Computational Design of Nanoclusters by Property-Based Genetic Algorithms: Tuning the Electronic Properties of $(TiO_2)_n$ Clusters

Saswata Bhattacharya[1], Benjamin H. Sonin[2], Christopher J. Jumonville[2],
Luca M. Ghiringhelli[1*], Noa Marom[2†]
[1] *Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany and*
[2] *Physics and Engineering Physics, Tulane University, New Orleans, Louisiana 70118, USA*
(Dated: January 22, 2015)

Supplemental Material

Contents

I. Computational details: 2
A. Implementation of *propery based cascade* Genetic Algorithm: 2
B. Validation and comparison of DFT-GA at *lower* and *higher-level* settings: 3
C. Post-process of the best fitted DFT-GA isomers using *higher-level* settings: 3
D. Validation of PBE0 delta-SCF energetics w.r.t G_0W_0@PBE0: 5

II. No. of dangling-O vs no. of Ti^{3+} in $(TiO_2)_n$ clusters: 5

III. Ti-O pair correlation function (PCF) of $(TiO_2)_n$ clusters: 9

IV. Low Gap ($\leq$ 5.5 eV) vs dangling-O and connectivity 13

V. Coordinates of all isomers at all sizes and comparison with previously published structures: 14

References 14

* corresponding author (LMG): ghiringhelli@fhi-berlin.mpg.de
† corresponding author (NM): nmarom@tulane.edu

I. COMPUTATIONAL DETAILS:

A. Implementation of *propery based cascade* Genetic Algorithm:

We investigate whether electronic properties other than total energy are associated with specific structural features and with the presence of active sites. This necessitates unbiased sampling of the potential energy surface (PES) of atomic clusters at the nano- and subnano-scale, we have implemented a suite of three massively-parallel cascade genetic algorithms (GA) schemes. The first scheme, as described in detail in Ref. [1, 2], is our recent implementation of a cascade genetic algorithm, where the minimized quantity is the total energy (Energy-GA, EGA), which is the conventional minimized quantity for atomic structures. For the second and third schemes, the fitness function is defined such that the largest VEA (vertical electron affinity) - viz., lowest VIP (vertical ionization potential) - is searched. As a test system, we have chosen $(TiO_2)_n$ clusters with n=2-10,15,20.

The strategy of a GA algorithm is to optimize structures by simulating an evolutionary process. First, local optimizations are performed for a randomly generated initial population of structures. The fitness of those structure is then evaluated with respect to a (scalar) property to be globally optimized. Here the property is total energy, VEA, VIP, in EGA, VEA-GA, VIP-GA schemes, respectively. Then two parent structures are selected using a roulette-wheel selection criterion [3] where the larger is the fitness value, the higher is the probability of getting selected. A trial structure is created by combining two fragments of the selected structures. In this mating step, the crossover operator takes care of combining two such fragments from the parent clusters. It is implemented as a modified version of the cut-and-splice crossover operator of Deaven and Ho. [4] Crossover generates a child, and following that mutation is introduced. Amongst many different possible mutation schemes, here we have only adopted a random translation between the two halves of the parent clusters, if atoms coming from the two different parents are too close after splicing of the two halves. After that the child structure is locally optimized (either using a forcefield or DFT with semi-local functional) and a new structure is obtained. To decide whether the newly found structure was already seen previously during the GA scan, after the local optimization we i) compare the total energy of the new structure with that of all the others seen before and ii) use a criterion based on the distances between all the atoms' pairs. We do this by constructing a coarse grained pair-correlation function (PCF) of the clusters, consisting of n bins conveniently spaced. (see details as in [2]). If the newly found structure was unseen before, we add this in the existing pool and the cycle is repeated until convergence is achieved. It should be noted here for any search performed in a multidimensional configuration space, it is generally almost impossible to guarantee that the global optimum (GO) is found. Therefore, we use an operative definition of convergence (see also Ref. [2]). After a transient initial number of GA cycles, in which the fittest structure (the running GO) changes rapidly, we reach a regime when only rarely a new running GO is found. In this regime, we always keep track of the total number of GA cycles needed to find the running GO and we make sure to run at least as many cycles further, before concluding that a new GO is unlikely be found and the running GO is confirmed as the searched GO structure. It should be noted here that to facilitate transitions between basins during the GA search, the population diversity is always maintained by occasionally selecting some unfit structure for mating. [2]

The GA algorithm is massively parallelized. The operation of selecting two structures from a genetic pool for the mating and the subsequent local optimization of the child, can be performed absolutely independently at any moment. Thus the scaling with respect to the replicas is almost linear, because the replicas remain independent for the most of the time and only at the beginning and at the end of each local optimization, the information is exchanged among the replicas. Therefore, first level of parallelization is performed within the FHI-aims code [5], by means of the MPI environment. [6] The second level of parallization is script based. In total $n \times p$ number of cores is equally divided into n replicas, where all the replicas perform independently a cycle of local optimizations each using p cores. The algorithm is, therefore, suitable for a very efficient parallelization. This second level of parallelization scales practically linearly with the number of cores, as there is no idling time between subsequent local-optimization cycles (a replica does not need to wait for the other replica before launching the next local optimization).

It's important to note that, the GA has the important advantage that the fitness function can be optimized to be any property of our interest, and not necessarily to be the conventionally used total or free energy.
We adopt our recent implementation (namely, cascade GA approach as in Ref. [1, 2]] to scan the configuration space of $(TiO_2)_n$ clusters. The term "cascade" refers to a multi-stepped procedure where successive steps employ higher level of theory and each of the next level takes information obtained at the immediate lower level. Therefore, an exhaustive GA pre-scanning of large number of possible structures is done by means of a computationally inexpensive, albeit rather sophisticated, reactive force field (ReaxFF). [7, 8] Structures found within a certain energy window from the energy global minimum (GM) are used as initial GA pool to start a fully quantum mechanical GA based

on density functional theory (DFT). The energy window depends on the size of the system (i.e.; number of atoms) and is usually taken as sufficiently large (4 eV) to generate a diverse initial pool for the DFT-based GA. We choose this window in such a way that the initial pool for DFT-GA is comprising a few tens (for a small system of ≤ 10 atoms) to several hundred structures (for a medium to large system ≥ 40 atoms). As shown in Ref. [2], it is generally unjustified to use only the low-energy structures found via a force field based GA as representative of the low-energy DFT structures, even if subsequent geometry optimization is performed using DFT. We strongly recommend, a complete DFT-based GA optimization should be performed with the structures generated by the force field based GA.

As the cascade starts to the DFT level, the balance between accuracy and efficiency is extremely important. The probability of selection for mating depends on the fitness function and the fitness function depends on the accuracy of the electronic properties evaluated at the DFT-level. It is, therefore, important to accurately calculate the electronic properties used to define the fitness function. It has been shown that the energy hierarchy of structures and electronic properties, such as the VEA/VIP, are sensitive to the employed DFT functional (or higher level theoretical approach). [2, 9] However, it would be computationally impractical to perform the whole DFT-GA scan at the highest level of theory. In view of this, the lower-accuracy steps of the cascade algorithm becomes important that provides considerable prescreening of structures, thereby saving a lot of computer time, without compromising the accuracy of the fitness function evaluation in the more promising regions of the PES.

All DFT calculations are performed with the all electron numeric atom-centered orbitals (NAO) code FHI-aims. [5] At the DFT level, the cascade flow continues with two sub-steps. Initial local optimizations are performed with the Perdew-Burke-Ernzerhof (PBE) [10, 11] functional using *lower-level settings*[16]. Structures that are similar to structures already in the pool or outside an energy window of 2 eV above the running GM are not considered for the next cascade step. However, the new structures found within 2 eV energy window from the running GM, are taken to the next level of cascade where single point total energy of the cluster is estimated using the PBE-based hybrid functional (PBE0). [12] This latter energy is used to calculate the fitness function of that cluster and is included in the GA-pool. For the electronic property based GAs at this stage the vertical VEA/VIP are evaluated from the total energy difference between the neutral species and the anion/cation, using PBE0 functional. These VEA/VIP values are used to estimate the fitness function. Based on Ref. [9], isomers with energy as high as 1.25 eV above the GM may have been observed in experiments, in which TiO_2 clusters were formed by laser vaporisation. Restricting the search window to 2 eV would not eliminate any isomer that can conceivably form in such experiments. Note that this 2 eV cut-off window is set at the PBE *lower-level settings* (at the DFT-cascade: step1) and after single point total energy calculation using PBE0 at the next cascade level (DFT-cascade: step2), we found that this window gets changed by 2.5 eV from the EGM. This is shown in Fig. 1 (bottom panel) of the main paper. In this figure, the relative energies, VEAs, and VIPs were obtained at the lower-level settings using PBE0 functional and this reflects the level of theory at which the fitness function is evaluated. The VEAs, VIPs, and gaps shown in Fig. 2-4 (main paper) were obtained at the higher-level settings using PBE0, reflecting the final post-processing step as explained in the section I(c).

B. Validation and comparison of DFT-GA at *lower* and *higher-level* settings:

It should be noted here that in Ref. [2] we have employed *higher-level* settings at the final cascade step of DFT-GA scan, after a sub-step performed at *lower-level* settings. On the contrary, in this present work we have performed the entire DFT-GA scan only at the *lower-level* settings and no *higher-level* settings are employed inside cascade steps of DFT-GA.[17] This is because in the present system the *hierarchy* of electronic properties of different isomers evaluated at the *lower-level* settings follows exactly similar *pattern* as that of evaluated at *higher-level* settings [see Fig. 1]. The *lower-level* settings give the correct hierarchy of the VEA/VIP, although the absolute values may change by 0.1-0.5 eV upon switching to the *higher-level* settings. We note that, for the GA, the *hierarchy* of the globally optimized quantity is more important than the absolute values. Moreover, we have verified for $(TiO_2)_4$ as a test case, that running the GA at the *higher-level* settings yields the same set of low-energy isomers as running the GA at the *lower-level* settings. Therefore, this enables us to make a considerable reduction of the computational effort by doing the entire DFT-GA at the *lower-level* settings. Obviously, if that would not have been the case [e.g. the system described in Ref. [1, 2]], we would have needed to employ *higher-level* settings inside the DFT-GA scan as explained in details in Ref. [2].

C. Post-process of the best fitted DFT-GA isomers using *higher-level* settings:

We have however noticed that the value of the electronic properties using *higher-level* settings differ by 0.1-0.5 eV as compared to the same found by *lower-level* settings. Therefore, in order to accurately calculate the VEA/VIP values

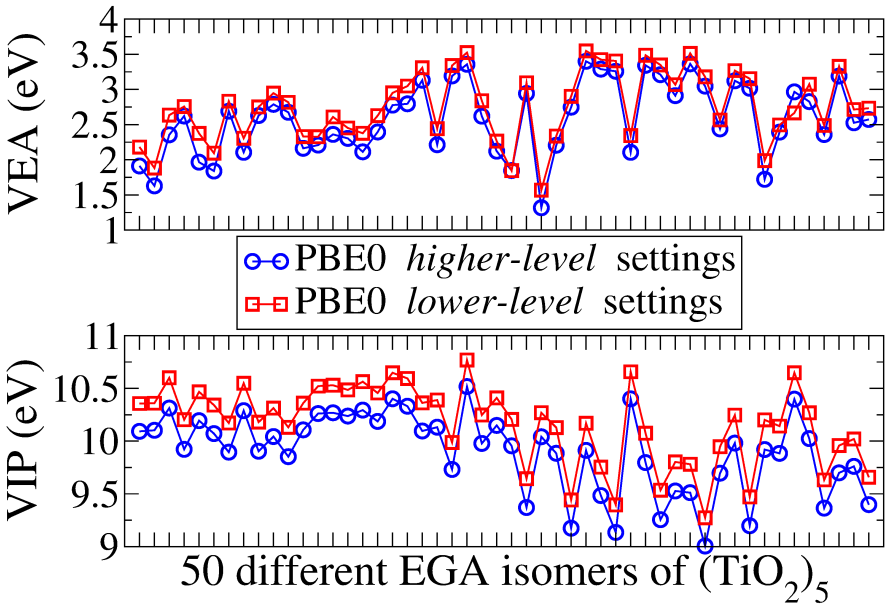


FIG. 1: *Hierarchy* of electronic properties (namely VEA, VIP) of $(TiO_2)_5$ EGM isomers (completely arbitrarily chosen) evaluated at *lower-level* and *higher-level* settings.

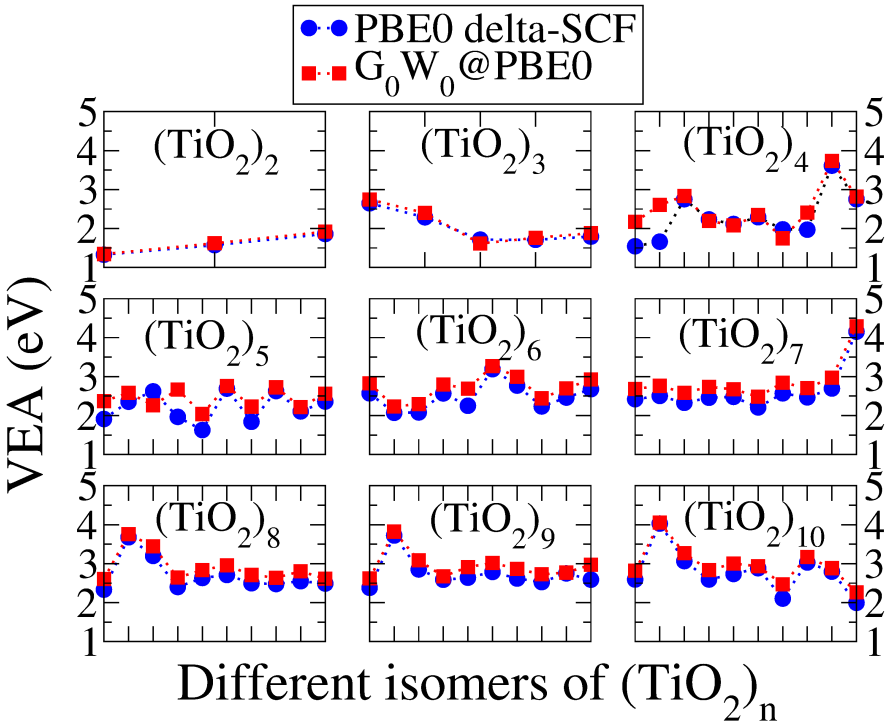


FIG. 2: Comparison of VEA values using G_0W_0@PBE0 vs PBE0 delta-SCF of randomly selected 10 different $(TiO_2)_n$ isomers.

of all different isomers, we have first optimized the structure using PBE (*higher-level* settings) and then two single point PBE0 calculation (*higher-level* settings) are performed at charge=0 and charge=-1/+1 (anion/cation). The VEA/VIP is therefore evaluated as the difference in total-energy between the neutral species and the anion/cation. We have named this methodological approach as PBE0 delta-SCF as explained in section I(d).

D. Validation of PBE0 delta-SCF energetics w.r.t G_0W_0@PBE0:

The VEA/VIP obtained from total energy differences with PBE0 at the *higher-level* settings are in good agreement with more accurate calculations based on many-body perturbation theory within the GW approximation, used in Ref. [9]. In addition, it has been shown in Ref. [13] that IPs and EAs calculated from DFT total energy differences exhibit similar size trends to GW calculations for bulk-like rutile TiO_2 nanocrystals. Thus, in order to understand the accuracy of VEA/VIP values as evaluated from PBE0 delta-SCF, we have randomly selected several isomers at all different sizes and compared with the VEA values as evaluated from many-body perturbation theory (e.g. G_0W_0@PBE0).

The absolute VEA values using G_0W_0@PBE0 are not always exactly matching with PBE0 delta-SCF values (with an estimated mean average error of 0.2 eV), but in all different sizes PBE0 delta-SCF qualitatively predicts similar energy profile as of G_0W_0@PBE0. It's clear that at all different sizes, PBE0 delta-SCF always provides the correct trends [see Fig. 2]. Note that the G_0W_0 calculations are performed using a large "tier4" basis set, while PBE0 delta-SCF calculations are done using "tier2" basis set. This validates the computational efficiency of our PES-search methodology.

II. NO. OF DANGLING-O VS NO. OF Ti^{3+} IN $(TiO_2)_n$ CLUSTERS:

It has been shown before that localization of the lowest unoccupied molecular orbital (LUMO) on a tri-coordinated Ti atom (namely, Ti^{3+} site) is often associated with a high vertical electron affinity (VEA). [9] But we found many structures without Ti^{3+} site that are having high VEA [see Fig. 3 below].

Therefore, we have given a comparative structural analysis of number of Ti^{3+} site vs number of dangling-O in the best 10 isomers found from EGA-search and VEA-GA search [see Fig. 4]. It's true that many isomers with high VEA are having this Ti^{3+} site especially at small cluster size, but with increasing size (n) of the cluster, many high-VEA isomers do not show this feature. In contrast, the correlation between having high VEA and increasing number of dangling-O in the structure holds at all cluster size (n) we have investigated.

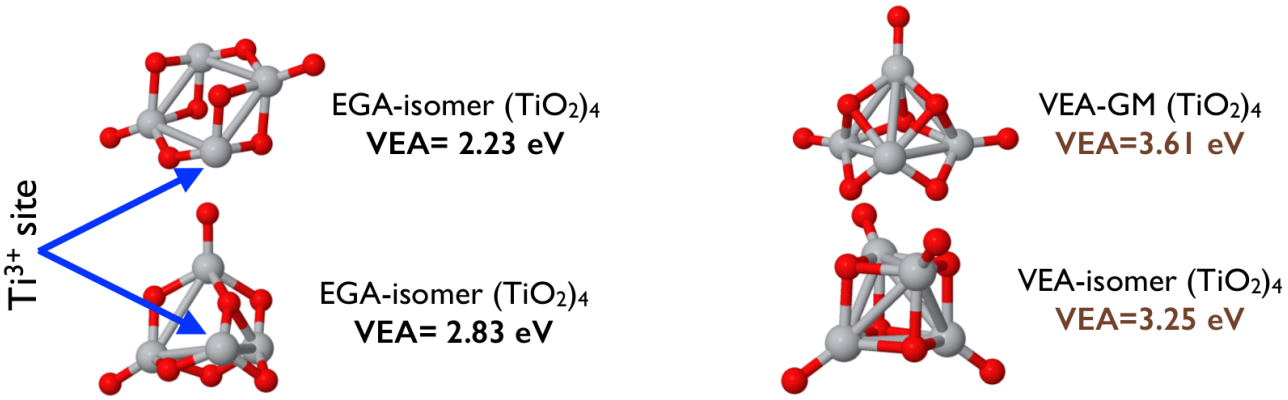
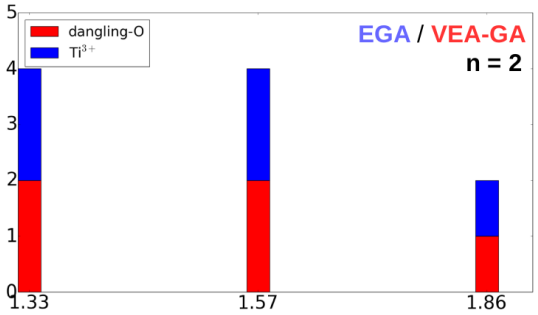
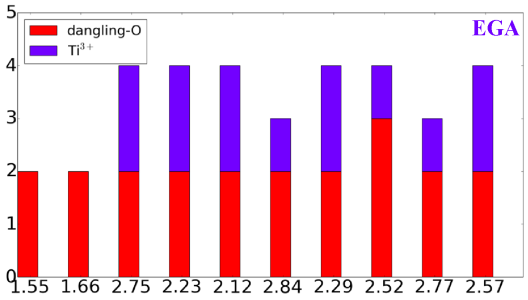
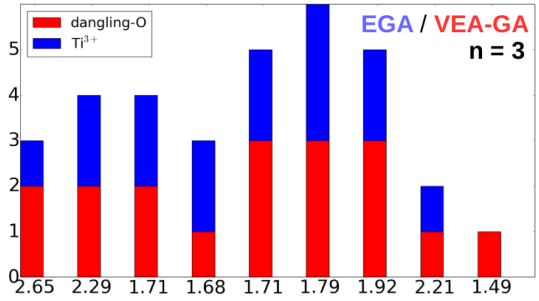


FIG. 3: The structures at the left-hand side have the Ti^{3+} site(s), while this feature is absent in the structures as in the right. However, the VEA of the later is larger than the former. The structures at the right-hand side have a larger number of dangling-O's, compared to the structures at the left-hand side. See more comparisons in Fig. 6.

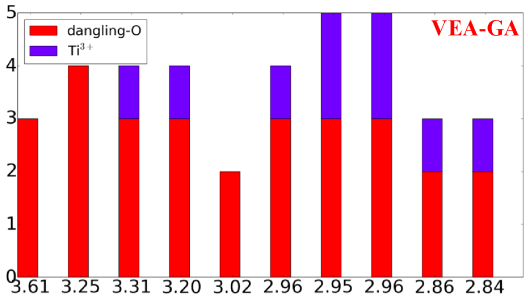
no. of dangling-O vs no. of Ti^{3+} in $(TiO_2)_n$ clusters



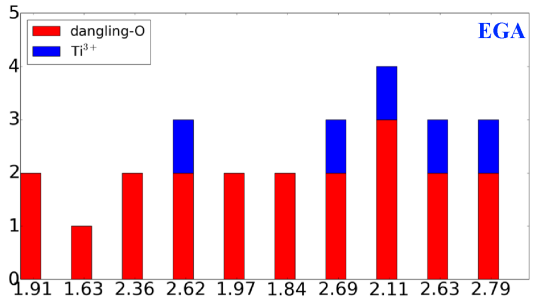


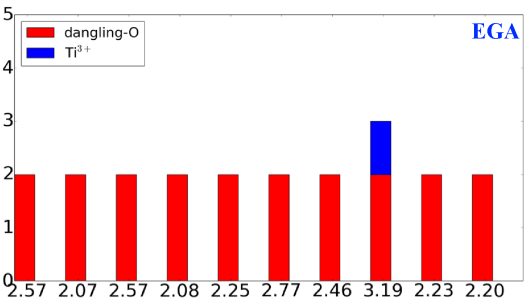


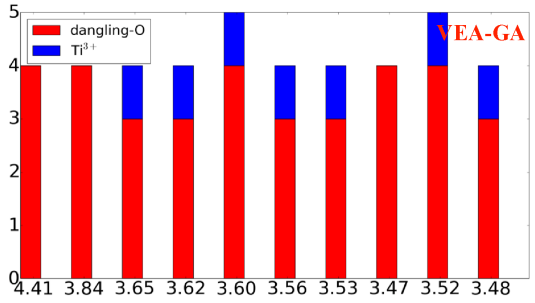
(a) $(TiO_2)_2$ & $(TiO_2)_3$



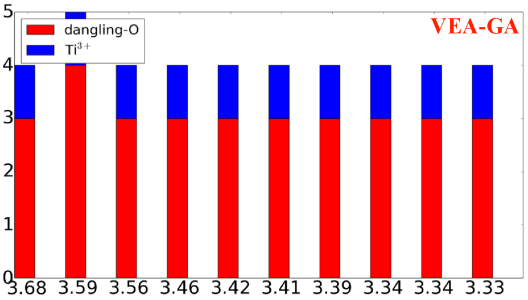
(b) $(TiO_2)_4$



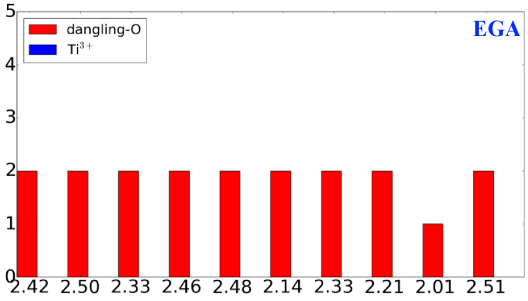


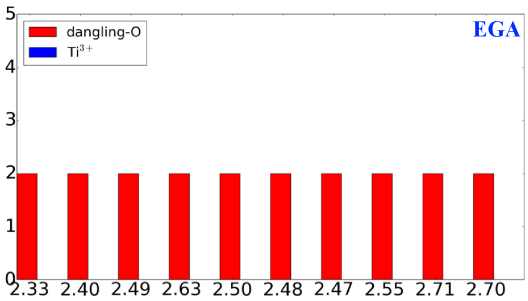


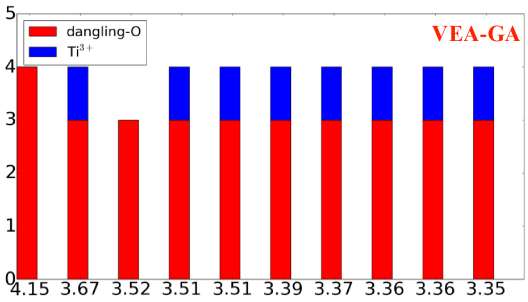
(c) $(TiO_2)_5$

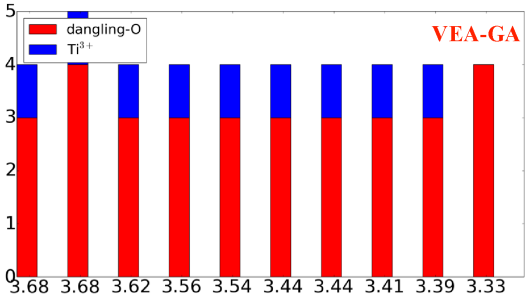


(d) $(TiO_2)_6$



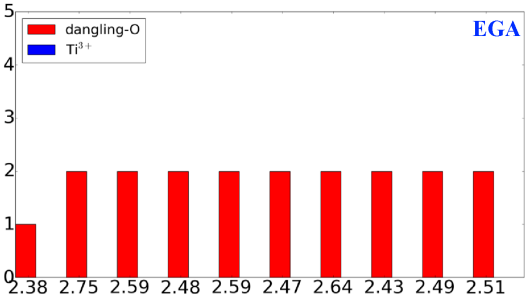


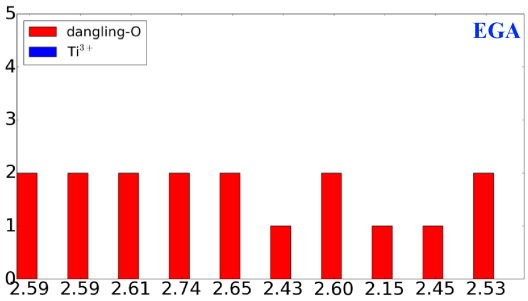


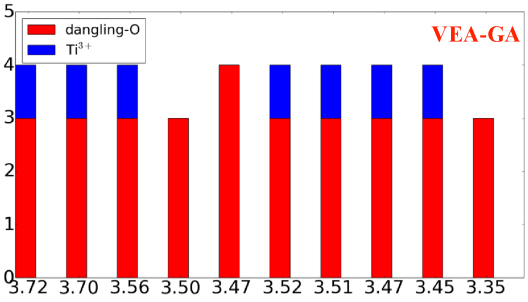


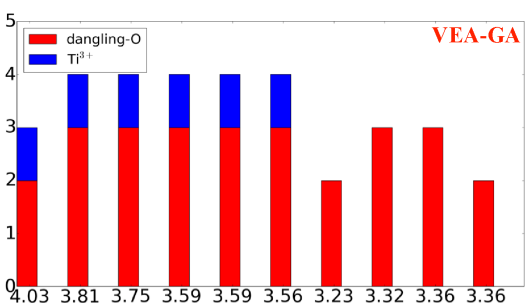
(e) $(TiO_2)_7$

(f) $(TiO_2)_8$



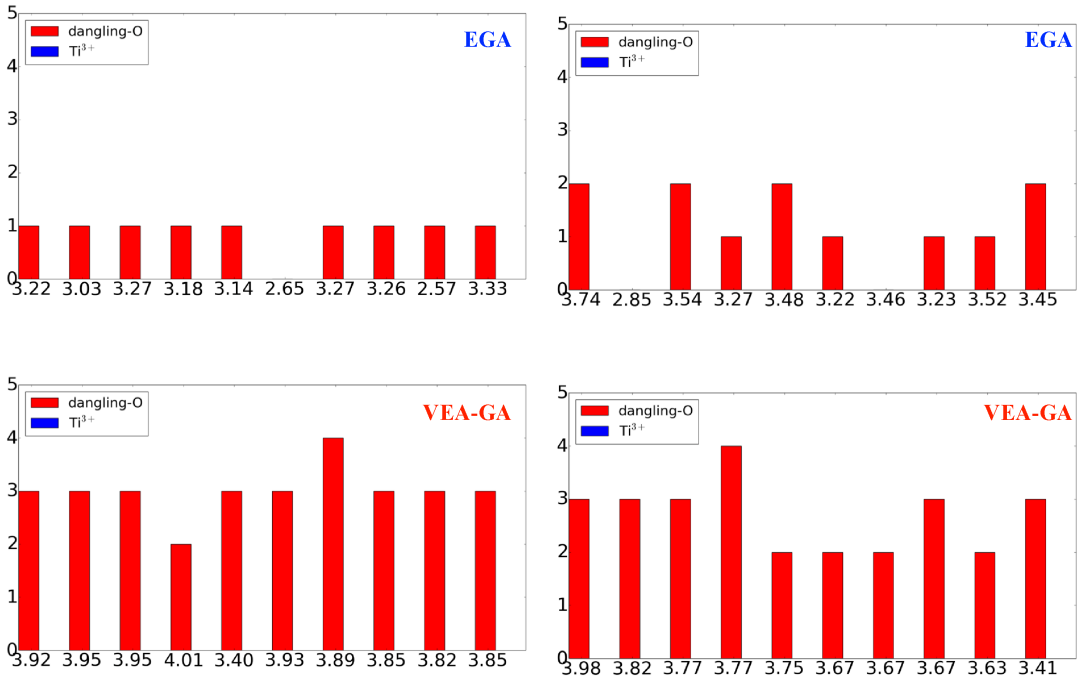






(g) $(TiO_2)_9$

(h) $(TiO_2)_{10}$



(i) $(TiO_2)_{15}$ (j) $(TiO_2)_{20}$

FIG. 4: We represent the structural features (namely, number of dangling-O, number of Ti^{3+} site) in the best 10 isomers of EGA and VEA-GA. In the x-axis we label the VEA value of the isomer in eV. In the Y-axis the red bars represent the number of dangling-O and on top of them, the blue bars represent number of Ti^{3+} site. Note that top of the red bar is the bottom of the blue bar. Thus if a red and blue bars end at 3 and 5 respectively, the number of dangling-Os are 3 and number of Ti^{3+} sites are (5 - 3 =) 2.

III. Ti-O PAIR CORRELATION FUNCTION (PCF) OF $(TiO_2)_n$ CLUSTERS:

We have plotted the Ti-O pair PCF of best 10 isomers (found by EGA and VEA-GA respectively) of the $(TiO_2)_n$ clusters. We have calculated all the Ti-O distances and divided them into 10 equally spaced bins to plot the histogram. The most significant difference in PCFs of the isomers found from VEA-GA and that of from EGA is the presence of the sharp peak (first peak in the PCFs as in Fig. 6) at 1.65 Å. This Ti-O bond length at 1.65 Å in the PCF represents the Ti-O bond of a dangling-O attached with a Ti-atom as shown in the Fig. 5 below. The intensity of this bond in PCF plot for VEA-GA-isomers is consistently higher than that of in EGA-isomers at all different sizes. This means all the isomers found by VEA-GA algorithm consists of more such dangling-Os in the structures.

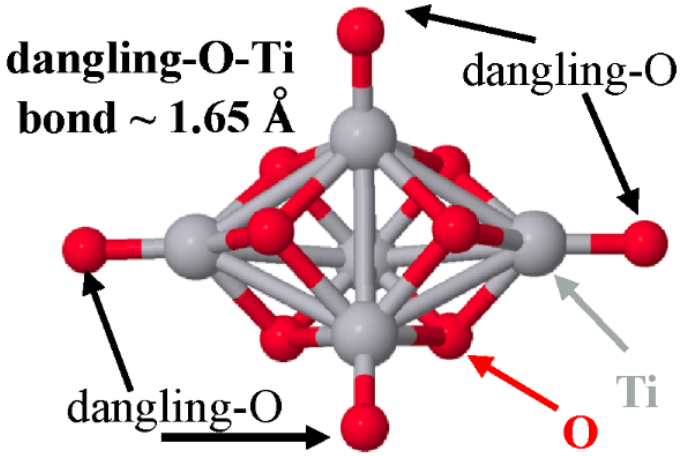
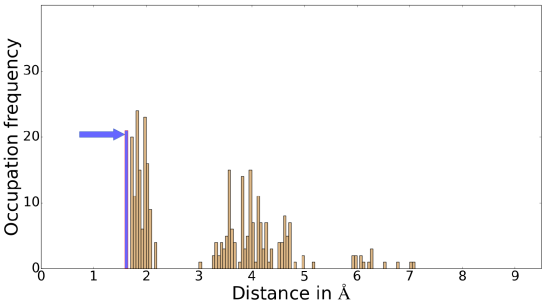
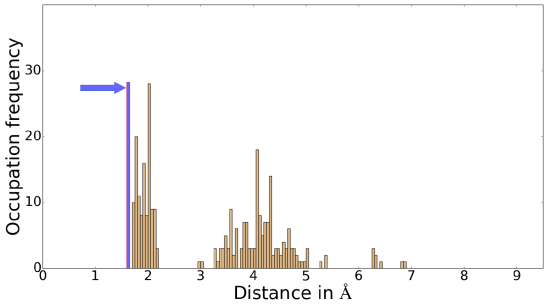


FIG. 5: Four dangling-Os in one VEA-GA isomer of $(TiO_2)_5$

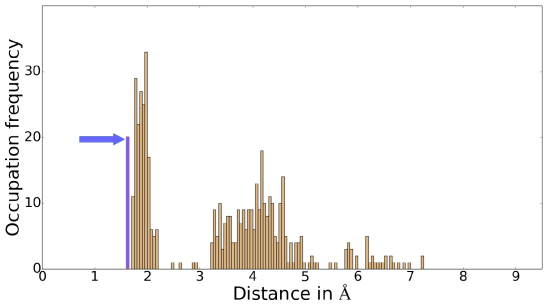
Ti-O PCF of $(TiO_2)_n$ clusters



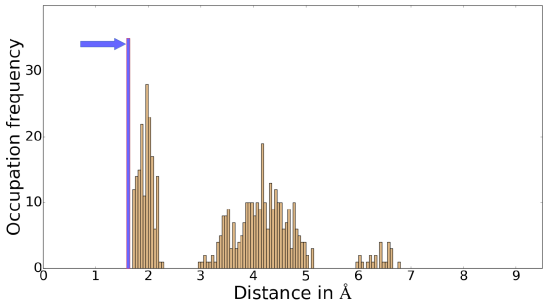
(a) EGA($(TiO_2)_4$)



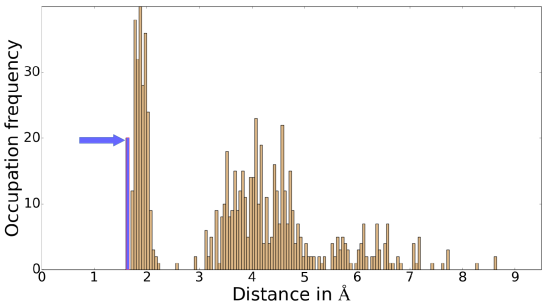
(b) VEA-GA($(TiO_2)_4$)



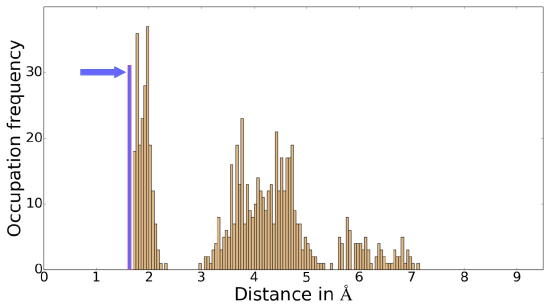
(c) EGA($(TiO_2)_5$)



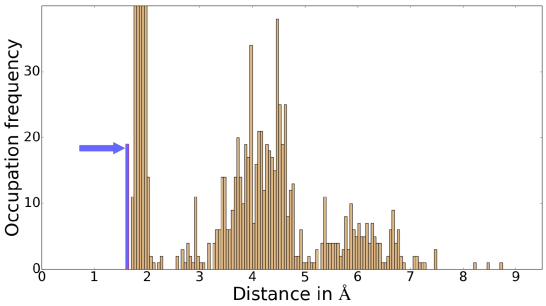
(d) VEA-GA($(TiO_2)_5$)



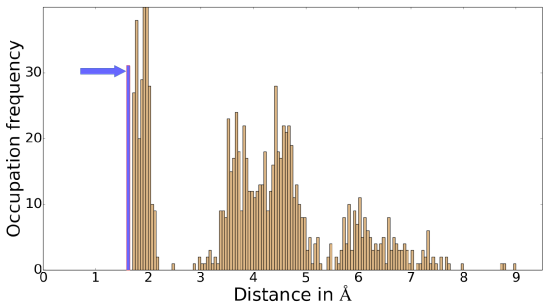
(e) EGA($(TiO_2)_6$)



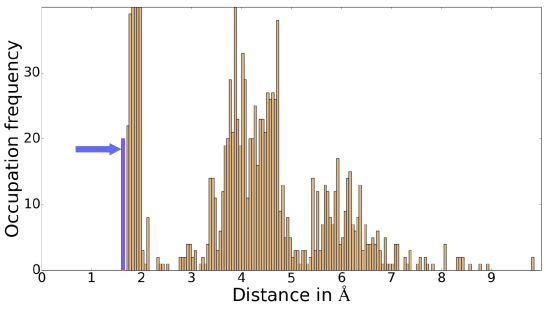
(f) VEA-GA($(TiO_2)_6$)



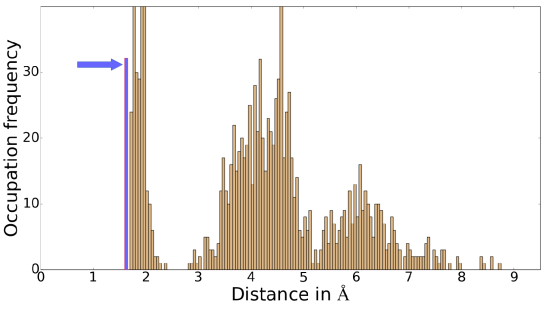
(g) EGA($(TiO_2)_7$)



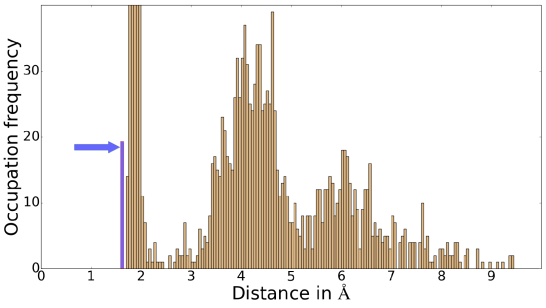
(h) VEA-GA($(TiO_2)_7$)



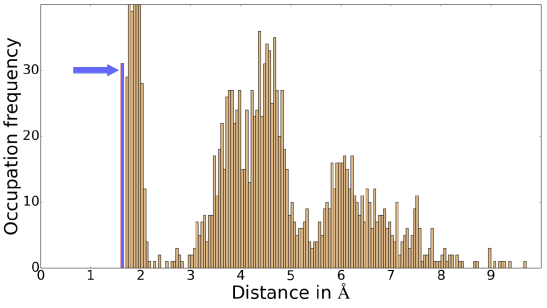
(i) EGA($(TiO_2)_8$)



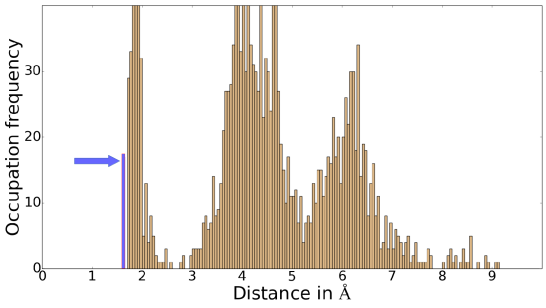
(j) VEA-GA($(TiO_2)_8$)



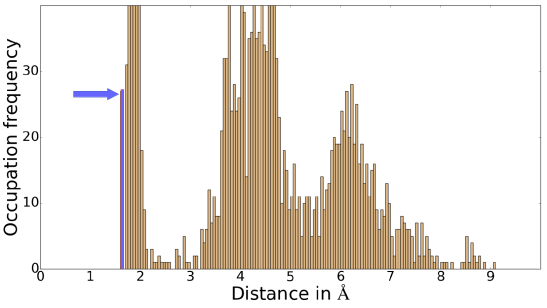
(k) EGA($(TiO_2)_9$)



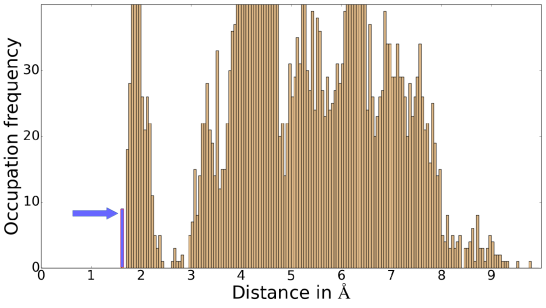
(l) VEA-GA($(TiO_2)_9$)



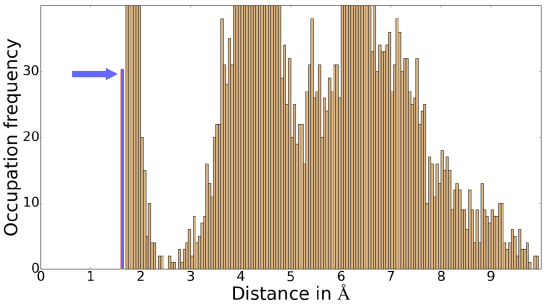
(m) EGA($(TiO_2)_{10}$)



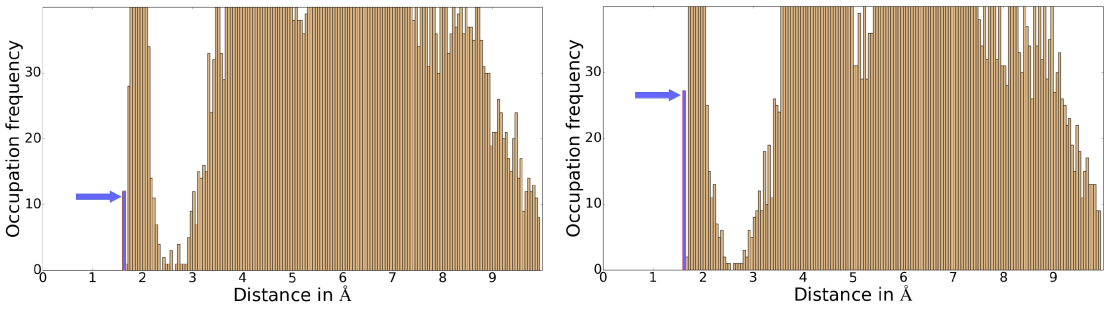
(n) VEA-GA($(TiO_2)_{10}$)



(o) EGA($(TiO_2)_{15}$)



(p) VEA-GA($(TiO_2)_{15}$)



(q) EGA($(TiO_2)_{20}$) (r) VEA-GA($(TiO_2)_{20}$)

FIG. 6: Ti-O PCF of $(TiO_2)_n$ clusters of the best 10 isomers found by EGA and VEA-GA. The striking difference is the presence of the sharp peak at 1.65 Å (the very first peak) in VEA-GA isomers. This peak is consistently higher in VEA-GA isomers as compared to EGA isomers at all studied sizes (e.g. n=4, 5,..., 10, 15, 20). This means all the VEA-GA isomers are having more Ti-O bonds of length 1.65 Å. This specific bond represents the dangling-O as shown in Fig. 5.

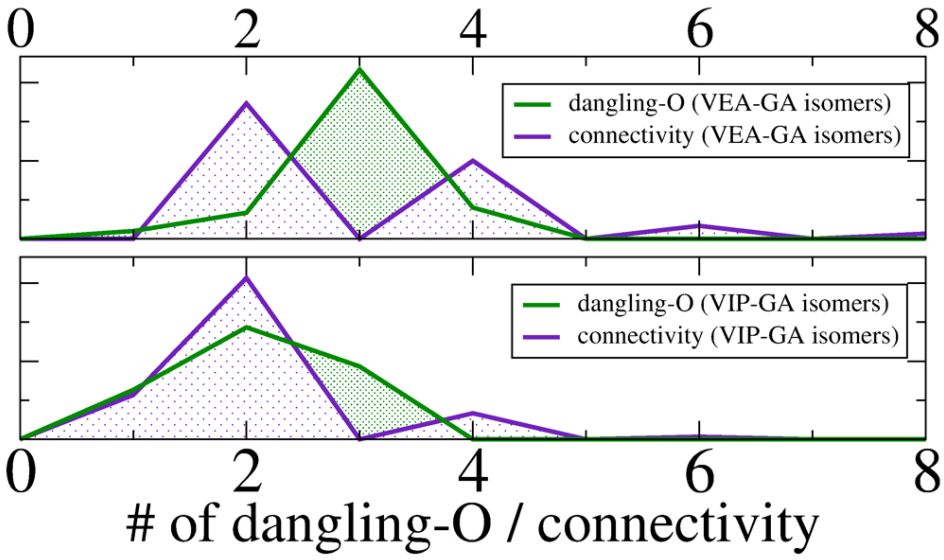


FIG. 7: Histograms of the abundance of isomers with a number of dangling-O atoms and low bond connectivity in the population of best isomers found by the VEA-GA and the VIP-GA.

IV. LOW GAP ($\leq$ 5.5 eV) VS DANGLING-O AND CONNECTIVITY

Fig.7 shows histograms of the abundance of isomers with a number of dangling-O atoms and low bond connectivity in the population of best isomers found by the VEA-GA and the VIP-GA. The VEA-GA selects for structures with a large number of dangling-O atoms, regardless of their connectivity, whereas the VIP-GA selects for structures with a low connectivity regardless of the number of dangling-O atoms. We therefore would expect clusters with a narrow gap to have a number of dangling-O atoms, two of which are close to each other. Several of the structures with gaps below 5.5 eV, found by the VEA/VIP-GA, have 2-3 dangling-O atoms and a bond connectivity of two. We note that more structures with a narrow gap and more definitive structure-property relations may be found by tailoring a GA to search specifically for this property. We have found several structures at different sizes with gaps $\leq$ 5.5 eV. On inspecting their structure, we have found approximately 2-3 dangling-O atoms with a bond connectivity of two. We think that more structures with a narrow gap and more definitive structure-property relations may be found by tailoring a GA to search specifically for this property. Following Table depicts those details (i.e. no. of dangling-O, connectivity) of different isomers that are having gap $\leq$ 5.5 eV.

Type of GA-scan	**Gap value (eV)**	**Connectivity:**	**# of dangling-O**
VIP-GA			
	5.34361	2	2
	4.93132	2	3
	5.41224	2	2
	4.89849	2	2
	5.29608	2	2
	5.2959	2	2
	5.4691	2	2
	5.31938	2	2
	5.2723	2	3
VEA-GA			
	4.96373	4	2
	5.34427	2	3
	5.25586	6	2
	5.38608	4	2
	5.36256	2	3

V. COORDINATES OF ALL ISOMERS AT ALL SIZES AND COMPARISON WITH PREVIOUSLY PUBLISHED STRUCTURES:

Our EGA was able to find all the low energy isomers reported previously [9, 14, 15] for all cluster sizes. For n=5 our choice of PBE0 as exchange-correlation functional and *higher-level* settings changes the ordering of the two most stable isomers, which are very close in energy, with respect to Ref. [9]. For n=10, we have found lower-energy structures than those previously reported. We have verified, at the PBE level, that the lowest energy, highest VEA, and lowest VIP isomers are stable minima in the sense of having no imaginary vibrational modes. All input and output files for the clusters described in this paper can be downloaded from `http://nomad-repository.eu`.

[1] S. Bhattacharya, S. V. Levchenko, L. M. Ghiringhelli, M. Scheffler, Phys. Rev. Lett. **111**, 135501 (2013).
[2] S. Bhattacharya, S. V. Levchenko, L. M. Ghiringhelli, M. Scheffler, New J. Phys. **16**, 123016 (2014).
[3] D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley: Reading, MA, (1989).
[4] D. M. Deaven, K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
[5] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, Comput. Phys. Commun. **180**, 2175 (2009).
[6] V. Havu, V. Blum, P. Havu, M. Scheffler, J. Comput. Phys. **228**, 8367 (2009).
[7] M. Raju, S.-Y. Kim, A. C. T. van Duin, K. A. Fichthorn, J. Phys. Chem. C **117**, 10558 (2013).
[8] M. Raju, A. C. T. van Duin, K. A. Fichthorn, Nano Lett. **14**, 1836 (2014).
[9] N. Marom, M. Kim, J. R. Chelikowsky, Phys. Rev. Lett. **108**, 106801 (2012).
[10] J. P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[11] J. P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. **78**, 1396 (1997).
[12] C. Adamo, V. Barone, J Chem Phys **110**, 6158 (1999).
[13] L. Hung, K. Baishya, and S. Ogut, Phys. Rev. B **90**, 165424 (2014).
[14] M. Calatayud, L. Maldonado,C. Minot, J. Phys. Chem. C **112**, 16087 (2008).
[15] O. A. Syzgantseva, P. Gonzalez-Navarrete, M. Calatayud, S. Bromley, C. Minot, J. Phys. Chem. C **115**, 15890 (2011).
[16] "light" integration grids and compact "tier 1" basis sets as explained in Ref. [5]
[17] The *lower-level* settings means a "light integration grids and compact "tier 1 basis sets and a *higher-level* settings means a "tight" integration grid and a bigger "tier2" basis sets as implemented in fhi-aims code [see details in Ref. [5]].