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,

$$\begin{aligned} s &= \frac{|\nabla n|}{2k_F n}, \quad 0 \leqslant s \leqslant \infty, \\ \varepsilon_x^{\mathrm{GGA}}(n,\nabla n) &= \varepsilon_x^{\mathrm{UEG}}(n) F_x[s(n,\nabla n)], \qquad (1)\\ E^{\mathrm{GGA\text{-}x}}[n,\nabla n] &= \int \varepsilon_x^{\mathrm{UEG}}(n) F_x[s(n,\nabla n)] d\mathbf{r}, \end{aligned}$$

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 $t(s)$:

$$\begin{aligned} t(s) &= \frac{2s^2}{4+s^2} - 1, \quad -1 \leqslant t \leqslant 1 \\ F_x^{\mathrm{GGA}}(s) &= \sum_m a_m B_m[t(s)], \\ E^{\mathrm{GGA\text{-}x}}[n,\nabla n] &= \sum_m a_m \int \varepsilon_x^{\mathrm{UEG}}(n) B_m[t(s)] d\mathbf{r} \qquad (2)\\ &= \sum_m a_m E_m^{\mathrm{GGA\text{-}x}}[n,\nabla n], \end{aligned}$$

where a_m are expansion coefficients, and $E_m^{\mathrm{GGA-x}}$ 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 $\phi(\mathbf{r},\mathbf{r}')$,

$$E^{\mathrm{nl\text{-}c}}[n] = \frac{1}{2}\int n(\mathbf{r})\phi(\mathbf{r},\mathbf{r}')n(\mathbf{r}')d\mathbf{r}\,d\mathbf{r}', \qquad (3)$$

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,

$$\begin{aligned} E_{xc} = &\sum_{m=0}^{M_x-1} a_m E_m^{\mathrm{GGA\text{-}x}} + \alpha_c E^{\mathrm{LDA\text{-}c}} \\ &+ (1-\alpha_c)E^{\mathrm{PBE\text{-}c}} + E^{\mathrm{nl\text{-}c}}, \qquad (4) \end{aligned}$$

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 $\omega^2 \rightarrow \infty$ 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

$$\boldsymbol{\Gamma}_{ij}^2 = \int_{-1}^{1} \frac{d^2B_i(t)}{dt^2}\frac{d^2B_j(t)}{dt^2}dt. \tag{6}$$

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 $\boldsymbol{\Gamma}$ 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 $\boldsymbol{\Gamma}$ 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 $\boldsymbol{\Gamma}^{-1}$. Ones are therefore inserted in the first two diagonal elements of $\boldsymbol{\Gamma}$ to avoid numerical issues. The solution vector $\mathbf{a}_i$ that minimizes C_i is written as

$$\mathbf{a}_i = \boldsymbol{\Gamma}^{-1}\left(\mathbf{X}_i'^T\mathbf{X}_i' + \mathbf{L}^2\omega_i^2\right)^{-1}\left(\mathbf{X}_i'^T\mathbf{y}_i + \omega^2\mathbf{L}^2\mathbf{a}_p'\right), \tag{7}$$

where $\mathbf{X}_i' = \mathbf{X}_i\boldsymbol{\Gamma}^{-1}$, $\mathbf{a}_p' = \boldsymbol{\Gamma}\mathbf{a}_p$, 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 $s \rightarrow \infty$. With the exchange basis transformation in Eq. (2), the prior for exchange is quite close to the PBE exchange enhancement factor. For $\omega^2 \rightarrow \infty$, 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.

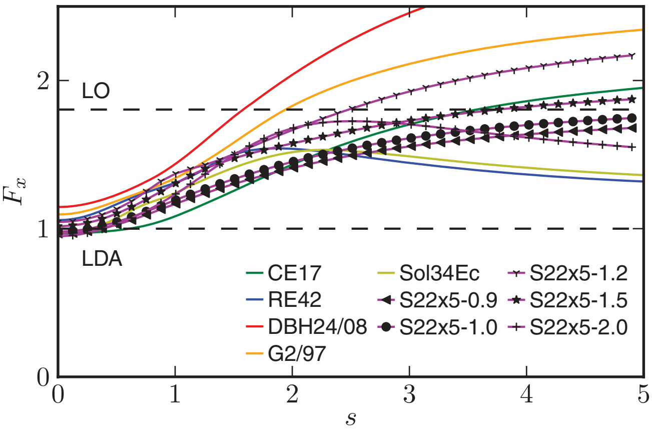

FIG. 1. (Color online) Exchange enhancement factors of the individually trained XC models listed in Table I.

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

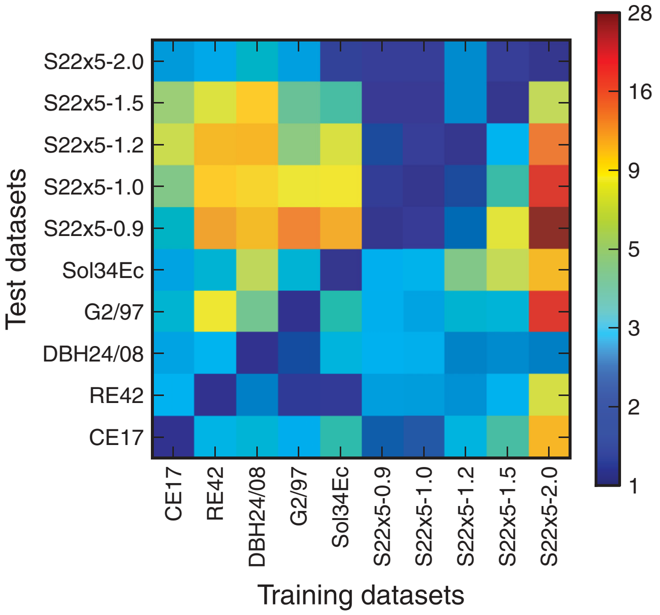

FIG. 2. (Color online) Color map of the relative standard deviations obtained when non-self-consistently applying the XC models found individually for each training data set, listed on the abscissa, to all 10 training data sets along the ordinate.

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

$$\text{rCost}[\,i\,] = \frac{C_i(\mathbf{b})}{C_i(\mathbf{a}_i)} = \frac{C_i(\mathbf{b})}{C_i^0} \geqslant 1. \tag{13}$$

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

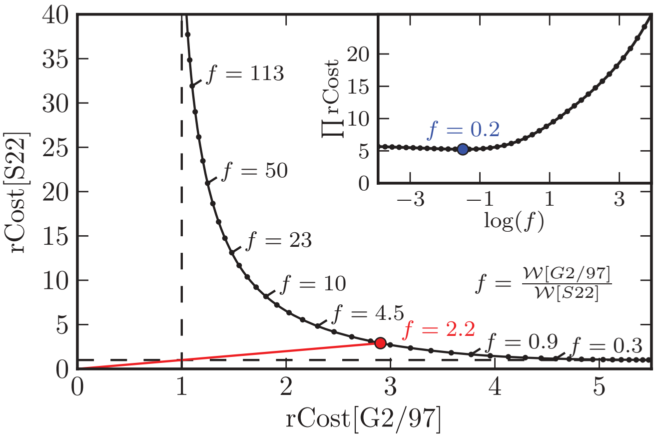

FIG. 3. (Color online) Main panel: XC model compromises between the G2/97 and S22x5-1.0 data sets illustrated in terms of relative costs (rCost) for both data sets when the weight fraction $f = \mathcal{W}[\text{G2/97}]/\mathcal{W}[\text{S22x5-1.0}]$ is varied and the summed cost function Eq. (10) is 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 rCost[G2/97] = rCost[S22x5-1.0] = 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 **c** yields a desired minimum of the product of costs. The cost function is written as

$$\Phi(\mathbf{a}) = \prod_i C_i(\mathbf{a})^{w_i}, \qquad (14)$$

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 **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 **a**, and solving

$$\sum_i \frac{w_i}{C_i}\frac{dC_i}{d\mathbf{a}} = 0. \qquad (15)$$

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

$$\mathbf{c} = \left(\sum_i \frac{w_i}{C_i}\mathbf{H}_i\right)^{-1}\left(\sum_i \frac{w_i}{C_i}\mathbf{H}_i\mathbf{a}_i\right), \qquad (16)$$

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 **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.

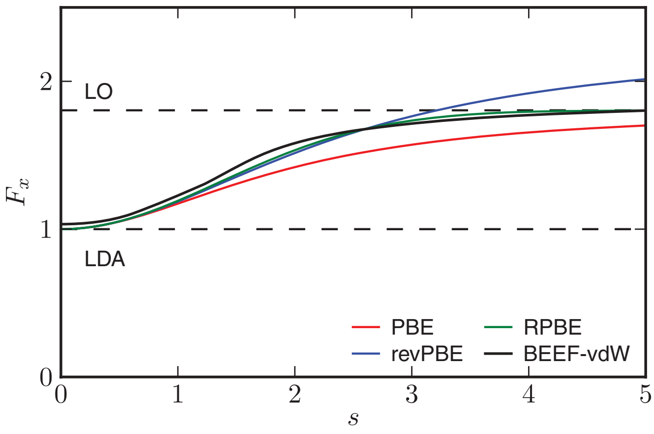

FIG. 4. (Color online) The BEEF–vdW exchange enhancement factor compared to those of a few standard GGA exchange functionals. The corresponding BEEF–vdW correlation functional is composed of 0.6 LDA, 0.4 PBE, and 1.0 nonlocal correlation.

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

$$\mathbf{H} = 2\sum_i^N \frac{w_i}{C_i(\mathbf{a}_p)} \mathbf{\Gamma}^{-1}\big(\mathbf{X}_i'^T\mathbf{X}_i' + \omega_i^2\mathbf{L}^2\big)\mathbf{\Gamma}^{-1^T}, \tag{18}$$

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

$$\Sigma' = \mathbf{V}^T\left(\sum_i^N \frac{w_i}{C_i(\mathbf{c})}\mathbf{X}_i'^T\mathbf{X}_i'\right)\mathbf{V}, \tag{20}$$

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, $M_{\rm eff}$, to be computed as

$$M_{\rm eff} = \sum_{m}^{M} \frac{{\Sigma_m'}^2}{{\Sigma_m'}^2 + \omega_{\rm eff}^2 L_m^2}. \tag{21}$$

Since $M_{\rm eff} = 7.11$ 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

$$\tau = 2\frac{C(\mathbf{c})}{M_{\rm eff}} \cdot \frac{N_{\rm tot}}{N_{\rm tot} - M_{\rm eff}}, \tag{22}$$

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

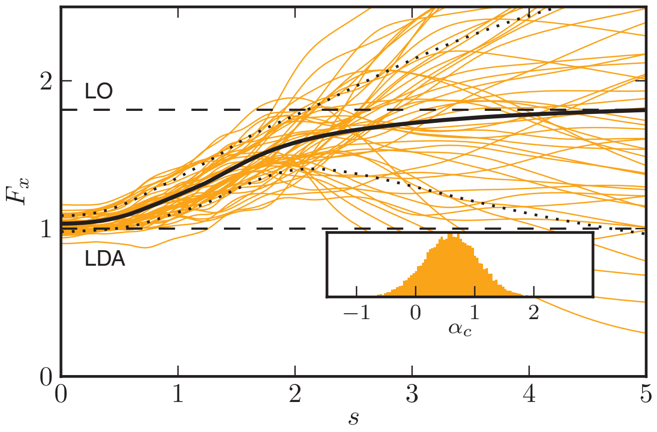

FIG. 5. (Color online) Bayesian ensemble of XC functionals around BEEF–vdW. Main panel: Black solid line is the BEEF–vdW exchange enhancement factor, while the orange lines depict $F_x(s)$ for 50 samples of the randomly generated ensemble. Dashed black lines mark the exchange model perturbations that 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 $\mathrm{CO} + \mathrm{H_2O} \rightarrow \mathrm{CO_2} + \mathrm{H_2}$ 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.

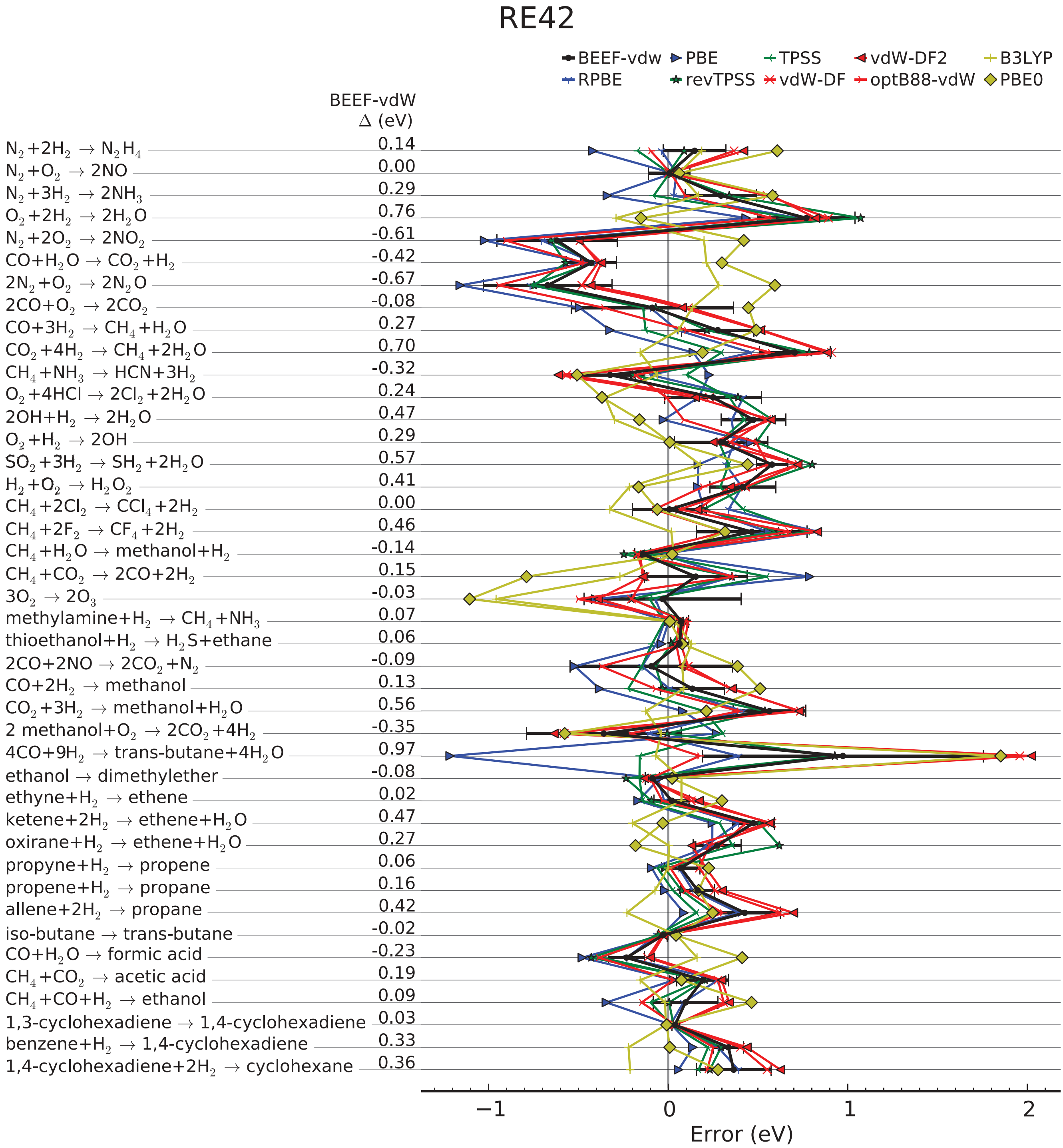

FIG. 6. (Color online) Deviations $\Delta = \Delta_r E^{\mathrm{DFT}} - \Delta_r E^{\mathrm{exp}}$ between the RE42 molecular reaction energies calculated using representative XC functionals and experiment. Color codes are BEEF–vdW: black; GGA: blue; MGGA: green; vdW–DF type: red; and hybrid: yellow. 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

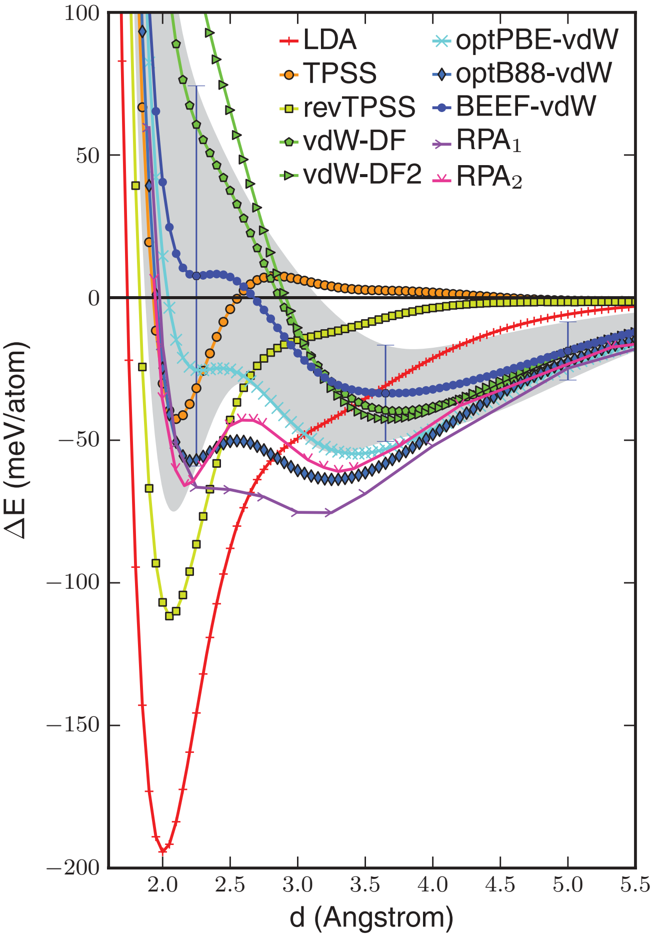

FIG. 7. (Color online) Potential-energy curves for graphene adsorption on the Ni(111) surface. Random phase approximation data are from Refs. 101 (RPA$_1$) and 102 (RPA$_2$). The gray area indicates the region spanned by the estimated standard deviations along the BEEF–vdW PEC.

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 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

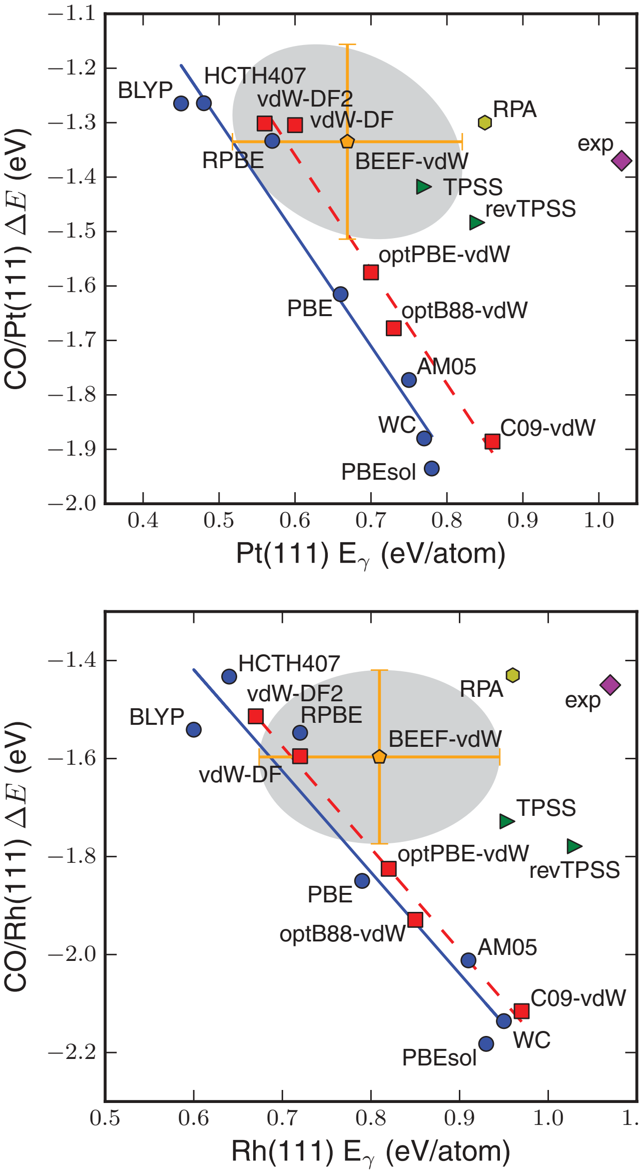

FIG. 8. (Color online) Atop CO chemisorption energies ΔE versus surface energies E_γ for Pt(111) and Ru(111). Red and blue lines are linear fits to GGA and vdW–DF type data points, respectively. MGGA data in green and yellow RPA data adapted from Ref. 119. Estimated standard deviations are indicated 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.

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

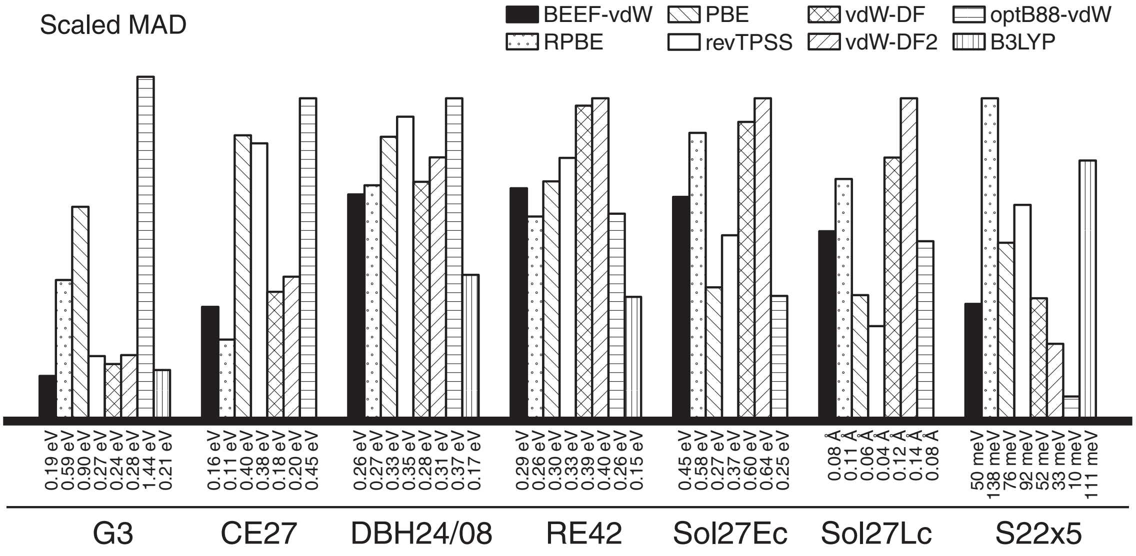

FIG. 9. Bar plot comparison of the accuracy of different density functionals in predicting various materials properties. For each data set, the bars illustrate proportionally scaled mean absolute deviations. The data sets are chosen to represent intramolecular bond energetics (G3), chemisorption energetics 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. VandeVondele. 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.

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 → 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 → 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 → 1,4-cyclohexadiene	−0.01
benzene $+ H_2 \rightarrow$ 1,4-cyclohexadiene	−0.01
1,4-cyclohexadiene $+ 2H_2 \rightarrow$ cyclohexane	−2.94

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 introduced as

$$E^{d}_{\mathrm{int}} := \varepsilon^{d}_{\mathrm{int}} \times \frac{E^{1.0}_{\mathrm{int}}}{\varepsilon^{1.0}_{\mathrm{int}}}, \tag{A4}$$

where E^{d}_{int} and $\varepsilon^{d}_{\mathrm{int}}$ denote modified and original S22x5 energies at the relative intermolecular distance d, respectively. For $E^{1.0}_{\mathrm{int}} = \varepsilon^{1.0}_{\mathrm{int}}$, Eq. (A4) obviously reduces to $E^{d}_{\mathrm{int}} = \varepsilon^{d}_{\mathrm{int}}$ for all distances. The obtained corrections to $\varepsilon^{d}_{\mathrm{int}}$ 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^{d}_{int} is calculated 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}:

$$E^{d}_{\mathrm{int}} = E^{d}_{0} - E^{d}_{1} - E^{d}_{2}. \tag{A5}$$

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.

TABLE XI. Corrections $E^{d}_{\mathrm{int}} - \varepsilon^{d}_{\mathrm{int}}$ 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	$E^{1.0}_{\mathrm{int}}/\varepsilon^{1.0}_{\mathrm{int}}$	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					

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.

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

$$\mathrm{ZPVE} = -\tfrac{9}{8} k_B \Theta_D. \tag{A6}$$

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).