

PAPER • OPEN ACCESS

## Main-group test set for materials science and engineering with user-friendly graphical tools for error analysis: systematic benchmark of the numerical and intrinsic errors in state-of-the-art electronic-structure approximations

To cite this article: Igor Ying Zhang *et al* 2019 *New J. Phys.* **21** 013025

View the [article online](#) for updates and enhancements.



**IOP** | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

**PAPER****Main-group test set for materials science and engineering with user-friendly graphical tools for error analysis: systematic benchmark of the numerical and intrinsic errors in state-of-the-art electronic-structure approximations****OPEN ACCESS**RECEIVED  
2 September 2018REVISED  
29 November 2018ACCEPTED FOR PUBLICATION  
10 December 2018PUBLISHED  
25 January 2019

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Igor Ying Zhang<sup>1,2,3,8</sup>, Andrew J Logsdail<sup>4</sup>, Xinguo Ren<sup>5</sup>, Sergey V Levchenko<sup>1,6,7</sup>, Luca Ghiringhelli<sup>1,8</sup> and Matthias Scheffler<sup>1</sup>**<sup>1</sup> Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany<sup>2</sup> Department of Chemistry and Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Fudan University, Shanghai 200433, People's Republic of China<sup>3</sup> MOE Key Laboratory of Computational Physical Science, Fudan University, Shanghai 200433, People's Republic of China<sup>4</sup> Cardiff Catalysis Institute, School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, United Kingdom<sup>5</sup> CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, People's Republic of China<sup>6</sup> Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, 3 Nobel Street, 143026 Moscow, Russia<sup>7</sup> National University of Science and Technology 'MISIS', 119049 Moscow, Russia<sup>8</sup> Author to whom any correspondence should be addressed.**E-mail:** [igor\\_zhangying@fudan.edu.cn](mailto:igor_zhangying@fudan.edu.cn) and [ghiringhelli@fhi-berlin.mpg.de](mailto:ghiringhelli@fhi-berlin.mpg.de)**Keywords:** density functional theory, test set, many-body perturbation theory, random-phase approximation**Abstract**

Understanding the applicability and limitations of electronic-structure methods needs careful and efficient comparison with accurate reference data. Knowledge of the quality and errors of electronic-structure calculations is crucial to advanced method development, high-throughput computations and data analyses. In this paper, we present a main-group test set for computational materials science and engineering (MSE), that provides accurate and easily accessible crystal properties for a hierarchy of exchange-correlation approximations, ranging from the well-established mean-field approximations to the state-of-the-art methods of many-body perturbation theory. We consider cohesive energy, lattice constant and bulk modulus of a set of materials that representatives for the first- and second-row elements and their binaries with cubic crystal structures and various bonding characters. A strong effort is made to achieve high numerical accuracy for cohesive properties as calculated using the local-density approximation (LDA), several generalized gradient approximations (GGAs), meta-GGAs and hybrids in *all-electron* resolution, and the second-order Møller–Plesset perturbation theory (MP2) and the random-phase approximation (RPA) both with frozen-core approximation based on *all-electron* Hartree–Fock, PBE and/or PBE0 references. This results in over 10 000 calculations, which record a comprehensive convergence test with respect to numerical parameters for a wide range of electronic-structure methods within the numerical atom-centered orbital framework. As an indispensable part of the MSE test set, a web site is established <http://mse.fhi-berlin.mpg.de>. This not only allows for easy access to all reference data but also provides user-friendly graphical tools for post-processing error analysis.

**1. Introduction**

First-principles electronic-structure calculations have become an indispensable complement to experiments in physics, chemistry, and materials science. Since the exact description of interacting electrons and nuclei is intractable for most systems, it is a perpetual challenge to find an 'optimal' approximation that reconciles computational accuracy, efficiency, and transferability across different chemical environments and

dimensionality. In quantum chemistry for atoms and molecules, test sets with accurate reference values for various relevant chemical and physical properties have been since long established [1–15]. These test sets play an instrumental role in the development of hierarchical electronic-structure approximations for both wavefunction theory (WFT) and density functional theory (DFT). In particular, they are needed for validating numerical implementations [16–19], investigating the basis set convergence [20–24], and benchmarking the intrinsic limitations of the various quantum-chemistry methods [1–6, 14, 15, 25–38]. For most existing test sets, the developers provide all essential information for each calculation, including molecular geometry, basis set, and/or code-specific numerical setting [39, 40].

Condensed-matter physics and materials science is lacking behind, so far, with respect to comparable benchmark datasets. Two types of accurate reference data are crucial. One is essentially the exact values, obtained from either precise experiments or high-level theoretical calculations. They are prerequisite for an unbiased benchmark of the intrinsic errors associated with, for example, the exchange-correlation (XC) approximations in DFT or more generally the treatment of relativistic effects. The other type of reference data can help to quantify the numerical error in calculated values using any chosen method, which might come from the basis set incompleteness, finite  $k$ -point sampling, or other approximations made in the numerical implementation. Unfortunately, for condensed-matter systems, neither of these reference data are easy to obtain.

The subsequent paper, which seeks to address some of these challenges, is structured as follows: after a detailed survey of the test sets that are widely used in quantum chemistry, as well as in materials science, we discuss the underlying challenges to obtain numerically accurate reference data in the latter. In section 2, a representative main group test set for materials science and engineering (MSE) is established. It is followed by an overview of a dedicated web site <http://mse.fhi-berlin.mpg.de>. We take three examples to demonstrate that the web site features a multi-mode access framework, versatile visualization, and a linear regression tool for post-processing data analysis. Section 4 presents the numerical strategies applied in this work to obtain the numerically well-converged data on a diverse levels of theory. We draw conclusions and give an outlook in section 5.

### 1.1. Test sets in quantum chemistry

The importance of reliable test sets for the success of quantum-chemistry methods was first realized by Pople and co-workers [1, 26, 41–44]. Along with the development of the Gaussian- $n$  theories [1, 26, 41–44], a hierarchy of extrapolating levels of correlation and basis sets have been developed to obtain increasingly accurate thermochemistry. A sequence of pioneering test sets for quantum chemistry were also developed and used for methodological validation. These test sets are now widely recognized as the  $G_n$  test sets, addressing atomization energies and other energetic properties of molecules with increasing numbers of accurate reference values obtained from experiments [1, 26, 41–44]. One of these test sets, G3/99 [26] generated in 1999, has been widely used in the development of density functional approximations (DFA) to describe covalent bonding in the main group molecules [1, 25–38]. Since then, many well-established test sets have been proposed for different, mainly ground-state properties of molecules and molecular processes [1–15, 26, 42]. For example, Hobza and co-workers designed the S22 set, comprising 22 non-covalent binding complexes of biologic relevance, which can be further divided into 7 systems for hydrogen bonds, 8 for dispersion bonds, and 7 for mixed bonds [7]. Moreover as one of the most popular standard benchmark sets, the BH76 set proposed by the Truhlar group consists of forward and reverse barrier heights of 19 hydrogen transfer reactions, 6 heavy atom transfer reactions, 8 nucleophilic substitution reactions, and 5 unimolecular and association reactions [4]. Most recently, the Grimme group compiled a comprehensive benchmark test set, GMTKN30, which includes 30 subsets collected from the literature, covering a large section of chemically relevant properties of the main-group molecules [14, 15].

It may appear plausible that reference data, i.e. the accurate values of the relevant chemical and physical properties of atoms and molecules, can be acquired from experiments [1–3, 6, 11–13, 26, 42], but the required accuracy is not always achievable. The influence of electron-vibrational coupling is often unclear, which can hamper the comparison with the directly computed values. Thus, reference data would ideally be created by accurate quantum chemistry methods such as the coupled-cluster (CC) method with single, double, and perturbative triple excitations, CCSD(T) [45–47], or single, double, triple and perturbative quadruple excitations, CCSDT(Q), or full configurational interaction (full-CI) theory extrapolated to the complete basis set (CBS) limit [48–50]. For the test sets aforementioned, the corresponding reference data are either carefully chosen from accurate experiments [1–3, 6, 11–13, 26, 42] or obtained from high-level first-principles calculations [4, 5, 7–10].

As the most popular choice in quantum chemistry, the atom-centered Gaussian-type orbital (GTO) basis sets provide well-converged total energies of atoms and molecules with a reasonable basis set size for the mean-

field methods, including the local-density approximation (LDA), generalized gradient approximations (GGAs) and hybrid functionals in DFT, and the Hartree–Fock method in WFT. For an all-electron approach and for heavier atoms (e.g.  $Z > 18$ ), i.e. when wave-functions oscillate more strongly at the core region, it is more convenient, efficient, and accurate to use numerical atom-centered orbital (NAO) basis sets [51, 52].

For advanced correlation methods which require unoccupied single-particle states, e.g. the second-order Møller–Plesset perturbation theory (MP2) [53], the random-phase approximation (RPA) [54–57], or CCSD(T) [45–47, 58], the slow basis set convergence is a more serious problem [23, 24, 59–62]. It can be ascribed to the inaccurate description of the electron–electron Coulomb cusps using smooth orbital product expansions [63, 64]. The so-called correlation consistent basis sets have been proposed [24, 59], which allow for an analytic extrapolation to the CBS limit. Alternatively, by introducing an explicit dependence on the inter-electronic distance into the wave-function (F12 strategies [64–68]), it is possible to treat the cusp explicitly and suppress the basis set incompleteness error at a finite basis set size. Both techniques were proposed to address the basis set incompleteness errors in advanced correlation methods for both atoms and molecules, and they have been demonstrated to be very successful for light elements (e.g.  $Z < 18$ ) and for ground-state properties [24, 59, 64, 68]. Therefore, accurate chemical or physical properties, e.g. reaction energies, reaction barrier heights and isomerization energies, can be obtained without concern for the numerical error cancellation originating from a given finite basis set. However, the quantum-chemistry test sets are mainly available only for the ground-state properties of light elements and small molecules, as it remains a challenge to obtain accurate reference data with advanced correlation methods for heavy elements (e.g.  $Z > 18$ ), excited states, and large systems.

## 1.2. Test sets in materials science

As mentioned earlier, in computational materials science, the situation is more complex and less developed than in the quantum chemistry of molecules, and there is an urgent need for accurate test sets to support the development of advanced DFAs for solids. Staroverov and co-workers considered lattice constants, bulk moduli, and cohesive energies of 18 solids, as well as jellium surface energies to benchmark the TPSS functional, a meta-generalized gradient approximation (meta-GGA) [69]. They also presented results for LDA, GGA PBE, and meta-GGA PKZB [70]. Heyd and Scuseria [71] used lattice constants and bulk moduli for 21 solids and band gaps for 8 semiconductors out of this set to investigate the screened Hartree–Fock-exchange properties in the HSE hybrid functional. Comparison was also made with LDA, PBE, and TPSS. This test set (or part of it) has been used to benchmark the HSE and HSEsol as implemented in VASP [72, 73], to understand the failure of B3LYP for solids [74], and to develop improved DFAs for solids [75–77]. A set of 60 cubic solids was used by Haas, Tran, and Blaha to compare the performance of different GGA functionals, LDA and TPSS in describing lattice constants and bulk moduli [78–80]. To compare the accuracy of different van der Waals (vdW) functionals applied to solids, Klimes and co-workers determined cohesive properties for a set of 23 bulk solids [81], based on a set of materials and properties used by Csonka and co-workers when testing the accuracy of GGA and meta-GGA functionals [82]. Recently, Tran *et al* performed an extensive test on the lattice constant, bulk modulus, and cohesive energy for a large set of XC approximations belonging to rungs 1–4 of ‘Jacob’s ladder of DFT’ [83], performed in a non-self-consistent way using PBE orbitals. The test set used contain 44 strongly and 5 weakly bound solids, most of which were cubic, except for hexagonal graphite and h-BN. These test sets of materials with different properties have already been playing an important role in developing and benchmarking the DFAs for solids. These test sets exclusively rely on reference values from experiment; no quasi-exact calculated results exist, which can be used as reference data for solids.

For reasons of comparability and consistency, reference values from theory provide the unique opportunity to compare calculations based on exactly the same atomic structure and exclude finite-temperature, zero-point vibration, relativistic and electron–vibration coupling effects. Recently, the ‘gold standard’ method in quantum chemistry, CCSD(T), has gained attention in materials science [84–89]. A significant step towards an exact description of solids was demonstrated by Booth *et al* by performing a full-CI quality calculation with the aid of the Quantum Monte Carlo (QMC) stochastic strategy [85]. However, the numerical difficulty and computational complexity for describing solids has strongly limited the applications of these highly accurate quantum chemistry methods to only model systems with very few electrons, basis functions, and  $\mathbf{k}$  points [84, 85, 88, 90]. Other QMC stochastic strategies, such as Variational Monte Carlo (VMC) [91] and Diffusion Monte Carlo [92] feature a low computational scaling with respect to the system size and a weak dependence on basis set. QMC has been demonstrated to be as accurate as aforementioned quantum chemistry methods, e.g. CCSD(T), if certain numerical issues could be considered properly, which includes the fixed-node error and the form and optimization of the trial function [93–95].

Unlike the molecular systems investigated by quantum chemistry, extended periodic materials are often simulated in reciprocal space with periodic boundary conditions (PBC) so as to fully leverage the periodic

symmetry. However, it also introduces extra numerical difficulties: the first Brillouin zone in reciprocal space must be sampled by a finite number of  $\mathbf{k}$  points, for which different kinds of  $\mathbf{k}$ -mesh have been proposed to provide an efficient sampling according to different periodic symmetries [96–100]. Basis set incompleteness error is another big issue: the atom-centered GTO-type basis sets that are dominant in quantum chemistry become cumbersome for condensed matter systems [101, 102]. In particular, large and especially diffuse GTO basis sets might cause severe ill-conditioning problems due to the lack of rigorous orthogonality for GTO basis functions. The plane-wave basis sets are a popular alternative choice for solids as they provide an intrinsically improvable description of the electronic orbitals with only one parameter: the plane-wave cutoff energy. However, the delocalized nature of plane waves makes them inefficient to describe the localized core electrons surrounding atomic nuclei. In practice, the pseudo-potential approximation, which removes the core electrons from explicit electronic-structure calculations, are often used together with the plane-wave basis sets in the projector augmented wave (PAW) framework [103, 104]. More recently, it has been demonstrated that the use of compact NAO basis sets [52, 105] or the linearized augmented plane-wave (LAPW) strategies [106] are able to converge the mean-field approximations in all-electron and full-potential resolution with relativistic effects explicitly included as opposed to implicitly as is the case for pseudo-potential methods.

Such diversity of methods leads to a need for a standard data set that is comparable among different numerical implementations. This is indeed the purpose of the  $\Delta$ -value concept introduced by Cottenier *et al* [107, 108], which is currently focusing on the implementation of the PBE method for solids. The numerically well-converged PBE results from the all-electron, full-potential LAPW-based program WIEN2k [108, 109] are taken as reference. To quantify the numerical errors in the description of equations of state, the  $\Delta$ -value is defined as the root-mean-squared (RMS) energy difference between the equations of state of two codes, averaged over an exhaustive test set of crystalline solids, containing all ground-state elemental crystals.

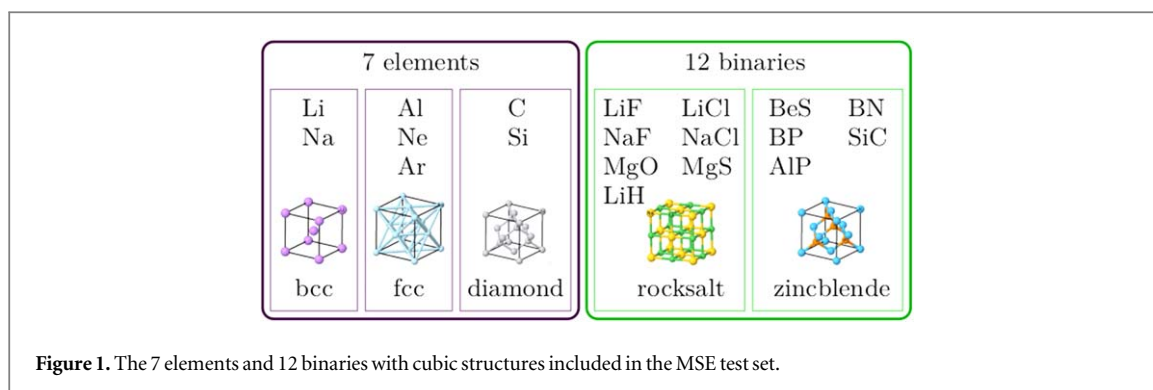
Despite the enormous success achieved by the (semi)-local DFA, e.g. LDA and GGA PBE, there are several notorious failures of these methods. For instance, there exist serious self-interaction errors and significant underestimation of vdW interactions for these functionals, which demand more sophisticated approximations. The meta-GGAs, in which the orbital kinetic energy density is added to the density functional evaluation, belong to a higher rung of the ‘Jacob’s Ladder in DFT’. Representative are the nonempirical TPSS [69] and the multiparameter empirical M06-L [110]. A recently proposed nonempirical meta-GGA, SCAN, which satisfies 17 known exact constraints appropriate to a semilocal functional [111], has been found to show a big step ahead in accuracy for manifold chemical and physical properties and has attracted increasing interest in computational materials science. The so-called hybrid functionals, e.g. HSE06 [112], incorporate the information of the occupied orbitals in Hartree–Fock-like ‘exact exchange’. The next level of complexity is to derive DFAs using the information of virtual orbitals. Two methods at such level, MP2 [53] and RPA [18, 34, 35, 54–56, 113], are state-of-the-art in computational materials science [34, 76, 77, 85, 87, 114–117]. The numerical errors in these methods can either be inherited from the aforementioned algorithms to solve the one-electron Kohn–Sham (or Hartree–Fock) equations, or arise from extra algorithms, such as the choice of the self-consistent Kohn–Sham orbitals for the post-processing evaluations [118], the resolution-of-identity technique to handle the two-electron four-center integrals [18, 119–121], and the localization approximations [114, 122, 123] to reduce the computational scaling in these advanced correlation methods.

In this context, it becomes imperative for computational materials science to have a representative test set with numerically well-converged reference values at various levels of theory. In spirit of the existing quantum-chemistry test sets and the  $\Delta$ -value concept, we introduce in the following our main group test set for MSE, coined MSE, which is based on results acquired using density functional methods from LDA, PBE, PBEsol, SCAN, TPSS, M06-L, HSE06; and state-of-the-art MP2 and RPA. The numerical convergence of these methods is investigated in terms of the total energy. Cohesive energies, lattice constants and bulk moduli are then reported. A comprehensive understanding of the numerical errors, particularly in MP2 and RPA, is discussed in order to aid the community’s pursuit of a numerically stable implementation of CCSD(T) and full-CI for the exact description of solids.

## 2. Main group test set for materials science and engineering (MSE)

In this project, we select 7 elemental solids and 12 binaries with cubic structure, as the first step in creating the MSE test set. As illustrated in figure 1, the set is composed of elements from the first and second rows of the periodic table, consisting of the body-center-cubic (bcc), face-center cubic (fcc), diamond, rocksalt, and zincblende structures. Thus, the set includes materials composed of main group elements with metallic, covalent, ionic, vdW, and mixed bonding characters.





The Fritz Haber Institute ‘*ab initio* molecular simulations’ (FHI-aims) electronic-structure package [52] is used to generate the numerically well-converged reference data in the MSE test set. The reasons for this choice are as follows:

- The numerical accuracy of the FHI-aims package has been shown to be equivalent to the accuracy of other high-quality codes [124]. FHI-aims, WIEN2K and exciting are indistinguishable in accuracy and at the top of all participating codes.
- A range of popular DFAs have been implemented in FHI-aims and can be used routinely for atoms, molecules, clusters, and periodic systems. Besides the conventional (semi)-local functionals, e.g. LDA, PBE, PBEsol, SCAN, TPSS and M06-L, a real-space implementation of the exact exchange operator in PBC by Levchenko *et al* has been demonstrated to allow for a practical use of hybrid functionals (including HSE06) and the Hartree–Fock method, as well, for both molecular and periodic systems [105]. Furthermore, as will be reported soon, a massively parallel, in-memory implementation of periodic MP2 and RPA methods has also been implemented in FHI-aims using canonical crystalline orbitals.
- FHI-aims employs numerically tabulated NAO basis sets [52]. The default basis sets were developed from a minimal basis that is composed of exact occupied orbitals for spherically symmetric free atoms [52]. Such minimal basis captures the essential core–electron behavior in the vicinity of atomic nuclei. Additional ‘tiers’ are defined in a step-wise minimization of the LDA total energies of symmetric dimers for all elements. These hierarchical basis sets, *tier-n* ( $n = 1-4$ ) for short, can provide the CBS total energies for mean-field approximations (LDA, PBE, PBEsol, SCAN, TPSS, M06-L and HSE06 in this project) in an *all-electron description*.
- To address the slow basis set convergence of advanced correlation methods (MP2 and RPA in this project), FHI-aims provides a sequence of NAO basis sets with *Valence Correlation Consistency* [24], namely NAO-VCC- $nZ$  with  $n = 2, 3, 4$  and 5. The basis set incompleteness error in the valence correlations of MP2 and RPA can be removed using two-point extrapolation schemes [24]. Unless otherwise stated, the MP2 and RPA calculations in this work are *valence-only* (frozen-core) using *all-electron* Hartree–Fock and PBE orbitals, respectively. Assuming complete convergence of the  $\mathbf{k}$ -mesh and basis sets, any discrepancy between our results and those obtained with a plane-wave basis should presumably originate from the error of the pseudo-potentials used in plane-wave approaches to generate the valence and virtual orbitals in the self-consistent procedure. Besides the minimal basis, the NAO-VCC- $nZ$  basis sets comprise an additional group of  $s, p$  hydrogen-like functions, named enhanced minimal basis [24]. For isolated molecules, such additional functions have been demonstrated to be useful to improve the description of valence correlations [24]; however, the densely packed nature of the condensed-matter systems largely alleviates the difficulty of saturating the  $s, p$  basis space for valence correlations. In this work, we exclude the enhanced minimal basis and re-optimize the NAO-VCC- $nZ$  basis sets, but do not change the name for simplicity (interested readers are referred to [24] for the optimization strategy in detail).

At present the NAO-VCC- $nZ$  basis sets in FHI-aims are only available from H to Ar [24]. As a consequence, the MSE test set is currently focused on light main group elements (see figure 1), with future intentions to expand to heavier elements. In this paper, we report the numerically well-converged results of MP2 and RPA for 14 selected materials, including 4 elemental solids (Ne and Ar in the fcc, and C and Si in the diamond structure) and 10 binaries (LiF, LiCl, LiH, MgO and MgS with rocksalt, BeS, BN, BP, SiC and AlP with zincblende structure). We snapshot in figure 2 the MP2 data currently available through the MSE web site. In the long term, the MSE test set shall be extended to include heavy elements and non-cubic structures including representatives for the

Group	Material	Structure	Method	$E_{\text{coh}}$ (eV)	$a_0$ (Å)	$B$ (GPa)	KGrid	Basis Set	CS	EOS	CutPot	ABS
10	Ne	fcc	MP2	0.01	4.499	1.0	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
18	Ar	fcc	MP2	0.09	5.150	3.7	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
6	C	diamond	MP2	8.08	3.541	454.8	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
14	Si	diamond	MP2	5.21	5.404	100.2	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
3-9	LiF	rocksalt	MP2	4.54	4.018	78.3	CKM(6,8)	CBS(2,3)	☑	☑	☑	☑
3-17	LiCl	rocksalt	MP2	3.70	5.035	38.7	CKM(6,8)	CBS(2,3)	☑	☑	☑	☑
8-12	MgO	rocksalt	MP2	5.53	4.240	156.7	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
12-16	MgS	rocksalt	MP2	4.39	5.182	84.0	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
4-16	BeS	zincblende	MP2	4.90	4.820	104.5	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
5-7	BN	zincblende	MP2	7.25	3.588	368.4	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
5-15	BP	zincblende	MP2	5.77	4.498	178.1	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
6-14	SiC	zincblende	MP2	7.01	4.350	230.1	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
13-15	AlP	zincblende	MP2	4.67	5.450	95.5	CKM(6,8)	CBS(3,4)	☑	☑	☑	☑
1-3	LiH	rocksalt	MP2	2.38	3.994	38.5	CKM(6,8)	CBS(2,3)	☑	☑	☑	☑

Figure 2. Snapshot of the MP2 data currently available from the MSE web site.

majority of systems of interest in MSE. To achieve this goal, it is prerequisite to establish a set of NAO basis sets for heavy elements ( $Z > 18$ ), which can provide a smooth numerical convergence of MP2, RPA and other advanced correlation methods towards the complete basis set limit.

Relativistic effects were investigated in our calculations as well, using the atomic zeroth order regular approximation (atomic ZORA) [52]. By performing a linear regression comparison, we confirm quantitatively that the relativistic effect has a negligible influence on the materials formed by light elements, regardless of any XC approximation that is used (see the following section for more details). Thus, we report the reference data without consideration of relativistic effects. As a side benefit, this approach excludes the numerical uncertainty arising from different relativistic methods [52]. Obviously, for heavy elements, the relativistic effect must be considered carefully, but this is beyond the scope of the present study.

### 3. The web site of the MSE test-set

A key feature of our MSE test set project is to establish a dedicated web site <http://mse.fhi-berlin.mpg.de>, which allows for easy access and analysis of the data.

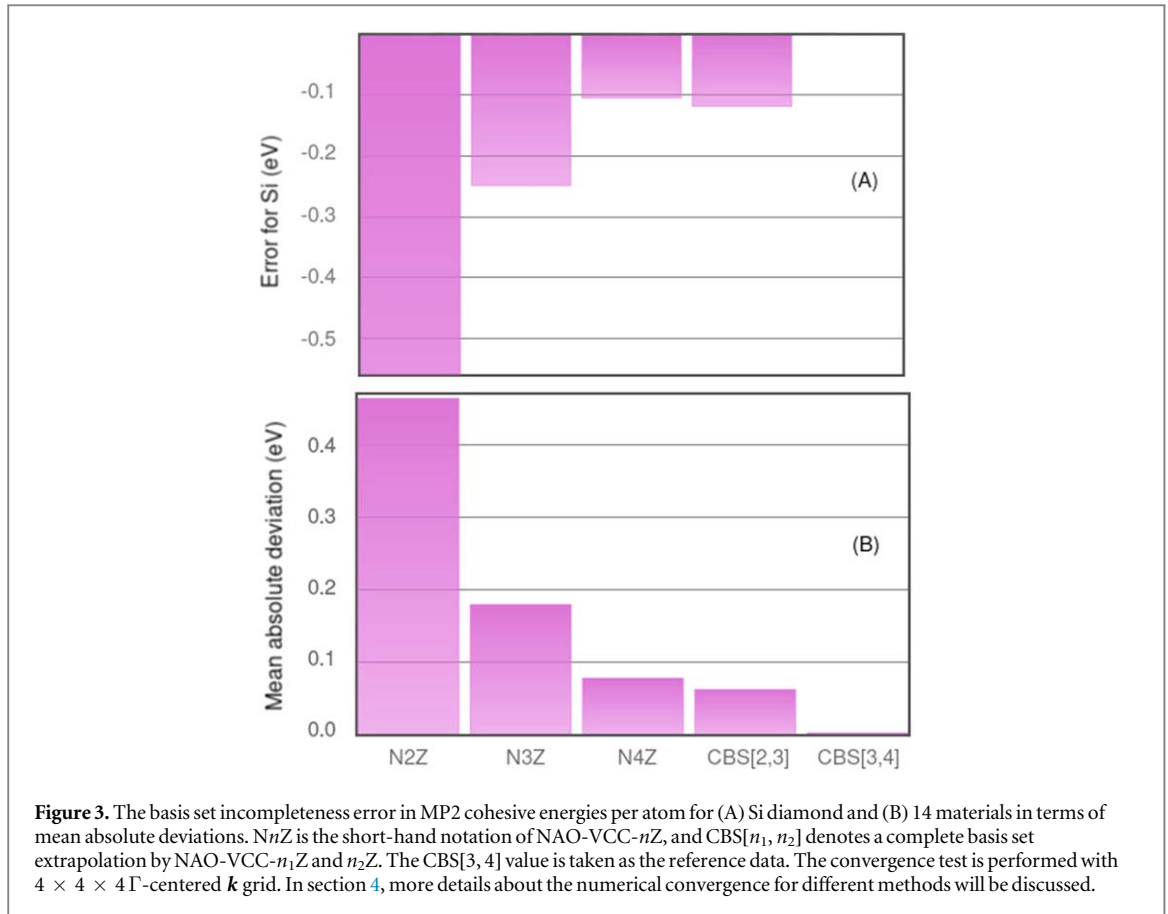
#### 3.1. Multi-mode access to the reference data

At the MSE home page, an overview of the MSE test set project is given along with a table that summarizes all materials available for a given DFA. The materials are sorted by their crystalline structures with a layout similar to figure 1. The default DFA is PBE; the table for other DFAs is obtained via a drop-down select box. A search engine allows to access a group of results for a given material, structure, or DFA, and/or combination of the above. Figure 2 shows a snapshot of the filtered table produced by the search engine, listing the MP2 reference data currently available. This multi-mode search framework, together with a well-organized data structure, guarantees easy access to the more than 10 000 calculations in the current MSE test set.

#### 3.2. Visualization

The MSE web site provides a quick and easy-to-use visual display of the test set data. For any individual reference value calculated for a given material and using a specific DFA, the crystal structure and the well-converged equation of state are displayed. If available, one can also visualize the numerical convergence towards this reference data; this includes convergence tests for basis set,  $\mathbf{k}$ -mesh, and internal numerical parameters. Furthermore, a statistic analysis of the basis set convergence for the whole test set can be visualized, which is derived using the mean absolute deviation (MAD).

As an example of this functionality, figure 3 presents the basis set incompleteness error in MP2 cohesive energies per atom for (A) Si diamond and (B) 14 materials representing covalent, ionic, vdW, and mixed bonding characters. Clearly, NAO-VCC- $nZ$  basis sets provide an improved description for advanced correlation methods with the increase of the basis set size, which allows for the extrapolation to the CBS limit from NAO-VCC-3Z and 4Z (i.e. CBS[3,4]). As a recommendation for practical use, we note that the basis set extrapolation from NAO-VCC-2Z and 3Z (CBS[23]) guarantees a near NAO-VCC-4Z quality, but with much less computational cost.



### 3.3. Linear regression

For the benchmark studies of the DFAs, the MAD and the root-mean-square deviation (RMSD) are often used to quantify the numerical errors in calculations. For a given group of materials with the number of materials of  $n$  and the targeted observables of  $\{Y_i\}$ , if the computational data are  $\{y_i\}$  with the errors distributed as  $x_i = Y_i - y_i$ , we then define the relevant MAD and RMSD as

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |x_i|; \quad \text{RMSD} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}. \quad (1)$$

Despite both MAD and RMSD being measures of accuracy, we note that larger errors have a disproportionately large effect on RMSD, making it more sensitive to outliers.

A linear regression by means of a least-squares fit allows us to separate the error into predictable (or systematic) and material-specific (or residual) parts [125, 126]. The resulting linear model

$$\hat{L}(y_i) = \beta y_i + \alpha \quad (2)$$

allows the analysis of the material-specific deviations  $\{\epsilon_i\}$  and the corresponding root-mean-square deviation (RMSD) as

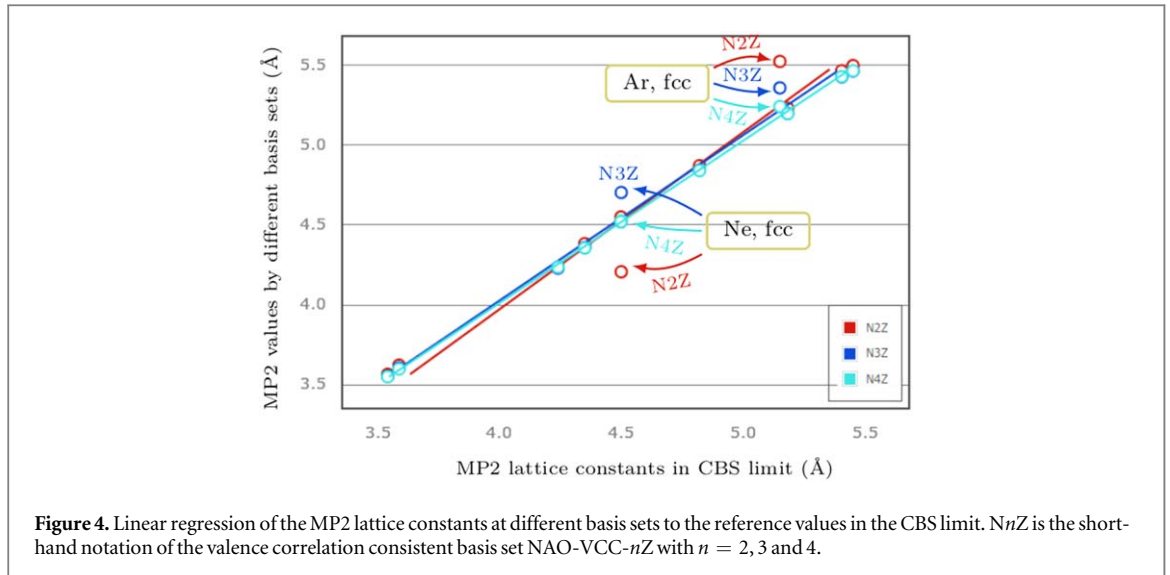
$$Y_i - \hat{L}(y_i) = \epsilon_i; \quad \overline{\text{RMSD}} = \sqrt{\frac{\sum_{i=1}^n \epsilon_i^2}{n}}. \quad (3)$$

The systematic error can be determined by the difference between RMSD and  $\overline{\text{RMSD}}$ . It was argued that the material-specific deviations represent a true measure of the inadequacy of the method or the numerical incompleteness of the basis set and  $\mathbf{k}$ -mesh [125, 126]. To make the most of this analysis method, we have equipped the MSE web site with a linear-regression analysis tool. In three examples, we will demonstrate how to gain insight into the numerical and intrinsic limitations of a state-of-the-art DFAs.

#### 3.3.1. Basis set convergence in the MP2 lattice constants

Slow basis set convergence is a well-documented problem for advanced correlation methods like MP2 and RPA. Figure 3 presents the basis set incompleteness errors in MP2 cohesive energies using the valence correlation consistent basis set in the NAO framework. Taking the MP2 results in the CBS limit as the reference, table 1 lists





**Figure 4.** Linear regression of the MP2 lattice constants at different basis sets to the reference values in the CBS limit.  $NnZ$  is the shorthand notation of the valence correlation consistent basis set NAO-VCC- $nZ$  with  $n = 2, 3$  and 4.

**Table 1.** RMSDs and material-specific  $\overline{\text{RMSD}}$  of the basis set incompleteness errors for the MP2 lattice constants. The values in parentheses exclude Ne and Ar. (unit: Å)

Method	RMSD	$\overline{\text{RMSD}}$
N2Z	0.148 (0.043)	0.126 (0.015)
N3Z	0.089 (0.017)	0.071 (0.009)
N4Z	0.032 (0.016)	0.021 (0.007)

the RMSDs and  $\overline{\text{RMSD}}$ s of the incompleteness errors for NAO-VCC- $nZ$  with  $n = 2, 3$ , and 4. Despite the basis set errors systematically decreasing with the increase of the basis set size, RMSD and  $\overline{\text{RMSD}}$  with NAO-VCC-3Z remain about 0.1 Å, which is unacceptable for real applications.

Figure 4 shows the linear regression analysis. It clearly suggests that there are two outliers, which are Ne and Ar with fcc structure. We note that Ne and Ar crystals are bonded by vdW forces, which are overestimated by the MP2 method; however it is an intrinsic error of the MP2 method and thus not addressed here. For isolated molecules, it has been discussed in previous literature [18, 24] that an accurate description of weak interactions using advanced DFAs needs either a counterpoise correction scheme or very large basis sets with diffuse basis functions to address the notorious basis set superposition error. As suggested by our results in this work, this conclusion is also true for solids governed largely by vdW interactions. Excluding these two cases from the linear regression analysis results in a much better quantitative accuracy (values in parentheses in table 1). The RMSD for NAO-VCC-3Z and  $\overline{\text{RMSD}}$  for NAO-VCC-2Z are below 0.02 Å, which clearly suggests that NAO-VCC-3Z or even 2Z with a systematic correction would be good enough to converge the MP2 lattice constants for strongly bonded systems, while the weak interactions must be treated with special care.

### 3.3.2. Starting point (SP) influence on meta-GGAs and RPA

The common practice of evaluating meta-GGAs, like TPSS, M06-L and SCAN, and advanced many-body perturbation methods, like RPA, is by performing an energy evaluation *a posteriori* using LDA, GGA or hybrid GGA orbitals [35, 83]. Meanwhile, the Hartree-Fock orbitals are used for the MP2 method generally in quantum chemistry. The self-consistent implementation of meta-GGAs has been realized in many computational packages, including FHI-aims. In the MSE test set, we provide accurate reference data of TPSS, M06-L and SCAN for both (a) self-consistent and (b) non-self-consistent using PBE orbitals, calculations. We also examine the starting point influence on RPA cohesive energies by using PBE and PBE0 orbitals. For simplicity and consistency, we use ‘DFA@SP’ in this discussion to denote which method is based on which starting point (SP) orbitals.

Table 2 shows the influence of the starting point on the cohesive energies per atom in terms of the RMSD. The linear regression was performed to extract the material-specific part of the deviations. Clearly, the starting point influence is mild for meta-GGAs: SCAN shows the smallest influence, leading to a RMSD of only 15 meV. Meanwhile, the linear regression analysis suggests that these errors are almost material-specific. The corresponding systematic errors (RMSD –  $\overline{\text{RMSD}}$ ) are less than 2 meV. In contrast, RPA is much more

	LDA	PBE	PBEsol	SCAN	TPSS	M06-L	HSE06	MP2	RPA	Exp
LDA		0.76	0.43	0.70	0.90	0.72	0.81	0.41	1.00	0.68
PBE	0.06		0.34	0.16	0.20	0.24	0.07	0.42	0.27	0.14
PBEsol	0.04	0.06		0.30	0.50	0.36	0.39	0.17	0.58	0.33
SCAN	0.15	0.12	0.13		0.22	0.11	0.16	0.33	0.31	0.09
TPSS	0.18	0.13	0.17	0.10		0.25	0.14	0.54	0.16	0.17
M06-L	0.24	0.19	0.22	0.10	0.12		0.23	0.35	0.34	0.26
HSE06	0.08	0.04	0.06	0.09	0.10	0.16		0.46	0.21	0.15
MP2	0.18	0.16	0.15	0.09	0.17	0.15	0.13		0.63	0.38
RPA	0.16	0.08	0.14	0.05	0.11	0.11	0.11	0.10		0.20
Exp	0.15	0.12	0.16	0.08	0.10	0.21	0.10	0.13	0.08	

**Figure 5.** Cross-over comparison of cohesive energies between different methods and experiment (Exp) in terms of the RMSD. The direct RMSDs are shown in the up triangular part, while the values in the lower triangular part are the material-specific errors ( $\overline{\text{RMSD}}$ ) after the linear regression. (unit: eV)

**Table 2.** RMSDs and material-specific  $\overline{\text{RMSD}}$  of the starting-point influence on cohesive energies per atom. ‘DFA@SP’ denotes which method is based on which starting point orbitals. ‘sc-DFA’ denotes the self-consistent study. (unit: eV)

Methods to compare	RMSD	$\overline{\text{RMSD}}$
sc-SCAN versus SCAN@PBE	0.015	0.015
sc-M06-L versus M06-L@PBE	0.049	0.043
sc-TPSS versus TPSS@PBE	0.035	0.033
RPA@PBE versus RPA@PBE0	0.119	0.030

sensitive to the choice of the starting point orbitals, though the influence is quite systematic: the material-specific error  $\overline{\text{RMSD}}$  is only about 30 meV. This small error indicates that a careful choice of the starting point orbitals can improve the RPA performance because of its systematic underestimation of cohesive energies [118] and weak interactions [34].

For meta-GGAs and RPA, the starting point influences on lattice constant and bulk modulus are similar: the material-specific deviations are mild, though the systematic deviation is quite large for RPA results. The readers interested in this topic can easily access the data and perform the linear regression online by themselves. It is also easy to investigate the influence of relativistic effects on different methods and properties in the same manner.

### 3.3.3. Cross-over comparison between different methods

Figure 5 summarizes the cross-over comparison of cohesive energies between different methods. RMSDs and the material-specific  $\overline{\text{RMSD}}$ s are shown in different subtables. In this comparison, RPA@PBE reference data were used and meta-GGAs were calculated self-consistently. The nine DFAs investigated cover all five rungs of the ‘Jacob’s ladder of DFT’. From the many-body perturbation theory point of view, MP2 and RPA consider the electronic correlations explicitly in the many-body interaction picture, while others are all mean-field approximations.

LDA and RPA are the density functional approximations on the lowest and highest rungs of the ‘Jacob’s ladder’, respectively, and they show the largest deviation in the cohesive energy calculations, leading to a RMSD of over 1 eV. In fact, LDA shows a large disagreement with all other methods, with the average RMSDs about 0.7 eV. This observation agrees with a widely accepted argument that the local density approximation derived from homogeneous electron gas completely misses the high order density derivative or response information

that is important for a proper description of chemical bonding. However, our linear regression analysis suggests that this error is quite systematic: the material-specific  $\overline{\text{RMSD}}$  is only 0.16 eV between LDA and RPA.

A similar analysis can be applied to study the intrinsic limitations in MP2 and RPA. It is well-known that the RPA method performs unsatisfactorily in the calculations of cohesive energy for solids [35, 37, 127, 128]. From the many-body perturbation point of view, this error is due to the lack of an infinite summation of the second-order exchange diagrams, which is necessary to eliminate the notorious one-electron self-interaction error in the RPA method [35, 37]. On the other hand, the MP2 method itself is exactly free from such self-interaction error, but completely ignores high-order perturbative diagrams. In consequence, a large RMSD of about 0.68 eV between MP2 and RPA cohesive energies is observed, which can be traced back to the dissimilarity of underlying physical models used in the two methods. Recently, much effort has been devoted to improve RPA and MP2 from different perspectives, but these approaches often exacerbate the computational complexity dramatically. Our linear regression analysis suggests that the difference between MP2 and RPA cohesive energies is also quite systematic, leading to a material-specific deviation  $\overline{\text{RMSD}}$  of only about 100 meV.

We also compared the results of different methods with experimental data that were corrected for thermal and zero-point vibrational effects [83, 107]. In-line with the observation from the above cross-over comparison, most of the errors in different theoretical approximations are quite systematic, leading to the material-specific errors  $\overline{\text{RMSD}}$ s in 9 methods are lower than 210 meV, with M06-L presenting the largest difference. The visualization of the linear regression analysis suggests that abnormally large errors in M06-L occur in metallic systems (Na and Al), which can be ascribed to an incorrect oscillation of the exchange enhancement factor of M06-L when approaching to the uniform electron gas limit [129]. Taking these two points out, the RMSD and  $\overline{\text{RMSD}}$  of M06-L are reduced to 131 meV and 95 meV, respectively.

To summarize this section, we have presented three examples that demonstrate the usage of our MSE web site. In the context of on-going innovation in computational materials science, driven by data technology, there is a growing awareness of the importance of effective data sharing and recycling. Here, we argue that a dedicated test-set web site should be about more than an easy, static access of the reference data. In order to liberate the power of test sets, of key importance is providing a friendly analysis interface that facilitate the users to play with the online data, repeat the observation in the original papers, and even gain new insights from the data by themselves.

## 4. Numerically well-converged reference data

Numerically well-converged reference data at various levels of theory are crucial to achieve an unbiased benchmark and discussion in the previous section. Due to the use of NAO basis sets, the electronic-structure problem is addressed using numerical integration in FHI-aims. The specific technical aspects, including the implementation of localized resolution-of-identity (called 'RI-LVL') method, are described in [24, 52, 121]. We have systematically examined the numerical convergence in terms of relevant numerical parameters, and also  $\mathbf{k}$ -mesh and basis set sizes. The results of mean-field approximations are included in the MSE web site (see section 3 for usage) and briefly summarized in the appendix.

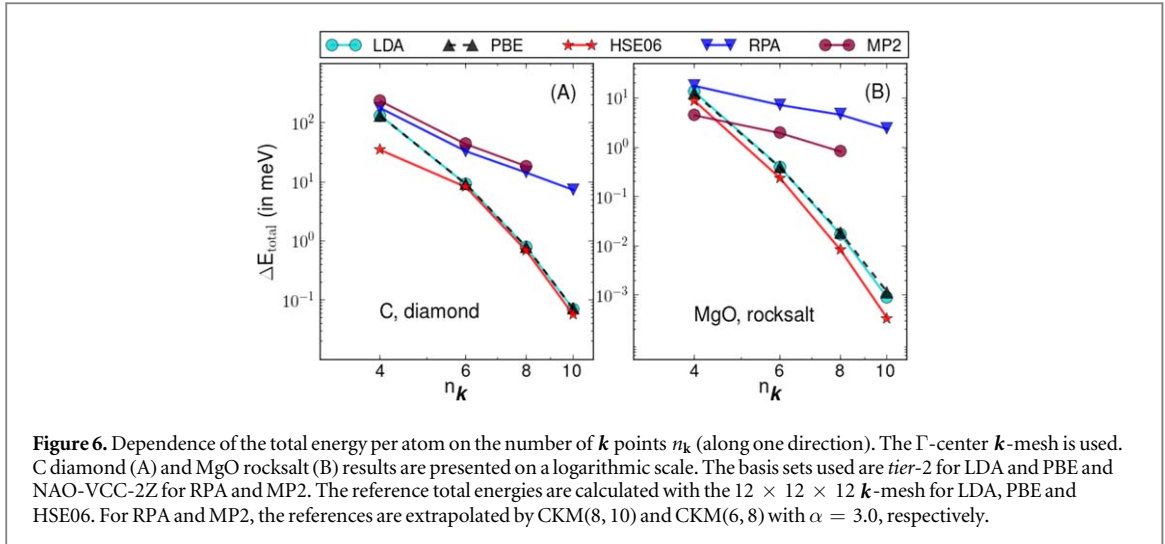
However, the periodic MP2 and RPA methods are state-of-the-art electronic-structure approaches for solid-state community, due to their greater expensive costs than mean-field approximations. Numerical convergence behavior is not yet fully examined for solids and thus, in this section, we introduce the numerical strategies applied to obtain the reference data of MP2 and RPA calculations in the MSE test set.

### 4.1. The $\mathbf{k}$ -mesh convergence

In the framework of the mean-field approximations, the evaluation of the XC energy needs only the electron density and/or occupied orbitals. However, the advanced correlation methods, i.e. MP2 and RPA, explicitly consider the excitations that are between occupied and virtual orbitals and that cross-over different  $\mathbf{k}$  points. Therefore, it is not surprising to see that MP2 and RPA calculations for solids have a much higher demand on the sampling of the first Brillouin zone. The  $\mathbf{k}$ -mesh convergence of the periodic MP2 method has been investigated with the  $\Gamma$ -center sampling strategy by Grüneis *et al* [76, 130]. In FHI-aims, we adopt a hybrid strategy to generate the Coulomb matrices in MP2 and RPA calculations, i.e. using the cut-Coulomb operator only for the  $\Gamma$  point and the full Coulomb operator for the rest of the  $\mathbf{k}$  points. For MP2 and RPA, this hybrid choice shows a faster  $\mathbf{k}$ -mesh convergence compared to the cut-Coulomb strategy used in HSE06 calculations [105].

Taking C diamond and MgO rocksalt as examples, figure 6 shows the  $\mathbf{k}$ -mesh incompleteness error of total energies per atom for five methods:

$$\Delta E_{\text{total}}[n_{\mathbf{k}}] = E_{\text{total}}[n_{\mathbf{k}}] - E_{\text{total}}[\infty], \quad (4)$$



**Figure 6.** Dependence of the total energy per atom on the number of  $\mathbf{k}$  points  $n_{\mathbf{k}}$  (along one direction). The  $\Gamma$ -center  $\mathbf{k}$ -mesh is used. C diamond (A) and MgO rocksalt (B) results are presented on a logarithmic scale. The basis sets used are *tier-2* for LDA and PBE and NAO-VCC-2Z for RPA and MP2. The reference total energies are calculated with the  $12 \times 12 \times 12$   $\mathbf{k}$ -mesh for LDA, PBE and HSE06. For RPA and MP2, the references are extrapolated by CKM(8, 10) and CKM(6, 8) with  $\alpha = 3.0$ , respectively.

where  $n_{\mathbf{k}}$  is the number of  $\mathbf{k}$  points in each direction ( $n_{\mathbf{k}} \times n_{\mathbf{k}} \times n_{\mathbf{k}}$ ). Our results reveals that the total energies per atom calculated at MP2 and RPA levels converge dramatically slower than those of LDA, PBE and HSE06. The  $\mathbf{k}$ -mesh incompleteness error in C diamond remains to be 15 meV for both MP2 and RPA total energies with a dense  $8 \times 8 \times 8$   $\mathbf{k}$ -mesh. We notice that other off-set evenly sampled  $\mathbf{k}$  meshes [96, 97], the so-called mean-value point strategy [98], and smearing sampling methods, such as the method of Methfessel and Paxton [99], the improved linear tetrahedron method [100], etc., have all been proposed to provide a better sampling with fewer (or even one)  $\mathbf{k}$  points, but at the semi-local DFT level of theory only. It is interesting to investigate the influence of these sampling strategies on advanced correlation methods.

The computational expense of the canonical MP2 (and RPA) scales as  $N_{\mathbf{k}}^3$  (and  $N_{\mathbf{k}}^2$ ), where  $N_{\mathbf{k}} = n_{\mathbf{k}}^3$  is the total grid number in a given  $\mathbf{k}$  mesh. For C diamond and NAO-VCC-2Z basis set (with 28 basis functions per unit cell), the MP2 calculation with the  $8 \times 8 \times 8$   $\mathbf{k}$ -mesh requires 10 hours to complete using 320 CPU cores of an Infiniband-connected Intel cluster with Intel Xeon E5-2680 v2 cores (2.8 GHz, 20 cores per node). The periodic MP2 implementation in FHI-aims is a  $\mathbf{k}$ -mesh-oriented parallelization in the framework of the message passing interface, guaranteeing efficient MP2 calculations with thousands of cores. However, with the optimistic assumption of a perfect parallel scalability, it would need over 2000 CPU cores to finish the MP2 calculation with a denser  $\mathbf{k}$ -mesh ( $10 \times 10 \times 10$ ) in a comparable amount of time. While the RPA calculations with such a dense  $\mathbf{k}$ -mesh and the NAO-VCC-2Z basis set can be carried out at a reasonable cost, the remaining  $\mathbf{k}$ -mesh error is about 7 meV for C diamond. The RPA calculations with denser  $\mathbf{k}$ -meshes and larger basis sets become infeasible as well.

A practical way to approach the complete  $\mathbf{k}$ -mesh limit (CKM) for advanced correlation methods is the two-point extrapolation in terms of an inverse relation between  $\Delta E_{\text{total}}[n_{\mathbf{k}}]$  and  $n_{\mathbf{k}}$ :

$$\begin{aligned} \Delta E_{\text{total}}[n_{\mathbf{k}}] &= A/n_{\mathbf{k}}^{\alpha} \\ \log(\Delta E_{\text{total}}[n_{\mathbf{k}}]) &= \log(A) - \alpha \log(n_{\mathbf{k}}) \end{aligned} \quad (5)$$

The exponential ‘ $\alpha$ ’ determines the speed of the  $\mathbf{k}$ -mesh convergence of different methods. On a logarithmic scale (see equation (5) and figure 6), ‘ $\alpha$ ’ is the negative value of the slope of  $\Delta E_{\text{total}}[n_{\mathbf{k}}]$  towards the CKM limit. Previously, the exponential  $\alpha = 3.00$  has been successfully used to extrapolated MP2 and CCSD(T) cohesive energies for several materials [85].

Table 3 shows the performance of such  $\mathbf{k}$ -mesh extrapolation for five selected materials (two elemental and three binary solids) from the MSE test set. The complete  $\mathbf{k}$ -mesh extrapolation from  $8 \times 8 \times 8$  and  $10 \times 10 \times 10$ , CKM(8, 10), are taken as the reference for the periodic RPA calculations. The deviations of the CKM(6, 8) values are smaller than 1.5 meV for all five materials and the MAD is only 1.1 meV. Our results confirm the validity of the practical extrapolation (equation (5)) with  $\alpha = 3.00$  for advanced methods provided that reasonable dense  $\mathbf{k}$ -meshes are utilized. However, a notable decrease of the performance can be observed in the combination of CKM(4, 6) and  $\alpha = 3.00$ . The MAD remains at about 8.0 meV (with the maximum error of 14 meV for C diamond) for the RPA total energies per atom. In this project, we optimize ‘ $\alpha$ ’ for the CKM(4, 6) extrapolation, and find that  $\alpha = 3.95$  improves the CKM(4, 6) extrapolation consistently, resulting in the MAD of only 1.6 meV.

The CKM(6, 8) values with  $\alpha = 3.00$  are chosen as the reference for MP2. Table 3 indicates that the CKM(4, 6) extrapolation is not good enough when using the default choice of  $\alpha = 3.00$ , and it can be effectively improved with an exponential value  $\alpha = 3.95$  as optimized for RPA. Our results reveal that advanced

**Table 3.** Errors of extrapolated RPA and MP2 total energies per atom for five materials (in meV). The complete  $\mathbf{k}$ -mesh (CKM) extrapolation is performed from the different combinations of  $\mathbf{k}$  grids, namely CKM( $n_{k_1}, n_{k_2}$ ), and with different exponentials  $\alpha$ . The mean absolute errors (MAE) listed are the deviations of  $E_{\text{total}} - E_{\text{total}}^{\text{Ref}}$ , where the reference  $E_{\text{total}}^{\text{Ref}}$  for RPA is the extrapolated value from  $8 \times 8 \times 8$  and  $10 \times 10 \times 10$ , CKM(8, 10), with  $\alpha = 3.00$ . For MP2, the reference is extrapolated by CKM(6, 8) and  $\alpha = 3.00$ . The mean absolute deviations (MADs) are shown for RPA and MP2 separately.

	Index in MSE	CKM(4, 6) $\alpha = 3.00$	CKM(4, 6) $\alpha = 3.95$	CKM(6, 8) $\alpha = 3.00$
RPA				
C	06	-13.8	-1.6	0.4
Si	07	-12.4	2.3	-1.3
MgO	12	1.5	2.4	1.4
BN	15	-5.7	1.0	1.3
AlP	18	-6.9	1.1	1.1
MAD		8.0	1.6	1.1
MP2				
C	06	-17.9	-1.9	
Si	07	-17.1	1.6	
MgO	12	0.5	0.7	
BN	15	-9.2	1.6	
AlP	18	-9.5	-0.2	
MAD		11.3	1.1	

correlation methods share a similar  $\mathbf{k}$ -mesh convergence behavior, and thus the CKM(6, 8) values with  $\alpha = 3.00$  are good enough to converge the error within 2.0 meV for insulators and semiconductors, which is chosen for both MP2 and RPA calculations in this work (see table 1 in supporting information).

We notice that the periodic MP2 and CCSD(T) results reported in the literature were often extrapolated using CKM(5, 6) or even sparser CKM(3, 4) with  $\alpha = 3.00$  [85]. It is an understandable compromise considering the computational expense with denser  $\mathbf{k}$ -meshes and larger basis sets for these methods; however, in order to achieve  $\mathbf{k}$ -mesh-converged values with the numerical uncertainty within 10 meV, an adapted exponential factor for CKM extrapolation appears necessary from our benchmarking.

#### 4.2. The basis set convergence

The NAO basis sets of NAO-VCC- $nZ$ , used in FHI-aims, hold the promise to provide numerically well-converged total energies for advanced correlation methods in frozen-core approximations. (see [24] for details). For MP2 and RPA calculations, the complete basis set (CBS) results are extrapolated using a two-point extrapolation formula [24, 60] based on NAO-VCC-3Z and NAO-VCC-4Z, namely CBS(3, 4):

$$E[\infty] = \frac{E[3]3^3 - E[4]4^3}{3^3 - 4^3} \quad (6)$$

Here  $E[n]$  is the RPA or MP2 total energy per atom using NAO-VCC- $nZ$  basis sets (with  $n = 3$  and 4). While the extrapolation from NAO-VCC-4Z and NAO-VCC-5Z shall further reduce the numerical uncertainty of the reference data [24], NAO-VCC-5Z is too expensive to be used for periodic MP2 and RPA calculations at present.

The  $\mathbf{k}$ -mesh convergence of MP2 and RPA methods tested in the previous section is carried out together with the smallest NAO-VCC-2Z basis sets (see figure 6). However, it would be impossible to perform MP2 calculations with an  $8 \times 8 \times 8$   $\mathbf{k}$ -mesh for NAO-VCC-4Z, and sometimes even a NAO-VCC-3Z calculation is too expensive, because of the unfavorable scaling of the MP2 method with respect to the number of  $\mathbf{k}$  points.

Taking C diamond as an example, table 4 shows the  $\mathbf{k}$ -mesh convergence of the MP2 total energies per atom for different basis sets. The most time consuming result in this table is the combination of NAO-VCC-3Z and the  $8 \times 8 \times 8$   $\mathbf{k}$ -mesh, which requires about 4 days using 320 CPU cores of an Infiniband-connected Intel cluster with Intel Xeon E5-2680 v2 cores (2.8 GHz, 20 cores per node). As illustrated in table 4, while the error of a  $4 \times 4 \times 4$   $\mathbf{k}$ -mesh varies with the basis set employed by several meV, it becomes almost independent of the choice of basis sets for  $6 \times 6 \times 6$  and  $8 \times 8 \times 8$  meshes. Ohnishi *et al* [131] and Grüneis *et al* [132] had a similar observation in MP2 and CCSD calculations using the GTO-type and plane-wave basis sets, respectively. In consequence, they concluded that the long-range behavior of the correlation energy depends mostly on the



**Table 4.** Errors in the MP2 total energies per atom using different basis sets for C diamond (in meV). The reference is the CKM(6, 8) extrapolated values with  $\alpha = 3.00$  (see equation (5)).

$k$ -mesh	NAO-VCC-2Z	NAO-VCC-3Z	NAO-VCC-4Z
$(4 \times 4 \times 4)$	117	114	115
$(6 \times 6 \times 6)$	22	22	22
$(8 \times 8 \times 8)$	9	9	9 <sup>a</sup>

<sup>a</sup> For NAO-VCC-4Z, the MP2 total energy per atom with the  $8 \times 8 \times 8$   $k$ -mesh is extrapolated using the energy difference between  $k$ -grids  $6 \times 6 \times 6$  and  $8 \times 8 \times 8$  with the NAO-VCC-3Z basis set.

low-lying excitations, and proposed their progressive down-sampling technique to approach the complete  $k$ -mesh and basis set limit for advanced correlation methods.

In this work, we adopt a similar down-sampling concept to estimate the MP2 total energies per atom with the NAO-VCC-4Z basis set and the  $8 \times 8 \times 8$   $k$ -grid:  $E_{\text{total}}^{4Z}[n_k = 8]$

$$E_{\text{total}}^{4Z}[n_k = 8] = E_{\text{total}}^{4Z}[n_k = 6] + \Delta E_{\text{total}}^{nZ}[n_k = 8] \quad (7)$$

using the energy change between  $6 \times 6 \times 6$  and  $8 \times 8 \times 8$   $k$ -meshes with the NAO-VCC-3Z (or 2Z) basis set  $\Delta E_{\text{total}}^{nZ}[n_k = 8]$  ( $n = 3$  or  $2$ ) of:

$$\Delta E_{\text{total}}^{nZ}[n_k = 8] = E_{\text{total}}^{nZ}[n_k = 8] - E_{\text{total}}^{nZ}[n_k = 6] \quad (8)$$

In summary, the reference data for the MP2 and RPA total energies (and also valence correlation energies) per atom are obtained by the combination strategy of CKM(6, 8), CBS(3, 4) and  $\Delta E_{\text{total}}^{nZ}[n_k = 8]$  with ( $n = 3$  or  $2$ ). The error bar is estimated to be 20 meV for these MP2 and RPA reference data. The numerical uncertainty in the combination strategy is dominated by the CBS(3, 4) extrapolation, i.e. 15 meV for CBS(3, 4), 2 meV for CKM(6, 8), and less than 1 meV for  $\Delta E_{\text{total}}^{3Z}[n_k = 8]$  or  $\Delta E_{\text{total}}^{2Z}[n_k = 8]$ , respectively.

#### 4.3. Lattice constant, bulk modulus, and cohesive energy

Based on the efforts discussed above to achieve the complete  $k$ -mesh and basis set convergence in terms of absolute total energy for various levels of theory, we next focus on the numerical convergence of MP2 and RPA calculations on lattice constants, bulk moduli and cohesive energies.

The post-processing MP2 and RPA correlations are frozen-core, but evaluated based on the Hartree–Fock, PBE and/or PBE0 orbitals in *all-electron* description. The basis set convergence of these properties are well-documented for (semi-) local and hybrid DFAs. However, the slow basis set convergence together with an unfavorable computational scaling makes it a big challenge to perform a systematic investigation of the basis set convergence of these materials' properties at MP2 and RPA levels. We present here the MP2 results of five selected materials from the MSE test sets in table 5. To the best of our knowledge, it is the first report of the basis set convergence of MP2 for condensed matter systems using NAO basis sets. The readers can easily access the MP2 and RPA data on convergence test for all 13 materials in our MSE web site.

Table 5 reveals that the basis set convergence is relatively fast in the MP2 calculations of lattice constants. NAO-VCC-3Z is good enough to provide a well-converged lattice constant (the convergence criterion is 0.01 Å) for all five materials (for details see also the first example in Section 3). In agreement with the previous investigations using plane-wave basis sets [76, 85] or a Gaussian and plane waves hybrid approach [117], the slow basis set convergence of MP2 cohesive energies is observed with NAO basis sets as well. Compared with CBS(3, 4) values, the basis set incompleteness errors at NAO-VCC-4Z still remain about 60–130 meV. Such slow convergence arises from the slower basis set convergence of the MP2 total energies in bulks than that in free atoms. It is well-documented in quantum chemistry that the accurate geometry information for the MP2 and other advanced correlation methods can be obtained with the basis sets of triple-zeta quality [133], but the converged atomization energy or other energy differences cannot be achieved with finite basis sets [41, 134, 135]. Our results confirm that this conclusion is also valid for solids. While the bulk moduli obtained at the NAO-VCC-4Z basis set are not well-converged, the basis set error remains about 3–7 GPa.

Inspecting table 5 also reveals the capability of NAO-VCC- $nZ$  to provide a consistently improvable description of cohesive properties. In general, the calculated MP2 cohesive energies increase with the cardinal index ( $n$ ) of NAO-VCC- $nZ$ . On the one hand, it allows for an accurate extrapolation to the CBS limit; and, on the other hand, the NAO-VCC-4Z values set up a quite rigorous lower bound for the converged values of MP2. In other words, any results smaller than the NAO-VCC-4Z values might be unreliable.

**Table 5.** MP2 cohesive energies per atom, lattice constants, and bulk moduli. All results have been extrapolated to the complete  $k$ -mesh limit using CKM(6, 8). The VASP results are listed as well.

	Basis set	$E_{coh}$ (eV)	$a_0$ (Å)	$B_0$ (GPa)
C	NAO-VCC-2Z	7.65	3.56	451
	NAO-VCC-3Z	7.81	3.55	454
	NAO-VCC-4Z	7.96	3.54	454
	CBS(3, 4)	8.08	3.54	454
	VASP <sup>a</sup>	7.97	3.55	450
	VASP <sup>b</sup>	8.04		
Si	NAO-VCC-2Z	4.62	5.44	98
	NAO-VCC-3Z	4.92	5.41	100
	NAO-VCC-4Z	5.07	5.41	100
	CBS(3, 4)	5.21	5.40	100
	VASP <sup>a</sup>	5.05	5.42	100
	VASP <sup>b</sup>			
BN	NAO-VCC-2Z	6.84	3.62	387
	NAO-VCC-3Z	7.01	3.59	392
	NAO-VCC-4Z	7.14	3.59	375
	CBS(3, 4)	7.25	3.58	368
	VASP <sup>a</sup>	7.12	3.61	395
	VASP <sup>b</sup>	7.15		
MgO	NAO-VCC-2Z	5.11	4.23	160
	NAO-VCC-3Z	5.39	4.23	156
	NAO-VCC-4Z	5.47	4.24	156
	CBS(3, 4)	5.53	4.24	156
	VASP <sup>a</sup>	5.35	4.23	153
	VASP <sup>b</sup>			
AlP	NAO-VCC-2Z	4.06	5.48	92
	NAO-VCC-3Z	4.41	5.45	94
	NAO-VCC-4Z	4.55	5.45	95
	CBS(3, 4)	4.67	5.45	95
	VASP <sup>a</sup>	4.32	5.46	93
	VASP <sup>b</sup>	4.63		

<sup>a</sup> In [76], the MP2 results with VASP were extrapolated to the CBS limit. The  $k$ -mesh was sampled by a composite scheme. The Hartree-Fock orbitals were generated in the pseudo-potential framework.

<sup>b</sup> In [85], the MP2 results with VASP were extrapolated by CKM(5, 6) with  $\alpha = 3$ . The Hartree-Fock orbitals were generated in the pseudo-potential framework.

The canonical MP2 method was implemented in VASP for periodic systems in 2009 [130]. Later, Grüneis *et al* calculated the MP2 cohesive properties of 13 materials in 2010 [76]. To approach the CBS limit in the framework of the projector-augmented-wave (PAW) method using plane-wave basis sets, they proposed an inverse relation between the MP2 correlation energy and the energy cutoff  $E_\chi$  that represents the overlap charge densities for the CBS extrapolation. Meanwhile, an efficient composite scheme was proposed to reduce the computational cost due to the unfavorable scaling of the  $k$ -point number. To be specific, the Hartree-Fock total energy is calculated with a dense  $10 \times 10 \times 10$   $\Gamma$ -centered  $k$ -mesh, but the MP2 second-order direct and exchange terms are evaluated with  $6 \times 6 \times 6$  and  $3 \times 3 \times 3$   $k$ -meshes, respectively. In 2013, Booth *et al* updated the MP2 cohesive energies of three materials [85], including C diamond, BN and AlP with zincblende structure. The main difference is to replace the composite scheme by an extrapolation scheme from  $5 \times 5 \times 5$  and  $6 \times 6 \times 6$ , i.e. CKM(5, 6) with  $\alpha = 3$ , to approach the converged  $k$ -mesh sampling. Table 5 also lists the canonical MP2 cohesive properties calculated with VASP. FHI-aims predicts very similar cohesive energies to the VASP values with the  $k$ -grid extrapolation scheme. The difference is about 40 meV for C diamond and AlP zincblende, and 100 meV for BN in zincblende. For MgO and Si, the discrepancy of over 170 meV between FHI-aims and VASP values shall be ascribed to the  $k$ -grid incompleteness of the composite scheme used in the previous VASP calculations [76].

## 5. Conclusions

We have accomplished the first step towards a representative material science and engineering (MSE) test set with well-defined and relevant cohesive and electronic properties, and reference values, obtained for a range of main-group materials using a hierarchy of the first-principles calculations. The accuracy of the reference values in the MSE test set is mainly determined by the well-defined numerical setting of each applied DFA. A strong effort has been made to provide numerically converged values from state-of-the-art theory, including periodic MP2, and RPA calculations in the CBS limit and the complete  $\mathbf{k}$ -space limit as well. At present, we provide 14 accurate data for MP2 and RPA.

A web site of the MSE test set, <http://mse.fhi-berlin.mpg.de>, is equipped with a multi-mode access framework, versatile visualization, and a linear regression tool for post-processing data analysis. We have demonstrated that these features dramatically assist the post-processing data analysis that is necessary to detect the numerical error in calculations and uncover the intrinsic limitations of DFAs. The presented paradigm for the test set construction is applicable to any new material and materials' property.

## Acknowledgments

The authors thank Fawzi Mohamed for his support in setting up the MSE web page, Steffen Kangowski for creating the layout of the MSE web page, and Norina A. Richter and Toktam Morshedloo for their contribution in the preparation stage of the MSE project. Work at Shanghai was supported by the 14th Recruitment Program of Young Professionals for IYZ. AJL is grateful to the Royal Society of Chemistry for the award of a Researcher Mobility Fellowship, and acknowledges the use of computing facilities provided by ARCCA at Cardiff University, HPC Wales, and the ARCHER high performance computing facilities, via membership of the UK HPC Materials Chemistry Consortium (EP/L000202). XR acknowledges the support from Chinese National Science Foundation (Grant number 11574283). This work received funding from the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement No. 676580, the NOMAD Laboratory CoE and No. 740233, ERC: TEC1P.

## Appendix. Numerically well-converged reference data for PBE and HSE06

### A.1. The $\mathbf{k}$ -mesh convergence

The  $\mathbf{k}$ -mesh convergence for LDA and PBE methods is investigated together with  $\Gamma$ -center (GC) and Monkhorst-Pack (MP) sampling strategies [136–138], both of which span the first Brillouin zone in an evenly spaced manner. The MP  $\mathbf{k}$ -mesh coincides with the GC one for an uneven number of grid points [137, 138]. In this project, we choose a set of  $\mathbf{k}$ -meshes with even numbers of grid points. Our observation is consistent with previous investigations where the MP  $\mathbf{k}$ -mesh shows faster convergence behavior than the GC  $\mathbf{k}$ -mesh for insulators and semiconductors at the semilocal density functional level of theory [137, 138]. In order to achieve 1 meV accuracy, the  $4 \times 4 \times 4$  MP grid is sufficient for rare-gas crystals (Ne and Ar with fcc structure), and the  $6 \times 6 \times 6$  MP grid for other ionic- and covalent-bond systems. To achieve the same accuracy, an  $8 \times 8 \times 8$  (sometimes even denser) grid is often necessary with the GC  $\mathbf{k}$ -mesh. However, the slow  $\mathbf{k}$ -mesh convergence in the calculation of metallic systems is more serious, and cannot be improved using the MP  $\mathbf{k}$ -mesh. Based on LDA, PBE, and three meta-GGAs convergence benchmarks, we summarize the  $\mathbf{k}$ -mesh setting in table A1, recommended for all semi-local DFAs investigated in this paper: LDA, PBE, PBEsol, TPSS, M06-L and SCAN.

A linear-scaling HSE06 method for condensed matter systems has been implemented in FHI-aims [105]. In order to fully utilize the sparsity of the density matrix in real space, this HSE06 implementation needs a projection of density matrix in a  $\Gamma$ -center  $\mathbf{k}$ -mesh to the corresponding Born-von-Karman (BvK) supercell by means of Fourier transformation. Furthermore, FHI-aims features a massively parallel, in-memory implementation of canonical MP2 and RPA methods for periodic systems. At present, HSE06, MP2 and RPA calculations can be performed only with the  $\Gamma$ -center  $\mathbf{k}$ -mesh in FHI-aims. MP2 and RPA calculations have been reported using  $\mathbf{k}$ -meshes centered at  $\Gamma$ -point for the PAW method and plane-wave basis sets [76, 77, 85, 87, 88, 132].

Table A1 lists the  $\mathbf{k}$ -meshes for converged HSE06 calculations: generally speaking, HSE06 shows a similar convergence behavior as LDA and PBE using the  $\Gamma$ -center  $\mathbf{k}$ -mesh. Notable differences happen only with very rough grids, e.g. in the rare-gas crystals (Ne and Ar) with  $2 \times 2 \times 2$  and ionic crystals with  $4 \times 4 \times 4$ . Such abnormality can be ascribed to an improper description of the integrable singularity of the Coulomb potential in reciprocal space [72, 139], which is mathematically equivalent to a slow decay of Coulomb interaction in condensed matter systems. In FHI-aims, a cut-Coulomb operator is used for HSE06 to generate the Coulomb matrices at every  $\mathbf{k}$ -point [105, 112]. However, a reasonably dense  $\mathbf{k}$ -mesh is still required. Taking C diamond

**Table A1.** The  $\mathbf{k}$ -meshes that are used for different methods. ‘GC’ denotes an evenly spaced  $\Gamma$ -center  $\mathbf{k}$ -mesh and ‘MP’ for the Monkhorst-Pack one. For mean-field approximations, the convergence criteria is 1 meV (or tighter) in the calculations of the total energy per atom. For MP2 and RPA, the complete  $\mathbf{k}$ -mesh extrapolation from  $6 \times 6 \times 6$  and  $8 \times 8 \times 8$ , denoted as CKM(6, 8) is adopted with  $\alpha = 3$  (see equation (5) in the main text).

	Structure	(semi-)local DFAs		Hybrid DFAs		MP2 and RPA	
		Type	$\mathbf{k}$ -points	Type	$\mathbf{k}$ -points	Type	$\mathbf{k}$ -points
Li	bcc	GC	$16 \times 16 \times 16$	GC	$16 \times 16 \times 16$	—	—
Na	bcc	GC	$16 \times 16 \times 16$	GC	$16 \times 16 \times 16$	—	—
Al	fcc	GC	$20 \times 20 \times 20$	GC	$20 \times 20 \times 20$	—	—
Ne	fcc	MP	$4 \times 4 \times 4$	GC	$6 \times 6 \times 6$	GC	CKM(6, 8)
Ar	fcc	MP	$4 \times 4 \times 4$	GC	$6 \times 6 \times 6$	GC	CKM(6, 8)
C	diamond	MP	$6 \times 6 \times 6$	GC	$10 \times 10 \times 10$	GC	CKM(6, 8)
Si	diamond	MP	$6 \times 6 \times 6$	GC	$10 \times 10 \times 10$	GC	CKM(6, 8)
LiF	rocksalt	MP	$6 \times 6 \times 6$	GC	$6 \times 6 \times 6$	GC	CKM(6, 8)
LiCl	rocksalt	MP	$6 \times 6 \times 6$	GC	$6 \times 6 \times 6$	GC	CKM(6, 8)
NaF	rocksalt	MP	$6 \times 6 \times 6$	GC	$6 \times 6 \times 6$	—	—
NaCl	rocksalt	MP	$6 \times 6 \times 6$	GC	$6 \times 6 \times 6$	—	—
MgO	rocksalt	MP	$6 \times 6 \times 6$	GC	$6 \times 6 \times 6$	GC	CKM(6, 8)
MgS	rocksalt	MP	$6 \times 6 \times 6$	GC	$8 \times 8 \times 8$	GC	CKM(6, 8)
BeS	zinblende	MP	$6 \times 6 \times 6$	GC	$8 \times 8 \times 8$	GC	CKM(6, 8)
BN	zinblende	MP	$6 \times 6 \times 6$	GC	$8 \times 8 \times 8$	GC	CKM(6, 8)
BP	zinblende	MP	$6 \times 6 \times 6$	GC	$10 \times 10 \times 10$	GC	CKM(6, 8)
SiC	zinblende	MP	$6 \times 6 \times 6$	GC	$10 \times 10 \times 10$	GC	CKM(6, 8)
AlP	zinblende	MP	$6 \times 6 \times 6$	GC	$8 \times 8 \times 8$	GC	CKM(6, 8)
LiH	rocksalt	MP	$6 \times 6 \times 6$	GC	$8 \times 8 \times 8$	GC	CKM(6, 8)

and MgO rocksalt as examples, figure 6 in the main text shows the  $\mathbf{k}$ -mesh incompleteness error of total energies per atom for five methods:

$$\Delta E_{\text{total}}[n_{\mathbf{k}}] = E_{\text{total}}[n_{\mathbf{k}}] - E_{\text{total}}[\infty], \quad (\text{A.1})$$

where  $n_{\mathbf{k}}$  is the number of  $\mathbf{k}$  points in each direction ( $n_{\mathbf{k}} \times n_{\mathbf{k}} \times n_{\mathbf{k}}$ ). Our results reveal that a  $\Gamma$ -center mesh with  $6 \times 6 \times 6$  grid points is enough for HSE06 to address this singularity issue, thus providing a similar  $\mathbf{k}$ -mesh sampling quality as LDA and PBE.

## A.2. The basis set convergence

The difficulty of using the GTO basis sets, the *de facto* choice in quantum chemistry, to extended systems has been studied extensively in different kinds of crystals mainly using the CRYSTAL program [101, 140, 141]. To quote the statement in their User’s Manual (Chapter 10) [102], ‘as a rule, extended atomic basis sets, or ‘triple-zeta’ type basis sets should be avoided (...), because the outer functions are too diffuse.’ On the other hand, despite the predominate use in calculations on periodic systems, the plane-wave basis set is ineffective to describe the localized core-electron states and usually needs to be used in conjunction with pseudo-potentials.

The NAO basis sets of FHI-aims [24, 52], including *tier-n* and NAO-VCC-*nZ*, hold the promise to provide numerically well-converged total energies for mean-field approximations in *all-electron* description, as the minimal basis of both *tier-n* and NAO-VCC-*nZ* is composed of the exact core and valence orbitals of spherically symmetric free atoms. In the meantime, a confining potential can be used to generate the primitive NAOs that are exactly localized in a certain region, so that any numerical instability caused by an unnecessary overlap between the diffuse atom-centered basis functions can be significantly suppressed (see [24, 52] for details).

Table A2 lists the basis set incompleteness errors in HSE06 total energies per atom using the *tier-n* series, which are the default choice for mean-field approximations in FHI-aims. The *tier-2* basis set is recommended, and *tier-3* is considered to be good enough to reach the CBS limit [52]. While the *tier-n* basis sets are optimized with respect to small molecules, i.e. symmetric dimers for all elements [52], our results demonstrate their transferability to the extended systems with diverse chemical environments. The average (maximum) error in *tier-2* and *tier-3* are about 8 meV (16 meV) and 2 meV (9 meV) in HSE06 total energies per atom of light elements’ materials. Note that the basis size of *tier-2* is slightly larger than Dunning’s GTO triple-zeta basis set cc-pVTZ [59], which has been recommended to be avoided for calculations on solids [102]. In our work, we do not encounter any self-consistent field (SCF) convergence problem with *tier-2* in all our calculations with GGAs and hybrid GGAs, including LDA, PBE, PBEsol, and HSE06. For meta-GGAs, the SCF convergence can be achieved with *tier-2* as well, but some code-specific numerical parameters must be set carefully. For instance, we should use a finer numerical integration grid for the vdW systems [142], and it is necessary to introduce a threshold for the kinetic-energy related variable  $\tau$ , which removes values that do not effect the total energy but do cause

**Table A2.** Basis set errors in HSE06 total energies per atom (in meV). The reference data are the lowest energies in two sequences of NAO basis sets, i.e. *tier-n* and NAO-VCC-*nZ*.

	<i>tier-n</i>				NAO-VCC- <i>nZ</i>			
	<i>n</i> = 1	<i>n</i> = 2	<i>n</i> = 3	<i>n</i> = 4	<i>n</i> = 2	<i>n</i> = 3	<i>n</i> = 4	<i>n</i> = 5
Li	17	4	0	—	55	18	1	0
Na	32	16	5	0 <sup>a</sup>	22	11	3	0
Al	16	4	0	—	50	13	1	—
Ne	13	11	0	—	16	6	4	1
Ar	16	11	9	0 <sup>a</sup>	16	7	2	0
C	46	2	0	0	67	23	2	—
Si	32	10	1	0	66	22	4	1
LiF	33	2	0	—	44	16	0	0 <sup>b</sup>
LiCl	23	7	2	—	42	15	2	0
NaF	23	7	2	—	44	14	2	0
NaCl	21	10	2	—	32	12	3	0
MgO	61	11	0	—	87	18	2	1 <sup>b</sup>
MgS	55	13	2	—	69	18	4	0
BeS	47	6	1	—	113	38	4	0
BN	53	2	0	—	115	17	1	—
BP	19	5	1	—	79	22	3	0 <sup>b</sup>
SiC	54	9	1	0	130	29	6	0 <sup>b</sup>
AlP	23	9	4	1	87	22	5	0
LiH	11	1	0	—	48	11	0	0 <sup>b</sup>
Average	31	8	2	—	62	18	3	—

<sup>a</sup> HSE06 results of Na bcc and Ar fcc are calculated using *tier-3+* which is *tier-3* plus 2 *s*-type and 2 *p*-type basis functions from NAO-VCC-3Z.

<sup>b</sup> For a binary crystal AB, a hybrid basis set with NAO-VCC-*n*<sub>1</sub>Z for A and NAO-VCC-*n*<sub>2</sub>Z for B is denoted as N(*n*<sub>1</sub>/*n*<sub>2</sub>)Z. These are N(4/5)Z for LiF, MgO and LiH rocksalts, BP zinblende, and N(5/4)Z for SiC.

singularities when calculating the potential in the NAO framework [143]. Both parameters have been tested extensively, with values chosen such that the errors these parameters introduce are in the sub-meV regime.

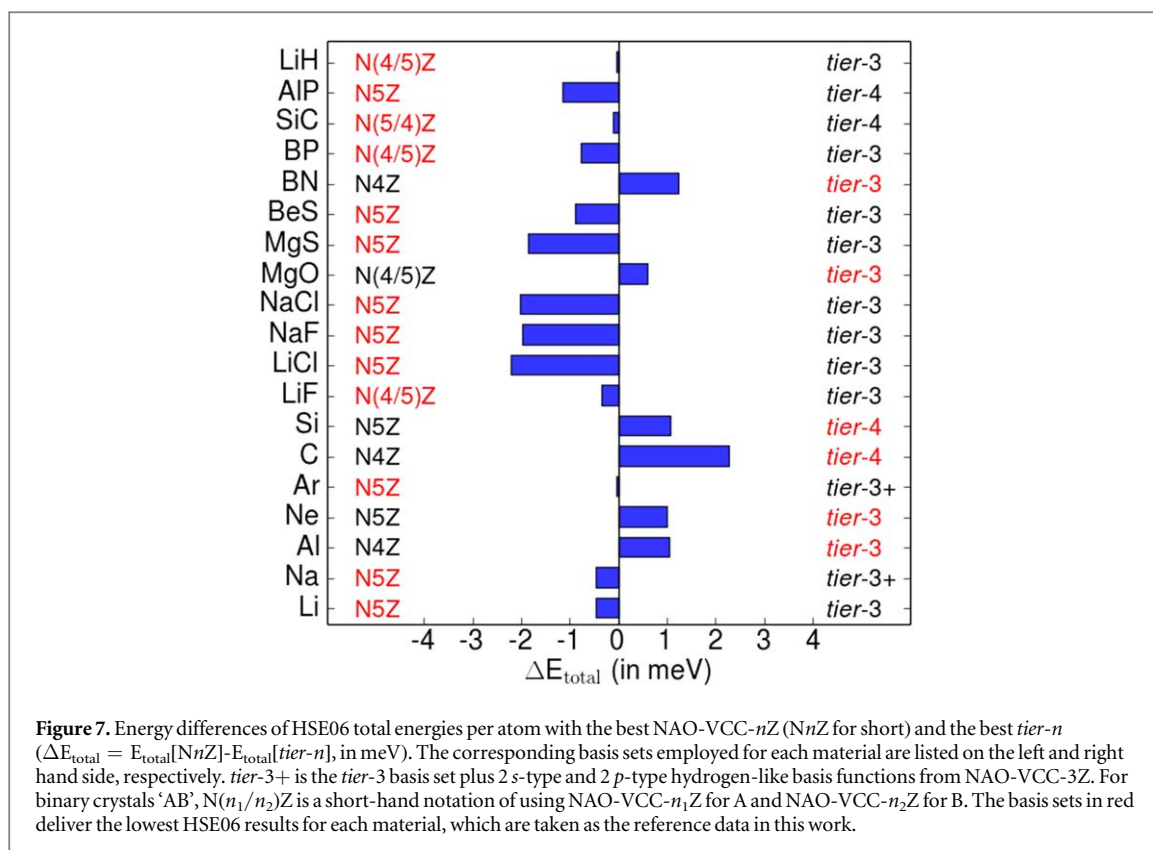
FHI-aims provides a larger *tier* basis set (*tier-4*) for some light elements. We notice that for the binaries with mainly ionic bonding characters (LiF—MgS), the SCF procedure with *tier-4* can fail. In this work, we report HSE06 total energies per atom using *tier-4* for C and Si with diamond structure, SiC and AlP with zinblende structure. Compared with *tier-3* results, a noticeable change can be observed only for AlP (about 3 meV). It suggests that *tier-3* is competent to provide numerically well-converged HSE06 total energies for most of materials. Concerning the two worst cases of *tier-3*, i.e. Na bcc and Ar fcc, the basis set incompleteness errors are about 5 meV and 9 meV, respectively. For Na and Ar, the *tier-4* basis sets are unavailable. We then introduce 2 *s*-type and 2 *p*-type hydrogen-like functions from NAO-VCC-3Z as an additional *s, p* group to the *tier-3*, denoted as *tier-3+*. Table A2 and figure 7 reveal that such *s, p* group effectively compensates for the basis set incompleteness in *tier-3*, resulting in the HSE06 total energies per atom with CBS quality for Na bcc and Ar fcc.

The NAO basis functions used in the *tier-n* series are apt to saturate the bonding region in the middle of two atoms, as they are optimized to minimize the LDA total energies of symmetric dimers [52]. In contrast, the sequence of NAO-VCC-*nZ* basis sets is determined by minimizing the RPA total energies of atoms [24], thus introducing more compact NAO basis functions than the *tier-n* series. With similar basis sizes, NAO-VCC-3Z delivers a larger basis set incompleteness error (18 meV on average) than *tier-2*. It indicates the necessity of introducing diffuse atom-centered basis functions in small basis sets to balance the aforementioned incompleteness in core and bonding regions, which renders a better performance for mean-field approximations. However, such discrepancy can be reduced with the increase of the index '*n*' in both sequences, the average error being only 2 meV and 3 meV for *tier-3* and NAO-VCC-4Z with similar basis sizes.

NAO-VCC-5Z is the largest NAO basis set in FHI-aims, and yields the lowest HSE06 total energies per atom for most of the materials in the test set. However, SCF convergence problem occurs at Al fcc, C diamond, LiF and MgO with rocksalt structure, BN, BP and SiC with zinblende structure. For binary crystals 'AB', we tried a hybrid basis set strategy N(*n*<sub>1</sub>/*n*<sub>2</sub>)Z, which is a short-hand notation of using NAO-VCC-*n*<sub>1</sub>Z for 'A' and NAO-VCC-*n*<sub>2</sub>Z for 'B'. Compared with NAO-VCC-4Z results, better HSE06 total energies per atom can be obtained without SCF convergence problem for LiF, MgO, BP and LiH with N(4/5)Z and for SiC with N(5/4)Z.

Figure 7 presents the differences of the HSE06 total energies per atom ( $\Delta E_{\text{total}}$ ) between the best NAO-VCC-*nZ* and *tier-n* basis sets. The discrepancy  $\Delta E_{\text{total}}$  is very small, about 1 meV on average. Considering that NAO-





VCC- $nZ$  and  $tier-n$  are designed for completely different purposes, such an agreement with each other clearly indicates that the CBS limit has been approached for the calculations of HSE06 total energies per atom in *all-electron* description.

Other (semi-)local DFAs share a similar basis set convergence as HSE06. The observation and discussion here are then transferable. In addition to HSE06, one can find a comprehensive basis set convergence test with LDA, PBE, TPSS, M06-L, and SCAN in the MSE web site. For PBEsol, the best NAO basis sets determined by PBE are used to calculate the reference data directly.

### A.3. Lattice constant, bulk modulus, and cohesive energy

To find the optimized lattice parameters for each material and each method, we calculate seven points within a range of  $\pm 5\%$  around the initial value of the lattice constant, and fit the respective (volume, energy) data points to the Birch-Murnaghan equation of state to obtain the optimized lattice constant. If all 7 lattice constants used to generate the equation of state lie within a range of  $\pm 7\%$  around the optimized value, the value is taken as the final, optimized lattice constant,  $a_0$ . Otherwise, this optimized lattice constant is used as a new starting point and more (volume, energy) data points are calculated, so that finally all materials' optimized lattice constants are obtained from an equation of state fitted to seven points with lattice constant values in a range of  $\pm 5\%$  to  $\pm 7\%$  around the final, optimized lattice constant  $a_0$ . We have carefully tested the stability of this procedure using the PBE functional. The obtained optimized lattice constants and bulk moduli are accurate within 0.02 Å and 0.1 GPa, respectively.

The cohesive energies  $E_{\text{coh}}^M$  of the materials are then calculated at the optimized lattice constant  $a_0$ :

$$E_{\text{coh}}^M = E^M / N_{\text{atom}} - \sum_{\text{atom}} E^{\text{atom}} / N_{\text{atom}}, \quad (\text{A.2})$$

where  $N_{\text{atom}}$  is the number of atoms in the unit cell,  $E^M$  the total energy of the unit cell for material  $M$ , and the sum is taken over the total energies  $E^{\text{atom}}$  of the constituent atoms in their spin-polarized symmetry-broken ground-state, i.e. no fractional occupancies.

## References

- [1] Curtiss LA, Redfern P C and Raghavachari K 2005 *J. Chem. Phys.* **123** 124107–12
- [2] Schultz N E, Zhao Y and Truhlar D G 2005 *J. Phys. Chem. A* **109** 4388–403
- [3] Schultz N E, Zhao Y and Truhlar D G 2005 *J. Phys. Chem. A* **109** 11127–43
- [4] Zhao Y, González-García N and Truhlar D G 2005 *J. Phys. Chem. A* **109** 2012–8

- [5] Zhao Y and Truhlar D G 2005 *J. Chem. Theory Comput.* **1** 415–32
- [6] Luo S, Averkiev B, Yang K R, Xu X and Truhlar D G 2014 *J. Chem. Theory Comput.* **10** 102–21
- [7] Jurečka P, Šponer J, Černý J and Hobza P 2006 *Phys. Chem. Chem. Phys.* **8** 1985–93
- [8] Ržáč J, Riley K E and Hobza P 2011 *J. Chem. Theory Comput.* **7** 2427–38
- [9] Karton A, Tarnopolsky A, Lamère J-F, Schatz G C and Martin J M L 2008 *J. Phys. Chem. A* **112** 12868–86
- [10] Bryantsev V S, Diallo M S, van Duin A C T and Goddard W A 2009 *J. Chem. Theory Comput.* **5** 1016–26
- [11] Jiang W, DeYonker N J, Determan J J and Wilson A K 2012 *J. Phys. Chem. A* **116** 870–85
- [12] Wu J M and Xu X 2008 *J. Chem. Phys.* **129** 164103
- [13] Wu J M and Xu X 2009 *J. Comput. Chem.* **30** 1424–44
- [14] Goerigk L and Grimme S 2010 *J. Chem. Theory Comput.* **6** 107–26
- [15] Goerigk L and Grimme S 2011 *J. Chem. Theory Comput.* **7** 291–309
- [16] Jung Y, Sodt A, Gill P M W and Head-Gordon M 2005 *Proc. Natl Acad. Sci. USA* **102** 6692–7
- [17] Hohenstein E G and Sherrill C D 2010 *J. Chem. Phys.* **133** 104107
- [18] Ren X, Rinke P, Blum V, Wieferink J, Tkatchenko A, Sanfilippo A, Reuter K and Scheffler M 2012 *New J. Phys.* **14** 053020
- [19] Mahler A and Wilson A K 2013 *J. Chem. Theory Comput.* **9** 1402–7
- [20] Zhang I Y, Luo Y and Xu X 2010 *J. Chem. Phys.* **133** 104105
- [21] Papajak E and Truhlar D G 2012 *J. Chem. Phys.* **137** 064110
- [22] Goldey M and Head-Gordon M 2012 *J. Phys. Chem. Lett.* **3** 3592–8
- [23] Eshuis H and Furche F 2012 *J. Chem. Phys.* **136** 084105
- [24] Zhang I Y, Ren X, Rinke P, Blum V and Scheffler M 2013 *New J. Phys.* **15** 123033
- [25] Curtiss L A, Raghavachari K, Redfern P C and Pople J A 1997 *J. Chem. Phys.* **106** 1063–79
- [26] Curtiss L A, Raghavachari K, Redfern P C and Pople J A 2000 *J. Chem. Phys.* **112** 7374–83
- [27] Xu X and Goddard W A 2004 *Proc. Natl Acad. Sci. USA* **101** 2673–7
- [28] Xu X and Goddard W A 2004 *J. Chem. Phys.* **121** 4068–82
- [29] Grimme S 2006 *J. Chem. Phys.* **124** 034108–16
- [30] Wu J M and Xu X 2007 *J. Chem. Phys.* **127** 214105
- [31] Zhang Y, Xu X and Goddard W A 2009 *Proc. Natl Acad. Sci. USA* **106** 4963–8
- [32] Zhang I Y, Xu X, Jung Y and Goddard W A 2011 *Proc. Natl Acad. Sci. USA* **108** 19896–900
- [33] Chai J D and Head-Gordon M 2009 *J. Chem. Phys.* **131** 174105
- [34] Ren X, Tkatchenko A, Rinke P and Scheffler M 2011 *Phys. Rev. Lett.* **106** 153003–4
- [35] Ren X, Rinke P, Scuseria G E and Scheffler M 2013 *Phys. Rev. B* **88** 035120
- [36] van Aggelen H, Yang Y and Yang W 2013 *Phys. Rev. A* **88** 030501
- [37] Bates J E and Furche F 2013 *J. Chem. Phys.* **139** 171103
- [38] Zhang I Y and Xu X 2013 *J. Phys. Chem. Lett.* **4** 1669–75
- [39] <http://thch.uni-bonn.de/tc/index.php?section=downloads&subsection=GMTKN30&lang=english>, (Accessed: 25 May 2017)
- [40] <https://comp.chem.umn.edu/db2015/>, (Accessed: 25 May 2017).
- [41] Pople J A, Head-Gordon M, Fox D J, Raghavachari K and Curtiss L A 1989 *J. Chem. Phys.* **90** 5622–9
- [42] Curtiss L A, Raghavachari K, Trucks G W and Pople J A 1991 *J. Chem. Phys.* **94** 7221–30
- [43] Curtiss L A, Raghavachari K, Redfern P C, Rassolov V and Pople J A 1998 *J. Chem. Phys.* **109** 7764–76
- [44] Curtiss L A, Redfern P C and Raghavachari K 2007 *J. Chem. Phys.* **126** 084108–12
- [45] Sinanoğlu O 1962 *J. Chem. Phys.* **36** 706
- [46] Purvis G D and Bartlett R J 1982 *J. Chem. Phys.* **76** 1910–8
- [47] Raghavachari K, Trucks G W, Pople J A and Head-Gordon M 1989 *Chem. Phys. Lett.* **157** 479–83
- [48] Booth G H, Thom A J W and Alavi A 2009 *J. Chem. Phys.* **131** 054106
- [49] Cleland D, Booth G H and Alavi A 2010 *J. Chem. Phys.* **132** 041103
- [50] Cleland D M, Booth G H and Alavi A 2011 *J. Chem. Phys.* **134** 024112
- [51] Delley B 1990 *J. Chem. Phys.* **92** 508–17
- [52] Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K and Scheffler M 2009 *Comput. Phys. Comm.* **180** 2175–96
- [53] Møller C and Plesset M S 1934 *Phys. Rev.* **46** 618–22
- [54] Bohm D and Pines D 1951 *Phys. Rev.* **82** 625–34
- [55] Pines D and Bohm D 1952 *Phys. Rev.* **85** 338–53
- [56] Bohm D and Pines D 1953 *Phys. Rev.* **92** 609–25
- [57] Furche F 2001 *Phys. Rev. B* **64** 195120
- [58] Koch H, Christiansen O, Kobayashi R, Jørgensen P and Helgaker T 1994 *Chem. Phys. Lett.* **228** 233–8
- [59] Dunning T H 1989 *J. Chem. Phys.* **90** 1007–23
- [60] Helgaker T, Klopper W, Koch H and Noga J 1997 *J. Chem. Phys.* **106** 9639–46
- [61] Wolinski K and Pulay P 2003 *J. Chem. Phys.* **118** 9497–503
- [62] Zhang I Y, Luo Y and Xu X 2010 *J. Chem. Phys.* **133** 104105
- [63] Tew D P, Klopper W and Helgaker T 2007 *J. Comput. Chem.* **28** 1307–20
- [64] Kong L, Bischoff F A and Valeev E F 2012 *Chem. Rev.* **112** 75–107
- [65] Kutzelnigg W and Klopper W 1991 *J. Chem. Phys.* **94** 1985–2001
- [66] Noga J, Kutzelnigg W and Klopper W 1992 *Chem. Phys. Lett.* **199** 497–504
- [67] Hättig C, Klopper W, Köhn A and Tew D P 2012 *Chem. Rev.* **112** 4–74
- [68] Ten-no S and Noga J 2012 *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2** 114–25
- [69] Tao J, Perdew J P, Staroverov V N and Scuseria G E 2003 *Phys. Rev. Lett.* **91** 146401
- [70] Staroverov V N, Scuseria G E, Tao J and Perdew J P 2004 *Phys. Rev. B* **69** 075102
- [71] Heyd J and Scuseria G E 2004 *J. Chem. Phys.* **121** 1187–92
- [72] Paier J, Marsman M, Hummer K, Kresse G, Gerber I C and Ángyán J G 2006 *J. Chem. Phys.* **124** 154709
- [73] Schimka L, Harl J and Kresse G 2011 *J. Chem. Phys.* **134** 024116
- [74] Paier J, Marsman M and Kresse G 2007 *J. Chem. Phys.* **127** 024103
- [75] Harl J and Kresse G 2009 *Phys. Rev. Lett.* **103** 056401
- [76] Grüneis A, Marsman M and Kresse G 2010 *J. Chem. Phys.* **133** 074107
- [77] Harl J, Schimka L and Kresse G 2010 *Phys. Rev. B* **81** 115126

- [78] Haas P, Tran F and Blaha P 2009 *Phys. Rev. B* **79** 085104
- [79] Haas P, Tran F and Blaha P 2009 *Phys. Rev. B* **79** 209902
- [80] Tran F, Laskowski R, Blaha P and Schwarz K 2007 *Phys. Rev. B* **75** 115131
- [81] Klimeš J, Bowler D R and Michaelides A 2011 *Phys. Rev. B* **83** 195131
- [82] Csonka G I, Perdew J P, Ruzsinszky A, Phillipsen P H T, Lebegue S, Paier J, Vydrov O A and Ángyán J G 2009 *Phys. Rev. B* **79** 155107
- [83] Tran F, Stelzl J and Blaha P 2016 *J. Chem. Phys.* **144** 204120
- [84] Shepherd J J and Grüneis A 2013 *Phys. Rev. Lett.* **110** 226401
- [85] Booth G H, Grüneis A, Kresse G and Alavi A 2013 *Nature* **493** 365–70
- [86] Shepherd J J, Henderson T M and Scuseria G E 2014 *J. Chem. Phys.* **140** 124102
- [87] Grüneis A 2015 *Phys. Rev. Lett.* **115** 066402
- [88] Grüneis A 2015 *J. Chem. Phys.* **143** 102817
- [89] McClain J, Sun Q, Chan G K-L and Berkelbach T C 2017 *J. Chem. Theory Comput.* **13** 1209–18
- [90] Shepherd J J, Booth G, Grüneis A and Alavi A 2012 *Phys. Rev. B* **85** 081103
- [91] Bressanini D and Reynolds P J 1999 *Adv. Chem. Phys.* **105** 37
- [92] Needs R J, Kent P R C, Porter A R, Towler M D and Rajagopal G 2002 *Int. J. Quantum Chem.* **86** 218–25
- [93] Grossman J C 2002 *J. Chem. Phys.* **117** 1434
- [94] Lester W A Jr., Mitas L and Hammond B 2009 *Chem. Phys. Lett.* **478** 1–10
- [95] Dubecký M, Mitas L and Jurečka P 2016 *Chem. Rev.* **116** 5188–215
- [96] Chadi D J and Cohen M L 1973 *Phys. Rev. B* **8** 5747
- [97] Marcus P P, Janak J F and Williams A R 1971 *Computational Methods in Band Theory* vol. 4 (New York: Plenum)
- [98] Baldereschi A 1973 *Phys. Rev. B* **7** 5212
- [99] Methfessel M and Paxton A T 1989 *Phys. Rev. B* **40** 4616
- [100] Blöchl P E 1994 *Phys. Rev. B* **49** 16223
- [101] Grüneis A and Heß B A 1998 *Theor. Chem. Accounts* **100** 253–63
- [102] Dovesi R et al 2013 *Crystal09 v2.0.1 User's Manual*
- [103] Hellmann H J. *Chem. Phys.* **3** 61
- [104] Schwerdtfeger P 2011 *ChemPhysChem* **12** 3143–55
- [105] Levchenko S V, Ren X, Wieferink J, Johanni R, Rinke P, Blum V and Scheffler M 2015 *Comput. Phys. Comm.* **192** 60–9
- [106] Singh D J and Nordström L 2006 *Planewaves, Pseudopotentials, and the LAPW Method* (New York: Springer)
- [107] Lejaeghere K, Van Speybroeck V, Van Oost G and Cottenier S 2014 *Crit. Rev. Solid State Mater. Sci.* **39** 1
- [108] <http://molmod.ugent.be/DeltaCodesDFT>, (Accessed: 17 August 2015).
- [109] Blaha P, Schwarz K, Madsen G K H, Kvasnicka D and Luitz J 1999 *Wien2k, Anaugmented Plane Wave + Local Orbitals Program for Calculating Crystal Properties, Karlheinz Schwartz, Techn.* (Vienna: Universität Wien)
- [110] Zhao Y and Truhlar D G 2008 *Theor. Chem. Acc.* **120** 215–41
- [111] Sun J, Ruzsinszky A and Perdew J P 2015 *Phys. Rev. Lett.* **115** 036402
- [112] Krukau A V, Vydrov O A, Izmaylov A F and Scuseria G E 2006 *J. Chem. Phys.* **125** 224106
- [113] Ren X, Rinke P, Joas C and Scheffler M 2012 *J. Mater. Sci.* **47** 7447–71
- [114] Pisani C, Maschio L, Casassa S, Halo M, Schütz M and Usvyat D 2008 *J. Comput. Chem.* **29** 2113–24
- [115] Pisani C, Schütz M, Casassa S, Usvyat D, Maschio L, Lorenz M and Erba A 2012 *Phys. Chem. Chem. Phys.* **14** 7615–28
- [116] Del Ben M, Hutter J and VandeVondele J 2012 *J. Chem. Theory Comput.* **8** 4177–88
- [117] Del Ben M, Hutter J and VandeVondele J 2013 *J. Chem. Theory Comput.* **9** 2654
- [118] Zhang I Y, Rinke P, Perdew J P and Scheffler M 2016 *Phys. Rev. Lett.* **117** 133002
- [119] Whitten J L 1973 *J. Chem. Phys.* **58** 4496
- [120] Dunlap B 2000 *J. Mol. Struct.: Theochem* **501** 221
- [121] Ihrig A C, Wieferink J, Zhang I Y, Ropo M, Ren X, Rinke P, Scheffler M and Blum V 2015 *New J. Phys.* **17** 093020
- [122] Ayala P Y, Kudin K N and Scuseria G E 2001 *J. Chem. Phys.* **115** 9698–707
- [123] Doser B, Zienau J, Clin L, Lambrecht D S and Ochsenfeld C 2010 *Z. Phys. Chem.* **224** 397–412
- [124] Lejaeghere K et al 2016 *Science* **351** aad3000
- [125] De Waele S, Lejaeghere K, Sluydts M and Cottenier S 2016 *Phys. Rev. B* **94** 235418
- [126] Pernot D P P, Civalieri B and Savin A 2015 *J. Phys. Chem. A* **119** 5288–304
- [127] Adrienn R and Perdew J P 2010 *J. Chem. Theory Comput.* **6** 127–34
- [128] Heßelmann A and Görling A 2010 *Mol. Phys.* **108** 359–72
- [129] Sun J, Xiao B, Fang Y, Haunschild R, Hao P, Ruzsinszky A, Csonka G I, Scuseria G E and Perdew J P 2013 *Phys. Rev. Lett.* **111** 106401
- [130] Marsman M, Grüneis A, Paier J and Kresse G 2009 *J. Chem. Phys.* **130** 184103
- [131] Ohnishi Y-Y and Hirata S 2010 *J. Chem. Phys.* **133** 034106
- [132] Grüneis A, Booth G H, Marsman M, Spencer J, Alavi A and Kresse G 2011 *J. Chem. Theory Comput.* **7** 2780–5
- [133] Wiberg K B 2004 *J. Comput. Chem.* **25** 1342–6
- [134] Kutzelnigg W and Morgan J D 1992 *J. Chem. Phys.* **96** 4484–508
- [135] Feller D and Sordo J A 2000 *J. Chem. Phys.* **113** 485–93
- [136] Monkhorst H J and Pack J D 1976 *Phys. Rev. B* **13** 5188–92
- [137] FHI aims team, *Fhi-aims User's Guide*, Fritz-Haber-Institut der Max-Planck-Gesellschaft, 2015.
- [138] VASP team VASP manual, Universitaet Wien, 2018
- [139] Spencer J and Alavi A 2008 *Phys. Rev. B* **77** 193110
- [140] Spackman M A and Mitchell A S 2001 *Phys. Chem. Chem. Phys.* **3** 1518–23
- [141] Champagne B, Jacquemin D, Gu F L, Aoki Y, Kirtman B and Bishop D M 2003 *Chem. Phys. Lett.* **373** 539–49
- [142] Johnson E R, Becke A D, David Sherrill C and DiLabio G A 2009 *J. Chem. Phys.* **131** 034111–7
- [143] Gräfenstein J, Izotov D and Cremer D 2007 *J. Chem. Phys.* **127** 214103