

## **General embedded cluster protocol for accurate modeling of oxygen vacancies in metal-oxides**

Benjamin X. Shi,<sup>1</sup> Venkat Kapil,<sup>1,2</sup> Andrea Zen,<sup>3,4</sup> Ji Chen,<sup>5</sup> Ali Alavi,<sup>6,1</sup> and Angelos Michaelides<sup>1, 7, 8</sup>

<sup>1</sup>*Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

<sup>2</sup>*Churchill College, University of Cambridge, Storey's Way, Cambridge CB3 0DS*

<sup>3</sup>*Dipartimento di Fisica Ettore Pancini, Università di Napoli Federico II, Monte S. Angelo, I-80126 Napoli, Italy*

<sup>4</sup>*Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, United Kingdom*

<sup>5</sup>*School of Physics, Peking University, Beijing, 100871, China*

<sup>6</sup>*Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany*

<sup>7</sup>*Department of Physics and Astronomy, University College London, Gower Street, London, WC1E 6BT, United Kingdom*

<sup>8</sup>*Thomas Young Centre and London Centre for Nanotechnology, 17-19 Gordon Street, London WC1H 0AH, United Kingdom*

(Dated: 7 March 2022)

The O vacancy (Ov) formation energy,  $E_{Ov}$ , is an important property of a metal-oxide, governing its performance in applications such as fuel cells or heterogeneous catalysis. These defects are routinely studied with density functional theory (DFT). However, it is well-recognized that standard DFT formulations (e.g. the generalized gradient approximation) are insufficient for modeling the Ov, requiring higher levels of theory. The embedded cluster method offers a promising approach to compute  $E_{Ov}$  accurately, giving access to all electronic structure methods. Central to this approach is the construction of quantum(-mechanically treated) clusters placed within suitable embedding environments. Unfortunately, current approaches to constructing the quantum clusters either require large system sizes, preventing application of high-level methods, or require significant manual input, preventing investigations of multiple systems simultaneously. In this work, we present a systematic and general quantum cluster design protocol that can determine small converged quantum clusters for studying the Ov in metal-oxides with accurate methods such as local coupled cluster with single, double and perturbative triple excitations [CCSD(T)]. We apply this protocol to study the Ov in the bulk and surface planes of rutile  $TiO_2$  and rocksalt  $MgO$ , producing the first accurate and well-converged determinations of  $E_{Ov}$  with this method. These reference values are used to benchmark exchange-correlation functionals in DFT and we find that all studied functionals underestimate  $E_{Ov}$ , with the average error decreasing along the rungs of Jacob's ladder. This protocol is automatable for high-throughput calculations and can be generalized to study other point defects or adsorbates.

## I. INTRODUCTION

Metal-oxides are a class of material with wide applications in fuel<sup>1</sup> and solar cells<sup>2</sup>, high- $k$  dielectrics<sup>3</sup>, and the catalysis industry<sup>4,5</sup>. As the most prevalent defect in metal-oxides, controlling the concentration of O vacancies (Ovs) in these systems underpins much of the major advances to their applications. The dominant quantity determining the Ov concentration is the Ov formation energy,  $E_{Ov}$ .

It is pivotal that  $E_{Ov}$  can be determined accurately. The Ov concentration can change by several orders of magnitude with small ( $\sim 0.1$  eV) changes in the  $E_{Ov}$  at a given temperature<sup>6</sup>, resulting in drastic changes in thermodynamic, electronic and optical properties of the metal-oxide. For example, such changes in the Ov concentration can change an insulating oxide into a photocatalyst<sup>7</sup> or metal<sup>8</sup>.  $E_{Ov}$  is also a measure of the reducibility of a metal-oxide system, with (thermal) reduction being a vital step in the thermochemical cycles for H<sub>2</sub>O and CO<sub>2</sub> splitting<sup>9,10</sup>. Similarly,  $E_{Ov}$  has also been shown to correlate well with key catalytic properties such as the adsorption and bond activation energies of various molecules<sup>11-13</sup> on metal-oxide surfaces.

The reliable experimental determination of  $E_{Ov}$  is very challenging as it depends sensitively on many factors<sup>14</sup> (e.g. Ov concentration, presence of dopants, and crystallite size). As such,  $E_{Ov}$  has been mostly computed through electronic structure modeling, particularly with density functional theory (DFT). However,  $E_{Ov}$  is highly sensitive to the approximation of the exchange-correlation (XC) functional in DFT, oftentimes leading to large disagreements in their predictions. For example, in the (110) rutile TiO<sub>2</sub> surface<sup>15</sup>,  $E_{Ov}$  can vary by more than 1.5 eV – around 50% of the absolute  $E_{Ov}$  – with similar discrepancies observed in MgO<sup>16</sup> and other metal-oxide systems<sup>15,17</sup>. Additionally, DFT with the generalized gradient approximation (GGA) XC functional has been shown to be inadequate for modeling Ovs in transition metal oxides such as rutile TiO<sub>2</sub>, predicting that the unpaired electrons produced during Ov formation are delocalized<sup>18</sup>, whilst hybrid functionals which incorporate exact exchange predict localized electrons on adjacent Ti sites<sup>19</sup>. There is a clear need for high-accuracy and well-converged reference values of the  $E_{Ov}$  for these systems.

The (electrostatic) embedded cluster approach<sup>20</sup> offers the potential to efficiently apply accurate methods to study the Ov. It limits explicit quantum-mechanical calculations to only a finite-sized quantum cluster, with the electrostatic interactions from the rest of the system approximated by point charges. Over the past few decades, there have been numerous applications of this approach to the Ov in metal-oxides<sup>21</sup>. These studies have ranged from applying the basic electro-

static embedding that has been described<sup>22–24</sup>, to more involved setups which utilize polarizable environments<sup>25–28</sup> (of varying degrees of sophistication) or couples the quantum cluster to the environment self-consistently via the “perturbed cluster” approach<sup>29</sup>. Regardless of the approach taken, the outstanding challenge is finding a quantum cluster that is sufficiently small, or even the smallest, which allows for inexpensive modeling at the reference level of theory whilst minimizing finite-size errors to converge results to the bulk (i.e. infinite size) limit. This process is difficult because the quantum cluster can take any chemical formula (i.e. size) and for each chemical formula, there can be many possible shapes, with widely differing convergence behaviors<sup>30,31</sup>.

To circumvent the complexities with searching the entire size and shape space, converged clusters are normally selected from a set of clusters generated from chosen design rubrics<sup>24,32</sup>. The most common rubrics are to keep quantum clusters stoichiometric and spherical<sup>16,33</sup>. Identifying suitable quantum clusters which follow these design principles require significant time investment and manual input, paired with lots of trial-and-error. As such, only a handful of clusters are normally created, making it difficult to affirm the quality of a cluster as well as study multiple crystal systems simultaneously. Additionally, recent work<sup>32,34</sup> has shown that these rubrics lead to poor convergence with cluster size.

There is growing interest in approaches which can lessen the manual labor involved. These approaches range from building clusters using layers (based on unit cell multipoles<sup>35</sup> or coordination spheres<sup>36</sup>) to using building blocks of either the unit cell<sup>31,34,37,38</sup> or fully-coordinated ions<sup>37,39</sup>. Whilst the quantum cluster series generated from these approaches have been shown to converge properties of an ideal crystal (e.g. NMR constants<sup>34</sup>, bandgaps<sup>37</sup> and optical spectra<sup>39,40</sup>), their extension to the study of surfaces or point defects such as the Ov is not clear and they still suffer from important deficiencies. For example, the positioning of the building blocks in the building block approach still requires manual definition. With the layered approach, the number of atoms within each layer increases significantly compared to the previous, providing little granularity in the sizes sampled. This leads to large converged quantum clusters that are not conducive for high accuracy calculations, which typically exhibit steep system size scaling.

To summarize, a systematic and general approach for designing a quantum cluster set which provides good granularity in the sizes and shapes being sampled whilst ensuring rapid convergence is currently lacking. In this work, we propose a quantum cluster design protocol (named SKZCAM after the authors’ initials) which achieves these qualities for computing accurate Ov formation energies in metal-oxides. The core enabling development is to put the control of the shape and

size of a cluster into a robust and flexible framework using the radial distribution function (RDF) to divide metal cations into “shells”. The O anion positions for each cluster arise naturally from the metal cation positions based on the criteria that all dangling bonds are removed from the metal cations. The result is a process that requires no manual intervention to generate clusters of high granularity for virtually any metal-oxide crystal system and surface termination, whilst converging rapidly with size.

This protocol is used to obtain small converged clusters for studying the  $E_{\text{Ov}}$  in the bulk and common surface planes of rocksalt MgO and rutile TiO<sub>2</sub>, two technologically relevant metal-oxide systems with properties that depend sensitively on the Ov. These clusters (involving fewer than 600 correlated electrons for all systems) enable the accurate local natural orbital (LNO-)CCSD(T)<sup>41–45</sup> method to be applied to compute  $E_{\text{Ov}}$ . We use our computed values to benchmark common DFT XC functionals, ranging from GGAs up to double-hybrids, and find that all studied XC functionals underestimate  $E_{\text{Ov}}$ , with this error decreasing, on average, along the rungs of Jacob’s ladder.

## II. COMPUTATIONAL DETAILS

We describe the embedded cluster and supercell approaches used to generate the necessary structures to evaluate  $E_{\text{Ov}}$  for bulk rocksalt MgO and its (100) surface as well as bulk rutile TiO<sub>2</sub> and its (110) surface in this section. This quantity is calculated as:

$$E_{\text{Ov}} = E[\text{D-MO}] - E[\text{P-MO}] + E[\text{O}] - \frac{1}{2}E_{\text{bind}}, \quad (1)$$

where  $E[\text{P-MO}]$  and  $E[\text{D-MO}]$  are the total energies of the pristine and defected metal-oxide systems respectively. The defected system is created from the pristine system by removing an O atom without further geometrical optimization. Both systems are treated in the closed-shell singlet state.  $E[\text{O}]$  is the total energy of an O atom in the (unrestricted) triplet state and  $E_{\text{bind}}$  is the molecular binding energy of an O<sub>2</sub> molecule, which has been computed for various levels of theory in Sec. S1 of the supplementary material. The final two terms sum up to (half) the energy of an O<sub>2</sub> molecule, thus defining  $E_{\text{Ov}}$  under O-rich conditions.

The definition given in Eq. 1 computes the *unrelaxed* O vacancy formation energy  $E_{\text{Ov}}$ . Whilst the use of unrelaxed Ov structures precludes direct comparison with experiment when relaxation effects are significant, it provides an upper bound to the relaxed  $E_{\text{Ov}}$  and a valid reference when comparing DFT values to high accuracy methods at the same geometry. In any case, relaxation

effects have been found to be negligible for rocksalt MgO<sup>46,47</sup>. Additionally, whilst this effect is significant for rutile TiO<sub>2</sub><sup>48</sup>, it is hard to quantify accurately due to a strong dependence on the spin-state (see Sec. II C) and chosen DFT XC functional (see Sec. S2 of the supplementary material), making it most appropriate to evaluate the *unrelaxed*  $E_{Ov}$  for the purposes of this work.

### A. Electrostatic embedded cluster calculations

The electrostatic embedded cluster approach, illustrated in Fig. 1 (a), used in this study features a quantum(-mechanical) cluster centred around the O vacancy. To model the long-range electrostatic potential from the rest of the material, this cluster is surrounded by a sphere and hemisphere of point charges of radius 30 and 40 Å for the bulk and surface respectively; formal point charges have been placed at the metal and O ion crystallographic positions. In the vicinity of the quantum cluster ( $< 7$  Å), the positive point charges are “capped” with the effective core potential (ECP) of the corresponding metal ion, taken from the Stuttgart/Cologne group<sup>49,50</sup>, to avoid electron leakage from the bonded O ions at the boundary of the quantum cluster. The chosen radii of the different regions (see Sec. S3 of the supplementary material) can converge  $E_{Ov}$  to 0.01 eV. The placement of the point charges and ECPs alongside the atoms of the quantum cluster were constructed using py-ChemShell<sup>51</sup>.

DFT calculations were performed in ORCA<sup>35</sup> version 5.0 and MRCC<sup>52</sup> 2020, with the latter interfaced to libXC<sup>53</sup>. The def2-SVP, def2-TZVPP and def2-QZVPP Weigend-Ahlrichs<sup>54</sup> basis sets were used throughout this paper, with the standard def2-JK<sup>55,56</sup> fitting basis set used for Coulomb and exchange integrals. Convergence tests indicate that the def2-SVP and def2-TZVPP basis sets exhibit errors of 0.4 and 0.02 eV w.r.t. the def2-QZVPP basis set (see Sec. S4 of the supplementary material).

Localized orbital correlated wave-function theory (cWFT) calculations were performed with the LNO-CCSD(T) and local Møller Plesset perturbation theory (LMP2) implementations of Nagy *et al.*<sup>44,57</sup> in MRCC, using the “Normal” LNO thresholds. The MP2 contribution to the B2PLYP<sup>58</sup> XC functional was also evaluated using the LMP2 implementation of MRCC. Complete basis set (CBS) extrapolation parameters for the def2-TZVPP and def2-QZVPP pair, CBS(TZVPP/QZVPP), taken from Neese and Valeev<sup>59</sup>, were used for the Hartree-Fock (HF) and correlation energy components of the cWFT total energy. Oxygen basis functions were placed at the Ov site and only valence electrons were correlated. Convergence tests indicate that these

settings can give accuracy to within 0.1 eV (see Sec. S4 B of the supplementary material). Deficiencies due to frozen-core or basis set size were accounted for through a correction computed on small tractable clusters featuring a "reduced frozen-core"<sup>60</sup> (e.g. Ne for Ti and He for Mg) and basis set involving cc-pwCVnZ<sup>61,62</sup> and aug-cc-pVnZ<sup>63</sup> on the metal and O ions respectively (see Sec. S5 of the supplementary material), which has been extrapolated<sup>59</sup> for  $n = T$  and  $Q$ . The def2- $n$ ZVPP-RI auxiliary basis sets<sup>64,65</sup> were used for the local cWFT calculations with the def2 basis sets, whilst the automatic auxiliary basis functions of Stoychev *et al.*<sup>66,67</sup> were generated for the correlation-consistent (cc) basis sets.

## B. Periodic supercell calculations

Periodic supercell calculations with DFT were performed in the Vienna *Ab initio* Simulation Package (VASP)<sup>68,69</sup>. These calculations serve to define the positions for constructing embedded cluster calculations as well as to provide reference  $E_{Ov}$  values. All systems used structures optimized at the R2SCAN<sup>70</sup> DFT level as these agree well with experimental lattice parameter values (see Sec. S6 of the supplementary material). The bulk rutile TiO<sub>2</sub> calculations employed a 192 atom ( $2\sqrt{2} \times 2\sqrt{2} \times 4$ ) supercell and the bulk rocksalt MgO calculations used a 64-atom ( $2 \times 2 \times 2$ ) supercell, both with a ( $2 \times 2 \times 2$ )  $\Gamma$ -centred Monkhorst-Pack  $k$ -point sampling. The (001) MgO surface and (110) TiO<sub>2</sub> surface calculations both employed an asymmetric five-layer slab with the top two layers allowed to relax to form the pristine surface. A  $p(2 \times 4)$  and ( $2 \times 2$ ) supercell was used for the TiO<sub>2</sub> and MgO surfaces respectively, each computed with a ( $2 \times 2 \times 1$ ) Monkhorst-Pack mesh and 12 Å of vacuum. For the TiO<sub>2</sub> surface, a correction to the  $p(2 \times 6)$  supercell size, at the PBE<sup>71</sup> level, was applied for the PBE0<sup>72</sup> hybrid DFT calculation (see Sec. S7 of the supplementary material). An energy cutoff of 500 eV was used for all four systems, with small core projector augmented wave (PAW) potentials on the metal cations, leaving 12 and 10 valence electrons for Ti and Mg respectively. The standard PAW potential was used on the O anion.

When constructing embedded cluster systems, the optimized bulk unit cells were repeated into supercells with dimensions larger than the embedded clusters to allow for them to be cleaved out. For the surface systems, the five-layer supercell slab was concatenated (along the surface normal direction) with an additional 25 layers (taken from the bulk) before being repeated into supercells for embedded cluster calculations.

### C. Spin-state of the Ov in Rutile TiO<sub>2</sub>

The *unrelaxed* Ov systems of rocksalt MgO and rutile TiO<sub>2</sub>, in both their bulk and common surface planes, all feature a well-controlled closed-shell singlet state<sup>26,46,47,73</sup>. Whilst this spin state is appropriate for the MgO systems (since relaxation effects are negligible), there are significant discrepancies regarding the spin-state of the relaxed Ov structures of rutile TiO<sub>2</sub> within the literature, both from experimental and computational studies.

Experimental results from electron paramagnetic resonance (EPR) and infrared spectroscopy have detected that the excess electrons, believed to originate from Ovs, localize on Ti ions of the system, forming a deep band gap state<sup>74–76</sup> and pointing towards an open-shell spin state<sup>77,78</sup>. However these experiments also conflict with the high mobilities predicted by electrical measurements<sup>79</sup>, suggesting a shallow n-type donor<sup>80</sup> which favors delocalized electrons.

Computational studies on the Ov of rutile TiO<sub>2</sub> have indicated various possible spin-states, ranging from the closed-shell singlet<sup>73,81</sup>, to the open-shell triplet<sup>26,82</sup> and singlet<sup>48,83,84</sup> states. For the open-shell states, the formation of polarons<sup>85</sup> (i.e. localized electrons on Ti ions) brings additional complications. Depending on the localization site of the polarons and degree of polaronic distortion, there can be a wide range of formation energies<sup>19,84,86</sup>.

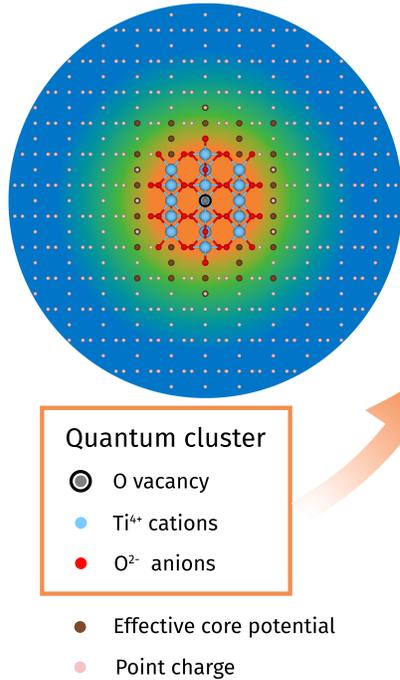
A recent benchmark study using CCSD(T) by Chen *et al.*<sup>81</sup> has found that the closed-shell singlet state is still the most stable state with the inclusion of relaxation effects. The validity of CCSD(T) for studying the Ov in TiO<sub>2</sub> was confirmed from preliminary full configuration interaction quantum Monte Carlo (FCIQMC) calculations, which reveal the single-reference character of the Ov state. However it should be noted that Chen *et al.* did not directly model polarons due to the large cluster sizes required<sup>32</sup>, which could potentially change the conclusions. We expect that the advances detailed in this work may provide the possibility to resolve these open questions/discrepancies for the TiO<sub>2</sub> system in future studies.

## III. RESULTS

### A. Quantum Cluster Design Protocol

Designing a systematic and general set of quantum clusters for metal-oxides is a complex process; we have identified and summarized the key challenges involved in Fig. 1 (b). Two highly important and interdependent factors are the size and shape of the quantum cluster. For metal-

(a) Embedded cluster approach



(b) Challenges with designing quantum cluster

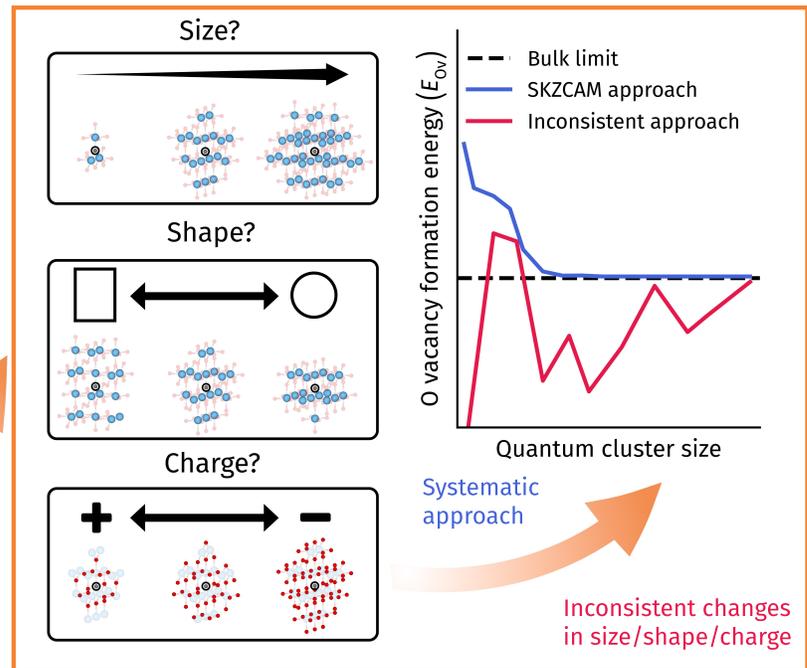


FIG. 1. (a) Schematic of the electrostatic embedding approach for bulk rutile  $\text{TiO}_2$ . The quantum cluster (in the orange region) is treated with the electronic structure theory of choice. It is surrounded by a spherical (hemispherical) field of point charges in the green and blue regions for a bulk (surface) system. In the vicinity (green region) of the quantum cluster, the point charges are replaced with effective core potentials to prevent spurious charge leakage out of the quantum cluster. Normally, converged quantum clusters are selected from a series of clusters constructed through chosen design rubrics. Panel (b) highlights the key challenges with designing such a series of quantum clusters for metal-oxides: (i) deciding what sizes (in terms of number of metal cations, visualized as light blue spheres) to sample; (ii) deciding how to arrange these metal cations (i.e. shape); (iii) and choosing the charge of each quantum cluster, which is largely controlled by the number of O anions (visualized as red spheres). Depending on the chosen rubrics for these three factors, convergence of the O vacancy formation energy  $E_{Ov}$  can be either systematic or inconsistent, as indicated by the blue and red lines in the schematic graph. The points representing the ‘Inconsistent approach’ were taken from calculations of stoichiometric, spherical-shaped quantum clusters with inconsistent changes in the number of Ti ions (i.e. size) and spatial arrangement of O anions (i.e. charge).

oxides, these two factors are predominantly controlled by the metal cations in the quantum cluster since these systems can be considered as formed from polyhedrons (e.g.  $\text{TiO}_6$  octahedron for  $\text{TiO}_2$ ) centered around the metal cations – highlighted in light blue for bulk rutile  $\text{TiO}_2$  in Fig. 1 (b). The choice of studied quantum cluster sizes is normally a manual process, often involving large arbitrary changes in sizes along the series. It is important that this process can be made systematic, whilst sampling clusters in small increments, so that small – even the smallest – converged clusters can be found for the property being studied, particularly when applying expensive cWFT methods. For a given size, the metal cations can take up virtually any spatial arrangement (i.e. shape) and  $E_{\text{Ov}}$  can vary significantly depending on the chosen shape, as shown in Sec. S8 A of the supplementary material. Under the electrostatic embedding scheme, the quantum cluster can be chosen to take up any charge. Thus, for a given metal cation configuration (i.e. size and shape), there can be virtually any number of O anions; if the ratio of O to Ti ions is more (less) than the stoichiometric ratio, then a negatively (positively) charged quantum cluster is formed. Here, the same challenges with deciding the number and spatial arrangement of O anions exist as the metal cations. If any of the factors discussed above are performed in an inconsistent manner, the convergence of the quantum cluster series becomes non-monotonic as shown by the red line of the schematic graph in Fig. 1 (b). In this work, we propose the SKZCAM approach for constructing a set of quantum clusters which shows fast and systematic convergence towards the bulk limit (blue line in the schematic graph of Fig. 1 (b)). We define a rigorous framework for deciding the metal cation configurations in the quantum cluster series, with the O anion configuration naturally arising based on a separate robust rubric.

The RDF of the number of metal cations as a function of distance from the Ov is used to generate the metal cation configurations in the quantum cluster series. Using bulk rutile  $\text{TiO}_2$  as an example in Fig. 2 (a), we see that the metal Ti cations arrange as shells (with the first three given distinct colors) of symmetry-related equidistant cations around the Ov, denoted by the gray sphere within a black circle. In an RDF plot, given at the bottom panel of Fig. 2 (b), these Ti cation shells will appear as distinct peaks. Starting from the first metal cation shell/peak found in the RDF plot, metal cation configurations of systematically increasing size can be created by adding subsequent metal cation shells, with the first six metal cation configurations for this series visualized in Fig. 2 (b). The RDF shells/peaks are completely controlled by the crystal structure, point defect site and surface termination of the system, requiring no manual input. Furthermore, it provides good granularity in the size (i.e. number) of metal cations sampled in the quantum cluster

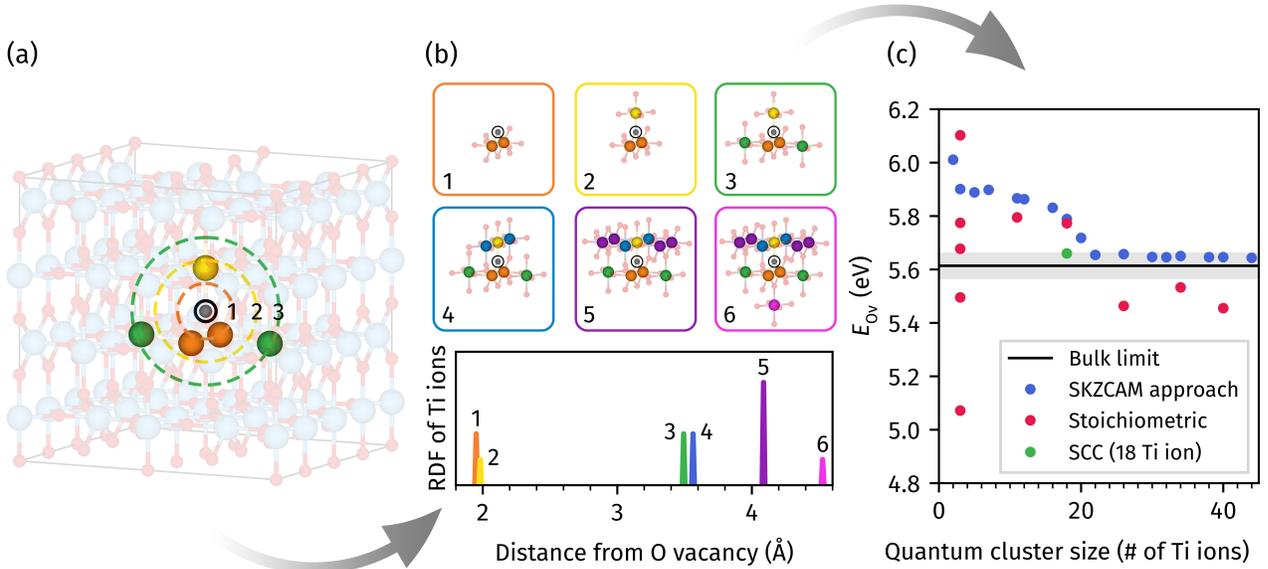


FIG. 2. Illustration of the SKZCAM approach for designing the quantum clusters described in this work, using bulk rutile  $\text{TiO}_2$  as an example. Around the gray O vacancy (Ov) outlined by a black circle, the Ti cations arrange as “shells” consisting of equidistant, symmetry related cations; we highlight the first three shells in (a) with unique colors. These shells appear as peaks in the radial distribution function (RDF)  $[4\pi r^2 g(r)]$  plot of Ti cations around the Ov in (b). Starting from the first shell, clusters of systematically increasing size can be generated by incorporating subsequent shells. We give the example of the first six quantum clusters generated in (b). In our approach, we choose O anions, illustrated by translucent red spheres, such that all dangling bonds around the Ti cations are removed. This combination leads to a systematic convergence of  $E_{\text{Ov}}$  at the PBE-DFT level towards the bulk limit with quantum cluster size, as indicated by the blue markers in (c). The red markers correspond to stoichiometric clusters created to have the same Ti cations as a subset of the SKZCAM approach clusters, but involving fewer O anions to meet the stoichiometric ratio. We observe poor non-monotonic convergence towards the bulk limit with this choice. The bulk limit in (c) was calculated from a supercell calculation described in Sec. II B, whilst embedded cluster calculations were performed with the def2-TZVPP basis set.

series. At the same time, a variety of shapes are studied, ranging from spherical to cuboidal, all whilst ensuring symmetry about the point defect is maintained.

The metal cations serve as the base for the subsequent O anion configuration in each cluster. It is most physical to consider only O sites which ensure at least one bond to a nearby metal cation, as shown as translucent red spheres for the quantum clusters in Fig. 2 (b). In principle,

any number of these O anion sites can be used; we take the unambiguous approach of placing O anions in all shown positions, ensuring no dangling bonds on the metal cations. This is the most chemically intuitive approach because it is equivalent to fully-coordinating all of the metal cations for bulk systems. For surface systems, some metal cations at the surface are not fully-coordinated due to the nature of the surface termination. As a result, the O anion configuration is completely determined by the metal cation configuration, which is in turn controlled by the RDF of metal cations around the point defect. We note that because the ratio of O anions to metal cations exceeds the stoichiometric ratio with this choice, the resulting quantum clusters are all negatively charged.

Beyond being fully systematic and general whilst providing good granularity, this SKZCAM approach also converges rapidly with cluster size. As shown by the blue markers in Fig. 2 (c),  $E_{Ov}$  becomes converged to within 0.05 eV of the bulk limit (illustrated by the gray error bars) for the 22 Ti ion cluster (consisting of the first ten RDF shells/peaks), with all subsequent clusters staying converged. As comparison, we have also constructed stoichiometric neutral clusters for a subset of the Ti cation configurations as shown by the red markers. These clusters are the most common type within the literature and we see a poor non-monotonic convergence towards the bulk limit, requiring clusters larger than the studied range of sizes. Notwithstanding the slow convergence, there is also significant ambiguity in constructing stoichiometric clusters for a given metal cation configuration since there are more possible O anion sites than allowed by stoichiometry. We give the  $E_{Ov}$  for 6 possible O anion configurations in the 3 Ti ion quantum cluster (visualized in Sec. S8 B) in Fig. 2 (c) - there is a wide range in  $E_{Ov}$  of 1.0 eV. For larger (stoichiometric) cluster sizes, there will be more possible O anion configurations; we have calculated  $E_{Ov}$  for only one such O anion configuration at larger cluster sizes based on including O anions closest to the Ov.

Whilst the converged 22 Ti ion cluster found is already quite small, it may be beneficial to search for smaller converged clusters, particularly for performing expensive cWFT calculations. The SKZCAM approach provides a robust framework for defining the shape of a quantum cluster based on metal cation shells, which serve as the building blocks of the cluster. The 22 Ti ion cluster consists of the first ten metal cation RDF shells/peak, with the removal of the furthest (tenth) shell leading to an unconverged 20 Ti ion cluster. By systematically removing closer RDF shells from the 22 Ti ion cluster (see Sec. S9 of the supplementary material), we identify a smaller 18 Ti ion cluster formed from removing the seventh metal cation shell (see green marker in Fig. 2 (c)). Removal of any subsequent shells from this 18 Ti ion cluster leads to large changes in  $E_{Ov}$ .

## B. Converged Clusters for the Ov

The SKZCAM approach outlined in the previous section is applied to create systematic sets of quantum clusters to study  $E_{Ov}$  at the DFT level and beyond for bulk rocksalt MgO and its (001) surface, as well as bulk rutile TiO<sub>2</sub> and its (110) surface in Fig. 3.

For the MgO systems and TiO<sub>2</sub> bulk, there is systematic convergence towards the bulk limit, approximated from a supercell calculation. We consider convergence reached at the point when  $E_{Ov}$  starts to plateau, with the cluster located at this point being the smallest converged cluster (SCC). We find an SCC with 38 Mg ions and 17 Mg ions for MgO bulk and surface respectively. Beyond these SCC sizes, the  $E_{Ov}$  of larger clusters are less than 0.02 eV from the bulk limit, as illustrated by the gray error bars. Although the 6 Mg ion quantum cluster for bulk MgO is within 0.02 eV of the bulk limit, this agreement is fortuitous at the DFT level because at higher levels of theory (LMP2), it differs by  $> 0.3$  eV from larger converged cluster sizes (see Sec. S11 A of the supplementary material). For TiO<sub>2</sub> bulk, the SCC with 18 Ti ions shows an error of 0.05 eV w.r.t. the bulk limit, with this error decreasing slowly with cluster size subsequently. This slow convergence could arise because the electrostatic embedding setup used here does not explicitly include (long-range) polarization effects; more sophisticated setups<sup>87–89</sup> including polarizable force fields beyond the quantum cluster could potentially improve this convergence.

The series of quantum clusters for the MgO systems were generated using the approach described in Section III of adding RDF shells of increasing distance from the Ov. We do not expect for there to be smaller converged clusters (in terms of number of Mg cations) beyond the sampled series for these systems due to their high degree of ionic bonding and cubic symmetry. For rutile TiO<sub>2</sub>, where there is a degree of directional covalent bonding<sup>90</sup> and anisotropy, there may be even smaller clusters. As discussed in Sec. III A, the SKZCAM approach provides the flexibility to find these smaller converged clusters, as has been done to find the 18 Ti ion SCC in TiO<sub>2</sub> bulk.

The (110) rutile TiO<sub>2</sub> surface requires a separate discussion due to its complex electronic structure. As seen in Sec. S12 of the supplementary material, the “noisy” nature of its convergence arises from well-behaved odd-even oscillations in  $E_{Ov}$  when Ti ions are added along specific crystallographic directions of the surface. Such odd-even size oscillations commonly appear in rutile TiO<sub>2</sub> surface<sup>91–93</sup> calculations. We find that the amplitude of these oscillations are correlated to the degree of self-interaction error in the electronic structure method, being weaker in PBE0 compared to PBE, and completely absent in methods (e.g. HF and LMP2) that do not suffer from

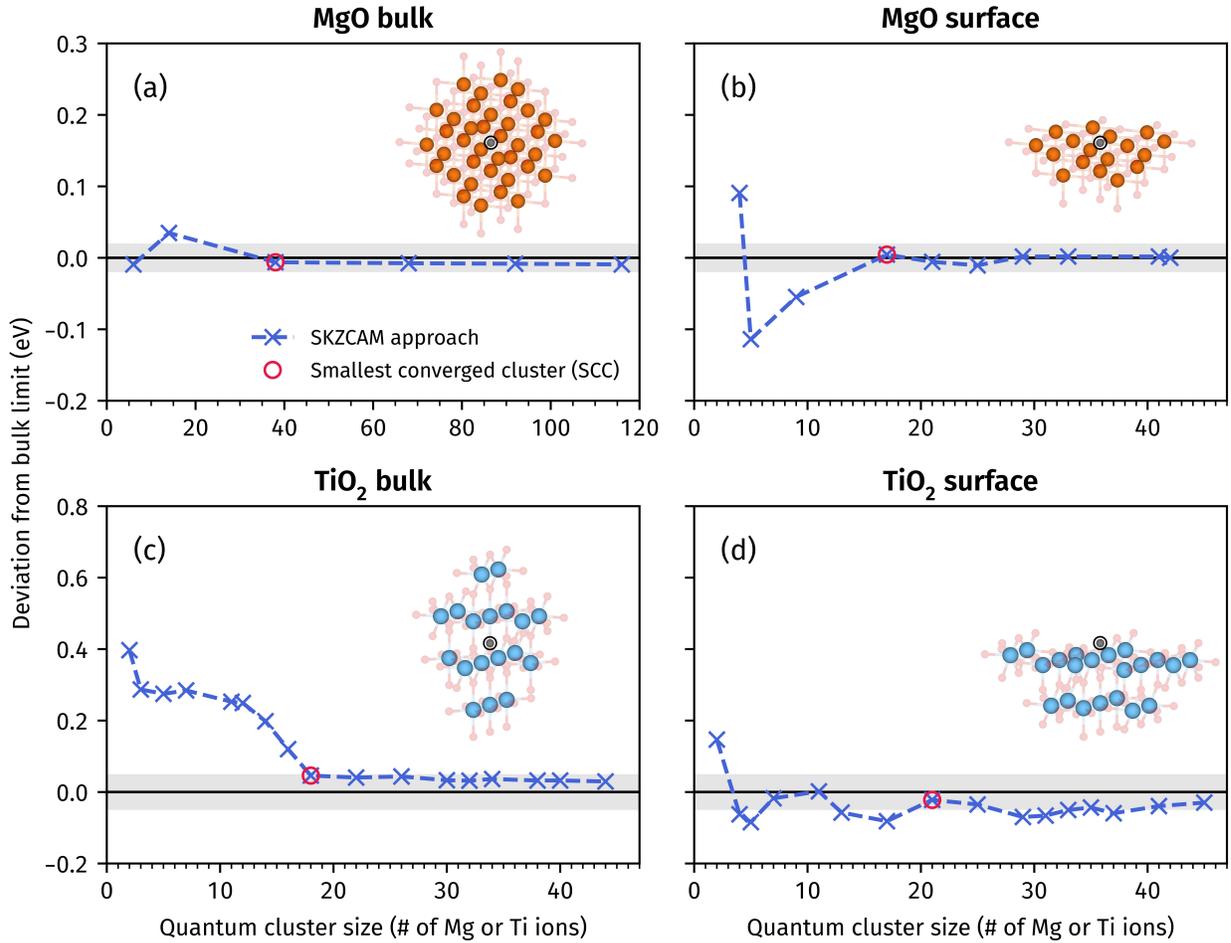


FIG. 3. The deviation from the bulk limit of the O vacancy formation energy ( $E_{Ov}$ ) as a function of quantum cluster size – generated via the SKZCAM approach – for (a) MgO bulk, (b) MgO surface, (c) TiO<sub>2</sub> bulk and (d) TiO<sub>2</sub> surface. The smallest converged cluster is marked with a red circle for each system and illustrated in each panel, where the light blue, orange and translucent red spheres correspond to Ti, Mg and O ions respectively. For the MgO systems, embedded cluster calculations were performed at the def2-SVP PBE-DFT level, with a correction to the def2-TZVPP basis to enable comparison to the bulk limit (see Sec. S10 of the supplementary material). The TiO<sub>2</sub> systems feature explicit def2-TZVPP calculations (to ensure smoother convergence with cluster size), with PBE and PBE0 used for the bulk and surface respectively. The bulk limit results were calculated using the supercell approach with the corresponding DFT level in each panel.

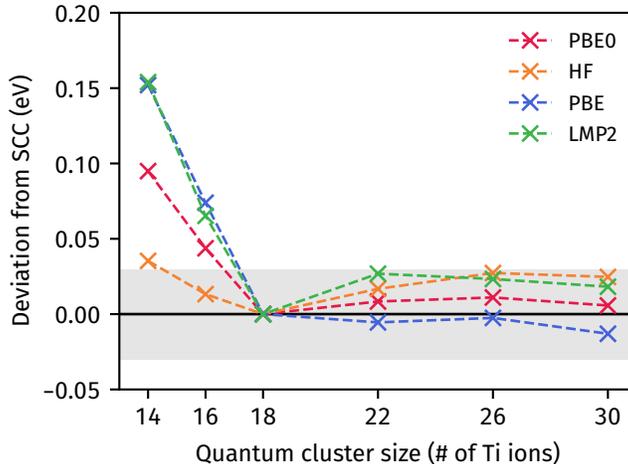


FIG. 4. The variation in the O vacancy formation energy ( $E_{Ov}$ ) with cluster size around the smallest converged cluster (SCC), containing 18 Ti ions, for bulk rutile  $TiO_2$  at the PBE, PBE0, HF and LMP2 levels of theory. The gray 0.03 eV error bar indicates the maximum level of deviation observed across the three levels of theory for sizes beyond the SCC. All calculations were performed with the def2-TZVPP basis set.

self-interaction error. The 21 Ti ion cluster was selected as the SCC on the basis of good agreement with the bulk limit at the PBE and PBE0 levels ( $< 0.03$  eV for the latter) as well as having reached the convergence plateau for HF and MP2 (see Sec. S11 B of the supplementary material). The gray 0.05 eV error bar in Fig. 3 (d) indicates the average error of the clusters larger than (and including) the chosen SCC.

We find evidence that the SCC determined at an appropriate DFT level also leads to converged clusters – not necessarily the smallest possible – at other levels of theory, including cWFT methods. As seen in Fig. 4 for bulk rutile  $TiO_2$ , the SCC (predicted from PBE calculations) shows small changes that are less than 0.03 eV (indicated by the gray error bars) in  $E_{Ov}$  compared to larger quantum clusters at all studied levels of theory, from HF to PBE0 and LMP2. We observe the same behavior in the MgO systems and  $TiO_2$  surface (see Sec. S11 of the supplementary material).

Whilst there is good agreement between all levels of theory in the quantum clusters larger than (and including) the SCC from the DFT calculation, their behavior can vary significantly for smaller unconverged clusters. To bypass the steep scaling of cWFT methods, it is common to employ the  $\Delta_{LL}^{HL}$  approach within the literature<sup>16,94,95</sup> to produce reference quantities. Here, the difference between a high-level (HL) and a low-level (LL) theory (e.g. DFT) for a small tractable

quantum cluster is added as a correction to the same low level theory computed at the bulk and basis set limits. The implicit underlying assumption is that the difference between the HL and LL methods stays the same regardless of cluster size. The differing size convergence of the various levels of theory suggests that such an assumption could exhibit uncontrolled errors if performed for arbitrarily small crystals.

### C. Reference Ov Formation Energy

Recent advances in localized orbital variants of CCSD(T) (e.g. LNO-CCSD(T)<sup>45</sup>, DLPNO-CCSD(T)<sup>96</sup>, PNO-LCCSD(T)<sup>97</sup>, etc.) have extended the remit of coupled cluster methods. To put these advances into perspective, for the def2-QZVPP basis set on a node equipped with 72 CPU cores, a PBE and PBE0 single-point calculation on the 17 Mg ion SCC for MgO surface took 1.5 and 2.5 hours respectively, whilst this time increased to only 7 hours for LNO-CCSD(T). Such a calculation would be far outside the reach of canonical CCSD(T), with the largest studied being a cluster with 6 Mg ions and 9 O anions<sup>16</sup>. These developments, alongside the identification of relatively small and converged clusters, enable us to compute Ov at the local CCSD(T) level with large basis sets. The O vacancy formation energy,  $E_{Ov}$ , computed with LNO-CCSD(T) in the O-rich limit for the four systems are  $7.68 \pm 0.15$  eV (MgO bulk),  $7.18 \pm 0.15$  eV (MgO (001) surface),  $6.39 \pm 0.15$  eV (TiO<sub>2</sub> bulk) and  $5.55 \pm 0.15$  eV (TiO<sub>2</sub> (110) surface). The decrease in  $E_{Ov}$  moving from bulk to surface can be expected due to the lowered coordination around the O anion sites on the surface.

Error bars have been added to the LNO-CCSD(T) values reported above. These have been conservatively estimated at 0.15 eV. This estimate comprises of errors arising due to: (i) basis set size and frozen core treatment; (ii) local approximation thresholds; and (iii) quantum cluster finite size errors. Tests on smaller clusters (see Sec. S4 B of the supplementary material) shows that the (i) CBS(TZVPP/QZVPP) basis and frozen core treatment (Ar on Ti and Ne on Mg) chosen give good agreement ( $\sim 0.1$  eV) w.r.t. a larger basis set and small frozen core (Ne on Ti and He on Mg). Based on the deviations observed in small clusters, our best estimates of  $E_{Ov}$  were corrected for the bias due to the basis set and frozen core treatment (see Sec. S5 of the supplementary material). The (ii) LNO threshold settings chosen have been validated against canonical CCSD(T) to give small errors of  $< 0.04$  eV (see Sec. S13 of supplementary). Finite size errors are expected to be less than 0.05 eV – the typical error found for DFT w.r.t. the bulk limit (in Fig. 3) as well as from

LMP2 calculations against larger clusters (see Sec. S11 of the supplementary material). Given the fast basis set convergence and all-electron nature of the embedded cluster DFT calculations, finite size errors (0.05 eV) are expected to be the dominant source of error.

#### D. Comparison of DFT Functionals

On top of enabling the accurate LNO-CCSD(T) method to be applied, the embedded cluster approach allows virtually all DFT XC functionals to be applied at low cost, including double-hybrids - normally computationally inaccessible for solid-state periodic DFT codes. We use our obtained LNO-CCSD(T) reference values to evaluate the performance for a range of DFT XC functionals, with their deviation from LNO-CCSD(T)  $E_{Ov}$  values plotted in Fig. 5. These errors, alongside the computed  $E_{Ov}$ , for all of the studied methods are summarized in Table I. For all 4 systems, the studied XC functionals underestimate  $E_{Ov}$  w.r.t. LNO-CCSD(T). This error decreases, on average, when using XC functionals on higher rungs, with meta-GGAs, hybrids and double-hybrids (DH) showing consistent improvement over the GGAs. The observed trends and variations of the XC functionals are quite consistent between the bulk and surface of the same system, but differs from one material to the next.

Out of the studied DFT XC functionals, the double-hybrid B2PLYP functional shows the best agreement to the LNO-CCSD(T) reference (mean absolute error, MAE, of 0.14 eV), with all points lying within the combined DFT and LNO-CCSD(T) error bars in Fig. 5. The  $\omega$ B97X<sup>98</sup> hybrid and SCAN<sup>99</sup> meta-GGA XC functionals give the next best performance, both with an MAE of 0.23 eV. In the hybrid functionals, B3LYP<sup>100</sup> also shows good (MAE of 0.32 eV) and consistent performance across the 4 systems. On the other hand, PBE0 as well as HSE06<sup>101</sup> – two common hybrid functionals for metal-oxide systems – give variable performance, having large errors (of around 0.7 eV) for the MgO systems, which lowers (to around 0.2 eV) in the TiO<sub>2</sub> systems. The GGAs all severely underestimate  $E_{Ov}$ , with MAE’s all exceeding 0.6 eV. The XC functionals in Fig. 5 are arranged based on decreasing MAE within their respective Jacob’s ladder rungs. XC functionals which incorporate the Becke-88 (B) exchange and Lee-Yang-Parr (LYP) correlation functionals, such as BLYP and B3LYP, are one of the top performers in their respective rungs, with B2PLYP being the best overall studied XC functional, as discussed previously.

Lower-level cWFT methods such as LMP2 and LNO-CCSD are automatically generated in any LNO-CCSD(T) calculation and they are also compared in Fig. 5. Both LMP2 and LNO-

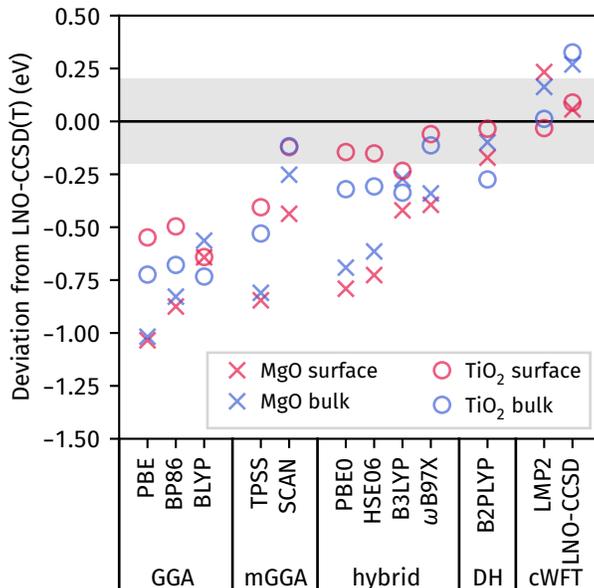


FIG. 5. Comparison of various DFT XC functionals, ranging from GGAs to double-hybrids, as well as LMP2 and LNO-CCSD to LNO-CCSD(T) reference calculations of the O vacancy formation energy ( $E_{Ov}$ ) in MgO bulk, MgO surface, TiO<sub>2</sub> bulk and TiO<sub>2</sub> surface. DFT calculations were performed at the def2-QZVPP level with the cWFT and double-hybrid (DH) calculations following the same procedure (discussed in the text) as the LNO-CCSD(T) calculations. The gray error bars are included to indicate the combined 0.2 eV error of the DFT (0.05 eV from finite size errors) and LNO-CCSD(T) (0.15 eV as discussed in Sec. III C) values.

CCSD show excellent agreement to LNO-CCSD(T), with MAEs of 0.11 and 0.19 eV respectively, and only the B2PLYP DFT XC functional has errors of a similar (small) size. As the errors of these three methods are all close to or within the error bars of the LNO-CCSD(T) values, it is not possible to ascertain which method performs better. We do find, however, that the good agreement of LMP2 appears to arise from fortuitous error cancellations, as elaborated in Sec. IV A.

## IV. DISCUSSION

The two key developments of this work are: (i) a protocol for obtaining small converged clusters for performing reference calculations of the Ov at high levels of theory; and (ii) the assessment of the accuracy of XC functionals for studying the  $E_{Ov}$ . It is important that both these developments are properly contextualized, either from physical theory or comparison to the literature. For

TABLE I.  $E_{\text{Ov}}$  (in eV) estimates at various levels of DFT XC functional approximations for the smallest converged clusters (SCCs) of rocksalt MgO and rutile TiO<sub>2</sub>, in their bulk and common surface planes. The final estimate of the B2PLYP, LMP2, LNO-CCSD and LNO-CCSD(T) methods are also given. These values are calculated using the O<sub>2</sub> binding energy of the corresponding electronic structure level (see Sec. S1 of the supplementary material) in Eq. 1. Errors are given w.r.t. the final LNO-CCSD(T) estimate. The computational details for these calculations are provided in the text.

	MgO bulk	Error	MgO surface	Error	TiO <sub>2</sub> bulk	Error	TiO <sub>2</sub> surface	Error	MAE
PBE	6.66	-1.02	6.15	-1.03	5.67	-0.72	5.00	-0.55	0.83
BP86	6.85	-0.83	6.31	-0.87	5.71	-0.68	5.06	-0.50	0.72
BLYP	7.11	-0.56	6.54	-0.64	5.66	-0.73	4.91	-0.64	0.64
TPSS	6.87	-0.81	6.34	-0.85	5.86	-0.53	5.15	-0.41	0.65
SCAN	7.43	-0.25	6.74	-0.44	6.27	-0.12	5.43	-0.12	0.23
PBE0	6.99	-0.69	6.39	-0.79	6.07	-0.32	5.41	-0.14	0.49
HSE06	7.06	-0.61	6.46	-0.73	6.08	-0.31	5.40	-0.15	0.45
B3LYP	7.41	-0.27	6.76	-0.42	6.05	-0.34	5.32	-0.23	0.32
$\omega$ B97X	7.34	-0.34	6.79	-0.39	6.28	-0.11	5.49	-0.06	0.23
B2PLYP	7.58	-0.10	7.01	-0.17	6.12	-0.27	5.52	-0.03	0.14
LMP2	7.84	0.16	7.41	0.23	6.40	0.01	5.52	-0.03	0.11
LNO-CCSD	7.95	0.27	7.24	0.06	6.72	0.33	5.64	0.09	0.19
LNO-CCSD(T)	7.68	–	7.18	–	6.39	–	5.55	–	
MAE		0.55		0.63		0.41		0.28	

the latter development, we will try to rationalize the observed trends in performances of the XC functionals in Sec. IV A. For the former, we will compare our LNO-CCSD(T) reference values to those in the literature in Sec. IV B.

### A. Origin of DFT Underestimation

Seeking to understand the relative performance of DFT XC functionals in complex systems, such as metal-oxides, is not straightforward. In particular,  $E_{\text{Ov}}$  is a quantity that depends on many factors. Nonetheless, we believe our results can reveal some useful insights on this property.

Systematic studies on a series of metal-oxides have shown that there is a correlation between the band gap and  $E_{\text{Ov}}$  in these systems<sup>48,102</sup>, as is confirmed by the larger  $E_{\text{Ov}}$  in MgO (with a PBE band gap of 4.83 eV in the bulk<sup>47</sup>) over TiO<sub>2</sub> (1.88 eV PBE band gap<sup>103</sup>) in our results. Physically, this correlation arises because the removal of an O atom leaves two electrons (originally occupying the O 2*p* band), which must redistribute by occupying an empty band from the conduction band<sup>10</sup>. This redistribution energy, and in turn  $E_{\text{Ov}}$ , would then be correlated with the size of the band gap of the crystal system.

Within the same crystal system, we find that the predicted DFT band gaps provides a rational basis for understanding the underestimation of  $E_{\text{Ov}}$  in most DFT XC functionals as well as the relative trends between the Jacob’s ladder rungs. More specifically, the errors in  $E_{\text{Ov}}$  w.r.t. reference methods should be related to the deviations of the XC functionals from the experimental or reference method band gap. Due to the presence of self-interaction error<sup>104</sup> (SIE), semilocal functionals (meta-GGAs and below) will underestimate the band gap, hence underestimating  $E_{\text{Ov}}$ , with meta-GGAs normally giving improved band gaps over GGAs<sup>105</sup>. Hybrid functionals correct for some of the SIE via the incorporation of exact exchange, with the possibility that band gaps are overestimated<sup>99,106</sup>. For both the rocksalt MgO and rutile TiO<sub>2</sub> systems, most of the studied hybrid functionals underestimate the band gap<sup>37,107</sup>, hence underestimating  $E_{\text{Ov}}$ .

The band gap can only serve as a general guide for the relative performance of DFT XC functionals and there are functionals which do not follow this trend, suggesting that there are other important factors which influence  $E_{\text{Ov}}$ . The key example is the PBE0 functional in rutile TiO<sub>2</sub>, which underestimates  $E_{\text{Ov}}$  despite predicting a band gap<sup>103</sup> of 4.05 eV that overestimates experimental electronic band gaps from photoemission experiments, typically in the range of 3.3–4.0 eV<sup>108–111</sup>. Additionally, despite predicting band gaps that are 0.6 and 1.0 eV higher than HSE06 for bulk MgO<sup>112,113</sup> and rutile TiO<sub>2</sub><sup>73,103</sup> respectively, these two functionals have very similar performances (both with MAE close to 0.5 eV) across the range of systems. Despite predicting worse band gaps<sup>114</sup>, the improved overall performance of the SCAN functional (with MAE of 0.23 eV) over many hybrid functionals is another example.

In GGAs, on top of the underestimated band gap due to the SIE, a major contribution to its errors also arises from a poor description of the binding energy of the O<sub>2</sub> molecule – they tend to predict strong overbinding (e.g. 0.50 eV/atom for PBE). It is common within the literature to correct for this error by replacing the GGA O<sub>2</sub> binding energy in Eq. 1 with the experimental binding energy of 5.22 eV<sup>115</sup>. With this change, the underestimation is decreased significantly,

TABLE II. The O vacancy formation energy ( $E_{\text{Ov}}$ ) in bulk MgO obtained through high-level theory methods and experiment from the literature and this study. The experiment and MP2 values were modified from their quoted values in the original papers to ensure the same definition of  $E_{\text{Ov}}$  as this paper. The difference between the reference method and PBE is also shown:  $E_{\text{Ov}}^{\text{Method}} - E_{\text{Ov}}^{\text{PBE}}$ , to enable comparison to the quoted Diffusion Monte Carlo (DMC) value. For values of  $E_{\text{Ov}}$  or  $E_{\text{Ov}}^{\text{Method}} - E_{\text{Ov}}^{\text{PBE}}$  which are not quoted in their original text, we fill the cell with “not available” (n/a). For further consistency with DMC and  $\Delta_{\text{PBE}}^{\text{CCSD(T)}}$  studies,  $E_{\text{Ov}}^{\text{PBE}}$  has been corrected for its overbinding by using the experimental binding energy of 5.22 eV<sup>115</sup> (see Sec. S14 of the supplementary material).

Method	Reference	$E_{\text{Ov}}^{\text{Method}}$ (eV)	$E_{\text{Ov}}^{\text{Method}} - E_{\text{Ov}}^{\text{PBE}}$ (eV)
Experiment	Kappers <i>et al.</i> <sup>118</sup>	9.29	n/a
DMC	Ertekin <i>et al.</i> <sup>47</sup>	n/a	0.5
$\Delta_{\text{PBE}}^{\text{CCSD(T)}}$	Richter <i>et al.</i> <sup>16</sup>	6.85	-0.09
MP2	Scorza <i>et al.</i> <sup>29</sup>	7.13	n/a
LNO-CCSD(T)	This work	7.68	0.52

with MAE decreases of the range of 0.3–0.5 eV for the GGAs (see Sec. S14 of the supplementary material). XC functionals on higher rungs are less impacted, with improvements in the MAE of less than 0.2 eV for the meta-GGAs and hybrids. We also find that the better performance of LMP2 over (the more accurate) LNO-CCSD method is fortuitous, arising from its overbinding of 0.22 eV/atom (in Sec. S1 of the supplementary material) and it exhibits a larger MAE of 0.34 eV relative to the 0.11 eV of LNO-CCSD when using the experimental binding energy.

## B. Comparison to Previous Work for MgO Bulk

As the prototypical system for studying the Ov in metal-oxides, bulk MgO has been subject to several studies involving high level reference methods<sup>16,29,47,116,117</sup> and out of the 4 studied systems, it is the only system, to our knowledge, where  $E_{\text{Ov}}$  has been experimentally determined<sup>118</sup>. Hence, bulk MgO makes for a good system to assess the accuracy of our obtained reference LNO-CCSD(T) values.

Experimental determination of  $E_{\text{Ov}}$  can be challenging due to the many experimental factors that can influence it. For MgO bulk, the single available value of 9.29 eV was obtained from

additive coloring experiments by Kappers *et al.*<sup>16,118</sup>. These experiments involve heating MgO crystals in Mg vapor under high temperatures and pressures. According to Smakula’s formula<sup>119</sup>, the measured optical absorption spectra at these temperatures allows for the determination of  $E_{Ov}$  from the relative Mg vapor and Ov concentrations. As seen in Table II, this value differs by  $> 1.5$  eV from our LNO-CCSD(T) values and other high-level methods. This discrepancy is likely attributed to the uncertainties arising in the experiment. For example, the oscillator strength obtained by Kappers *et al.* differs significantly ( $> 70$  %) compared to a previous experiment<sup>120</sup>. Richter *et al.*<sup>16</sup> have also attributed this discrepancy to thermal equilibrium not being reached in the experiment. Additionally, some of the assumptions in Smakula’s formula can be questionable for ionic solids such as MgO<sup>119</sup>. Beyond experimental uncertainties, some of this discrepancy can also arise due to the neglect of temperature effects in the static LNO-CCSD(T) calculations.

Given the above considerations, it is more appropriate to compare high-level references available for the  $E_{Ov}$  in MgO bulk. Ertekin and Grossman<sup>47</sup> have evaluated  $E_{Ov}$  with DMC for MgO bulk, finding the DMC value to be 0.5 eV higher than PBE. These  $E_{Ov}$  values were computed under Mg-rich conditions, where  $E_{Ov}$  is defined as:

$$E_{Ov} = E[\text{D-MO}] - E[\text{P-MO}] + E[\text{MgO}] - E[\text{Mg}], \quad (2)$$

with  $E[\text{MgO}]$  and  $E[\text{Mg}]$  being the total energy of bulk MgO and Mg per formula unit respectively. Compared to the O-rich limit in Eq. 1, this definition will not suffer from the poor O<sub>2</sub> binding description in  $E_{\text{bind}}$ . If the O<sub>2</sub> overbinding (0.50 eV/atom) is corrected in the PBE  $E_{Ov}$ , our LNO-CCSD(T) value also becomes 0.52 eV higher than PBE, agreeing with DMC values.  $\Delta_{\text{PBE}}^{\text{CCSD(T)}}$  embedded cluster calculations by Richter *et al.*<sup>16</sup> have obtained an  $E_{Ov}$  value of 6.85 eV, which is 0.83 eV smaller than our LNO-CCSD(T) estimate. In Sec. S15 of the supplementary material, we show that 0.2 eV of this difference can be attributed to differing lattice parameters and inclusion of structural relaxation around the Ov by performing  $\Delta_{\text{PBE}}^{\text{LNO-CCSD(T)}}$  calculations on the same quantum clusters as Richter *et al.*. Remaining differences could arise from the use of differing embedding environments, which our work (Fig. 4 and Sec. S11 A of the supplementary material), alongside others<sup>81</sup>, have shown cWFT methods to be highly sensitive to, requiring careful convergence. The reference  $E_{Ov}$  value from the  $\Delta_{\text{PBE}}^{\text{CCSD(T)}}$   $E_{Ov}$  study was also found to be smaller than the PBE (with overbinding corrected) value by 0.09 eV, which goes against trends observed from high-level calculations on other metal-oxide systems<sup>121,122</sup> as well as the DMC study of Ertekin *et al.*. Explicit MP2 calculations on a converged cluster<sup>29</sup> give a  $E_{Ov}$  of 7.13 eV, closer to our LNO-

CCSD(T) value. The difference w.r.t. our LNO-CCSD(T) or LMP2 values arise due to the use of a small (double-zeta quality) basis set; larger basis sets or CBS extrapolations can increase  $E_{\text{Ov}}$  significantly by  $> 0.5$  eV w.r.t. a double-zeta basis set for MgO bulk (see Sec. S4 of the supplementary material).

To our knowledge, the only available high level reference calculation in the  $\text{TiO}_2$  system comes from a DFT+GW study<sup>123</sup>, with the large number of electrons in Ti making application of methods such as canonical CCSD(T) highly expensive<sup>81</sup>. Additionally, DMC has not been applied to study the Ov in  $\text{TiO}_2$  potentially because the inclusion of a  $3d$  transition metal brings additional complications with the starting trial wave-function<sup>124</sup> and the need to validate its pseudopotentials<sup>125,126</sup>.

## V. CONCLUSION

In this work, we discuss a systematic and general approach (named SKZCAM after the authors' initials) for designing small converged quantum clusters for studying the O vacancy with the electrostatic embedded cluster method. When combined with localized orbital correlated wave-function theory methods such as LNO-CCSD(T), this approach allows for accurate determination of the Ov formation energy ( $E_{\text{Ov}}$ ) at a computationally tractable cost. We applied this approach to compute  $E_{\text{Ov}}$  values for the bulk and common surface planes of rocksalt MgO and rutile  $\text{TiO}_2$  systems. Comparison of these reference values to common DFT XC functionals shows that the studied XC functionals underestimate  $E_{\text{Ov}}$  for all studied systems. We observe general improvements in the XC functional errors as Jacob's ladder is ascended, which we find can be correlated to the improvements in the predicted band gaps of the XC functionals. Of the XC functionals studied, the double-hybrid B2PLYP functional gives the best performance, with a mean absolute error within the error bars of the reference calculation. Other BLYP-based functionals, such as BLYP and B3LYP are also found to perform well within their respective Jacob's ladder rungs, alongside the meta-GGA SCAN and the hybrid  $\omega$ B97X functionals.

Although this work focuses on the Ov formation energy, the simple and intuitive nature of the outlined protocol should allow its application to other chemical problems, including molecular adsorption, spectroscopic quantities or complex (ternary) metal-oxide systems. Furthermore, the approach can be automated, making it amenable for integration into existing high-throughput calculation frameworks. Such high-throughput frameworks<sup>127</sup> for point defects or molecular adsorbates are highly desirable. For example, it can be used to screen for target applications in

catalysis<sup>11,102</sup> or to produce large reference databases to validate current XC functionals<sup>128</sup>. Beyond the applications, we have defined a rigorous and well-founded framework for controlling the shape of the embedded quantum clusters. This framework can be valuable in not only defining the quantum cluster in electrostatic embedded cluster approaches, but also in other embedding approaches<sup>20,129–132</sup>, where multiple clusters, corresponding to different levels of theory, may have to be defined simultaneously.

## **SUPPLEMENTARY MATERIAL**

See the supplementary material for a detailed compilation of the obtained results as well as further data and analysis to support the points made throughout the text.

## **ACKNOWLEDGMENTS**

We are grateful for: resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service ([www.csd3.cam.ac.uk](http://www.csd3.cam.ac.uk)), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council ([www.dirac.ac.uk](http://www.dirac.ac.uk)); computational resources granted by the UCL Myriad and Kathleen High Performance Computing Facility (Myriad@UCL and Kathleen@UCL), and associated support service; computational support from the UK Materials and Molecular Modelling Hub, which is partially funded by EPSRC (EP/P020194 and EP/T022213); and computational support from the UK national high performance computing service, ARCHER 2. Both the UK Materials and Molecular Modelling Hub and ARCHER 2 access was obtained via the UKCP consortium and funded by EPSRC grant ref EP/P022561/1.

## **DATA AVAILABILITY**

The data that supports the findings of this study are available within the article and its supplementary material. The input and output files associated with this study and all analysis can be found on GitHub at [benshi97/Data\\_Embedded\\_Cluster\\_Protocol](https://github.com/benshi97/Data_Embedded_Cluster_Protocol) and [Binder](#). The scripts used to generate the quantum clusters with the SKZCAM approach is also provided in a separate GitHub repository at [benshi97/SKZCAM-Protocol](https://github.com/benshi97/SKZCAM-Protocol) and [Binder](#).

## REFERENCES

- <sup>1</sup>S. B. Adler, *Chem. Rev.* **104**, 4791–4844 (2004).
- <sup>2</sup>M. Grätzel, *Nature* **414**, 338–344 (2001).
- <sup>3</sup>B. Wang, W. Huang, L. Chi, M. Al-Hashimi, T. J. Marks, and A. Facchetti, *Chem. Rev.* **118**, 5690–5754 (2018).
- <sup>4</sup>A. Ruiz Puigdollers, P. Schlexer, S. Tosoni, and G. Pacchioni, *ACS Catal.* **7**, 6493–6513 (2017).
- <sup>5</sup>R. Asahi, T. Morikawa, T. Ohwaki, K. Aoki, and Y. Taga, *Science* **293**, 269–271 (2001).
- <sup>6</sup>S. Kim, S. N. Hood, J.-S. Park, L. D. Whalley, and A. Walsh, *J. Phys. Energy* **2**, 036001 (2020).
- <sup>7</sup>X. Pan, M.-Q. Yang, X. Fu, N. Zhang, and Y.-J. Xu, *Nanoscale* **5**, 3601–3614 (2013).
- <sup>8</sup>D. Lee, B. Chung, Y. Shi, G.-Y. Kim, N. Campbell, F. Xue, K. Song, S.-Y. Choi, J. P. Podkaminer, T. H. Kim, P. J. Ryan, J.-W. Kim, T. R. Paudel, J.-H. Kang, J. W. Spinuzzi, D. A. Tenne, E. Y. Tsybal, M. S. Rzechowski, L. Q. Chen, J. Lee, and C. B. Eom, *Science* (2018).
- <sup>9</sup>B. Meredig and C. Wolverton, *Phys. Rev. B* **80**, 245119 (2009).
- <sup>10</sup>A. M. Deml, V. Stevanović, C. L. Muhich, C. B. Musgrave, and R. O’Hayre, *Energy Environ. Sci.* **7**, 1996–2004 (2014).
- <sup>11</sup>Y. Hinuma, T. Toyao, T. Kamachi, Z. Maeno, S. Takakusagi, S. Furukawa, I. Takigawa, and K.-i. Shimizu, *J. Phys. Chem. C* **122**, 29435–29444 (2018).
- <sup>12</sup>G. Kumar, S. L. J. Lau, M. D. Krcha, and M. J. Janik, *ACS Catal.* **6**, 1812–1821 (2016).
- <sup>13</sup>H.-Y. Su, X. Ma, K. Sun, C. Sun, Y. Xu, and F. Calle-Vallejo, *Chem. Sci.* **11**, 4119–4124 (2020).
- <sup>14</sup>F. Gunkel, D. V. Christensen, Y. Z. Chen, and N. Pryds, *Appl. Phys. Lett.* **116**, 120505 (2020).
- <sup>15</sup>M. V. Ganduglia-Pirovano, A. Hofmann, and J. Sauer, *Surf. Sci. Rep.* **62**, 219–270 (2007).
- <sup>16</sup>N. A. Richter, S. Siculo, S. V. Levchenko, J. Sauer, and M. Scheffler, *Phys. Rev. Lett.* **111**, 045502 (2013).
- <sup>17</sup>J. P. W. Wellington, B. E. Tegner, J. Collard, A. Kerridge, and N. Kaltsoyannis, *J. Phys. Chem. C* **122**, 7149–7165 (2018).
- <sup>18</sup>C. Di Valentin, G. Pacchioni, and A. Selloni, *Phys. Rev. Lett.* **97**, 166803 (2006).
- <sup>19</sup>T. Shibuya, K. Yasuoka, S. Mirbt, and B. Sanyal, *J. Phys.: Condens. Matter* **24**, 435504 (2012).
- <sup>20</sup>N. Bernstein, J. R. Kermode, and G. Csányi, *Rep. Prog. Phys.* **72**, 026501 (2009).
- <sup>21</sup>G. Pacchioni, *J. Chem. Phys.* **128**, 182505 (2008).
- <sup>22</sup>T. Bredow and G. Pacchioni, *Chem. Phys. Lett.* **355**, 417–423 (2002).
- <sup>23</sup>J. Carrasco, N. Lopez, C. Sousa, and F. Illas, *Phys. Rev. B* **72**, 054109 (2005).

- <sup>24</sup>S. C. Ammal and A. Heyden, *J. Chem. Phys.* **133**, 164703 (2010).
- <sup>25</sup>P. V. Sushko, A. L. Shluger, and C. R. A. Catlow, *Surf. Sci.* **450**, 153–170 (2000).
- <sup>26</sup>D. Berger, H. Oberhofer, and K. Reuter, *Phys. Rev. B* **92**, 075308 (2015).
- <sup>27</sup>V. A. Nasluzov, E. A. Ivanova, A. M. Shor, G. N. Vayssilov, U. Birkenheuer, and N. Rösch, *J. Phys. Chem. B* **107**, 2228–2241 (2003).
- <sup>28</sup>K. M. Neyman, C. Inntam, A. V. Matveev, V. A. Nasluzov, and N. Rösch, *J. Am. Chem. Soc.* **127**, 11652–11660 (2005).
- <sup>29</sup>E. Scorza, U. Birkenheuer, and C. Pisani, *J. Chem. Phys.* **107**, 9645–9658 (1997).
- <sup>30</sup>X. Xu, H. Nakatsuji, X. Lü, M. Ehara, Y. Cai, N. Q. Wang, and Q. E. Zhang, *Theor. Chem. Acc.* **102**, 170–179 (1999).
- <sup>31</sup>X. Lü, X. Xu, N. Wang, and Q. Zhang, *Int. J. Quantum Chem.* **73**, 377–386 (1999).
- <sup>32</sup>M. Kick and H. Oberhofer, *J. Chem. Phys.* **151**, 184114 (2019).
- <sup>33</sup>D. O. Scanlon, C. W. Dunnill, J. Buckeridge, S. A. Shevlin, A. J. Logsdail, S. M. Woodley, C. R. A. Catlow, M. J. Powell, R. G. Palgrave, I. P. Parkin, G. W. Watson, T. W. Keal, P. Sherwood, A. Walsh, and A. A. Sokol, *Nat. Mater.* **12**, 798–801 (2013).
- <sup>34</sup>A. Dittmer, G. L. Stoychev, D. Maganas, A. A. Auer, and F. Neese, *J. Chem. Theory Comput.* **16**, 6950–6967 (2020).
- <sup>35</sup>F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, *J. Chem. Phys.* **152**, 224108 (2020).
- <sup>36</sup>D. Krasikov, V. Mikhailin, and A. Scherbinin, *Phys. Procedia* **2**, 559–565 (2009).
- <sup>37</sup>A. Dittmer, R. Izsák, F. Neese, and D. Maganas, *Inorg. Chem.* **58**, 9303–9315 (2019).
- <sup>38</sup>E. B. Kadossov, K. J. Gaskell, and M. A. Langell, *J. Comput. Chem.* **28**, 1240–1251 (2007).
- <sup>39</sup>P. S. Bagus, M. J. Sassi, and K. M. Rosso, *J. Chem. Phys.* **151**, 044107 (2019).
- <sup>40</sup>P. Liao and E. A. Carter, *J. Phys. Chem. C* **115**, 20795–20805 (2011).
- <sup>41</sup>Z. Rolik and M. Kállay, *J. Chem. Phys.* **135**, 104111 (2011).
- <sup>42</sup>Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, and M. Kállay, *J. Chem. Phys.* **139**, 094105 (2013).
- <sup>43</sup>P. R. Nagy and M. Kállay, *J. Chem. Phys.* **146**, 214106 (2017).
- <sup>44</sup>P. R. Nagy, G. Samu, and M. Kállay, *J. Chem. Theory Comput.* **14**, 4193–4215 (2018).
- <sup>45</sup>P. R. Nagy and M. Kállay, *J. Chem. Theory Comput.* **15**, 5275–5298 (2019).
- <sup>46</sup>R. Orlando, R. Millini, G. Perego, and R. Dovesi, *J. Mol. Catal. A: Chem.* **119**, 253–262 (1997).
- <sup>47</sup>E. Ertekin, L. K. Wagner, and J. C. Grossman, *Phys. Rev. B* **87**, 155210 (2013).

- <sup>48</sup>Z. Helali, A. Jedidi, O. A. Syzgantseva, M. Calatayud, and C. Minot, *Theor. Chem. Acc.* **136**, 100 (2017).
- <sup>49</sup>P. Fuentealba, L. von Szentpaly, H. Preuss, and H. Stoll, *J. Phys. B: Atom. Mol. Phys.* **18**, 1287–1296 (1985).
- <sup>50</sup>M. Dolg, U. Wedig, H. Stoll, and H. Preuss, *J. Chem. Phys.* **86**, 866–872 (1987).
- <sup>51</sup>Y. Lu, M. R. Farrow, P. Fayon, A. J. Logsdail, A. A. Sokol, C. R. A. Catlow, P. Sherwood, and T. W. Keal, *J. Chem. Theory Comput.* **15**, 1317–1328 (2019).
- <sup>52</sup>M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégyel, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and Á. Ganyecz, *J. Chem. Phys.* **152**, 074107 (2020).
- <sup>53</sup>S. Lehtola, C. Steigemann, M. J. T. Oliveira, and M. A. L. Marques, *SoftwareX* **7**, 1–5 (2018).
- <sup>54</sup>F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
- <sup>55</sup>F. Weigend, *J. Comput. Chem.* **29**, 167–175 (2008).
- <sup>56</sup>F. Weigend, *Phys. Chem. Chem. Phys.* **8**, 1057–1065 (2006).
- <sup>57</sup>P. R. Nagy, G. Samu, and M. Kállay, *J. Chem. Theory Comput.* **12**, 4897–4914 (2016).
- <sup>58</sup>S. Grimme and F. Neese, *J. Chem. Phys.* **127**, 154116 (2007).
- <sup>59</sup>F. Neese and E. F. Valeev, *J. Chem. Theory Comput.* **7**, 33–43 (2011).
- <sup>60</sup>G. Bistoni, C. Riplinger, Y. Minenkov, L. Cavallo, A. A. Auer, and F. Neese, *J. Chem. Theory Comput.* **13**, 3220–3227 (2017).
- <sup>61</sup>K. A. Peterson and T. H. Dunning, *J. Chem. Phys.* **117**, 10548–10560 (2002).
- <sup>62</sup>N. B. Balabanov and K. A. Peterson, *J. Chem. Phys.* **123**, 064107 (2005).
- <sup>63</sup>R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.* **96**, 6796–6806 (1992).
- <sup>64</sup>F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, *Chem. Phys. Lett.* **294**, 143–152 (1998).
- <sup>65</sup>A. Hellweg, C. Hättig, S. Höfener, and W. Klopper, *Theor. Chem. Acc.* **117**, 587–597 (2007).
- <sup>66</sup>G. L. Stoychev, A. A. Auer, and F. Neese, *J. Chem. Theory Comput.* **13**, 554–562 (2017).
- <sup>67</sup>S. Lehtola, *J. Chem. Theory Comput.* **17**, 6886–6900 (2021).
- <sup>68</sup>G. Kresse and J. Furthmüller, *Comput. Mater. Sci.* **6**, 15–50 (1996).
- <sup>69</sup>G. Kresse and J. Furthmüller, *Phys. Rev. B* **54**, 11169–11186 (1996).
- <sup>70</sup>J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, *J. Phys. Chem. Lett.* **11**, 8208–8215 (2020).
- <sup>71</sup>J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- <sup>72</sup>C. Adamo and V. Barone, *J. Chem. Phys.* **110**, 6158–6170 (1999).

- <sup>73</sup>A. Janotti, J. B. Varley, P. Rinke, N. Umezawa, G. Kresse, and C. G. Van de Walle, *Phys. Rev. B* **81**, 085212 (2010).
- <sup>74</sup>C. M. Yim, C. L. Pang, and G. Thornton, *Phys. Rev. Lett.* **104**, 036806 (2010).
- <sup>75</sup>M. A. Henderson, W. S. Epling, C. H. F. Peden, and C. L. Perkins, *J. Phys. Chem. B* **107**, 534–545 (2003).
- <sup>76</sup>M. Setvin, C. Franchini, X. Hao, M. Schmid, A. Janotti, M. Kaltak, C. G. Van de Walle, G. Kresse, and U. Diebold, *Phys. Rev. Lett.* **113**, 086402 (2014).
- <sup>77</sup>A. M. Stoneham, J. Gavartin, A. L. Shluger, A. V. Kimmel, D. M. Ramo, H. M. Rønnow, G. Aepli, and C. Renner, *J. Phys.: Condens. Matter* **19**, 255208 (2007).
- <sup>78</sup>S. Yang, L. E. Halliburton, A. Manivannan, P. H. Bunton, D. B. Baker, M. Klemm, S. Horn, and A. Fujishima, *Appl. Phys. Lett.* **94**, 162114 (2009).
- <sup>79</sup>E. Yagi, R. R. Hasiguti, and M. Aono, *Phys. Rev. B* **54**, 7945–7956 (1996).
- <sup>80</sup>M. K. Nowotny, T. Bak, and J. Nowotny, *J. Phys. Chem. B* **110**, 16270–16282 (2006).
- <sup>81</sup>J. Chen, N. A. Bogdanov, D. Usvyat, W. Fang, A. Michaelides, and A. Alavi, *J. Chem. Phys.* **153**, 204704 (2020).
- <sup>82</sup>G. Mattioli, F. Filippone, P. Alippi, and A. Amore Bonapasta, *Phys. Rev. B* **78**, 241201 (2008).
- <sup>83</sup>P. Deák, B. Aradi, and T. Frauenheim, *Phys. Rev. B* **86**, 195206 (2012).
- <sup>84</sup>N. A. Deskins, R. Rousseau, and M. Dupuis, *J. Phys. Chem. C* **115**, 7562–7572 (2011).
- <sup>85</sup>C. Franchini, M. Reticcioli, M. Setvin, and U. Diebold, *Nat. Rev. Mater.* **6**, 560–586 (2021).
- <sup>86</sup>S. Chrétien and H. Metiu, *J. Phys. Chem. C* **115**, 4696–4705 (2011).
- <sup>87</sup>D. Berger, A. J. Logsdail, H. Oberhofer, M. R. Farrow, C. R. A. Catlow, P. Sherwood, A. A. Sokol, V. Blum, and K. Reuter, *J. Chem. Phys.* **141**, 024105 (2014).
- <sup>88</sup>A. Kubas, D. Berger, H. Oberhofer, D. Maganas, K. Reuter, and F. Neese, *J. Phys. Chem. Lett.* **7**, 4207–4212 (2016).
- <sup>89</sup>J. Buckeridge, K. T. Butler, C. R. A. Catlow, A. J. Logsdail, D. O. Scanlon, S. A. Shevlin, S. M. Woodley, A. A. Sokol, and A. Walsh, *Chem. Mater.* **27**, 3844–3851 (2015).
- <sup>90</sup>D. Koch and S. Manzhos, *J. Phys. Chem. Lett.* **8**, 1593–1598 (2017).
- <sup>91</sup>T. Bredow, L. Giordano, F. Cinquini, and G. Pacchioni, *Phys. Rev. B* **70**, 035419 (2004).
- <sup>92</sup>K. J. Hameeuw, G. Cantele, D. Ninno, F. Trani, and G. Iadonisi, *J. Chem. Phys.* **124**, 024708 (2006).
- <sup>93</sup>L.-M. Liu, C. Zhang, G. Thornton, and A. Michaelides, *Phys. Rev. B* **82**, 161415 (2010).
- <sup>94</sup>A. D. Boese and J. Sauer, *Phys. Chem. Chem. Phys.* **15**, 16481–16493 (2013).

- <sup>95</sup>T. Tsatsoulis, F. Hummel, D. Usvyat, M. Schütz, G. H. Booth, S. S. Binnie, M. J. Gillan, D. Alfè, A. Michaelides, and A. Grüneis, *J. Chem. Phys.* **146**, 204108 (2017).
- <sup>96</sup>C. Riplinger and F. Neese, *J. Chem. Phys.* **138**, 034106 (2013).
- <sup>97</sup>Q. Ma and H.-J. Werner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1371 (2018).
- <sup>98</sup>J.-D. Chai and M. Head-Gordon, *J. Chem. Phys.* **128**, 084106 (2008).
- <sup>99</sup>J. Sun, A. Ruzsinszky, and J. P. Perdew, *Phys. Rev. Lett.* **115**, 036402 (2015).
- <sup>100</sup>A. D. Becke, *J. Chem. Phys.* **98**, 5648–5652 (1993).
- <sup>101</sup>J. Heyd, G. E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **118**, 8207–8215 (2003).
- <sup>102</sup>A. M. Deml, A. M. Holder, R. P. O’Hayre, C. B. Musgrave, and V. Stevanović, *J. Phys. Chem. Lett.* **6**, 1948–1953 (2015).
- <sup>103</sup>L. Chiodo, J. M. García-Lastra, A. Iacomino, S. Ossicini, J. Zhao, H. Petek, and A. Rubio, *Phys. Rev. B* **82**, 045207 (2010).
- <sup>104</sup>J. L. Bao, L. Gagliardi, and D. G. Truhlar, *J. Phys. Chem. Lett.* **9**, 2353–2358 (2018).
- <sup>105</sup>P. Borlido, J. Schmidt, A. W. Huran, F. Tran, M. A. L. Marques, and S. Botti, *npj Comput. Mater.* **6**, 1–17 (2020).
- <sup>106</sup>P. Borlido, T. Aull, A. W. Huran, F. Tran, M. A. L. Marques, and S. Botti, *J. Chem. Theory Comput.* **15**, 5069–5079 (2019).
- <sup>107</sup>F. Viñes, O. Lamiel-García, K. Chul Ko, J. Yong Lee, and F. Illas, *J. Comput. Chem.* **38**, 781–789 (2017).
- <sup>108</sup>R. G. Breckenridge and W. R. Hosler, *Phys. Rev.* **91**, 793–802 (1953).
- <sup>109</sup>Y. Tezuka, S. Shin, T. Ishii, T. Ejima, S. Suzuki, and S. Sato, *J. Phys. Soc. Jpn.* **63**, 347–357 (1994).
- <sup>110</sup>P. J. Hardman, G. N. Raikar, C. A. Muryn, G. van der Laan, P. L. Wincott, G. Thornton, D. W. Bullett, and P. A. D. M. A. Dale, *Phys. Rev. B* **49**, 7170–7177 (1994).
- <sup>111</sup>S. Rangan, S. Katalinic, R. Thorpe, R. A. Bartynski, J. Rochford, and E. Galoppini, *J. Phys. Chem. C* **114**, 1139–1147 (2010).
- <sup>112</sup>A. J. Garza and G. E. Scuseria, *J. Phys. Chem. Lett.* **7**, 4165–4170 (2016).
- <sup>113</sup>D. Fritsch, B. J. Morgan, and A. Walsh, *Nanoscale Res. Lett.* **12**, 19 (2017).
- <sup>114</sup>Y. Zhang, J. W. Furness, B. Xiao, and J. Sun, *J. Chem. Phys.* **150**, 014105 (2019).
- <sup>115</sup>D. Feller and K. A. Peterson, *J. Chem. Phys.* **110**, 8384–8396 (1999).
- <sup>116</sup>A. Gallo, F. Hummel, A. Irmeler, and A. Grüneis, *J. Chem. Phys.* **154**, 064106 (2021).

- <sup>117</sup>P. Rinke, A. Schleife, E. Kioupakis, A. Janotti, C. Rödl, F. Bechstedt, M. Scheffler, and C. G. Van de Walle, *Phys. Rev. Lett.* **108**, 126404 (2012).
- <sup>118</sup>L. A. Kappers, R. L. Kroes, and E. B. Hensley, *Phys. Rev. B* **1**, 4151–4157 (1970).
- <sup>119</sup>D. L. Dexter, *Phys. Rev.* **101**, 48–55 (1956).
- <sup>120</sup>Y. Chen, J. L. Kolopus, and W. A. Sibley, *Phys. Rev.* **186**, 865–870 (1969).
- <sup>121</sup>J. A. Santana, J. T. Krogel, J. Kim, P. R. C. Kent, and F. A. Reboredo, *J. Chem. Phys.* **142**, 164705 (2015).
- <sup>122</sup>J. A. Santana, J. T. Krogel, P. R. C. Kent, and F. A. Reboredo, *J. Chem. Phys.* **147**, 034701 (2017).
- <sup>123</sup>A. Malashevich, M. Jain, and S. G. Louie, *Phys. Rev. B* **89**, 075205 (2014).
- <sup>124</sup>J. Kolorenč, S. Hu, and L. Mitas, *Phys. Rev. B* **82**, 115108 (2010).
- <sup>125</sup>J. T. Krogel, J. A. Santana, and F. A. Reboredo, *Phys. Rev. B* **93**, 075143 (2016).
- <sup>126</sup>A. Annaberdiyev, G. Wang, C. A. Melton, M. C. Bennett, L. Shulenburger, and L. Mitas, *J. Chem. Phys.* **149**, 134108 (2018).
- <sup>127</sup>Y. Kumagai, N. Tsunoda, A. Takahashi, and F. Oba, *Phys. Rev. Materials* **5**, 123803 (2021).
- <sup>128</sup>J. Řezáč and P. Hobza, *Chem. Rev.* **116**, 5038–5071 (2016).
- <sup>129</sup>S. J. R. Lee, M. Welborn, F. R. Manby, and T. F. Miller, *Acc. Chem. Res.* **52**, 1359–1368 (2019).
- <sup>130</sup>T. Schäfer, F. Libisch, G. Kresse, and A. Grüneis, *J. Chem. Phys.* **154**, 011101 (2021).
- <sup>131</sup>K. Yu, F. Libisch, and E. A. Carter, *J. Chem. Phys.* **143**, 102806 (2015).
- <sup>132</sup>B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay, *J. Chem. Phys.* **145**, 064107 (2016).