (2003) J Chem Phys 119(23):12129
26. Lide DR (eds) CRC handbook of chemistry and physics, 90th edn (Internet Version 2010). CRC Press/Taylor and Francis, Boca Raton
27. Kittel C (2005) Introduction to solid state physics, 8th edn. Wiley, New York
28. Klimes J, Bowler DR, Michaelides A (2011) Phys Rev B 83:195131
29. Wöll C (2001) J Synchrotron Radiat 8:129
30. Syomin D, Kim J, Ellison GB, Koel BE (2001) J Phys Chem B 105(35):8387
31. Weiss K, Gebert S, Wühn M, Wadepohl H, Wöll C (1998) J Vac Sci Technol A 16(3):1017
32. Rockey TJ, Yang M, Dai HL (2006) J Phys Chem B 110(40):19973
33. Xi M, Yang MX, Jo SK, Bent BE, Stevens P (1994) J Chem Phys 101(10):9122
34. Netzer FP (1991) Langmuir 7(11):2544
35. Dudde R, Frank KH, Koch EE (1990) Surf Sci 225(3):267
36. Caputo R, Prascher BP, Staemmler V, Bagus PS, Wöll C (2007) J Phys Chem A 111(49):12778

Paper III

Density Functionals for Surface Science: Exchange–correlation Model Development with Bayesian Error Estimation

J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, and K. W. Jacobsen

Physical Review B **85**, 235149 (2012)

Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation

Jess Wellendorff,[1,*] Keld T. Lundgaard,[1] Andreas Møgelhøj,[1,2] Vivien Petzold,[1] David D. Landis,[1] Jens K. Nørskov,[2,3] Thomas Bligaard,[2,3] and Karsten W. Jacobsen[1]

[1]*Center for Atomic-Scale Materials Design (CAMD), Department of Physics, Building 307, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*

[2]*SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA*

[3]*Department of Chemical Engineering, Stanford University, Stanford, California 94305, USA*

(Received 20 March 2012; revised manuscript received 14 May 2012; published 27 June 2012)

A methodology for semiempirical density functional optimization, using regularization and cross-validation methods from machine learning, is developed. We demonstrate that such methods enable well-behaved exchange-correlation approximations in very flexible model spaces, thus avoiding the overfitting found when standard least-squares methods are applied to high-order polynomial expansions. A general-purpose density functional for surface science and catalysis studies should accurately describe bond breaking and formation in chemistry, solid state physics, and surface chemistry, and should preferably also include van der Waals dispersion interactions. Such a functional necessarily compromises between describing fundamentally different types of interactions, making transferability of the density functional approximation a key issue. We investigate this trade-off between describing the energetics of intramolecular and intermolecular, bulk solid, and surface chemical bonding, and the developed optimization method explicitly handles making the compromise based on the directions in model space favored by different materials properties. The approach is applied to designing the Bayesian error estimation functional with van der Waals correlation (BEEF–vdW), a semilocal approximation with an additional nonlocal correlation term. Furthermore, an ensemble of functionals around BEEF–vdW comes out naturally, offering an estimate of the computational error. An extensive assessment on a range of data sets validates the applicability of BEEF–vdW to studies in chemistry and condensed matter physics. Applications of the approximation and its Bayesian ensemble error estimate to two intricate surface science problems support this.

DOI: 10.1103/PhysRevB.85.235149 PACS number(s): 71.15.Mb, 31.15.eg, 68.43.−h

I. INTRODUCTION

Kohn-Sham density functional theory[1,2] (KS–DFT) is a widely celebrated method for electronic-structure calculations in physics, chemistry, and materials science.[3,4] Indeed, modern DFT methods have proven valuable for elucidating mechanisms and fundamental trends in enzymatic and heterogeneous catalysis,[5–13] and computational design of chemically active materials is now within reach.[14–17] Successful use of DFT often relies on accurate but computationally tractable approximations to the exact density functional for the exchange-correlation (XC) energy. The generalized gradient approximation (GGA) is very popular due to a high accuracy-to-cost ratio for many applications, but suffers from a range of shortcomings. Thus, common GGA functionals are well suited for computing many important quantities in chemistry and condensed matter physics, but appear to be fundamentally unable to accurately describe the physics and chemistry of a surface at the same time.[18] Moreover, van der Waals (vdW) dispersion interactions are not accounted for by GGAs,[19] and spurious self-interaction errors can be significant.[20–22] The interest in applying DFT to more and increasingly complex problems in materials science is not likely to decrease in the years to come. Much effort is therefore devoted to improve on current density functional approximations.

The five-rung "Jacob's ladder" of Perdew[23] represents a classification of the most popular density functional methods. Each rung adds new ingredients to the density functional approximation (DFA), and so should enable better approximations, but also adds to the computational cost. In order of increasing complexity, the ladder consists of the local spin-density approximation[1] (LDA), GGA, meta-GGA (MGGA), hyper-GGA, and finally the generalized random phase approximation (RPA). The LDA uses only the local density as input, while rungs 2 and 3 introduce semilocal dependence of the density (GGA) and the KS orbitals (MGGA).[24] Hyper-GGAs introduce nonlocal dependence of the occupied KS orbitals in the exact exchange energy density, and fifth-rung approximations calculate correlation energies from the unoccupied KS orbitals. The latter is computationally heavy, but RPA-type methods are the only DFAs in this five-rung hierarchy that can possibly account for vdW dispersion between nonoverlapped densities.[24]

The failure of lower-rung DFAs in capturing dispersion forces has spurred substantial developments in recent years.[19] Such interactions are spatially nonlocal in nature, and several different approaches to add "vdW terms" to lower-rung DFAs now exist.[25–28] The vdW–DF nonlocal correlation[25] is a particularly promising development in this field. It is a fully nonlocal functional of the ground-state density, and has proven valuable in a wide range of sparse matter studies.[29] However, the vdW–DF and vdW–DF2 (Ref. 30) methods yield much too soft transition-metal crystal lattices,[31,32] and the correct choice of GGA exchange functional to use in vdW–DF type calculations is currently investigated.[30,32–34] One approach to choosing GGA exchange is comparison to Hartree-Fock exchange[35,36] and consideration of the behavior of the exchange functional in the limit of large density

gradients.[35] Where does the vdW–DF approximation belong in a hierarchy such as Jacob's ladder? In terms of computational complexity, the method contains fully nonlocal density-density information without explicit use of the KS orbitals. From this point of view, it should fit between rungs 3 and 4, and we assign it here to rung 3.5. Note that nonlocal exchange approximations, designed to partially mimic exact exchange at a reduced computational cost, have recently been proposed[37,38] as belonging to a rung 3.5.

Put in simple terms, two paradigms for developing density functionals are dominant: that of constraint satisfaction by reduction[24] and that of fitting to empirical data.[39–42] Both have contributed greatly to the success of DFT. Reductionists impose constraints based on analytic properties of the exact density functional, and strive for nonempirical functionals that fulfill as many constraints as possible on each rung of Jacob's ladder. Empirically oriented DFA developers use experimental or high-level theoretical training data to optimize the DFA description of one or more materials properties. Reduction is arguably the most systematic approach to density functional development, and has had a significant impact on the field of KS–DFT. However, choices are often made as to what types of physics and chemistry the DFA should describe well.[43,44] The empirical approach is fundamentally a matter of explicitly making these choices, and parametrize an XC model to suit personal preferences for computational performance. This makes overfitting the training data and transferability of the optimized DFA to systems and materials properties not contained in the training data a central issue.[24]

The risk of overfitting was realized early on by Becke and others.[40,45] Using polynomial expansions of GGA exchange and correlation in least-squares-fitting procedures, polynomial orders above four were found to yield increasingly oscillatory and unphysical XC functionals, that is, "a transition to mathematical nonsense."[45] Nevertheless, semiempirical DFAs containing many parameters have been constructed[42,46,47] with little attention to the overfitting issue. Transferability of a DFA parametrization depends not only on the degree of overfitting to a single set of molecular or condensed matter properties, but also on how many physically different properties the approximate model was trained on. Optimizing XC parametrizations to several different properties naturally leads to a "competition" between data sets in determining the model solution, i.e., an XC model compromise. Implicitly acknowledging this, each data set is often assigned more or less arbitrary weights.[46,47] In our view, such an approach is not guaranteed to yield the optimum model compromise.

In this study, we apply machine-learning methods to avoid the above-mentioned pitfalls of semiempirical density functional development. Regularization of a very flexible polynomial GGA exchange expansion is at the heart of the developed approach. We furthermore investigate the characteristics of XC model compromises in a GGA + vdW model space, and formulate and apply an explicit principle for how an XC model trade-off should be composed. Using several training data sets of quantities representing chemistry, solid state physics, surface chemistry, and vdW dominated interactions, the Bayesian error estimation functional with van der Waals (BEEF–vdW) exchange-correlation model is generated. The three most important aspects of semiempirical DFA design are thus considered in detail: data sets, model space, and model selection. The developed approach furthermore leads to an ensemble of functionals around the optimum one, allowing an estimate of the computational error to be calculated. Lastly, BEEF–vdW is evaluated on systems and properties partly not in the training sets, and is also applied in two small surface science studies: calculating potential-energy curves for graphene adsorption on the Ni(111) surface, and investigation of the correlation between theoretical chemisorption energies and theoretical surface energies of the substrate.

II. DATA SETS

Several sets of energetic and structural data describing bonding in chemical and condensed matter systems are used throughout this study. These data sets are either adapted from literature or compiled here from published works, and are briefly presented in the following. Additional information is found in the Appendix.

(a) *Molecular formation energies*. The G3/99 (Ref. 48) molecular formation enthalpies of Curtiss and co-workers represent intramolecular bond energetics. Experimental room-temperature heats of formation are extrapolated to 0 K, yielding 223 electronic-only static-nuclei formation energies. The G2/97 (Ref. 49) set of 148 formation energies is a subset of G3/99.

(b) *Molecular reaction energies*. Molecular formation energies lend themselves well to compilation of gas-phase reaction energies. The RE42 data set of 42 reaction energies involves 45 different molecules from G2/97.

(c) *Molecular reaction barriers*. The DBH24/08 (Ref. 50) set of Zheng *et al.*, comprising 12 forward and 12 backward benchmark barriers, is chosen to represent gas-phase reaction barriers.

(d) *Noncovalent interactions*. The S22 (Ref. 51) and S22x5 (Ref. 52) sets of intermolecular interaction energies of noncovalently bonded complexes calculated at the coupled-cluster level with single, double, and perturbative triple excitations [CCSD(T)] were compiled by Hobza and co-workers. Particularly, the S22 set has become popular for assessment[34,53–58] and parametrization[30,33,47,54,59,60] of density functional methods for vdW–type interactions. The S22x5 set consists of potential-energy curves (PECs) for each S22 complex, with interaction energies at relative interaction distances d of 0.9, 1.0, 1.2, 1.5, and 2.0 as compared to S22, totaling 110 data points. For convenience, this study divides S22x5 into five subsets according to interaction distance, e.g., "S22x5-0.9."
The accuracy of the original S22 and S22x5 energies have certain deficiencies, so the revised S22x5-1.0 energies of Takatani *et al.*[61] are used instead. The remaining (nonequilibrium) data points on each CCSD(T) PEC are correspondingly corrected according to the difference between original and revised S22x5-1.0 energies, as elaborated on in the Appendix.

(e) *Solid state properties*. Three sets of 0-K experimental solid state data are used, here denoted Sol34Ec, Sol27LC, and Sol27Ec. The first comprises cohesive energies of 34 period 2–6 bulk solids in fcc, bcc, diamond, and hcp lattices. Zero-point phonon effects have not been corrected for. Conversely, the Sol27LC and Sol27Ec sets contain lattice constants and

cohesive energies, respectively, of 27 cubic lattices, both corrected for zero-point vibrational contributions.

(f) *Chemisorption on solid surfaces*. The CE17 and CE27 data sets comprise experimental reaction energies for chemisorption of simple molecules on the (111), (100), and (0001) facets of late transition-metal surfaces at low coverage. The CE17 set is a subset of CE27.

III. COMPUTATIONAL DETAILS

Self-consistent density functional calculations are performed using GPAW,[62–64] a real-space grid implementation of the projector augmented-wave method.[65] The ASE (Refs. 64 and 66) package provides a convenient interface to GPAW. Grid-point spacings of 0.16 Å are employed for high-quality computations of simple properties such as molecular binding energies. Properties of bulk solids are calculated using somewhat denser grids with a spacing of 0.13 Å. Real-space structure relaxation is applied to the G3/99 molecules and CE27 chemisorption systems with 0.05 eV/Å as the criterion of maximum force on each relaxing atom. Molecular and single-atomic systems are placed in a box with at least 7 Å vacuum to the box boundaries, except for the S22x5 complexes for which the vacuum width is 10 Å. Further details on the computational procedure employed are found in the Appendix.

IV. MODEL SPACE

The GGA exchange energy density $\varepsilon_x^{\mathrm{GGA}}(n,\nabla n)$ is conveniently expressed in terms of the exchange energy density of the uniform electron gas $\varepsilon_x^{\mathrm{UEG}}(n)$ and an exchange enhancement factor $F_x(s)$, depending on the local density as well as its gradient through the reduced density gradient s,

$$\begin{aligned} s &= \frac{|\nabla n|}{2k_F n}, \quad 0 \leqslant s \leqslant \infty, \\ \varepsilon_x^{\mathrm{GGA}}(n,\nabla n) &= \varepsilon_x^{\mathrm{UEG}}(n) F_x[s(n,\nabla n)], \\ E^{\mathrm{GGA\text{-}x}}[n,\nabla n] &= \int \varepsilon_x^{\mathrm{UEG}}(n) F_x[s(n,\nabla n)] d\mathbf{r}, \end{aligned} \tag{1}$$

where $n = n(\mathbf{r})$, $k_F = (3\pi^2 n)^{1/3}$ is the Fermi wave vector of the UEG, and $E^{\mathrm{GGA-x}}$ is the semilocal GGA exchange energy.

In this study, a highly general exchange model space is obtained by expanding the GGA exchange enhancement factor in a basis of M_x Legendre polynomials $B_m[t(s)]$ of orders 0 to $M_x - 1$ in a transformed reduced density gradient, denoted $t(s)$:

$$\begin{aligned} t(s) &= \frac{2s^2}{4+s^2} - 1, \quad -1 \leqslant t \leqslant 1 \\ F_x^{\mathrm{GGA}}(s) &= \sum_m a_m B_m[t(s)], \\ E^{\mathrm{GGA\text{-}x}}[n,\nabla n] &= \sum_m a_m \int \varepsilon_x^{\mathrm{UEG}}(n) B_m[t(s)] d\mathbf{r} \\ &= \sum_m a_m E_m^{\mathrm{GGA\text{-}x}}[n,\nabla n], \end{aligned} \tag{2}$$

where a_m are expansion coefficients, and $E_m^{\mathrm{GGA\text{-}x}}$ is the exchange energy corresponding to the Legendre basis function B_m. The polynomial basis is constructed such that the boundary limits $t = [-1,1]$ are zero for all $m > 1$ basis functions. Therefore, these limits are determined by the order 0 and 1 basis functions only.

Semilocal approximations to electron correlation effects beyond GGA exchange are not easily cast in terms of a single variable, such as s. The correlation model space is chosen to be a linear combination of the Perdew-Burke-Ernzerhof (PBE) (Ref. 67) semilocal correlation functional, purely local Perdew-Wang[68] LDA correlation, and vdW–DF2 (Ref. 30) type nonlocal correlation. The latter is calculated from a double integral over a nonlocal interaction kernel $\phi(\mathbf{r},\mathbf{r}')$,

$$E^{\mathrm{nl\text{-}c}}[n] = \frac{1}{2}\int n(\mathbf{r})\phi(\mathbf{r},\mathbf{r}')n(\mathbf{r}')d\mathbf{r}\,d\mathbf{r}', \tag{3}$$

which is evaluated using the fast Fourier transformation method of Román-Pérez and Soler,[69] implemented in GPAW as described in Ref. 70.

In total, the XC model space consequently consists of GGA exchange expanded in Legendre polynomials as well as local, semilocal, and nonlocal correlation,

$$\begin{aligned} E_{xc} = &\sum_{m=0}^{M_x-1} a_m E_m^{\mathrm{GGA\text{-}x}} + \alpha_c E^{\mathrm{LDA\text{-}c}} \\ &+ (1-\alpha_c)E^{\mathrm{PBE\text{-}c}} + E^{\mathrm{nl\text{-}c}}, \end{aligned} \tag{4}$$

where $M_x = 30$, and the total number of parameters is $M = M_x + 1 = 31$.

None of the commonly imposed constraints on GGA exchange are invoked, e.g., the LDA limit of $F_x(s)$ and recovery of the correct gradient expansion for slowly varying densities, nor the Lieb-Oxford (LO) bound[71,72] for large electron density gradients. However, as seen from Eq. (4), the sum of LDA and PBE correlation is constrained to unity.

V. MODEL SELECTION

Choices are made when developing a semiempirical density functional. These are both explicit and implicit choices pertaining to what the functional is to be designed for, that is, for the selection of an optimum exchange-correlation model that captures the materials properties of main interest when applying the approximation. This study aims to explicate the choices, and to develop a set of principles for the model selection process. These principles are used to guide the inevitable compromise between how well significantly different quantities in chemistry and condensed matter physics are reproduced by an incomplete XC model space. Development of an XC functional is in this approach divided into two steps. First an individual model selection for a number of data sets is carried out, and subsequently a simultaneous model selection is made, compromising between the individual fits.

A. Individual materials properties

1. *Regularizing linear models*

Model training is formulated in terms of finding the expansion coefficient vector that minimizes a cost function without overfitting the data. This may be viewed as determining the optimum trade-off between bias and variance of the model.[73] The cost function contains two terms: a squared error term and

a regularization term. One simple regularization suitable for varying the bias-variance ratio is one that "penalizes" the cost function for model solutions that differ from a suitably chosen prior solution. This effectively removes sloppy[74] eigenmodes of the cost function by adding curvature to all modes, and thereby limits the effective number of parameters in the model solution. As the regularization strength is continuously decreased from infinity towards zero, the model parameters that minimize the cost function are allowed to differ increasingly from the prior solution. In a sufficiently large model space, the solution that reproduces the data best without overfitting is in general found for intermediate regularization strength. A slightly more elaborate regularization is used in this study, as outlined later on.

Finding the optimum model is then a matter of determining the optimum regularization strength. This may be done by minimizing the estimated prediction error (EPE) for varying regularization strength. The EPE provides a statistical estimate of the validity of a model outside the space of its training data, and can be obtained by a large variety of resampling methods. We obtain it using bootstrap resampling.[75] Even though common error quantities, such as the standard deviation (STD), will in general decrease for regularization strengths smaller than that which minimizes the EPE, the corresponding model solutions are likely to be increasingly overfitted. Minimizing the EPE and not the STD is therefore preferred for determining well-behaved XC functionals.

2. Details of the procedure

The standard Tikhonov regularization method[73] is chosen to control overfitting. A cost function for the ith data set is therefore defined as

$$C_i(\mathbf{a}) = (\mathbf{X}_i\mathbf{a} - \mathbf{y}_i)^2 + \omega^2\Gamma^2(\mathbf{a} - \mathbf{a}_p)^2, \tag{5}$$

where $\mathbf{X}_i$ is a data matrix, $\mathbf{a}$ the coefficient vector, $\mathbf{y}_i$ a target vector of training data, ω^2 the regularization strength, Γ is denoted the Tikhonov matrix, and the prior vector $\mathbf{a}_p$ is the origo for regularization, i.e., the model solution for $\omega^2 \to \infty$ and thus the model space reference point for regularization.

In accordance with Eq. (4), the data matrix consists of XC contributions to a materials property for each system in the ith data set from the M basis functions. These are evaluated non-self-consistently on revised PBE (RPBE) (Ref. 76) densities. The target vector contains the target XC contribution to each quantity in the set. The Tikhonov matrix is defined from a smoothness criterion on the basis functions. The exchange part of Γ is the overlap of the second derivative of the exchange basis functions with respect to the transformed reduced density gradient

$$\mathbf{\Gamma}^2_{ij} = \int_{-1}^{1} \frac{d^2B_i(t)}{dt^2}\frac{d^2B_j(t)}{dt^2}dt. \tag{6}$$

Defined this way, the Tikhonov matrix directly penalizes the integrated squared second derivative of the exchange fit for finite regularization strength. This can be understood as penalizing a measure of nonsmoothness of the fitted exchange enhancement factor. In effect, the $\mathbf{\Gamma}$ matrix scales the regularization strength acting on each exchange basis function, such that higher-order basis functions are suppressed when minimizing the cost function. This leads to a model selection preference for solution vectors with small coefficients for higher-order polynomials, unless they are essential for obtaining a satisfactory fit. Physically, it is very reasonable to require $F_x(s)$ to be a smooth and preferably injective function of s, and significantly nonsmooth exchange solutions have been shown to degrade transferability of fitted exchange functionals to systems outside the training data.[77] The correlation part of Γ has one in the diagonal and zeros in the off-diagonal elements. Since $\mathbf{\Gamma}$ acts in the transformed $t(s)$ space, the transformation in Eq. (2) causes the regularization penalty on exchange to be strongest in the large-s regime, where information from the data matrix about the optimum behavior of $F_x(s)$ is expected to be scarce.[76,78]

In order to minimize the cost function in Eq. (5), it is transformed by $\mathbf{\Gamma}^{-1}$. Ones are therefore inserted in the first two diagonal elements of $\mathbf{\Gamma}$ to avoid numerical issues. The solution vector $\mathbf{a}_i$ that minimizes C_i is written as

$$\mathbf{a}_i = \mathbf{\Gamma}^{-1}\left(\mathbf{X}_i'^T\mathbf{X}_i' + \mathbf{L}^2\omega_i^2\right)^{-1}\left(\mathbf{X}_i'^T\mathbf{y}_i + \omega^2\mathbf{L}^2\mathbf{a}_p'\right), \tag{7}$$

where $\mathbf{X}_i' = \mathbf{X}_i\mathbf{\Gamma}^{-1}$, $\mathbf{a}_p' = \mathbf{\Gamma}\mathbf{a}_p$, and $\mathbf{L}^2$ is the identity matrix with zeros in the first two diagonal elements. Singular value decomposition of $\mathbf{X}_i'^T\mathbf{X}_i'$ is used to calculate the inverse matrix. The LDA and PBE correlation coefficients in the XC model are constrained to be between 0 and 1, implying $\alpha_c \in [0,1]$ for the correlation coefficient in Eq. (4). In the cases that this is not automatically fulfilled, it is enforced by recalculating the solution while fixing α_c to the nearest bound of the initial solution.

The exchange part of the prior vector is chosen as the linear combination of the order 0 and 1 polynomial basis functions that fulfills the LDA limit at $s = 0$ and the LO bound for $s \to \infty$. With the exchange basis transformation in Eq. (2), the prior for exchange is quite close to the PBE exchange enhancement factor. For $\omega^2 \to \infty$, we therefore nearly recover PBE exchange, while lower regularization strengths allow increasingly nonsmooth variations away from this prior solution. The optimum model is expected to include at least some semilocal correlation,[31] so the origo of correlation is $\alpha_c = 0.75$.

As previously mentioned, the optimum regularization is found by minimizing the estimated prediction error for varying ω^2. Bootstrap resampling of the data matrix with the .632 estimator[75,79] is used. It is defined as

$$\mathrm{EPE}_{.632} = \sqrt{0.368 \cdot \widehat{\mathrm{err}} + 0.632 \cdot \widehat{\mathrm{Err}}}, \tag{8}$$

where $\widehat{\mathrm{err}}$ is the variance between the target data and the prediction by the optimal solution $\mathbf{a}_i$, and $\widehat{\mathrm{Err}}$ measures the variance on samples of data to which solutions were not fitted in the resampling. Both are determined as a function of ω^2, and $\widehat{\mathrm{Err}}$ is given by

$$\widehat{\mathrm{Err}} = \frac{1}{N_\mu}\sum_\mu \frac{1}{N_{s|\mu\notin s}}\sum_{s|\mu\notin s}(\mathbf{x}_\mu\mathbf{b}_s - y_\mu)^2, \tag{9}$$

where μ is an entry in the data set, N_μ the number of data points, s a bootstrap sample of N_μ data points, and $N_{s|\mu\notin s}$ the number of samples not containing μ. The parentheses calculate the difference between the prediction $\mathbf{x}_\mu\mathbf{b}_s$ of the data point μ by the best-fit coefficient vector $\mathbf{b}_s$ and the μth target value y_μ.

TABLE I. Model selection results of individually training the XC model of Eq. (4) to 10 different data sets. M_{eff} is the effective number of parameters in a model [see Eq. (21)]. The $s = 0$ and $s \to \infty$ limits of the obtained exchange enhancement factors are also shown. MSD, MAD, and STD are mean signed, mean absolute, and standard deviation, respectively, all in meV. Note that these are non-self-consistent results.

	α_c	M_{eff}	$F_x(0)$	$F_x(\infty)$	MSD	MAD	STD
CE17	0.90	4.7	0.97	2.15	−10	96	116
RE42	1.00	4.2	1.06	1.21	19	168	207
DBH24/08	0.00	3.7	1.14	3.14	1	116	142
G2/97	0.27	7.2	1.10	2.53	−13	109	149
Sol34Ec	0.00	7.7	0.97	1.25	−4	168	208
S22x5-0.9	0.81	3.2	0.96	1.68	0	9	11
S22x5-1.0	0.82	3.1	0.98	1.87	0	8	10
S22x5-1.2	0.40	5.7	1.04	2.38	0	4	6
S22x5-1.5	0.85	4.0	1.02	1.91	−1	3	4
S22x5-2.0	1.00	3.3	0.95	1.37	2	3	3

The best-fit solution is found by minimizing the cost function with the data in sample s only.

In the bootstrap resampling procedure, 500 randomly generated data samples are selected independently for each ω^2. The regularization strength that minimizes the .632 estimator is found by a smooth fitting of the slightly scattered estimator plot near the minimum. To properly regularize the S22x5 subsets with long interaction distances, a condition $F_x(s = \infty) \geqslant 1$ is enforced.

3. *Individually trained XC models*

Table I and Fig. 1 show details and statistics for the optimized XC models obtained when the procedure outlined above is applied to molecular, solid state, surface chemical, and vdW dominated energetics. Each model is therefore trained on a single materials property only, and their features differ significantly.

The DBH24/08 set appears to favor GGA exchange that substantially violates the LDA limit [$F_x(0) = 1.14$] along with inclusion of full PBE correlation ($\alpha_c = 0$; no LDA correlation). The model furthermore overshoots the LO bound $F_x^{\text{LO}} = 1.804$ significantly [$F_x(\infty) = 3.14$]. The XC model optimized to the G2/97 set shows similar trends with respect to GGA exchange and PBE correlation, but is less extreme.

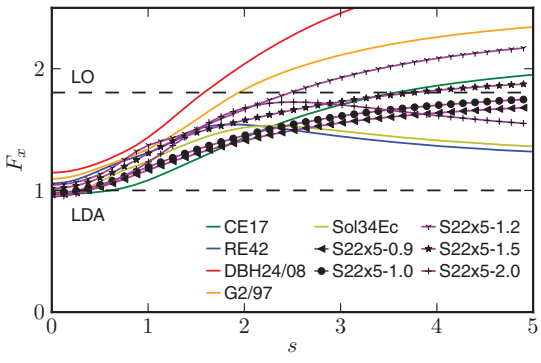

FIG. 1. (Color online) Exchange enhancement factors of the individually trained XC models listed in Table I.

In the other end of the spectrum is the model optimized to the Sol34Ec cohesive energies. These favor GGA exchange starting out slightly below $F_x = 1$, then reaching a maximum at $s \approx 2$, and finally declining slowly towards $F_x = 1.25$. Best agreement with experimental cohesive energies is found with full PBE correlation in addition to nonlocal correlation. The occurrence of a maximum in the exchange enhancement factor should, however, not be overemphasized. It has been shown[76,78] that only small GGA exchange contributions to chemical and solid state binding energetics can be attributed to reduced density gradients above 2.5. In the region of large s, where the smoothness criterion on exchange is strongly enforced, the regularization term in the cost function [Eq. (5)] will therefore be dominant in determining the solution for such systems. The regularization may therefore well determine the behavior of $F_x(s)$ for large density gradients.

For the remaining data sets in Table I, the optimized XC models appear reasonable, with all exchange enhancement factors starting out near the LDA limit. It is illustrative to investigate how the XC models perform for data sets on which they were not trained. The standard deviation is a natural measure of performance. Defining the relative standard deviation rSTD on some data set with some XC model, as the STD obtained by that model divided by the STD of the model that was fitted to that data set, rSTD is a measure of transferability. Figure 2 shows a color map of the rSTD for all 10 training data sets with all 10 trained models. The diagonal from bottom left to top right is, by definition, ones. In a background of blue and yellow-green squares, the map features two distinct areas of mostly reddish squares. To the far right, the S22x5-2.0 model yields rSTD > 5 for all other sets than DBH24/08, and rSTD ≈ 28 for S22x5-0.9. Furthermore, a 5×4 square in the top left corner illustrates that XC models trained on chemical or solid state data sets perform significantly worse on vdW–type energetics than models fitted to the latter. It is also interesting to see that the S22x5-2.0 rSTDs are largely unaffected by changing XC models. With little or no density-density overlap between many of the S22x5-2.0 complexes, the constant nonlocal correlation in all 10 models is likely the main XC contribution to intermolecular binding.

In summary, the deviation statistics in Table I illustrate that the XC model space considered here most certainly spans the

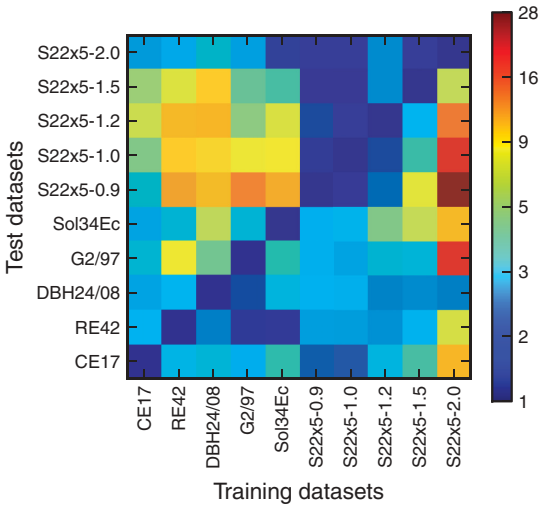

FIG. 2. (Color online) Color map of the relative standard deviations obtained when non-self-consistently applying the XC models found individually for each training data set, listed on the abscissa, to all 10 training data sets along the ordinate.

model degrees of freedom necessary to obtain well-performing density functionals with smooth exchange enhancement factors and sound correlation components. However, a high degree of transferability between the data sets should not be expected for several of the models.

B. Several materials properties

Fundamentally, a compromise has to be made between how well different materials properties are reproduced by the same semiempirical density functional. This is expressed as a compromise between how well the functional quantitatively performs on different training data sets. What the compromise should be can only be determined by the target applications of the functional, and one challenge is to make this choice as explicit as possible. This section presents one route towards a methodology for optimizing an XC model to simultaneously describe several different materials properties. First, the nature of the model compromise is illustrated for the case of simultaneously fitting two data sets using a summed cost function with varying weights on the two sets. However, in the end, a product cost function is found more convenient for determining the optimum weights according to the directions in model space favored by different data sets.

1. Model compromise

Consider first the problem of simultaneously fitting two data sets, and let the model compromise be described through the total cost function, given as the sum of the two individual cost functions:

$$\Lambda(\mathbf{a}) = \mathcal{W}_1 C_1(\mathbf{a}) + \mathcal{W}_2 C_2(\mathbf{a}), \tag{10}$$

where $\mathcal{W}_i$ is a weight on data set i. The coefficient vector solution $\mathbf{b}$ that minimizes $\Lambda(\mathbf{a})$ is found by setting the derivative to zero: Since the summed cost function is quadratic in $\mathbf{a}$, as the individual cost functions C_i are, it may be expressed in terms of the individual solutions $\mathbf{a}_i$ as

$$\Lambda(\mathbf{a}) = \sum_{i=1,2} \mathcal{W}_i \left(C_i^0 + \frac{1}{2}(\mathbf{a}-\mathbf{a}_i)^T \mathbf{H}_i (\mathbf{a}-\mathbf{a}_i) \right), \tag{11}$$

where $C_i^0 = C_i(\mathbf{a}_i)$ is the minimized cost of data set i, and $\mathbf{H}_i$ is the Hessian of $C_i(\mathbf{a})$. The minimizing solution $\mathbf{b}$ is thus found from the individual solutions $\mathbf{a}_i$ as

$$\mathbf{b} = \left(\sum_{i=1,2} \mathcal{W}_i \mathbf{H}_i \right)^{-1} \left(\sum_{i=1,2} \mathcal{W}_i \mathbf{H}_i \mathbf{a}_i \right). \tag{12}$$

However, a principle for guiding the choice of weights is needed.

Let us consider establishing a compromise based on explicit principles. The regularized cost functions for each training data set $C_i(\mathbf{a})$ contain information of the costs associated with deviating from the individually found model solutions $\mathbf{a}_i$ along all directions in model space. The individual costs all increase when moving away from $\mathbf{a}_i$ due to deterioration of the fits, increased overfitting, or a combination of both. Define now the relative cost for each data set, rCost[i], as the individual cost for set i evaluated at the compromising solution $\mathbf{b}$ relative to the individual cost at $\mathbf{a}_i$, hence

$$\mathrm{rCost}[\,i\,] = \frac{C_i(\mathbf{b})}{C_i(\mathbf{a}_i)} = \frac{C_i(\mathbf{b})}{C_i^0} \geqslant 1. \tag{13}$$

Thus defined, the relative cost for each training data set is a simple measure of how unfavorable it is for each data set to be fitted by the compromising solution $\mathbf{b}$ instead of the individual solutions $\mathbf{a}_i$.

The main panel of Fig. 3 illustrates XC model compromises between the G2/97 and S22x5-1.0 data sets. The curve maps out the relative costs on both data sets obtained from model solutions $\mathbf{b}$ when systematically varying the weights in $\Lambda(\mathbf{a})$. The weight fraction f is introduced (see caption of Fig. 3). A wide range of poorly compromising models can obviously be produced, sacrificing a lot of relative cost on one set while gaining little on the other. However, if both materials properties represented by the two data sets are considered important, the optimum weightening is somewhere midway between the asymptotic extrema.

The inset in Fig. 3 shows how the product of the relative costs varies with f. To the right along the abscissa, where the fraction increasingly favors the G2/97 set, the rCost product increases rapidly. To the left, the increase is much smaller, but a minimum is located in-between. At least one intermediate minimum is always present since the slopes in the two asymptotic regions are $-\infty$ and 0, respectively. This property is induced by the variational property around the two original minima of the individual cost functions. Similar conclusions apply to any combination of two or more data sets that do not favor the same directions in the incomplete model space.

We find in general that the condition of minimizing the product of relative costs is well suited for choosing cost function weights for arbitrary numbers of training data sets, if the aim is a general-purpose model. This condition, which

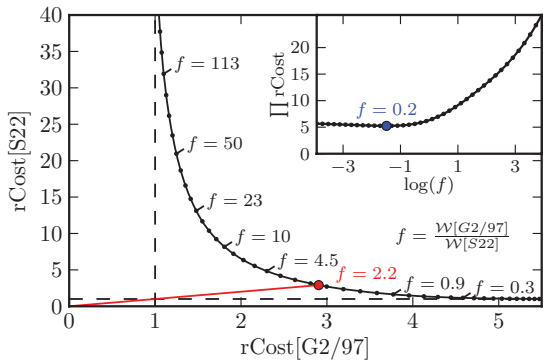

FIG. 3. (Color online) Main panel: XC model compromises between the G2/97 and S22x5-1.0 data sets illustrated in terms of relative costs (rCost) for both data sets when the weight fraction $f = \mathcal{W}[\mathrm{G2/97}]/\mathcal{W}[\mathrm{S22x5\text{-}1.0}]$ is varied and the summed cost function Eq. (10) is minimized. A range of compromising solutions are obtained, many of which are essentially fitting one data set only (rCost ≈ 1) while sacrificing model performance on the other (rCost $\gg 1$). A red dot marks the point of equal rCost. The fact that an XC model with $\mathrm{rCost[G2/97]} = \mathrm{rCost[S22x5\text{-}1.0]} = 1$ is not obtainable illustrates the necessity of a model compromise. Inset: The product of relative costs display a minimum (blue dot) for a certain weight fraction.

is identical to minimizing the product of costs, is applied henceforth.

2. Product cost function

A product cost function for arbitrary numbers of training data sets is here defined, such that the minimizing solution $\mathbf{c}$ yields a desired minimum of the product of costs. The cost function is written as

$$\Phi(\mathbf{a}) = \prod_i C_i(\mathbf{a})^{w_i}, \tag{14}$$

where w_i is a constant weight, and C_i is again an individual cost function. The constant weight is an important feature of $\Phi(\mathbf{a})$ since it allows inclusion of training data sets which are perceived significantly less important than others. It is thus chosen from personal preferences given the purpose of the functional, and we shall see that $\mathbf{c}$ minimizes the product of costs *given* this choice.

For the case of two data sets, the stationary point between the two individual solutions in model space is found by differentiating the logarithm of $\Phi(\mathbf{a})$ with respect to $\mathbf{a}$, and solving

$$\sum_i \frac{w_i}{C_i}\frac{dC_i}{d\mathbf{a}} = 0. \tag{15}$$

Using the method outlined above, the model solution that minimizes $\Phi(\mathbf{a})$ is found in terms of the individual solutions as

$$\mathbf{c} = \left(\sum_i \frac{w_i}{C_i}\mathbf{H}_i\right)^{-1}\left(\sum_i \frac{w_i}{C_i}\mathbf{H}_i\mathbf{a}_i\right), \tag{16}$$

where $C_i = C_i(\mathbf{c})$, and w_i simply scales the individual costs. We see that this solution corresponds to letting $\mathcal{W}_i$ in Eq. (11) equal w_i/C_i. Thus, minimizing the product of costs has introduced a natural weight C_i^{-1}, while w_i still leave room for deliberately biasing the model solution.

From here on, the product solution is therefore used to find the desired XC model solution: Since C_i is evaluated at $\mathbf{c}$, the optimum solution is found iteratively, using C_i^{-1} as an iterator while searching for a converged minimum of the product cost function, given the constant weights w_i.[80]

3. BEEF–vdW density functional

The BEEF–vdW exchange-correlation functional was designed using the set of weights w listed in Table II. In principle, these should all equal one, however, correlations between some of the data sets have led us to lower the constant weight on some of them: Since the RE42 set is based on G2/97 molecules, the data in RE42 are correlated with some of the data in G2/97. Both weights were therefore lowered to 0.5. The same reasoning applies to the S22x5 subsets, where the same complexes are found in all the five sets, albeit at different interaction distances. A weight of $1/5 = 0.2$ on each S22x5 subset would therefore be natural, but for reasons of performance of the final functional, constant weights of 0.1 were chosen. The prior vector was the same for the combined functional as for the individual models.

The resulting model compromise is also tabulated in Table II, showing the effective data-set weight w/C, rCost, and rSTD for all data sets used in model training. It is clearly seen that especially the S22x5-0.9 interaction energies are hard to fit simultaneously with the other data sets within the XC model space employed here: The relative cost for the set is high, allowing the model to adapt mostly to the other data sets by lowering w/C for this set. This is furthermore reflected in the rSTD of 5.4, indicating that the BEEF–vdW performance on this data set is significantly worse than obtained in the individual fit to the S22x5-0.9 systems reported in Table I. Even so, the remaining S22x5 subsets appear to share XC

TABLE II. The BEEF–vdW model compromise. The effective weight in determining the XC model solution is w/C for each data set, as iteratively found from minimizing the product cost function [Eq. (14)]. The relative standard deviation (rSTD) is the ratio of the STD at the BEEF–vdW compromise to the STD at the regularized individual solution in Table I. The relative costs (rCost) are defined similarly, but includes regularization [see Eq. (13)].

	w	w/C	rCost	rSTD
CE17	1.0	1.80	1.7	1.3
RE42	0.5	0.62	2.5	1.8
DBH24/08	1.0	0.65	4.9	2.3
G2/97	0.5	0.62	2.6	1.6
Sol34Ec	1.0	0.43	7.5	2.8
S22x5-0.9	0.1	0.01	28.6	5.4
S22x5-1.0	0.1	0.04	9.1	2.9
S22x5-1.2	0.1	0.09	3.5	2.1
S22x5-1.5	0.1	0.08	4.1	2.1
S22x5-2.0	0.1	0.18	1.8	1.5

TABLE III. Expansion coefficients a_m for the BEEF–vdW Legendre exchange basis functions of order m. The correlation mixing parameter, α_c in Eq. (4), is 0.6001664769.

m	a_m	m	a_m
0	$1.516501714 \times 10^{0}$	15	$-8.018718848 \times 10^{-4}$
1	$4.413532099 \times 10^{-1}$	16	$-6.688078723 \times 10^{-4}$
2	$-9.182135241 \times 10^{-2}$	17	$1.030936331 \times 10^{-3}$
3	$-2.352754331 \times 10^{-2}$	18	$-3.673838660 \times 10^{-4}$
4	$3.418828455 \times 10^{-2}$	19	$-4.213635394 \times 10^{-4}$
5	$2.411870076 \times 10^{-3}$	20	$5.761607992 \times 10^{-4}$
6	$-1.416381352 \times 10^{-2}$	21	$-8.346503735 \times 10^{-5}$
7	$6.975895581 \times 10^{-4}$	22	$-4.458447585 \times 10^{-4}$
8	$9.859205137 \times 10^{-3}$	23	$4.601290092 \times 10^{-4}$
9	$-6.737855051 \times 10^{-3}$	24	$-5.231775398 \times 10^{-6}$
10	$-1.573330824 \times 10^{-3}$	25	$-4.239570471 \times 10^{-4}$
11	$5.036146253 \times 10^{-3}$	26	$3.750190679 \times 10^{-4}$
12	$-2.569472453 \times 10^{-3}$	27	$2.114938125 \times 10^{-5}$
13	$-9.874953976 \times 10^{-4}$	28	$-1.904911565 \times 10^{-4}$
14	$2.033722895 \times 10^{-3}$	29	$7.384362421 \times 10^{-5}$

model space with the data sets representing formation and rupture of interatomic bonds to a significantly greater extent. Thus, accurate description of the balance of strong and weak interactions in the S22x5-0.9 complexes is nearly incompatible with at least one of the other sets of materials properties, when demanding well-behaved exchange and correlation functionals in the present model space.

Table III lists the BEEF–vdW expansion coefficients. The correlation functional consists of 0.6 LDA, 0.4 PBE, and 1.0 nonlocal correlation. The qualitative shape of the BEEF–vdW exchange enhancement factor is shown in Fig. 4, with $s = 0$ and $s \to \infty$ limits of 1.034 and 1.870, respectively. Thus, BEEF–vdW exchange does not exactly obey the LDA limit for $s = 0$, but is 3.4% higher. The enhancement factor is above most GGA exchange functionals up to $s \approx 2.5$, from where it approaches the LO bound with a small overshoot in the infinite limit.

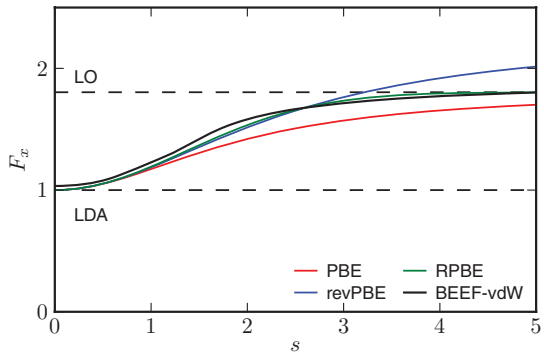

FIG. 4. (Color online) The BEEF–vdW exchange enhancement factor compared to those of a few standard GGA exchange functionals. The corresponding BEEF–vdW correlation functional is composed of 0.6 LDA, 0.4 PBE, and 1.0 nonlocal correlation.

The lack of exact fulfillment of the LDA limit for exchange indicates a conflict between this limit, the training data, and the employed preference for smooth exchange models. The G2/97 and DBH24/08 chemical data sets are found to give particular preference to exchange enhancement models with $F_x(0) \approx 1.1$, and enforcing $F_x(0) = 1.0$ for these sets leads to severely nonsmooth exchange solutions for $s \to 0$. Similar behavior was found in Ref. 77. Note that MGGA approximations are able to achieve exchange models with $F_x(0) \neq 1.0$ for densities different from the UEG, while still obeying the LDA limit for UEG-like densities. The BEEF–vdW F_x also has small "bump" at $s \approx 1.3$. This is not essential to the quality of the model and is not expected to harm its transferability. However, completely removing such features requires overly strong regularization.

VI. ENSEMBLE ERROR ESTIMATION

A normal DFT calculation does not provide any information about the uncertainty of the result from using an approximate XC functional. One method to obtain an estimate of the uncertainty is performing several calculations using different functionals, and observe the variations in the prediction of the quantity of interest. Another more systematic approach is to use an ensemble of functionals designed to provide an error estimate, as discussed in Ref. 81. This method is applied to the BEEF–vdW model, and the adaptation is briefly presented here.

Inspired by Bayesian statistics,[73] we define a probability distribution P for the model parameters $\mathbf{a}$ given the model θ and training data D:

$$P(\mathbf{a}|\theta D) \sim \exp[-C(\mathbf{a})/\tau], \tag{17}$$

where $C(\mathbf{a})$ is the cost function, and τ is a cost "temperature." Given the data D, a model perturbation $\delta\mathbf{a}$ has a certain probability associated with it, and this defines an ensemble of different XC functionals. The temperature is to be chosen such that the spread of the ensemble model predictions of the training data reproduces the errors observed when using BEEF–vdW self-consistently. This approach to constructing the probability distribution is closely related to the maximum entropy principle.[77,82]

The ensemble is defined through a Hessian scaled with the temperature. The Hessian is calculated directly from

$$\mathbf{H} = 2\sum_{i}^{N} \frac{w_i}{C_i(\mathbf{a}_p)} \mathbf{\Gamma}^{-1} \left(\mathbf{X}_i'^{T} \mathbf{X}_i' + \omega_i^2 \mathbf{L}^2 \right) \mathbf{\Gamma}^{-1^T}, \tag{18}$$

where the sum is over training data sets. The temperature is related to the effective number of parameters in the model, calculated from the effective regularization

$$\omega_{\text{eff}}^2 = \sum_{i}^{N} \frac{w_i}{C_i(\mathbf{c})} \omega_i^2, \tag{19}$$

where ω_i^2 are the regularization strengths for the individual data sets. Additionally, diagonalization of the combined square of the transformed data matrix

$$\mathbf{\Sigma}' = \mathbf{V}^T \left(\sum_{i}^{N} \frac{w_i}{C_i(\mathbf{c})} \mathbf{X}_i'^{T} \mathbf{X}_i' \right) \mathbf{V}, \tag{20}$$

where Σ' contains the eigenvalues along the diagonal and $\mathbf{V}$ the eigenvectors, allows the effective number of parameters left in the model after regularization, M_{eff}, to be computed as

$$M_{\text{eff}} = \sum_{m}^{M} \frac{{\Sigma'_m}^2}{{\Sigma'_m}^2 + \omega_{\text{eff}}^2 L_m^2}. \tag{21}$$

Since $M_{\text{eff}} = 7.11$ in the BEEF–vdW model compromise, more than 75% of the initially 31 model degrees of freedom have been suppressed by regularization.

The temperature calculation is slightly modified from the method in Ref. 81 in order to construct an unbiased error estimation. This reflects that a larger error is expected when BEEF–vdW is applied to systems not included in the training data sets. The temperature is therefore calculated as

$$\tau = 2\frac{C(\mathbf{c})}{M_{\text{eff}}} \cdot \frac{N_{\text{tot}}}{N_{\text{tot}} - M_{\text{eff}}}, \tag{22}$$

where N_{tot} is the total number of systems in the training sets. The second term is close to unity since $N_{\text{tot}} \gg M_{\text{eff}}$. An ensemble matrix is now found as

$$\mathbf{\Omega}^{-1} = \tau\, \mathbf{H}^{-1}, \tag{23}$$

with eigenvalues $\mathbf{w}^2_{\Omega^{-1}}$ and eigenvectors $\mathbf{V}_{\Omega^{-1}}$.

Finally, using an ensemble of k vectors $\mathbf{v}_k$, each of length M with elements randomly drawn from a normal distribution of zero mean and variance one, the BEEF–vdW ensemble coefficient vectors $\mathbf{a}_k$ are calculated from

$$\mathbf{a}_k = \mathbf{V}_{\Omega^{-1}} \cdot \mathbf{1}\mathbf{w}_{\Omega^{-1}} \cdot \mathbf{v}_k. \tag{24}$$

The BEEF–vdW ensemble matrix is provided in the Supplemental Material.[83]

An illustration of the BEEF–vdW ensemble is shown in Fig. 5. For each data point in each data set, this ensemble may be applied non-self-consistently to BEEF–vdW electron densities. The standard deviation of the ensemble predictions of a quantity is then the ensemble estimate of the BEEF–vdW standard deviation on that quantity. The exchange enhancement ensemble expands after $s \approx 2$, where most of the chemistry and solid state physics have already happened.[76,78]

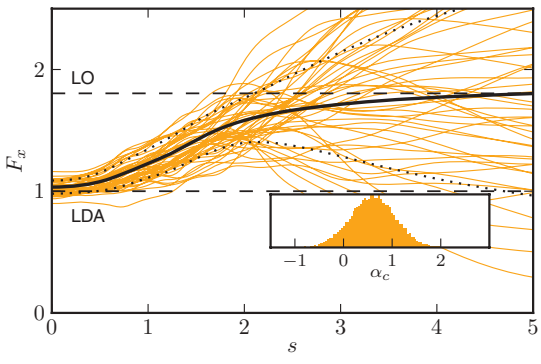

FIG. 5. (Color online) Bayesian ensemble of XC functionals around BEEF–vdW. Main panel: Black solid line is the BEEF–vdW exchange enhancement factor, while the orange lines depict $F_x(s)$ for 50 samples of the randomly generated ensemble. Dashed black lines mark the exchange model perturbations that yield DFT results ± 1 standard deviation away from BEEF–vdW results. The inset shows a histogram of the distribution of correlation parameters in an ensemble containing 20 000 samples. The distribution is centered around $\alpha_c = 0.6$.

TABLE IV. Comparison of self-consistent BEEF–vdW standard deviations to those predicted by the ensemble of functionals around BEEF–vdW. All energies in meV.

	BEEF–vdW	Ensemble estimate
CE17	143	209
RE42	372	253
DBH24	331	144
G2/97	242	312
SolEc34	576	436
s22x5-0.9	171	197
s22x5-1.0	94	181
s22x5-1.2	36	137
s22x5-1.5	8	67
s22x5-2.0	5	18

The predictive performance of the ensemble has been evaluated using 20 000 ensemble functionals. In practice, however, a few thousand ensemble functionals suffice for well-converged error estimates at a negligible computational overhead. Estimated standard deviations on the training data sets are compared to those from self-consistent calculations in Table IV. The ensemble performance on the data-set level should be assessed in combination with observing the error predictions on a system-to-system basis. Figure 6 illustrates the BEEF–vdW ensemble error estimates for the RE42 molecular reaction energies, and compares BEEF–vdW results to those of other functionals. Similar figures for more data sets are found in the Supplemental Material.[83]

On the data-set level, the overall predictive performance of the ensemble is satisfactory. The ensemble standard deviations in Table IV are slightly overestimated for the G2/97, CE17, and S22x5-0.9 data sets, while the ensemble underestimates the errors for RE42, DBH24/08, and Sol34Ec. For the remaining S22x5 subsets, the error estimates are too large.

Importantly, Fig. 6 illustrates strengths and weaknesses of the present approach to error estimation. Many of the reaction energies are accurately reproduced by BEEF–vdW, and the ensemble estimates a relatively small error on those data. However, some of the reactions for which BEEF–vdW yields larger errors are assigned too small error bars. The water-gas shift reaction $CO + H_2O \rightarrow CO_2 + H_2$ is one of these. The reason for this is indicated by the fact that all tested GGA, MGGA, and vdW–DF–type functionals yield nearly identical reaction energies for this reaction. One simply has to move rather far in XC model space to find a functional that predicts a reaction energy significantly different from the BEEF–vdW result. This causes the ensemble to underestimate the actual error for that reaction. Since the hybrid functionals appear to break the overall trends observed for the lower-rung functionals in Fig. 6, inclusion of exact exchange in the model space might remedy such limitations of the BEEF–vdW functional and its Bayesian ensemble.

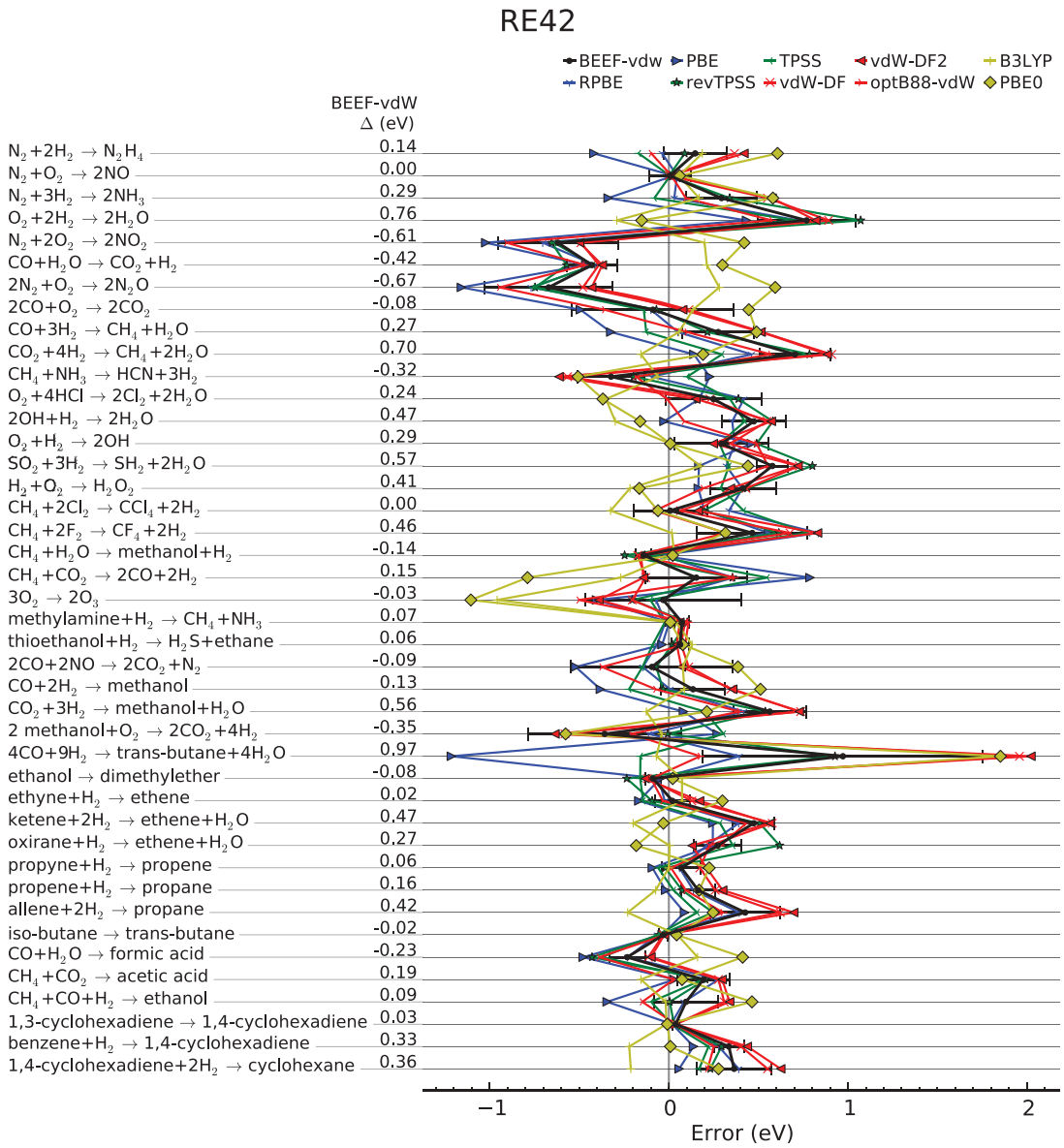

FIG. 6. (Color online) Deviations $\Delta = \Delta_r E^{\mathrm{DFT}} - \Delta_r E^{\mathrm{exp}}$ between the RE42 molecular reaction energies calculated using representative XC functionals and experiment. Color codes are BEEF–vdW: black; GGA: blue; MGGA: green; vdW–DF type: red; and hybrid: yellow. BEEF–vdW ensemble error estimates are indicated by horizontal error bars. The numbers in the middle column are self-consistent BEEF–vdW deviations from experiment.

VII. BENCHMARKS

The following is a comparative assessment of BEEF–vdW and a selection of literature XC functionals of the LDA, GGA, MGGA, vdW–DF, and hybrid types. These are listed in Table V. The benchmark data sets used are the six sets to which BEEF–vdW was trained, except Sol34Ec, as well as the G3-3, CE27, Sol27Ec, and Sol27LC data sets. The latter sets were introduced in Sec. II. Statistics on deviations of computed quantities from experimental or high-level theoretical references are reported for each density functional in terms of the mean signed (MSD), mean absolute (MAD), and standard deviation (STD). The sign

TABLE V. A selection of density functionals at the LDA (1), GGA (2), MGGA[a] (3), vdW–DF (3.5), and hybrid[b] (4) rungs of Jacob's ladder.

	Type	Target[c]	Ref.
LDA	1		68
PBE	2	General	67
RPBE	2	Chemistry	76
BLYP	2	Chemistry	87, 88
HCTH407	2	Chemistry	46
PBEsol	2	Solid state	43
WC	2	Solid state	89
AM05	2	Solid state	90
TPSS	3	General	91
revTPSS	3	General	44
vdW-DF	3.5	vdW	25
vdW-DF2	3.5	vdW	30
optPBE-vdW	3.5	vdW	33
optB88-vdW	3.5	vdW	33
C09-vdW	3.5	vdW	34
B3LYP	4	Chemistry	92
PBE0	4	Chemistry	93

[a]Attempts to apply the M06-L (Ref. 47) MGGA were unsuccessful due to convergence issues for a wide range of systems from almost all considered data sets. Note that problematics of evaluating MGGA potentials, especially for the M06 family of functionals, are discussed in recent literature (Refs. 84–86).

[b]Hybrid functionals have not been applied to extended systems.

[c]Should be understood as a very general characterization of the main target of a functional, and does not consider underlying principles of design.

convention is

$$\text{deviation} = \text{DFT} - \text{reference}. \tag{25}$$

Computed deviations for all systems in all data sets considered are tabulated in the Supplemental Material,[83] which also provides the raw DFT data.

All data are furthermore available online in the Computational Materials Repository (CMR).[95] The repository contains all information about the individual DFT calculations which form the basis for the results presented here, including atomic configurations and GPAW specific parameters. Access to search, browse, and download these data is provided through the CMR web interface.[96]

A. Molecular formation energies

The G2/97 and G3/99 thermochemical test sets have become standards for validating density functional methods, and the present calculations are well in line with published benchmark data[94] for these sets. Statistics are reported in Table VI. Considering first G2/97, the LDA grossly overestimates the molecular formation energies. Significant improvements are found with GGAs, where XC functionals designed to capture molecular energetics (RPBE, BLYP, HCTH407) yield STDs below 0.5 eV, while those targeted at solid state properties (PBEsol, WC, AM05) perform significantly worse: their MSDs are large and negative, indicating severe overbinding. The TPSS and revTPSS MGGA approximations perform quite well on this set.

Turning to the vdW–DF variants, good description of the G2/97 formation energies is also found for vdW–DF and vdW–DF2. This, however, is not the case for the optPBE–vdW, optB88–vdW, and C09–vdW functionals, for which the GGA exchange components are optimized with vdW dominated energetics in mind. This approach apparently leads to intramolecular overbinding, as previously noted in Ref. 31.

For comparison, Table VI also includes statistics for the B3LYP and PBE0 hybrids. As the wide application of hybrid XC functionals in the quantum chemistry community suggests, B3LYP and PBE0 accurately describe molecular bond energetics, and the B3LYP parametrization is found to be the best DFA for the G2/97 data set. Table VI furthermore shows that also the BEEF–vdW functional performs very well in predicting molecular formation energies. With a MAD of 0.16 eV, BEEF–vdW is highly accurate on the G2/97 thermochemical set, and even outperforms the PBE0 hybrid on these systems.

Now, let us switch attention to the G3-3 set of 75 molecules, which the BEEF–vdW model was not trained on. For most XC functionals tested here, the average deviations on G3-3 are larger than on G2/97. It is, however, noteworthy that TPSS, revTPSS, vdW–DF, and vdW–DF2 are exceptions to this trend. Benchmarking BEEF–vdW on G3-3 validates its good performance in predicting molecular bond energetics. This conclusion is underlined by the BEEF–vdW deviation statistics on the full G3/99 compilation. With a MAD of 0.19 eV, it is the most accurate DFA tested on G3/99, closely followed by B3LYP. Both MGGA functionals as well as vdW–DF and vdW–DF2 also perform well on this set.

B. Molecular reaction energies

The last column of Table VI summarizes deviation statistics for the RE42 data set. Even though the reaction energies are derived from the G2/97 formation energies, the reaction energies appear difficult to capture accurately with GGA, MGGA, and vdW–DF type functionals. None of them yield a STD less than 0.3 eV. The B3LYP hybrid proves significantly more accurate in this respect. Interestingly, the optPBE–vdW and optB88–vdW functionals, which both severely overestimate the G2/97 formation energies, prove as reliable for calculating gas-phase reaction energies as the best GGA (RPBE), and compare well to TPSS and BEEF–vdW.

C. Chemisorption on solid surfaces

Deviation statistics for the CE17 and CE27 data sets are reported in the first two columns of Table VII. The BEEF–vdW model was trained on CE17, while CE27 contains 10 extra entries, mostly covering dissociative H_2 chemisorption on late transition-metal surfaces. With MADs $\geqslant 0.7$ eV, LDA and the GGAs designed for solid state applications are clearly overbinding simple adsorbates to solid surfaces (negative MSDs). The RPBE, BLYP, and HCTH407 functionals are significantly more reliable for calculation of chemisorption energies, RPBE performing best with a MAD of 0.11 eV for both CE17 and CE27. Also, vdW–DF and vdW–DF2 yield

TABLE VI. Deviation statistics on the G2/97, G3-3, and G3/99 thermochemical data sets, as well as the RE42 set of molecular reaction energies. All energies in eV.

	G2/97 (148)			G3-3 (75)			G3/99 (223)			RE42 (42)		
Method	MSD	MAD	STD	MSD	MAD	STD	MSD	MAD	STD	MSD	MAD	STD
LDA	−3.69	3.69	4.27	−8.35	8.35	8.78	−5.25	5.25	6.16	−0.55	1.06	1.62
PBE	−0.64	0.68	0.84	−1.32	1.32	1.48	−0.87	0.90	1.10	−0.08	0.30	0.42
RPBE	0.25	0.40	0.51	0.94	0.96	1.13	0.48	0.59	0.78	0.11	0.26	0.34
PBEsol	−1.69	1.70	2.00	−3.94	3.94	4.14	−2.45	2.45	2.90	−0.29	0.48	0.73
BLYP	0.00	0.32	0.43	0.57	0.62	0.76	0.19	0.42	0.56	0.16	0.29	0.37
AM05	−1.77	1.78	2.07	−4.00	4.00	4.19	−2.52	2.52	2.96	−0.21	0.41	0.62
WC	−1.24	1.26	1.51	−2.86	2.86	3.03	−1.79	1.80	2.14	−0.24	0.43	0.65
HCTH407	0.09	0.26	0.35	0.48	0.55	0.65	0.22	0.36	0.47	0.06	0.27	0.35
TPSS	−0.22	0.28	0.33	−0.26	0.29	0.33	−0.24	0.28	0.33	0.06	0.25	0.32
revTPSS	−0.21	0.28	0.34	−0.24	0.26	0.31	−0.22	0.27	0.33	0.16	0.33	0.43
vdW–DF	−0.10	0.24	0.33	0.18	0.24	0.32	−0.01	0.24	0.33	0.24	0.39	0.52
vdW–DF2	−0.15	0.28	0.39	0.11	0.26	0.36	−0.06	0.28	0.38	0.24	0.40	0.54
optPBE–vdW	−0.84	0.85	0.98	−1.72	1.72	1.82	−1.14	1.14	1.32	0.06	0.27	0.35
optB88–vdW	−1.04	1.04	1.20	−2.22	2.22	2.34	−1.44	1.44	1.68	0.02	0.26	0.34
C09–vdW	−1.55	1.55	1.80	−3.55	3.55	3.72	−2.22	2.22	2.61	−0.11	0.33	0.45
B3LYP[a]	0.05	0.14	0.19	0.36	0.37	0.41	0.15	0.21	0.28	−0.05	0.15	0.22
PBE0[a]	−0.10	0.21	0.28	−0.40	0.44	0.55	−0.20	0.29	0.39	0.13	0.33	0.47
BEEF-vdW	−0.02	0.16	0.24	0.19	0.25	0.31	0.05	0.19	0.27	0.14	0.29	0.37

[a]B3LYP and PBE0 data adapted from Ref. 94.

MADs of 0.20 eV of less on CE27, while the two MGGAs overbind on average. Again, a significant overbinding is found for the three exchange-modified vdW–DF flavors. Lastly, it is seen from the CE17 column in Table VII that BEEF–vdW is among the DFAs offering most accurate predictions of chemisorption energies of simple adsorbates on solid surfaces. Since much of this accuracy is retained when moving to CE27, good transferability is expected when applying BEEF–vdW to other types of surface processes involving rupture and formation of chemical bonds.

D. Molecular reaction barriers

The DBH24/08 reaction barrier heights belong to a class of systems for which a fraction of exact exchange is known to increase computational accuracy significantly over GGAs.[22,97] This is supported by the DBH24/08 data in Table VII, where the two hybrids clearly outperform the lower-rung XC functionals. Considering the corresponding statistics for BEEF–vdW as well as for the individual DBH24/08 XC model reported in Table I, where a MAD of 0.12 eV was obtained, it is

TABLE VII. Deviation statistics on the CE17 and CE27 chemisorption energies, DBH24/08 reaction barriers, and the S22x5 interaction energies of noncovalently bonded complexes. All energies in eV, except S22x5, which is in meV.

	CE17 (17)			CE27 (27)			DBH24/08 (24)			S22x5 (110)		
Method	MSD	MAD	STD	MSD	MAD	STD	MSD	MAD	STD	MSD	MAD	STD
LDA	−1.34	1.34	1.39	−1.33	1.33	1.42	−0.58	0.58	0.73	−50	62	110
PBE	−0.42	0.42	0.44	−0.40	0.40	0.43	−0.33	0.33	0.43	76	76	132
RPBE	−0.02	0.11	0.13	0.00	0.11	0.14	−0.27	0.27	0.34	138	138	227
PBEsol	−0.85	0.85	0.87	−0.85	0.85	0.89	−0.44	0.44	0.56	38	53	85
BLYP	−0.04	0.13	0.16	0.02	0.15	0.18	−0.33	0.33	0.39	140	140	218
AM05	−0.70	0.70	0.73	−0.69	0.69	0.73	−0.41	0.41	0.53	99	99	157
WC	−0.76	0.76	0.78	−0.76	0.76	0.80	−0.41	0.41	0.52	56	63	105
HCTH407	0.11	0.17	0.22	0.15	0.20	0.30	−0.19	0.21	0.31	115	116	218
TPSS	−0.32	0.32	0.37	−0.34	0.34	0.41	−0.35	0.35	0.41	100	100	162
revTPSS	−0.38	0.38	0.43	−0.38	0.38	0.45	−0.35	0.35	0.41	92	92	141
vdW–DF	−0.05	0.12	0.14	0.04	0.18	0.22	−0.27	0.28	0.34	39	52	87
vdW–DF-2	−0.04	0.13	0.18	0.07	0.20	0.26	−0.30	0.31	0.37	31	33	61
optPBE–vdW	−0.39	0.39	0.42	−0.31	0.35	0.40	−0.33	0.33	0.41	−4	21	29
optB88–vdW	−0.52	0.52	0.56	−0.44	0.45	0.52	−0.37	0.37	0.45	3	10	15
C09–vdW	−0.78	0.78	0.81	−0.73	0.73	0.79	−0.41	0.41	0.50	−5	12	18
B3LYP							−0.17	0.17	0.21	111	111	180
PBE0							−0.13	0.15	0.19	71	71	124
BEEF–vdW	−0.08	0.12	0.14	−0.01	0.16	0.19	−0.26	0.26	0.33	42	50	88

clear that the BEEF–vdW model has moved significantly away from the part of model space favored by gas-phase reaction barrier heights. Nevertheless, BEEF–vdW is among the best nonhybrid functionals for such quantities.

E. Noncovalent interactions

The last column of Table VII lists deviation statistics for the S22x5 interaction energies. As previously found in several studies[30,33,53,59] of the original S22 data set, vdW dominated interactions are well described by vdW–DF type density functionals, especially those with an optimized exchange component. With MADs of 20 meV or less over all 110 points on the 22 potential-energy curves, the optPBE–vdW, optB88–vdW, and C09–vdW functionals prove highly accurate in this respect. The vdW–DF2 functional also captures vdW interactions well, but the positive MSD signifies that most of the deviations from the CCSD(T) reference energies stem from underbinding. For vdW–DF and BEEF–vdW, this is even more pronounced. None of the tested MGGA or hybrid DFAs convincingly capture vdW interactions. Only the most weakly gradient enhancing GGAs (PBEsol, WC, AM05) yield reasonable statistics. Taking into account the appreciable LDA overbinding of the S22x5 complexes, what appears to be GGA functionals capturing long-ranged dispersion is more likely a case of getting it right for the wrong reasons.

TABLE VIII. Detailed statistics on the deviations of calculated S22x5 interaction energies from CCSD(T) benchmarks using van der Waals density functionals in all five points along the intermolecular potential-energy curve. Mean signed and mean absolute deviations are in meV. Mean signed relative (MSRD) and mean absolute relative (MARD) deviations are also listed. Negatively signed deviation means overbinding on average.

Method	MSD	MAD	MSRD	MARD
$d = 0.9$				
vdW–DF	140	140	198%	198%
vdW–DF2	99	99	143%	143%
optPBE–vdW	29	31	28%	35%
optB88–vdW	17	19	26%	26%
C09–vdW	−13	21	−13%	35%
BEEF–vdW	136	137	214%	214%
$d = 1.0$				
vdW–DF	70	71	20%	25%
vdW–DF2	43	44	13%	15%
optPBE–vdW	−1	20	−9%	13%
optB88–vdW	5	13	3%	6%
C09–vdW	−3	13	1%	6%
BEEF–vdW	72	74	20%	28%
$d = 1.2$				
vdW–DF	4	32	−16%	23%
vdW–DF2	5	13	−2%	7%
optPBE–vdW	−25	28	−29%	30%
optB88–vdW	−4	13	−6%	9%
C09–vdW	−3	13	−8%	11%
BEEF–vdW	6	27	−12%	18%
$d = 1.5$				
vdW–DF	−13	15	−39%	40%
vdW–DF2	2	4	4%	6%
optPBE–vdW	−20	20	−44%	44%
optB88–vdW	−3	6	−12%	13%
C09–vdW	−6	11	−26%	28%
BEEF–vdW	−5	6	−13%	14%
$d = 2.0$				
vdW–DF	−4	4	−20%	20%
vdW–DF2	5	5	34%	34%
optPBE–vdW	−5	5	−20%	21%
optB88–vdW	1	2	3%	8%
C09–vdW	−2	2	−13%	15%
BEEF–vdW	2	3	27%	28%

For completeness, Table VIII shows detailed S22x5 statistics for vdW–DF variants and BEEF–vdW. Although performing reasonably well on S22x5 as a whole, the vdW–DF, vdW–DF2, and BEEF–vdW functionals underestimate the intermolecular binding energies at shortened binding distances $d = 0.9$. Also, at $d = 1.0$ the exchange-modified vdW–DF flavors offer a better description, but the difference between the two groups is much reduced. Concerning computational accuracy, the vdW–DF2 MSD of 43 meV and MAD of 44 meV for S22x5-1.0 obtained here compare very well to the MSD and MAD of 40 and 41 meV, respectively, found in a recent study[59] for a revised S22 data set.

F. Solid state properties

Table IX reports a summary of deviation statistics for calculations of lattice constants (Sol27LC) and cohesive energies (Sol27Ec). The lattice constant statistics are in clear favor of the PBEsol, AM05, WC, and revTPSS functionals. Their standard deviations are small and the MSDs are close to 0 Å. On average, however, these remarkably accurate predictions of equilibrium crystal volumes come at the price of overestimated cohesive energies.

The picture is opposite for vdW–DF and vdW–DF2. Lattice constants are overestimated and more so than with any other XC functional tested, vdW–DF2 yielding a standard deviation of 0.18 Å. Furthermore, those two DFAs notably underestimate cohesive energies. The less repulsive exchange functionals of the modified vdW–DF variants lead in general to statistics similar to those of PBE and TPSS for the two materials properties in question. These findings closely match those reported in recent studies[32,78,98–100] assessing the performance of GGA, MGGA, and vdW–DF type XC functionals for solid state properties.

Benchmarking finally BEEF–vdW, we find in Table IX that it performs reasonably well for cohesive energies and lattice constants, though still predicting softer crystal lattices than the optimized vdW–DF variants. With BEEF–vdW, these two bulk materials properties are, however, significantly closer to agreement with experiments than predictions by vdW–DF, vdW–DF2, and most of the GGAs designed mainly for chemistry.

VIII. APPLICATIONS

Two applications of BEEF–vdW to problems of current interest in the surface science community are here presented: graphene adsorption on the close-packed Ni(111) surface, and the trends observed when applying lower-rung density functionals in calculations of the binding energy of CO to

TABLE IX. Deviation statistics for the Sol27Ec cohesive energies (eV/atom) and Sol27LC lattice constants (Å). Zero-point vibrational effects have been removed from both experimental data sets.

	Sol27Ec (27)			Sol27LC (27)		
Method	MSD	MAD	STD	MSD	MAD	STD
LDA	0.89	0.89	1.08	−0.07	0.07	0.10
PBE	−0.10	0.27	0.38	0.05	0.06	0.07
RPBE	−0.54	0.58	0.71	0.11	0.11	0.13
PBEsol	0.43	0.45	0.63	−0.01	0.03	0.04
BLYP	−0.79	0.80	0.89	0.11	0.11	0.14
AM05	0.25	0.36	0.51	0.01	0.03	0.04
WC	0.37	0.41	0.57	0.00	0.03	0.04
HCTH407	−0.59	0.67	0.82	0.08	0.10	0.14
TPSS	0.08	0.27	0.36	0.05	0.05	0.08
revTPSS	0.31	0.37	0.50	0.03	0.04	0.07
vdW–DF	−0.54	0.60	0.72	0.12	0.12	0.14
vdW–DF2	−0.58	0.64	0.75	0.12	0.14	0.18
optPBE–vdW	−0.12	0.27	0.38	0.06	0.08	0.10
optB88–vdW	0.01	0.25	0.36	0.04	0.08	0.09
C09–vdW	0.42	0.43	0.59	0.01	0.05	0.06
BEEF–vdW	−0.37	0.45	0.59	0.08	0.08	0.11

Pt(111) and Rh(111) substrates as well as the surface energy of those substrates.

A. Graphene adsorption on Ni(111)

The remarkable electronic properties of monolayer graphene[103–105] and its potential application in electronics technology[104,106] motivate investigation of the interactions between graphene sheets and metallic surfaces. The nature of graphene adsorption on metals is highly metal dependent,[107,108] some surfaces binding graphene only weakly and others forming strong covalent bonds to the carbon sheet. The Ni(111) surface belongs to the latter group, graphene forming a (1×1) overlayer at a graphene-metal distance of $d = 2.1$ Å.[109] Furthermore, a band gap is induced in graphene upon adsorption, underlining the strong hybridization responsible for changing the electronic structure of the carbon sheet.[110,111]

Several theoretical studies have investigated the graphene/Ni(111) potential-energy curve, with mixed results.[112–118] However, based on RPA calculations, it is by now established that this particular adsorption process is a delicate competition between strong interactions close to the surface and vdW forces further from the surface.[101,102] Figure 7 shows calculated PECs for graphene adsorption on Ni(111) using LDA, MGGA, and vdW–DF type density functionals, as well as BEEF–vdW. Computational details are given in the Appendix. Additionally, two sets of RPA data are shown for comparison, indicating that graphene adsorption on Ni(111) is characterized by a physisorption minimum at $d = 3.0$–3.5 Å and a chemisorbed state at $d \approx 2.2$ Å, the latter in good agreement with experiments.[109] However, as previously found,[101,102,116,117] rung 1–3 DFAs, as well as vdW–DF and vdW–DF2, fail to simultaneously describe both qualitative features. Conversely, the optPBE–vdW and optB88–vdW PECs are increasingly closer to RPA data. The BEEF–vdW PEC shows qualitatively similar features, but the

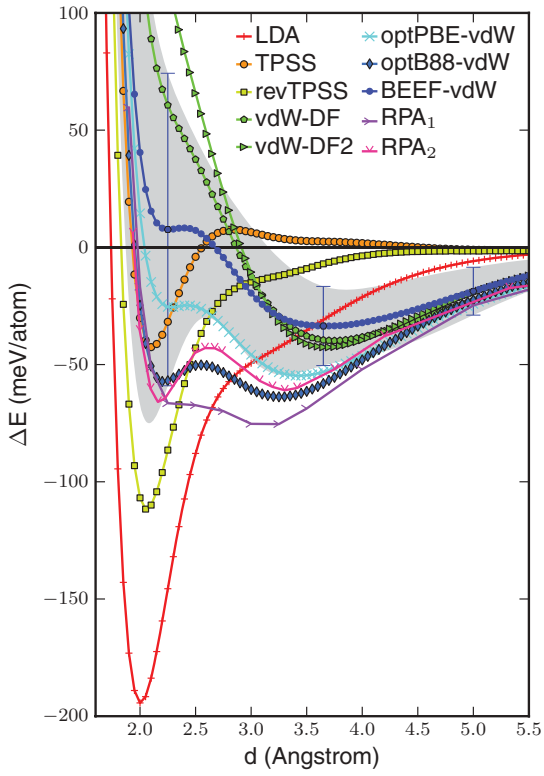

FIG. 7. (Color online) Potential-energy curves for graphene adsorption on the Ni(111) surface. Random phase approximation data are from Refs. 101 (RPA_1) and 102 (RPA_2). The gray area indicates the region spanned by the estimated standard deviations along the BEEF–vdW PEC.

local minimum at $d = 2.25$ Å is very shallow and yields a positive adsorption energy.

Figure 7 also shows ensemble error estimates along the BEEF–vdW PEC. Especially two aspects of these are of interest. First of all, the error bars do not straddle the zero line for large graphene-metal distances, indicating that confidence in the presence of a physisorption minimum is high. Second, the error bars enlarge notably at smaller distances from $d = 2.6$ Å and inwards, reflecting that these BEEF–vdW data points are associated with a significantly larger uncertainty. Recalling how the ensemble error estimate is designed (Sec. VI), the error estimates indicate that the graphene/Ni(111) PEC is very sensitive to the choice of XC functional in the chemically interesting range. Put differently, the ensemble suggests that we should not trust the BEEF–vdW prediction of a positive PEC for $d < 2.7$ Å as a definite result, as the estimated errors are simply too large in this region of the PEC.

B. Surface chemistry and stability

Chemisorption energies of molecules on surfaces are obviously important quantities in heterogeneous catalysis and surface science. However, accurate computation of surface energies E_γ can be critical as well since minimization of surface energy is a driving force determining the morphology and composition of surfaces, interfaces, and nanoparticles.[123] GGA density functionals, however, often underestimate E_γ, and the GGAs yielding most accurate surface energies also vastly overbind molecules to surfaces.[119] It thus appears that accurate computation of chemisorption energies on a surface as well as the stability of that surface is not possible with the same GGA approximation, underscoring a fundamental incompleteness of the GGA XC model space.

The issue is here investigated for vdW–DF variants and BEEF–vdW. Figure 8 shows atop chemisorption energies of CO on Pt(111) and Rh(111) against surface energies of those substrates, calculated using GGA, MGGA and vdW–DF type functionals, and BEEF–vdW with error estimation. These are compared to RPA results and experimental data. As previously reported,[119,124] the GGA data points fall along an approximately straight line, which is significantly offset from the experimental data, thus illustrating the issue discussed above. This is here shown to be the case for vdW–DF variants also: The dashed vdW–DF lines are parallel to the solid GGA lines, and are only slightly offset from the latter, especially for Rh(111). The vdW–DF and vdW–DF2 data points are quite close to RPBE. Larger surface energies are found with the exchange-modified vdW–DF variants, albeit at the expense of overestimated chemisorption energies. Note that such a correlation should be expected from Tables VII and IX and a linear relation between E_γ and the solid cohesive energy.[123]

Although BEEF–vdW contains the vdW–DF2 nonlocal correlation functional as an essential component, the former predicts larger surface energies than the latter without sacrificing accuracy of the CO-metal binding energy. We expect that this ability of BEEF–vdW to "break" the vdW–DF line is due to the expanded GGA model space as compared to vdW–DF, the latter of which pairs nonlocal correlation with LDA correlation. Significant inclusion of semilocal correlation in vdW–DF type calculations was also found in Ref. 31 to

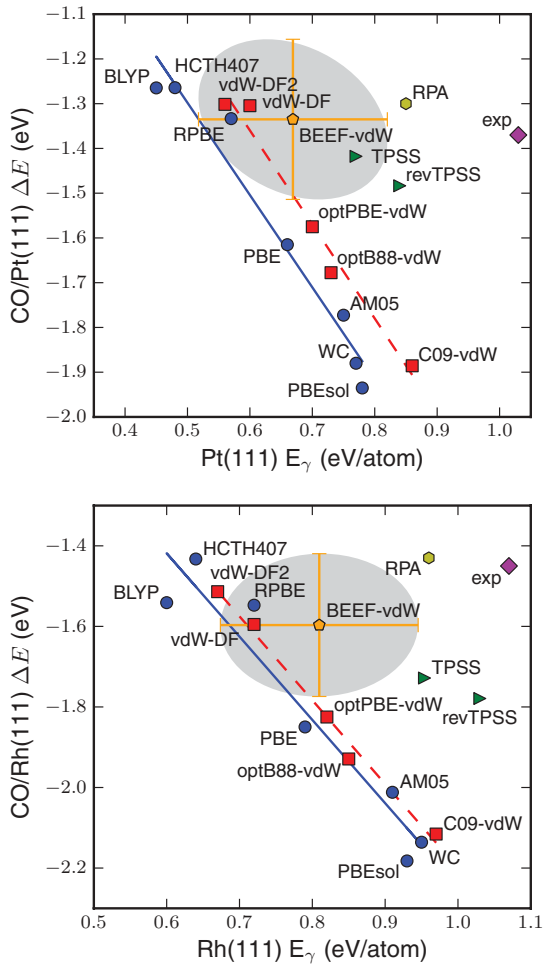

FIG. 8. (Color online) Atop CO chemisorption energies ΔE versus surface energies E_γ for Pt(111) and Ru(111). Red and blue lines are linear fits to GGA and vdW–DF type data points, respectively. MGGA data in green and yellow RPA data adapted from Ref. 119. Estimated standard deviations are indicated by error bars around the orange BEEF–vdW data points. All points $(E_\gamma, \Delta E)$ inside the gray areas are within one standard deviation from the BEEF–vdW point for both quantities. Experimental surface energies from liquid-metal data (Refs. 120 and 121), and experimental CO chemisorption energies from Ref. 122.

broadly improve accuracy for several materials properties. The BEEF–vdW error estimates furthermore appear very reasonable. The experimental CO chemisorption energies are straddled for both Pt(111) and Rh(111), and the error estimates along E_γ almost fill out the gap between the GGA lines to the left and the RPA and C09–vdW surface energies to the right. Lastly, it is seen from the green TPSS and revTPSS data points in Fig. 8, as also reported in Ref. 124, that the third rung of Jacob's ladder may offer the possibility of quite

accurate surface energies with only moderately overbound surface adsorbates.

IX. DISCUSSION

The presented approach to semiempirical DFA development fundamentally considers XC functionals as more or less accurate models of the exact density functional. From this point of view, the XC model space expansion and model selection procedure are essential, as are data sets for calibrating or benchmarking XC models. The concept of an ensemble of model solutions is intrinsic to the present model selection procedure. The cost function for a single data set has both weak (sloppy) and strong (important) eigenmodes in a sufficiently flexible model space. Regularization is used to suppress the weak modes in order to facilitate a physically sensible model and maximize transferability. The regularized ensemble thus contracts around the strong modes, and the optimum model can, to some extent, be regarded an average of the ensemble solutions. Without Tikhonov regularization of exchange, all XC approximations obtained in this work would have 31 parameters and wildly oscillating GGA exchange solutions, corresponding to a least-squares fit of an order-30 polynomial in the reduced density gradient. Instead, well-behaved models with 3–8 effective parameters are obtained.

It is important to note that model selection is intricately connected to the model space. The reduced density gradient transformation $t(s)$ defines the expansion of GGA exchange. It thereby also determines how hard the regularization punishes nonsmoothness in different regions of s space, as well as how the exchange part of the prior solution transforms to s space. As previously stated, the prior is the origo for the XC model solution. Many different priors may be chosen, but we find it convenient that it transforms to a reasonable exchange approximation. Then, decreasing regularization from infinity towards zero leads to increasingly nonsmooth variations away from this initial guess.

The linear combination correlation model space of local, semilocal, and nonlocal correlation was anticipated[31] to enable highly accurate calculations for several, if not all, of the data sets considered. The individually trained models in Table I confirm this, some sets favoring full LDA correlation in addition to nonlocal ditto, other sets preferring full PBE correlation, while most sets are fitted best by a combination of both. The corresponding exchange functionals are also significantly different, so the sets of strong eigenmodes for the regularized cost functions are very materials property dependent. We argue here that explicitly considering transferability among different materials properties is important for producing a single DFA composed of the most important modes for the combined data sets, that is, the optimum model compromise must be found.

One approach to this task is minimizing a weighted sum of the individual cost functions. This is somewhat similar to weighted training functions used in least-squares-fitting procedures, but with the critically important addition of regularization. The summed cost function is elegantly minimized using the individual solutions only, but gives no information regarding how the weights should be chosen. Clearly, an XC model trade-off is inevitable, so the weights should be the ones yielding an optimum compromise. For just two data sets, a wide range of poor choices of weights can be made, and the complexity of this choice increases with the number of data sets. In line with the statistical approach taken in the bulk of this work, we believe that such choice should not be made based on experience or intuition alone. Rather, a systematic methodology for locating one or more points in XC model space, where a well-behaved and properly compromising solution resides, is desirable. The condition of minimizing the product of relative costs for each data set is a reasonable requirement for the model solution, philosophically as well as in practice: The condition essentially states that if changing the solution vector $\mathbf{a}$ to $\mathbf{a} + \delta\mathbf{a}$ gains a larger relative reduction in cost on one materials property than is lost in total on all other properties considered, then $\mathbf{a} + \delta\mathbf{a}$ is preferred.

Extensive benchmarking of BEEF–vdW against popular GGA, MGGA, vdW–DF type, and hybrid XC functionals shows that the developed methodology is able to produce truly general-purpose XC approximations. Results are summarized in Fig. 9, where error statistics for representative functionals on gas-phase chemical, surface chemical, solid state, and vdW dominated data sets are illustrated by bars. The BEEF–vdW model compromise is indeed a very agreeable one. For none of the data sets is the average BEEF–vdW error among the largest, while several other functionals are highly biased towards certain types of materials properties. This is especially true for vdW–DF2 and optB88–vdW, displaying severely erroneous description of binding energetics for bulk solids and molecules, respectively. Furthermore, the figure shows an overall performance equivalence of BEEF–vdW and the original vdW–DF for gas-phase and surface chemical properties, although the former more accurately predicts bonding in the solid state. Further testing of the functional might, however, prove interesting. Systems such as ionic solids, semiconductors, and transition-metal complexes are not included in the present benchmark, nor are the BEEF–vdW predictions of molecular ionization potentials and electron affinities tested. This will be addressed in future work.

We emphasize the strengths and weaknesses of the BEEF–vdW ensemble error estimate. The ensemble functionals are based on a probability distribution for the model parameters, which limits the ensemble to the BEEF–vdW model space only. This space is incomplete in the sense that it can not accommodate a physically reasonable XC model yielding zero error on all systems in all data sets considered, hence the model trade-off. The BEEF–vdW computational errors are in general reasonably well estimated, but the energetics of certain systems is rather insensitive to the choice of XC approximation within the GGA, MGGA, and vdW–DF type model spaces. This leads to relatively small error estimates for these systems, even though the actual computational error may be substantial.

Meanwhile, we find BEEF–vdW and the Bayesian ensemble highly useful in surface science related applications. The fact that BEEF–vdW appears to yield more accurate surface energies than GGA or vdW–DF type XC approximations of similar accuracy for adsorbate-surface bond strengths is very promising. The error estimate proves very useful in this case, even though the kinetic energy density of MGGA type functionals may be needed in the model space if the surface energy error bars are to span the experimental data. This again illustrates that the ensemble does not give information beyond

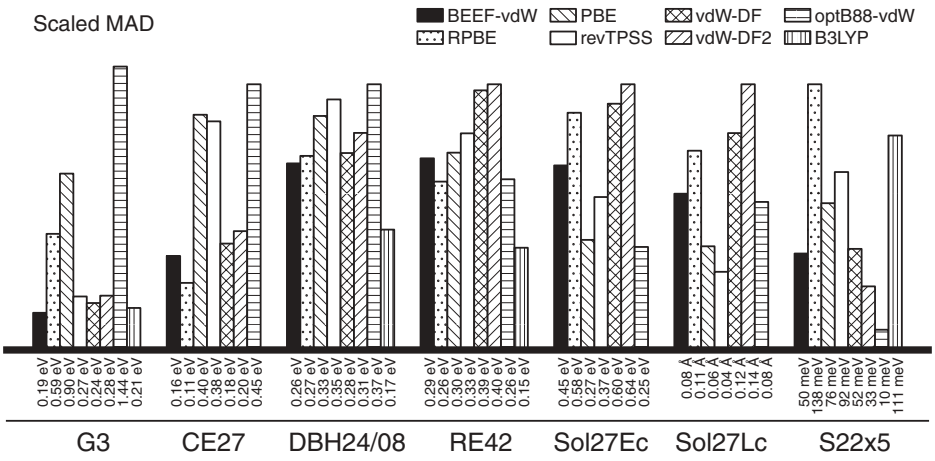

FIG. 9. Bar plot comparison of the accuracy of different density functionals in predicting various materials properties. For each data set, the bars illustrate proportionally scaled mean absolute deviations. The data sets are chosen to represent intramolecular bond energetics (G3), chemisorption energetics of molecules on surfaces (CE27), molecular reaction barrier heights (DBH24/08), molecular reaction energies (RE42), bulk solid cohesive energies (Sol27Ec) and lattice constants (Sol27LC), and interaction energies of noncovalently bonded complexes (S22x5). B3LYP calculations were not performed for bulk solids nor the extended CE27 systems.

its model space, as it is solely based on it. However, the error estimate carries important information in the BEEF–vdW study of graphene adsorption on Ni(111). The PEC is qualitatively wrong in the region of chemical bonding for this intricate case of "solid state adsorption," and the estimated errors indeed indicate that this part of the BEEF–vdW PEC is poorly determined. BEEF–vdW calculations can therefore not predict with any confidence whether graphene should form chemical bonds to the Ni(111) substrate in a low-temperature experiment. It is encouraging that the ensemble is able to capture this.

X. SUMMARY AND CONCLUSIONS

We have presented and evaluated a machine-learning-inspired approach to semiempirical density functional development. Focus has been on general applicability of the resulting density functional to both strong and weak interactions in chemistry and condensed matter physics, including surface chemistry. Transferability and avoiding overfitting are thus key issues, leading the presented methodology to rely primarily on (1) a variety of data sets chosen to represent vastly different interactions and bonding situations, (2) a very flexible XC model space expansion at a computationally feasible GGA + vdW level of approximation, and (3) XC model selection procedures designed to "tame" the flexible model space and yield XC approximations which properly compromise between describing different types of physics and chemistry.

To conclude, we have shown that regularization and cross-validation methods are very useful for semiempirical density functional development in highly flexible model spaces. It is furthermore clear that computationally efficient general-purpose functionals, targeted at accurately describing several physically and chemically different materials properties, necessarily must compromise between those properties in an incomplete XC model space. However, the optimum model trade-off is not easily found from simple intuition. A simple but powerful principle for determining the position in model space of a properly compromising XC approximation is therefore formulated.

Application of the developed methodology has yielded the BEEF–vdW density functional, and a benchmark of BEEF–vdW against popular GGA, MGGA, vdW–DF type, and hybrid XC functionals for energetics in chemistry and condensed matter physics has been conducted. This benchmark validates BEEF–vdW as a general-purpose XC approximation, with a reasonably reliable description of van der Waals forces and quantitatively accurate prediction of chemical adsorption energies of molecules on surfaces, while avoiding large sacrifices on solid state bond energetics. This should make it a valuable density functional for studies in surface science and catalysis.

Furthermore, an error estimation ensemble of functionals around BEEF–vdW comes out naturally of the developed fitting methodology. The ensemble is designed to provide an easily obtainable estimate of the XC approximation error. It is based on a probability distribution for the XC model parameters, and has been applied in the BEEF–vdW benchmark and qualitative assessments for molecular surface adsorption, surface energies, and graphene adsorption on Ni(111).

Finally, the methods developed here should lend themselves well to other XC model spaces also, including the MGGA level of theory or self-interaction correction schemes.

ACKNOWLEDGMENTS

The authors are grateful for the helpful suggestions and comments by N. Marzari, F. Abild-Pedersen, and J. Vande-Vondele. This work was supported by the Danish Center for

Scientific Computing. The Center for Atomic-Scale Materials Design is sponsored by the Lundbeck Foundation. The work at SUNCAT/SLAC has been supported by the US Department of Energy, Basic Energy Sciences.

APPENDIX: DETAILS OF DATA SETS AND COMPUTATIONS

1. G2/97 and G3/99

In accordance with the procedure of Ref. 49, the G3/99 formation enthalpies are corrected for thermal and vibrational contributions using thermal corrections and zero-point energies from Refs. 49 and 94. The G3/99 set is divided into three subsets denoted G3-1, G3-2, and G3-3 comprising 55, 93, and 75 molecules, respectively. The G3-1 and G3-2 subsets constitute G2/97. The G3-3 subset contains a significant fraction of larger carbon-rich molecules as compared to G2/97.

Theoretical G3/99 formation energies $\Delta_f E$ are calculated from the difference between molecular and atomic total energies as

$$\Delta_f E = E_M - \sum_A E_A, \tag{A1}$$

where A runs over all atoms in the molecule M, while E_M and E_A are ground-state molecular and atomic total energies at 0 K, respectively.

2. RE42

The 42 molecular reaction energies $\Delta_r E$ of the RE42 set are listed in Table X. Theoretical reaction energies are calculated from the total energies of G2/97 molecules after full geometry relaxation as

$$\Delta_r E = \sum_P E_P - \sum_R E_R, \tag{A2}$$

where the sums run over reactant (R) and product (P) molecules.

3. DBH24/08

Forward (V_f) and backward (V_b) benchmark reaction barriers from high-level theory or experiments are adapted from Ref. 50. Ground- and transition-state molecular geometries determined from quadratic configuration interaction calculations with single and double excitations (QCISD) are from Ref. 136. Density functional barrier heights are computed from the transition-state total electronic energy (E_{ts}) and the initial (E_i) and final (E_f) state total energies as

$$V_f = E_{ts} - E_i, \quad V_b = E_{ts} - E_f. \tag{A3}$$

4. S22x5

The original S22 publication[51] from 2006 reported CCSD(T) interaction energies of 22 noncovalently bonded complexes with extrapolation to the complete basis-set (CBS) limit. However, different basis sets were used for small and large complexes. Geometries were determined from MP2 or CCSD(T) calculations. Later works[61,137] have revised the S22 interaction energies, employing larger and identical basis sets for all complexes without changing the geometries. For the larger complexes, the reported basis-set effects are significant, so we use here the CCSD(T)/CBS energies of Takatani *et al.*[61] as the current best estimate of the true S22 interaction energies.

The S22x5 (Ref. 52) CCSD(T)/CBS potential-energy curves were reported more recently. The computational protocol was, however, not updated from that used for S22, so we expect the aforementioned interaction-energy inaccuracies to persist for S22x5. In order to shift the equilibrium point on each PEC to the revised S22 energies, and approximately correct the remaining data points, a modification of the (possibly) slightly inaccurate S22x5 CCSD(T) interaction energies is here

TABLE X. Gas-phase molecular reactions and reaction energies (in eV) constituting the RE42 data set. The experimental reaction energies are compiled from the G2/97 static-nuclei formation energies. $\Delta_r E < 0$ means exothermic.

Reaction	$\Delta_r E$
$N_2 + 2H_2 \rightarrow N_2H_4$	0.41
$N_2 + O_2 \rightarrow 2NO$	1.88
$N_2 + 3H_2 \rightarrow 2NH3$	−1.68
$O_2 + 2H_2 \rightarrow 2H_2O$	−5.45
$N_2 + 2O_2 \rightarrow 2NO_2$	0.62
$CO + H_2O \rightarrow CO_2 + H_2$	−0.31
$2N_2 + O_2 \rightarrow 2N_2O$	1.57
$2CO + O_2 \rightarrow 2CO_2$	−6.06
$CO + 3H_2 \rightarrow CH_4 + H_2O$	−2.80
$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$	−2.50
$CH_4 + NH_3 \rightarrow HCN + 3H_2$	3.32
$O_2 + 4HCl \rightarrow 2Cl_2 + 2H_2O$	−1.51
$2OH + H_2 \rightarrow 2H_2O$	−6.19
$O_2 + H_2 \rightarrow 2OH$	0.74
$SO_2 + 3H_2 \rightarrow SH_2 + 2H_2O$	−2.62
$H_2 + O_2 \rightarrow H_2O_2$	−1.68
$CH_4 + 2Cl_2 \rightarrow CCl_4 + 2H_2$	0.19
$CH_4 + 2F_2 \rightarrow CF_4 + 2H_2$	−8.60
$CH_4 + H_2O \rightarrow$ methanol $+ H_2$	1.33
$CH_4 + CO_2 \rightarrow 2CO + 2H_2$	3.11
$3O_2 \rightarrow 2O_3$	2.92
methylamine $+ H_2 \rightarrow CH_4 + NH_3$	−1.15
thioethanol $+ H_2 \rightarrow H_2S +$ ethane	−0.71
$2CO + 2NO \rightarrow 2CO_2 + N_2$	−7.94
$CO + 2H_2 \rightarrow$ methanol	−1.48
$CO_2 + 3H_2 \rightarrow$ methanol $+ H_2O$	−1.17
2 methanol $+ O_2 \rightarrow 2CO_2 + 4H_2$	−3.11
$4CO + 9H_2 \rightarrow$ trans-butane $+ 4H_2O$	−9.00
ethanol $\rightarrow$ dimethylether	0.53
ethyne $+ H_2 \rightarrow$ ethene	−2.10
ketene $+ 2H_2 \rightarrow$ ethene $+ H_2O$	−1.92
oxirane $+ H_2 \rightarrow$ ethene $+ H_2O$	−1.56
propyne $+ H_2 \rightarrow$ propene	−2.00
propene $+ H_2 \rightarrow$ propane	−1.58
allene $+ 2H_2 \rightarrow$ propane	−3.64
iso-butane $\rightarrow$ trans-butane	0.08
$CO + H_2O \rightarrow$ formic acid	−0.39
$CH_4 + CO_2 \rightarrow$ acetic acid	0.28
$CH_4 + CO + H_2 \rightarrow$ ethanol	−0.91
1,3-cyclohexadiene $\rightarrow$ 1,4-cyclohexadiene	−0.01
benzene $+ H_2 \rightarrow$ 1,4-cyclohexadiene	−0.01
1,4-cyclohexadiene $+ 2H_2 \rightarrow$ cyclohexane	−2.94

introduced as

$$E_{\rm int}^{d} := \varepsilon_{\rm int}^{d} \times \frac{E_{\rm int}^{1.0}}{\varepsilon_{\rm int}^{1.0}}, \tag{A4}$$

where $E_{\rm int}^{d}$ and $\varepsilon_{\rm int}^{d}$ denote modified and original S22x5 energies at the relative intermolecular distance d, respectively. For $E_{\rm int}^{1.0} = \varepsilon_{\rm int}^{1.0}$, Eq. (A4) obviously reduces to $E_{\rm int}^{d} = \varepsilon_{\rm int}^{d}$ for all distances. The obtained corrections to $\varepsilon_{\rm int}^{d}$ are listed in Table XI. The maximum correction of 11.4% amounts to 25.6 meV for the indole-benzene complex in a stacked geometry, while the total mean signed correction to all the 110 interaction energies is 0.1 meV.

The modified CCSD(T) interaction energies are used throughout this study for the S22x5 data set and subsets. Each density functional interaction energy $E_{\rm int}^{d}$ is calculated as the difference between the total electronic energy of the interacting complex E_0^d and those of its two isolated molecular constituents E_1^d and E_2^d:

$$E_{\rm int}^{d} = E_0^d - E_1^d - E_2^d. \tag{A5}$$

TABLE XI. Corrections $E_{\rm int}^{d} - \varepsilon_{\rm int}^{d}$ to the S22x5 interaction energies in Ref. 52 computed from Eq. (A4). Reported statistics are most negative (min), most positive (max), mean signed (msc), and mean absolute (mac) interaction energy correction at each distance. Furthermore, the total mean signed (MSC) and total mean absolute (MAC) energy corrections over all 110 energies are reported in the bottom rows. All energies in meV.

		Relative interaction distance d				
Complex	$E_{\rm int}^{1.0}/\varepsilon_{\rm int}^{1.0}$	0.9	1.0	1.2	1.5	2.0
1	−1.0%	−1.0	−1.3	−1.0	−0.5	−0.1
2	−1.0%	−1.9	−2.2	−1.8	−1.0	−0.4
3	−1.1%	−8.0	−9.1	−7.6	−4.5	−1.8
4	−1.1%	−6.5	−7.3	−6.1	−3.7	−1.6
5	−1.1%	−9.2	−10.0	−8.4	−5.1	−2.2
6	−1.8%	−11.8	−13.0	−10.8	−6.4	−2.5
7	−2.3%	−14.7	−16.0	−13.0	−7.3	−2.5
8	0.0%	0.0	0.0	0.0	0.0	0.0
9	−1.2%	−0.4	−0.8	−0.4	−0.1	0.0
10	3.2%	1.5	2.1	1.6	0.7	0.2
11	6.8%	0.4	8.3	5.7	1.6	0.2
12	6.9%	5.1	13.5	9.0	2.9	0.6
13	1.3%	3.8	5.6	3.6	1.4	0.4
14	11.4%	10.5	25.6	17.8	5.3	0.5
15	4.6%	15.9	24.3	16.4	6.5	1.8
16	−1.4%	−0.7	−0.9	−0.7	−0.3	−0.1
17	−0.6%	−0.8	−0.9	−0.7	−0.4	−0.1
18	1.3%	1.1	1.3	1.0	0.5	0.2
19	−0.7%	−1.2	−1.3	−1.1	−0.6	−0.2
20	3.2%	3.1	3.9	3.1	1.6	0.5
21	2.1%	4.5	5.2	4.4	2.5	1.0
22	−0.6%	−1.6	−1.8	−1.5	−0.9	−0.4
min	−2.3%	−14.7	−16.0	−13.0	−7.3	−2.5
max	11.4%	15.9	25.6	17.8	6.5	1.8
msc	1.2%	−0.5	1.1	0.4	−0.4	−0.3
mac	2.5%	4.7	7.0	5.3	2.4	0.8
MSC	0.1					
MAC	4.0					

TABLE XII. Experimental solid-state properties of 27 cubic bulk solids. The ZPAE exclusive Sol27LC 0-K lattice constants a_0 (Å) are adapted from Ref. 98. 0-K Sol27Ec cohesive energies E_c (eV/atom) from Ref. 125 are corrected for ZPVE contributions. Strukturbericht symbols are indicated in parentheses for each solid. A1: fcc, A2: bcc, A3: hcp, A4: diamond.

	Sol27LC	Sol27Ec	
Solid	a_0	E_c	ZPVE[a]
Li (A2)	3.451	1.66	0.033
Na (A2)	4.209	1.13	0.015
K (A2)	5.212	0.94	0.009
Rb (A2)	5.577	0.86	0.005
Ca (A1)	5.556	1.86	0.022
Sr (A1)	6.040	1.73	0.014
Ba (A2)	5.002	1.91	0.011
V (A2)	3.024	5.35	0.037
Nb (A2)	3.294	7.60	0.027
Ta (A2)	3.299	8.12	0.023
Mo (A2)	3.141	6.86	0.044
W (A2)	3.160	8.94	0.039
Fe (A2)	2.853	4.33	0.046
Rh (A1)	3.793	5.80	0.047
Ir (A1)	3.831	6.98	0.041
Ni (A1)	3.508	4.48	0.044
Pd (A1)	3.876	3.92	0.027
Pt (A1)	3.913	5.86	0.023
Cu (A1)	3.596	3.52	0.033
Ag (A1)	4.062	2.97	0.022
Au (A1)	4.062	3.83	0.016
Pb (A1)	4.912	2.04	0.010
Al (A1)	4.019	3.43	0.041
C (A4)	3.544	7.59	0.216
Si (A4)	5.415	4.69	0.063
Ge (A4)	5.639	3.89	0.036
Sn (A4)	6.474	3.16	0.019

[a]ZPVE corrections are calculated according to Eq. (A6) using Debye temperatures from Ref. 125.

Computational accuracy is enhanced by keeping all atoms in the molecular fragments in the same positions in the box as those atoms have when evaluating the total energy of the complex.

5. Sol27LC and Sol27Ec

It was recently shown[78] that removal of thermal and zero-point contributions to experimentally determined lattice constants and bulk moduli may be important when benchmarking density functional methods. Experimental zero Kelvin lattice constants and cohesive energies (E_c) contain zero-point vibrational contributions, leading to zero-point anharmonic expansion (ZPAE) of the lattice and zero-point vibrational energy (ZPVE) contributions to E_c. As discussed in Ref. 138, an estimate of the ZPVE may be obtained from the Debye temperature Θ_D of the solid according to

$$\text{ZPVE} = -\tfrac{9}{8} k_B \Theta_D. \tag{A6}$$

The vibrational contribution is subtracted from the cohesive energy, leading to increased stability of the crystal towards

atomization. The same reference derived a semiempirical estimate of the ZPAE contribution to the volume of cubic crystals. A recent study[18] calculating the ZPAE from first principles largely validates this approach. The Sol27LC and Sol27Ec sets of zero Kelvin lattice constants and cohesive energies of 27 fcc, bcc, and diamond structured bulk solids are appropriately corrected for zero-point phonon effects. Details are given in Table XII.

Density functional computation of total energies of the extended bulk solids is done using a 16×16×16 **k**-point mesh for sampling reciprocal space of the periodic lattice and 0.1 eV Fermi smearing of the electron occupation numbers. Calculations for bulk Fe, Ni, and Co are spin polarized. The cohesive energy for a given crystal lattice constant a is calculated from

$$E_c = E_A - E_B, \tag{A7}$$

where E_A is the total energy of the free atom and E_B is the bulk total energy per atom. By this definition, the equilibrium cohesive energy of a stable solid is a positive quantity. Equilibrium lattice constants of cubic crystals a_0 are determined from fitting the stabilized jellium equation of state (SJEOS, Ref. 138) to cohesive energies sampled in five points in a small interval around the maximum of the $E_c(a)$ curve.

6. CE17 and CE27

The CE17 and CE27 data are derived from temperature-programed desorption experiments or from microcalorimetry, most often at low coverage. The 27 chemisorption energies have been critically chosen from literature with emphasis on reliability as well as covering a reasonably wide range of substrates and adsorbates. All data are listed in Table XIII along with details regarding adsorption mode, adsorption site, and references.

Most of the CE27 surface reactions are molecular adsorption processes at 0.25 ML coverage. In that case, the chemisorption energy is computed according to

$$\Delta E = E_{AM} - E_M - xE_A, \tag{A8}$$

where E_{AM} is the total electronic energy of the adsorbate A on metal surface M, and E_A and E_M total energies of the isolated adsorbate and metal surface, respectively. The constant x equals 1 for molecular adsorption and N_2 dissociation on Fe(100), while $x = \frac{1}{2}$ for the dissociative H_2 chemisorption reactions. In the case of NO dissociation on Ni(100) at 0.25 ML coverage, the chemisorption energy is

$$\Delta E = E_{AM} + E_{BM} - 2E_M - E_{AB}, \tag{A9}$$

where AB is the NO molecule.

With these definitions of chemisorption energies, we consider extended surface slab models with 2×2 atoms in each layer and five layers in total. The slab models are periodic in the surface plane and a vacuum width of 20 Å separates periodically repeated slabs perpendicularly to the surface planes. Calculations involving Fe, Ni, and Co are spin polarized. Well-converged chemisorption energies are obtained using a 10 × 10 × 1 **k**-point mesh and a real-space grid spacing around 0.16 Å. The self-consistently determined lattice constant of the slab solid obviously determines the xy dimensions of the slab simulation cell. Since the number of real-space grid points employed in each direction is discrete, a grid spacing of exactly 0.16 Å in the x and y directions is rarely possible for slab calculations. Instead, it may be slightly smaller or larger, which should not affect the computational accuracy significantly. During structure relaxations, the bottom two layers of the 2 × 2 × 5 slab models are fixed in the bulk structure as found from bulk calculations.

TABLE XIII. The 27 experimental reaction energies ΔE for chemisorption on late transition-metal surfaces constituting the CE27 data set. The somewhat smaller CE17 data set is a subset of CE27. Reactions in CE17 are marked with a "⋆". All chemisorption energies are in eV per adsorbate at a surface coverage of 0.25 ML, except where otherwise noted. The adsorption mode is indicated by "m" (molecular) or "d" (dissociative), along with the adsorption site. Chemisorption energies for O have been evaluated as $\frac{1}{2}\{\Delta E(O_2) - E_b(O_2)\}$ with $E_b(O_2) = 118$ kcal/mol (Ref. 126) for the dioxygen bond energy.

		Mode	Site	ΔE	Reference(s)
CO/Ni(111)	⋆	m	fcc	−1.28	122
CO/Ni(100)		m	hollow	−1.26	127
CO/Rh(111)	⋆	m	top	−1.45	122
CO/Pd(111)	⋆	m	fcc	−1.48	122
CO/Pd(100)	⋆	m	bridge	−1.60	127–130
CO/Pt(111)	⋆	m	top	−1.37	122
CO/Ir(111)	⋆	m	top	−1.58	122
CO/Cu(111)	⋆	m	top	−0.50	122
CO/Co(0001)	⋆	m	top	−1.20	122
CO/Ru(0001)	⋆	m	top	−1.49	122
O/Ni(111)	⋆	m	fcc	−4.95	130
O/Ni(100)	⋆	m	hollow	−5.23	130
O/Rh(100)	⋆	m	hollow	−4.41	130
O/Pt(111)	⋆	m	fcc	−3.67	131
NO/Ni(100)	⋆	d	hollow	−3.99	127
NO/Pd(111)	⋆	m	fcc	−1.86	132
NO/Pd(100)	⋆	m	hollow	−1.61	133
NO/Pt(111)		m	fcc	−1.45	131
N_2/Fe(100)[b]		d	hollow	−2.3	134
H_2/Pt(111)	⋆	d	fcc	−0.41	135
H_2/Ni(111)		d	fcc	−0.98	135
H_2/Ni(100)		d	hollow	−0.93	135
H_2/Rh(111)		d	fcc	−0.81	135
H_2/Pd(111)		d	fcc	−0.91	135
H_2/Ir(111)		d	fcc	−0.55	135
H_2/Co(0001)		d	fcc	−0.69	135
H_2/Ru(0001)[c]		d	fcc	−1.04	135

[a] ΔE is the average of −1.58, −1.67, −1.69, and −1.45 eV.
[b] The coverage of atomic nitrogen is 0.5 ML.
[c] ΔE is the average of −0.83 and −1.24 eV, both from Ref. 135.

7. Graphene adsorption on Ni(111)

Adsorption of graphene on Ni(111) was modeled using a 1 × 1 × 5 surface slab, a Ni(fcc) lattice constant of 3.524 Å as determined with the PBE density functional, and 20 Å vacuum width. The top three atomic layers were fully relaxed with PBE using a grid spacing of 0.15 Å and a (20 × 20 × 1) **k**-point mesh. Carbon atoms were placed in atop and fcc adsorption sites, respectively.

*jesswe@fysik.dtu.dk

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[2] W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).
[3] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999).
[4] E. A. Carter, Science **321**, 800 (2008).
[5] B. Hammer and J. K. Nørskov, in *Impact of Surface Science on Catalysis*, Advances in Catalysis, Vol. 45, edited by B. C. Gates and H. Knözinger (Academic, New York, 2000), p. 71.
[6] M. J. Field, J. Comput. Chem. **23**, 48 (2002).
[7] J. Greeley, J. K. Nørskov, and M. Mavrikakis, Annu. Rev. Phys. Chem. **53**, 319 (2002).
[8] J. K. Nørskov, T. Bligaard, A. Logadottir, S. Bahn, L. B. Hansen, M. V. Bollinger, H. S. Bengaard, B. Hammer, Z. Sljivancanin, M. Mavrikakis, Y. Xu, S. Dahl, and C. J. H. Jacobsen, J. Catal. **209**, 275 (2002).
[9] H. Grönbeck, Top. Catal. **28**, 59 (2004).
[10] K. M. Neyman and F. Illas, Catal. Today **105**, 2 (2005).
[11] R. A. Friesner and V. Guallar, Annu. Rev. Phys. Chem. **56**, 389 (2005).
[12] J. K. Nørskov, M. Scheffler, and H. Toulhoat, MRS Bull. **31**, 669 (2006).
[13] H. M. Senn and W. Thiel, Angew. Chem. Int. Ed. **48**, 1198 (2009).
[14] M. Neurock, J. Catal. **216**, 73 (2003).
[15] J. Greeley and M. Mavrikakis, Nat. Mater. **3**, 810 (2004). .
[16] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, Nat. Chem. **1**, 37 (2009).
[17] J. K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard, Proc. Natl. Acad. Sci. USA **108**, 937 (2011).
[18] L. Schimka, J. Harl, and G. Kresse, J. Chem. Phys. **134**, 024116 (2011).
[19] C. D. Sherrill, J. Chem. Phys. **132**, 110902 (2010).
[20] O. A. Vydrov, G. E. Scuseria, and J. P. Perdew, J. Chem. Phys. **126**, 154109 (2007).
[21] I. Dabo, A. Ferretti, N. Poilvert, Y. Li, N. Marzari, and M. Cococcioni, Phys. Rev. B **82**, 115121 (2010).
[22] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Chem. Rev. **112**, 289 (2012).
[23] J. P. Perdew and K. Schmidt, in *Density Functional Theory and its Application to Materials*, AIP Conference Proceedings, Vol. 577, edited by V. Van Doren, C. Van Alsenoy, and P. Geerlings (AIP, New York, 2001), p. 1.
[24] J. P. Perdew, A. Ruzsinszky, J. M. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, J. Chem. Phys. **123**, 062201 (2005).
[25] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004).
[26] A. D. Becke and E. R. Johnson, J. Chem. Phys. **124**, 014104 (2006).
[27] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).
[28] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[29] D. C. Langreth, B. I. Lundqvist, S. D. Chakarova-Kack, V. R. Cooper, M. Dion, P. Hyldgaard, A. Kelkkanen, J. Kleis, L. Kong, S. Li, P. G. Moses, E. Murray, A. Puzder, H. Rydberg, E. Schröder, and T. Thonhauser, J. Phys.: Condens. Matter **21**, 084203 (2009).
[30] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Phys. Rev. B **82**, 081101 (2010).
[31] J. Wellendorff and T. Bligaard, Top. Catal. **54**, 1143 (2011).
[32] J. Klimes, D. R. Bowler, and A. Michaelides, Phys. Rev. B **83**, 195131 (2011).
[33] J. Klimes, D. R. Bowler, and A. Michaelides, J. Phys.: Condens. Matter **22**, 022201 (2010).
[34] V. R. Cooper, Phys. Rev. B **81**, 161104(R) (2010).
[35] E. D. Murray, D. C. Lee, and K. Langreth, J. Chem. Theory Comput. **5**, 2754 (2009).
[36] F. O. Kannemann and A. D. Becke, J. Chem. Theory Comput. **5**, 719 (2009).
[37] B. G. Janesko, J. Chem. Phys. **133**, 104103 (2010).
[38] B. G. Janesko and A. Aguero, J. Chem. Phys. **136**, 024111 (2012).
[39] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[40] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998).
[41] H. L. Schmider and A. D. Becke, J. Chem. Phys. **108**, 9624 (1998).
[42] T. Van Voorhis and G. E. Scuseria, J. Chem. Phys. **109**, 400 (1998).
[43] J. 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).
[44] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun, Phys. Rev. Lett. **103**, 026403 (2009).
[45] A. D. Becke, J. Chem. Phys. **107**, 8554 (1997).
[46] A. D. Boese and N. C. Handy, J. Chem. Phys. **114**, 5497 (2001).
[47] Y. Zhao and D. G. Truhlar, J. Chem. Phys. **125**, 194101 (2006).
[48] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **112**, 7374 (2000).
[49] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **106**, 1063 (1997).
[50] J. Zheng, Y. Zhao, and D. G. Truhlar, J. Chem. Theory Comput. **5**, 808 (2009).
[51] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[52] L. Grafova, M. Pitonak, J. Rezac, and P. Hobza, J. Chem. Theory Comput. **6**, 2365 (2010).
[53] A. Gulans, M. J. Puska, and R. M. Nieminen, Phys. Rev. B **79**, 201105(R) (2009).
[54] F. O. Kannemann and A. D. Becke, J. Chem. Theory Comput. **6**, 1081 (2010).
[55] L. A. Burns, A. Vazquez-Mayagoitia, B. G. Sumpter, and C. D. Sherrill, J. Chem. Phys. **134**, 084107 (2011).
[56] F. Hanke, J. Comput. Chem. **32**, 1424 (2011).
[57] L. Goerigk and S. Grimme, Phys. Chem. Chem. Phys. **13**, 6670 (2011).
[58] Y. Zhao and D. G. Truhlar, Theor. Chem. Account. **120**, 215 (2008).
[59] O. A. Vydrov and T. Van Voorhis, J. Chem. Phys. **133**, 244103 (2010).
[60] S. Grimme, S. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[61] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **132**, 144104 (2010).
[62] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).
[63] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dulak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen,

G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, J. Phys.: Condens. Matter **22**, 253202 (2010).

[64]The GPAW and ASE codes are available as parts of the CAMPOS software: [http://www.camd.dtu.dk/Software].

[65]P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[66]S. R. Bahn and K. W. Jacobsen, Comput. Sci. Eng. **4**, 56 (2002).

[67]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[68]J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[69]G. Román-Pérez and J. M. Soler, Phys. Rev. Lett. **103**, 096102 (2009).

[70]J. Wellendorff, A. Kelkkanen, J. J. Mortensen, B. I. Lundqvist, and T. Bligaard, Top. Catal. **53**, 378 (2010).

[71]E. H. Lieb and S. Oxford, Int. J. Quantum Chem. **19**, 427 (1981).

[72]J. P. Perdew, in *Electronic Structure of Solids '91*, Vol. 17, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991), p. 11.

[73]C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. (Springer, New York, 2006).

[74]K. S. Brown and J. P. Sethna, Phys. Rev. E **68**, 021904 (2003).

[75]T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York, 2009).

[76]B. Hammer, L. B. Hansen, and J. K. Nørskov, Phys. Rev. B **59**, 7413 (1999).

[77]V. Petzold, T. Bligaard, and K. W. Jacobsen, Top. Catal. (to be published).

[78]G. I. Csonka, J. P. Perdew, A. Ruzsinszky, P. H. T. Philipsen, S. Lebegue, J. Paier, O. A. Vydrov, and J. G. Angyan, Phys. Rev. B **79**, 155107 (2009).

[79]B. Efron, J. Am. Stat. Assoc. **78**, 316 (1983).

[80]If one or more of the individual cost functions have very strong modes, several stationary points of the product cost function can exist. In this case, one must carefully determine all stationary points, and select the one which represents the best compromise. For the subsequent fitting of the BEEF–vdW, this was a minor issue.

[81]J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen, J. K. Nørskov, J. P. Sethna, and K. W. Jacobsen, Phys. Rev. Lett. **95**, 216401 (2005).

[82]D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial* (Oxford University Press, Oxford, UK, 2006).

[83]See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.85.235149 for (1) figures comparing the performance of representative XC functionals at system level for all data sets considered in benchmarking, and including BEEF–vdW ensemble error estimates, (2) tables of those quantities, which are also used for the benchmark statistics, (3) comma-separated values (CSV) files containing the raw DFT data for benchmarking, and (4) the BEEF–vdW $\mathbf{\Omega}^{-1}$ ensemble matrix in CSV format. All the CSV files contain header information.

[84]E. R. Johnson, R. A. Wolkow, and G. A. DiLabio, Chem. Phys. Lett. **394**, 334 (2004).

[85]E. R. Johnson, A. D. Becke, C. D. Sherrill, and G. A. DiLabio, J. Chem. Phys. **131**, 034111 (2009).

[86]S. E. Wheeler and K. N. Houk, J. Chem. Theory Comput. **6**, 395 (2010).

[87]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[88]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[89]Z. Wu and R. E. Cohen, Phys. Rev. B **73**, 235116 (2006).

[90]R. Armiento and A. E. Mattsson, Phys. Rev. B **72**, 085108 (2005).

[91]J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).

[92]P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).

[93]C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[94]V. N. Staroverov, G. E. Scuseria, J. M. Tao, and J. P. Perdew, J. Chem. Phys. **119**, 12129 (2003).

[95]D. D. Landis, J. S. Hummelshøj, S. Nestorov, J. Greeley, M. Dułak, T. Bligaard, J. K. Nørskov, and K. Jacobsen, Comput. Sci. Eng. **99** (2012), doi:10.1109/MCSE.2012.16.

[96]The CMR web-interface is found at [http://cmr.fysik.dtu.dk].

[97]A. D. Boese and J. M. L. Martin, J. Chem. Phys. **121**, 3405 (2004).

[98]P. Haas, F. Tran, and P. Blaha, Phys. Rev. B **79**, 085104 (2009).

[99]J. 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).

[100]P. Hao, Y. Fang, J. Sun, G. I. Csonka, P. H. T. Philipsen, and J. P. Perdew, Phys. Rev. B **85**, 014111 (2012).

[101]T. Olsen, J. Yan, J. J. Mortensen, and K. S. Thygesen, Phys. Rev. Lett. **107**, 156401 (2011).

[102]F. Mittendorfer, A. Garhofer, J. Redinger, J. Klimes, J. Harl, and G. Kresse, Phys. Rev. B **84**, 201401(R) (2011).

[103]K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).

[104]A. K. Geim and K. S. Novoselov, Nat. Mater. **6**, 183 (2007).

[105]A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).

[106]F. Schwierz, Nat. Nanotechnol. **5**, 487 (2010).

[107]J. Wintterlin and M.-L. Bocquet, Surf. Sci. **603**, 1841 (2009).

[108]Y. S. Dedkov and M. Fonin, New J. Phys. **12**, 125004 (2010).

[109]Y. Gamo, A. Nagashima, M. Wakabayashi, M. Teria, and C. Oshima, Surf. Sci. **374**, 61 (1997).

[110]A. Grüneis and D. V. Vyalikh, Phys. Rev. B **77**, 193401 (2008).

[111]A. Varykhalov, J. Sánchez-Barriga, A. M. Shikin, C. Biswas, E. Vescovo, A. Rybkin, D. Marchenko, and O. Rader, Phys. Rev. Lett. **101**, 157601 (2008).

[112]G. Bertoni, L. Calmels, A. Altibelli, and V. Serin, Phys. Rev. B **71**, 075402 (2005).

[113]G. Kalibaeva, R. Vuilleumier, S. Meloni, A. Alavi, G. Ciccotti, and R. Rosei, J. Phys. Chem. B **110**, 3638 (2006).

[114]G. Giovannetti, P. A. Khomyakov, G. Brocks, V. M. Karpan, J. van den Brink, and P. J. Kelly, Phys. Rev. Lett. **101**, 026803 (2008).

[115]M. Fuentes-Cabrera, M. I. Baskes, A. V. Melechko, and M. L. Simpson, Phys. Rev. B **77**, 035405 (2008).

[116]M. Vanin, J. J. Mortensen, A. K. Kelkkanen, J. M. Garcia-Lastra, K. S. Thygesen, and K. W. Jacobsen, Phys. Rev. B **81**, 081408 (2010).

[117]I. Hamada and M. Otani, Phys. Rev. B **82**, 153412 (2010).

[118]C. Gong, G. Lee, B. Shan, E. M. Vogel, R. M. Wallace, and K. Cho, J. Appl. Phys. **108**, 123711 (2010).

[119]L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse, Nat. Mater. **9**, 741 (2010).

[120]W. R. Tyson and W. A. Miller, Surf. Sci. **62**, 267 (1977).

[121]L. Vitos, A. V. Ruban, H. L. Skriver, and J. Kollar, Surf. Sci. **411**, 186 (1998).

[122]F. Abild-Pedersen and M. P. Andersson, Surf. Sci. **601**, 1747 (2007).

[123] I. Chorkendorff and J. W. Niemantsverdriet, *Concepts of Modern Catalysis and Kinetics*, 2nd ed. (Wiley-VCH, Weinheim, 2007).
[124] J. Sun, M. Marsman, A. Ruzsinszky, G. Kresse, and J. P. Perdew, Phys. Rev. B **83**, 121410(R) (2011).
[125] C. Kittel, *Indtroduction to Solid State Physics*, 8th ed. (Wiley, New York, 2005).
[126] H. B. Gray, *Chemical Bonds: An Introduction to Atomic and Molecular Structure* (University Science Books, Mill Valley, California, 1994).
[127] W. A. Brown, R. Kose, and D. A. King, Chem. Rev. **98**, 797 (1998).
[128] H. Conrad, G. Ertl, J. Koch, and E. E. Latta, Surf. Sci. **43**, 462 (1974).
[129] R. J. Behm, K. Christmann, and G. Ertl, J. Chem. Phys. **73**, 2984 (1980).
[130] Q. Ge, R. Kose, and D. A. King, Adv. Catal. **45**, 207 (2000).
[131] V. Fiorin, D. Borthwick, and D. A. King, Surf. Sci. **603**, 1360 (2009).
[132] R. D. Ramsier, Q. Gao, H. N. Neergaard Waltenburg, K.-W. Lee, O. W. Nooij, L. Lefferts, and J. T. Yates Jr., Surf. Sci. **320**, 209 (1994).
[133] Y. Y. Yeo, L. Vattuone, and D. A. King, J. Chem. Phys. **106**, 1990 (1997).
[134] F. Bozso, G. Ertl, M. Grunze, and M. Weiss, J. Catal. **49**, 18 (1977).
[135] K. Christmann, Surf. Sci. Rep. **9**, 1 (1988).
[136] J. Zheng, Y. Zhao, and D. G. Truhlar, J. Chem. Theory Comput. **3**, 569 (2007).
[137] R. Podeszwa, K. Patkowski, and K. Szalewicz, Phys. Chem. Chem. Phys. **12**, 5974 (2010).
[138] A. B. Alchagirov, J. P. Perdew, J. C. Boettger, R. C. Albers, and C. Fiolhais, Phys. Rev. B **63**, 224115 (2001).