Enhancing DFT-based energy landscape exploration by coupling quantum mechanics and static modes

Lionel Foulon, Anne Hémeryck, Georges Landa, Marie Brut

PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Enhancing DFT-based energy landscape exploration by coupling Quantum Mechanics and Static Modes†

Lionel Foulon,[a] Anne Hémeryck,*[a] Georges Landa,[a] and Marie Brut*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Unravelling the atomic scale diffusion that can occur at the surface, at the interface or into the bulk is challenging: multi-scale modelling approach usually requires intensive prospective calculations and moreover huge human investment. In this article, the Static Mode (SM) approach is coupled with Quantum Mechanics (QM) calculations in order to guide the exploration of the energy landscape, by optimizing the choice of events that are significant for the evolution of the system. SM enable the determination of the strain field of a set of atoms submitted to external and localized stresses, like atomic displacements. Here, we present a workflow based on the systematic SM exploration, with the objective to reduce both exploration time and human load when used with ab initio level calculations. The QMSM coupling allows to screen, score and select relevant directions that are further used to initiate and study diffusion in atomic systems. The most relevant deformations are next refined and relaxed with DFT calculations. In this paper, the overall QMSM approach is described and we discuss its use for the identification of atomic diffusion in two different systems of interest: grafting of a molecule on an oxide surface and studying dynamical behavior of point-defect in bulk crystalline material.

1 Introduction

Multi-scale modeling has become a strategic tool for understanding, optimizing and designing new materials with improved performance in many industrial fields, from microelectronics to aeronautics and pharmacology. By bridging a wide range of time and length scales, simulations provide insights into the structural, electronic and kinetic properties of materials. However, despite their effectiveness in guiding researchers and engineers in their developments, the associated human cost remains too high to allow their systematic use. The current challenge is therefore to propose more suitable and predictive tools based on atomic granularity, while allowing to access the macroscopic properties of the materials of interest.

At this level, the main lock is the exhaustive exploration of the energy landscape, i.e. the identification of the atomic events to be taken into account in higher scale models. The different local atomic configurations and the associated energy barriers must be identified through *ab initio* approaches. Among them, Density Functional Theory (DFT)[1] is extensively used as a good compromise between efficiency and accuracy. Nevertheless, it requires a human investment that is not adapted to the exploration of complex Potential Energy Surfaces (PES) as needed in materials engineering. Several other methodologies exist to build a library of events such as free-energy methods or trajectory-based methods, all of them being costly in terms of resources and computational time. We can notably cite Molecular Dynamics (MD) and associated biased methods like umbrella sampling[2], metadynamics[3], Activation-Relaxation Technique (ART)[4]. Metadynamics and umbrella sampling use collective variables and sample the space to explore the PES. They precisely investigate a restricted region defined by the user and require a good knowledge of the system. ART and MD are less restrictive about *a priori* knowledge of the system and allow to explore a larger area, which, in turn, increases the simulation time. Thus, the challenge is to find a compromise between the computational cost and the accessible area to explore. Significant efforts are also deployed to make this search more automated and exhaustive[5–7].

In this challenging context, we propose to couple the Static Mode approach (SM)[8] with DFT calculations to guide the latter and reduce the user intervention. The Static Mode method was initially designed to compute all possible deformations of a macromolecular assembly subject to external stresses, like intermolecular interactions. In such systems, like biomolecules, accounting for full-atom flexibility is a pre-requisite to access functional properties. Nevertheless, their PES is too complex to be explored in an exhaustive way and within an acceptable computational time through conventional approaches like MD. The SM

[a] *LAAS-CNRS, Université de Toulouse, CNRS, UPS, INSA, Toulouse, France.*
* E-mails: anne.hemeryck@laas.fr, marie.brut@laas.fr
† Electronic Supplementary Information (ESI) available. See DOI: 10.1039/cXCP00000x/

method does not aim at providing the complete pathway of a conformational change. However, it constitutes a simplified approach that enables the systematic determination of the strain field of a set of atoms submitted to an external and localized stress. SM are therefore an accurate indicator of the propensity of a system to accommodate a restraint.

SM calculation is based on the Hessian matrix, which is generated from an equilibrium conformation obtained after submitting the system to a total energy minimization through any model, quantum or classical, as chosen by the user. A Static Mode corresponds to the deformation field of a system submitted to the displacement of a given atom in a given direction x, y, z of the cartesian space. For a system containing N atoms, $3N-6$ SM can be computed, stored and combined to further map the deformations induced by a specific perturbation. This post-treatment procedure, based on pre-calculated deformations, allows to apply single or multiple stresses while immediately computing the induced conformational response. Such perturbations can be designed by the user to address specific questions, as discussed later. In this manner, the SM approach was validated and applied to various biological systems, SM being used to simulate protein folding upon electrostatic interactions[9], active site stability upon mutations[10], conformational change upon ligand binding[11], molecular properties upon functionalization[12], to name but a few. Through these different cases, SM predictive nature has been consolidated and shown to be associated with a very low computational cost, then offering a versatile range of possibilities.

In this paper, our approach combining Quantum Mechanics and Static Modes (QMSM) is presented. We propose to use the Static Modes to guide DFT calculations and save time in exploring the energy landscape by optimizing the choice of events that are significant for the evolution of the system. In the following, we describe this coupling and present its application to different systems that have already been treated in the literature, thus allowing a qualitative and quantitative comparison of the results. We first explore different grafting modes of a model DNA strand on an alumina surface, while in a second time, we focus on the diffusion of point defects in crystalline silicon.

2 Methods

2.1 The QMSM coupling

The QMSM procedure is summarized in Fig. 1 and presented here.

Step 1 - SM calculation. As previously explained, we start from an equilibrium configuration obtained from geometry optimization (here by DFT calculations). From this relaxed configuration and the associated Hessian matrix, SM are computed within the harmonic approximation. In this frame, the Hessian matrix elements H_{ij}, *i.e.* the second derivatives of the energy with respect to the displacements x_i, x_j can be used to express the force vector $F_i = -\sum_j H_{ij} x_j (1)$. Next, a constraint is imposed on the l-th atom by considering a given displacement x_l resulting from the application of the associated force F_l, which acts as a reaction force.

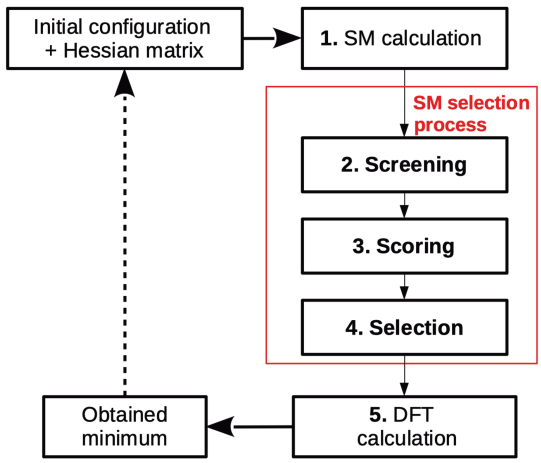


Fig. 1 Presentation of the QMSM workflow. SM are calculated (step 1) and used to screen the strain fields induced by imposed atomic displacements (step 2). The associated deformations are further scored and ranked (step 3). The most relevant SM are then selected (step 4) and used to guide DFT calculations in search of new conformations (step 5).

Solving the set of equations (1) finally provides a vector containing the $3N$ atomic displacements imposed by the constraint x_l, which is always set to 1 Å so that the resulting deformation field is normalized to this elementary stress. (see[8] for more details about the SM calculation). This vector, called a Static Mode, corresponds to the collective response of the system to a single stress. This operation is performed for stresses imposed to each atom, in each direction x, y, z, allowing to build the complete SM data bank.

Step 2 - Screening. In the following step, SM are used to screen the deformations induced by the systematic displacement of each atom n in all directions of a discrete space. To define these directions, we use a sphere mesh centered on each atom of interest, each mesh point being used to create an atomic displacement. With a sufficiently fine mesh, we can admit that we explore all possible directions. Each constrained displacement $\vec{d}$ then allows the determination of the induced strain field M_i, through a simple linear combination of the Static Modes. Using m_n, a $3N \times 3$ matrix containing the three SM of the atom n, the resulting deformation can be written as $M_i = \sum_{j=0}^{3} m_{n_{i,j}} d_j$. The vector M_i can be further applied to the starting structure to obtain a new set of coordinates. This process is systematically achieved for each atom of the system, leading to screen all possible deformations.

Step 3 - Scoring. Those deformations are then characterized with a geometric or energetic criterion defined by the user, depending on the response to be scored and ranked. This step is more detailed through the description of an example in the next section.

Step 4 - Selection. Depending on the scores, *i.e.* the criterion values obtained in the previous step, the best ranked deformations are selected to be used as inputs for subsequent DFT calculations, in search of new configurations.

Step 5 - DFT calculations. As detailed later in the article, DFT calculations can be performed in two different ways, based on two types of information provided by the SM selection process:

i) SM strain field can be applied to the structure, which is further relaxed in search of a new minimum; ii) the optimal stress $\vec{d}$, generating this deformation, can be used as a guide for a constrained calculation, applying a drag method for instance. Both are tested and discussed. Let us finally specify that SM calculation is run through an independent package. The energy model used upstream to generate the starting structure, and downstream for the post-treatment, can be fully chosen by the user.

2.2 Step by step procedure

The QMSM procedure is now illustrated on a model system. As shown on Fig. 2 (a), we have chosen a molecule composed of 12 atoms and evaluated its internal rotation along the axis defined by atoms 11 and 12. To this end, we monitored and optimized the variations of the dihedral angle defined by atoms 9-11-12-10, and referred to as θ.

Steps 1-2. Once the SM are obtained, the first step consists in screening all the deformations induced by constrained atomic displacements. To this end, for each atom to be constrained, we apply the following procedure in a systematic way. For the sake of clarity, we chose to focus on atom 11, which induces the largest θ angle variation. To sample all possible stresses, a sphere mesh centered on this atom is created, as shown on Fig. 2 (a). Mesh parameters can be set by the user (600 equidistributed points are used here). Combining the three SM associated with atom 11, we compute the strain fields induced by the displacement of atom 11 toward each mesh point, the magnitude of each displacement being fixed to 1.00 Å.

Step 3. The following step consists in characterizing the 600 deformed structures obtained from step 2. To this end, a criterion allowing a quantitative scoring needs to be defined. This criterion is a scalar value calculated from the geometry of the deformed system. Multiple criteria can be envisaged depending on the study objectives. As an example, one can estimate the energetic cost of the deformation to consider it as favorable or not. In the present example, we choose the variation of the θ angle from its initial value ($\sim$ -180 °). The corresponding variations are represented on Fig. 2 (a) for each point of the mesh, blue and red colors correspond to large and small variations, respectively.

Step 4. The next step consists in selecting the best ranked deformations, that is, the deformations associated with the largest θ angle variation. The stress applied on atom 11, which induces the largest rotation (50 °, see Fig. 2(c)), is represented with a blue arrow on Fig. 2 (b). The optimized stresses obtained for each atom through the same process are also shown on this figure, together with the associated dihedral variation on Fig. 2 (c).

Step 5. Finally, the resulting strain field (shown with green arrows on Fig. 2 (d)) associated to atom 11 displacement (blue arrow) can be used as an input for subsequent calculations, for example, an energy minimization in search of a new conformer. The relaxation of the deformed structure allows to reach a new minimum energy structure, shown on Fig. 2 (e). The overall process can be iterated from the new geometry if needed.

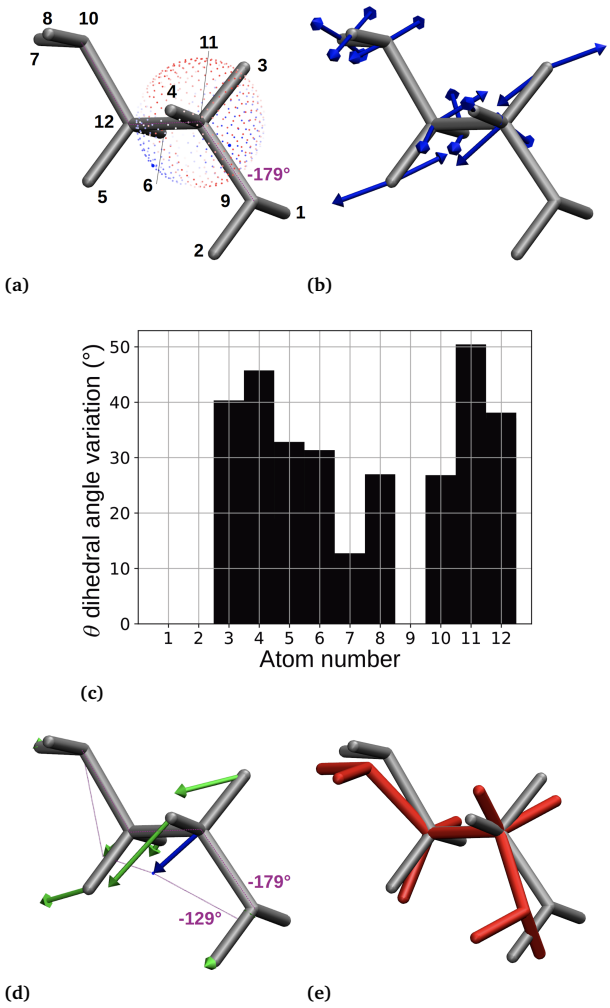


Fig. 2 (a) Representation of the model system composed of 12 atoms. A spherical mesh grid centered on atom 11 is used to define all possible stresses applied on this atom and screen the induced deformations. The associated variations of the θ angle are shown with a red to blue color scale, for increasing values. (b) Optimized stresses obtained from the screening applied to each atom and associated with the largest θ angle variations are shown with blue arrows. (c) Resulting variations of the θ angle for each optimized stresses. The largest variation is obtained for a stress applied on atom 11. (d) Strain field (green arrows) resulting from the optimized stress applied on atom 11 (blue arrow). (e) In red, the new configuration obtained from the energy minimization of the deformed structure shown in d.

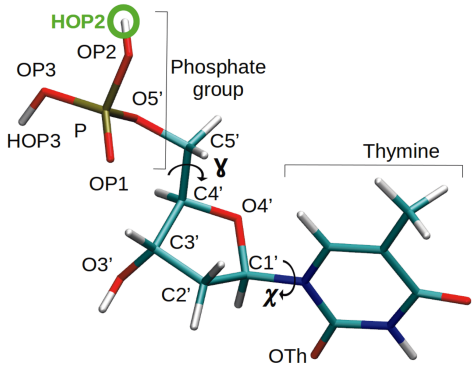
3 Grafting of model DNA strand on an oxide surface

Direct and tailored assembly of organic and/or biological molecules through the control of their interaction and grafting on surfaces is of technological interest. Such assemblies find applications in many industrial domains as adhesion[13,14], energy[15–17], sensor[18,19], health[20–22], environment[23,24]. To achieve such complex architectures, several crucial features must be ensured: is it possible to graft the molecules on these surfaces? what are the nature and the strength of the bond between the organic molecule and the substrate? which functional groups are the most effi-

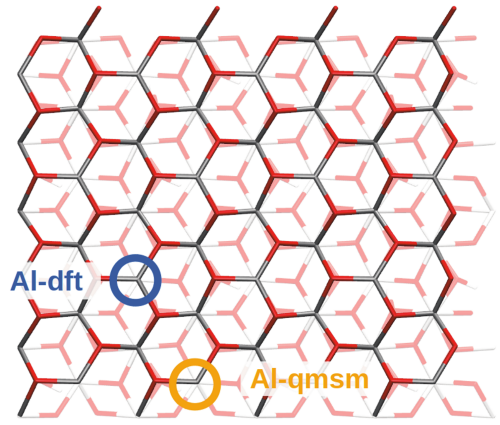
cient? is assembly well-ordered or chaotic? if so, which are the most relevant grafting modes and how to reach them? or how to prevent unwanted configurations? etc.

Facing such fundamental issues and complexity of the molecule-substrate interactions, characterizing grafting modes in details at the atomic scale might be regarded as the first step in mitigation strategies to help for a controlled design.

Numerous theoretical studies are carried out on the adsorption of molecules on surface[25–29]. Even if those calculations are valuable and accurate, most of them are limited to static calculations using *ab initio* relaxation, simply characterizing the most favorable grafted, *i.e.* adsorbed configurations. These calculations claim to be exhaustive, however they do not take into account the modified flexibility of the molecules or the loss of flexibility of these molecules once their configuration has been modified, whereas their intrinsic ability to change their conformation has changed, particularly in the presence of an anchoring point.



(a)



(b)

Fig. 3 (a) Model of DNA strand, a 2'-deoxythymidine-5'-monophosphate (dTMP). HOP2 atom is highlighted for its central role in this study. (b) Top view of the Al_2O_3 (0001) surface. Red, light blue, dark blue, white, brown and silver colors represent oxygen, carbon, nitrogen, hydrogen, phosphorus and aluminium atoms respectively. Labelling of the distinct anchoring points on Al_2O_3 (0001) reached during the adsorption and following DFT-driven method (Al-dft) or free exploration QMSM method (Al-qmsm) is given in (b).

In such a context, our previous study investigated the grafting of a model DNA strand on alumina surface (Al_2O_3) with the goal to achieve smart tailored assembly of nano-objects using the DNA complementarity[30]. The model DNA strand, a 2'-deoxythymidine-5'-monophosphate (dTMP) exhibiting all DNA chemical moieties, and the top-view of the Al_2O_3 surface are shown on Fig. 3. Several adsorption modes have been identified using a thorough study coupling IR measurements and DFT calculations, each of them differing from the anchoring point(s) between the dTMP and the Al_2O_3 (0001) surface. In the DFT study, a static approach was applied by screening and mapping all the possible configurations adopted by the dTMP, as obtained after relaxation of the isolated dTMP molecule as a function of the Al_2O_3 surface topology. No modification (distortion, rotation) of the dTMP was considered when positioning the dTMP above the surface, the configurations were obtained by positioning the dTMP perfectly on top of the anchoring points. Moreover, the possibility to reach some of the most stable configurations from pre-adsorbed state to another one was not discussed. Two of the adsorbed configurations obtained in Ref. [30] are shown on Fig. 4 (A) and (C), named Up and Down-DFT respectively. Those configurations differ by the number of bonds formed with the alumina surface. In the case of Up configuration, three O-Al bonds are formed between the dTMP (thymine and sugar) and the surface. In the case of Down-DFT, a fourth bond is formed between the phosphate group that is grafted to the surface involving OP1 and Al-dft atoms, while HOP2 also binds to the surface, forming a H-bond with OP2 atom.

In the following, we use the QMSM coupling to investigate the grafting of the phosphate group on the alumina surface starting from the configuration Up (Fig. 4 (A)). More information on DFT calculations in given in the Computational Detail section. Two main strategies are applied to unravel the mechanisms responsible for the grafting of phosphate group:

(i) knowing the final configuration that must be reached as a target configuration Down-DFT (Fig. 4 (C)). The procedure to carry out a driven QMSM exploration to target a known configuration is detailed in section 3.1.

(ii) using only QMSM coupling from configuration Up to explore unknown path without pre-knowledge of the final configuration. In section 3.2, depicted as a free exploration using QMSM, a final configuration shown in Fig. 4 (H) is obtained and named Down-QMSM.

3.1 Driven QMSM exploration to reach a target configuration

As postulated by the DFT study[30], both adsorbed configurations Up and Down-DFT exist. Considering the similarity between these configurations, we assume that it exists a diffusion path linking them: the configuration Down-DFT can be reached from configuration Up when grafting OP1 atom from the phosphate group on Al-dft atom of Al_2O_3 surface (see Fig. 3).

To explicitly reach the target Down-DFT configuration (Fig. 4 (C)), we consider all the possible deformations calculated with the SM that allow reducing the distance between OP1

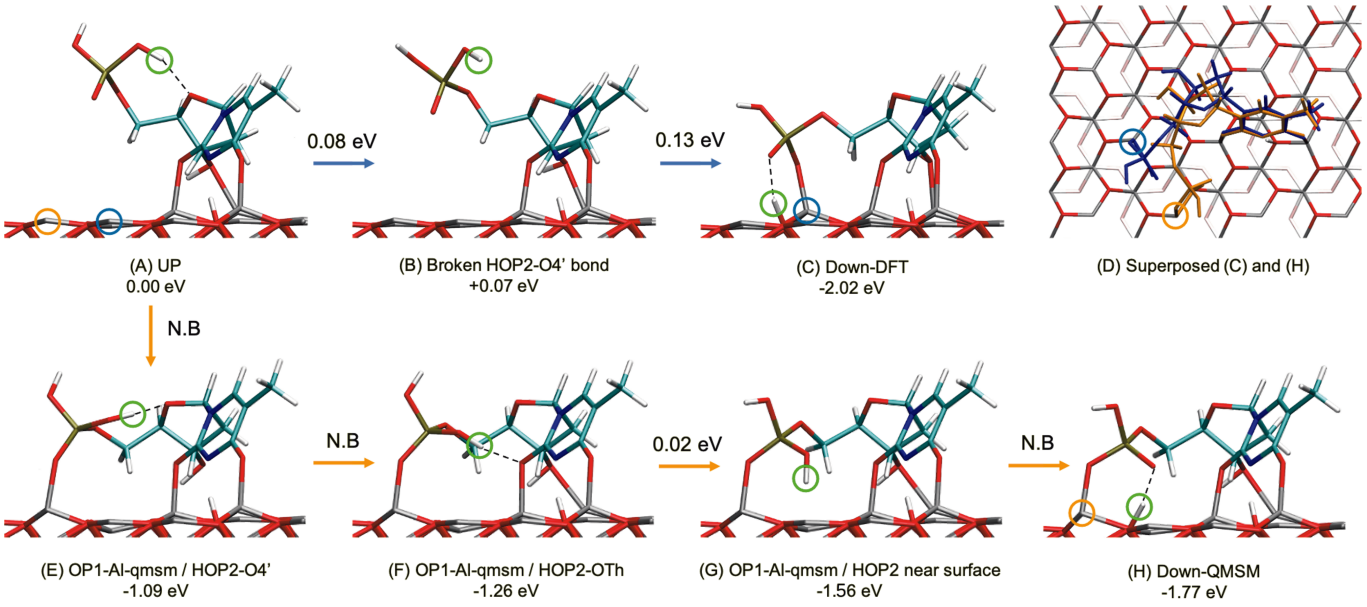


Fig. 4 Adsorption of dTMP on Al_2O_3(0001) surface using QMSM: a driven exploration is shown in blue and a free exploration in orange. Side-view snapshots of the stable and metastable configurations along the MEP are provided from (A) to (C) and (E) to (H). (A) Up: relaxed adsorbed configuration obtained with DFT. (B) and (C): Driven exploration from Up to Down-DFT - Blue path: (B) Intermediate where the original H-bond between HOP2 and O4' atoms is broken and (C) Down-DFT: adsorbed state relaxed with DFT where the phosphate group is grafted to the Al-dft surface atom. (E) to (H): Free exploration from Up to Down-QMSM - Orange path: (E), (F) and (G) show the first, second and third intermediate configuration, respectively. (H) Down-QMSM: relaxed adsorbed configuration obtained through the free exploration where the phosphate group is grafted to the Al-qmsm surface atom. (D) Top-view of the Al_2O_3 (0001) surface where Down-DFT (blue) and Down-QMSM (orange) configurations are superposed, highlighting the difference in their adsorbed states. The activation barriers are given above the arrow between two configurations and the energy of each configuration is given below each compared to (A).

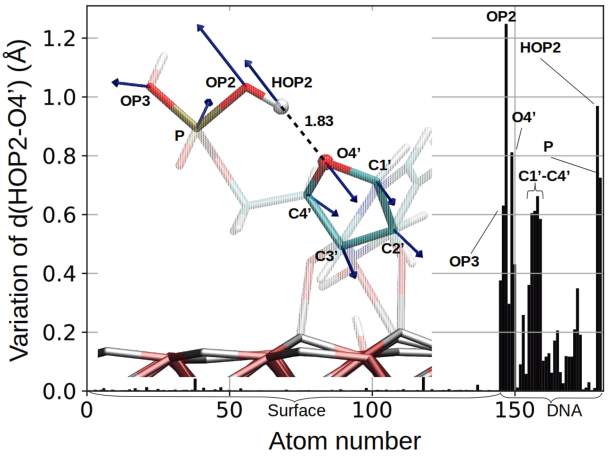
and Al-dft atoms to directly form the target bond d(OP1–Al-dft). All optimized stresses with respect to this criterion are provided in Supp. Mat. S1 (a)†. The five most favorable ones are investigated and the associated strain fields are also provided in Supp. Mat. S1 (b) to (f)†. None of these applied deformations is able to relax toward the target configuration: once deformed, and independently of the stress applied up to 2.00 Å, the system gets back to its initial configuration for three of them, or reaches a new final configuration that will be discussed later. Looking closely at the strain fields and the associated structural relaxations, we clearly identify that the origin of the deadlock is the H-bond existing between the HOP2 atom from the phosphate group and the O4' atom from the sugar (d(HOP2-O4') = 1.83 Å, Fig. 5 (a)), avoiding the phosphate group to dive toward the surface.

Consequently, we investigate, from the SM, the stresses able to break the H-bond existing between O4' and HOP2: the criterion relative to QMSM coupling is set as elongating the distance d(HOP2-O4'). All optimized stresses calculated using SM method are provided in Fig. 5 (a). The deformation that allows to break the HOP2-O4' bond (see Fig. 5 (b)) is obtained by applying the SM associated with the OP2 atom, as revealed in Fig. 5 (a). Once this deformation is applied and the system relaxed, a new minimum is reached, where the HOP2-O4' bond is broken, compared to the isolated relaxed configuration (Fig. 4 (B)). In such configuration, the system is less stable by 0.07 eV compared to configuration Up.

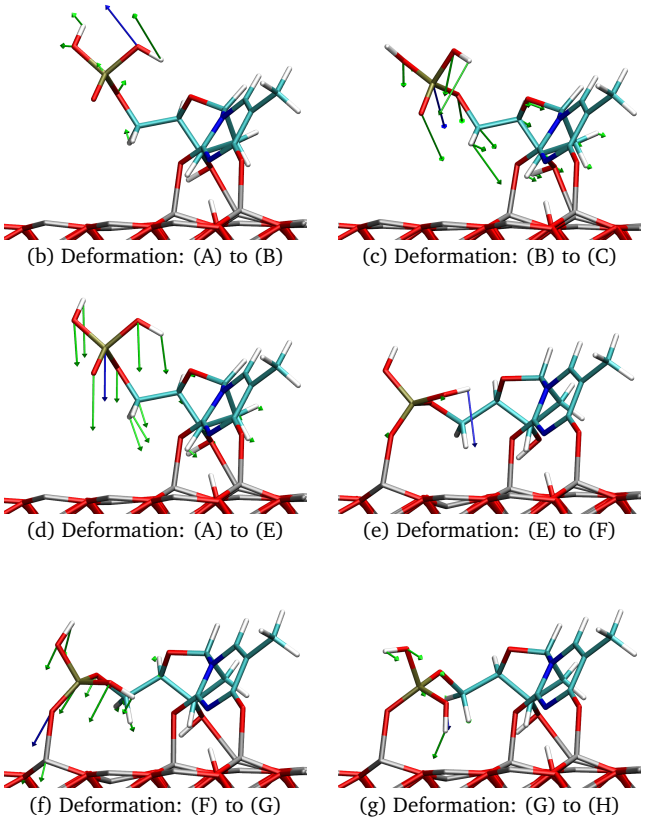
QMSM is finally applied again on the new identified intermediate to select the relevant deformation allowing the previous criterion to reach a new minimum, *i.e.* reducing d(OP1–Al-dft). All optimized stresses from (B) configuration are given in Supp. Mat. S2†. Once more, from this configuration (B), it appears difficult for the system to evolve and modify its configuration toward the target. The system deformation due to the SM application on P atom is shown in Fig. 5 (c), forcing the system to move toward Down-DFT, which finally relaxes in that configuration. This last diffusion stabilizes the system by -2.02 eV.

Finally, knowing the existence of this necessary intermediate minimum along the Minimum Energy Path (MEP), we are able to run NEB calculations to determine the activation barrier of the two-step diffusion. The full MEP is detailed on Fig. 6 (blue curve). From the configuration Up (A) to the first intermediate (B), where HOP2-O4' is broken, an activation barrier of 0.08 eV is obtained. The second activation barrier to rotate the phosphate group from intermediate (B) to configuration Down-DFT is large of 0.13 eV.

As a critical discussion, one might argue that a simple NEB calculation could have been run between the two configurations Up and Down-DFT, as usually performed. In such an approach, a linear interpolation would have been applied between the two known states, resulting in a bad description of the actual MEP of the diffusion. This approach would also miss the intermediate state corresponding to the breaking of H-bond, which finally appears to be a key step of the grafting process. Here, QMSM permits to identify this key step mechanism that unlocks the pre-adsorbed Up configuration.



(a) Optimized stresses



(b) Deformation: (A) to (B)

(c) Deformation: (B) to (C)

(d) Deformation: (A) to (E)

(e) Deformation: (E) to (F)

(f) Deformation: (F) to (G)

(g) Deformation: (G) to (H)

Fig. 5 (a) Variations of the distance d(HOP2-O4') induced by the stresses optimized on each atom of the configuration Up. The arrow length is proportional to the criterion magnitude. (b) and (c) refer to the strain field (green arrows) associated with the optimized stress (blue arrow) relative to the driven exploration (blue path in Fig. 4). (d), (e), (f) and (g) refer to the strain field associated with the optimized stress relative to the free exploration (orange path in Fig. 4).

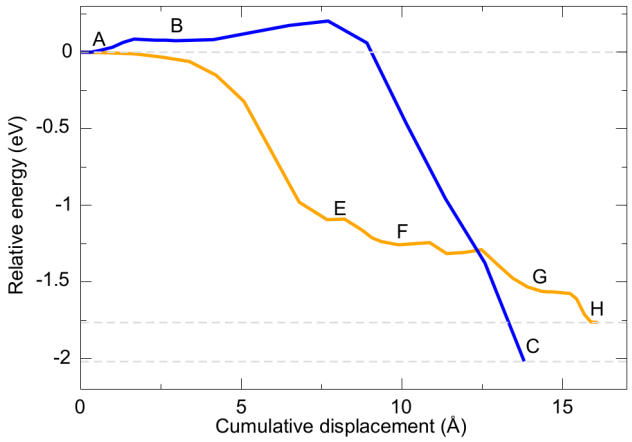


Fig. 6 Minimum energy paths obtained with driven (blue path) and free (orange path) explorations for the grafting of dTMP on $Al_2O_3(0001)$ surface using QMSM. Letters refer to configurations of Fig. 4. The non-labelled minimum appearing between F and G configurations corresponds to the OH group rotation and is not further discussed in the article.

Much more, it appears that the actual deformation allowing H-bond to break is not so simple: clearly influenced by the final result, the intuitive move that the user would have applied as a first approximation, would have been to drive the system going in the direction of the surface, or to repel the two atoms in the exact opposite direction of the actual bond HOP2-O4'. It is finally revealed with the QMSM approach that it is better to apply a stress in the upward direction on the OP2 atom (see Fig. 5 (b)), going in the opposite direction of the surface plane. This last upward motion is the mechanism that makes the PO3 group rotate and prepare for adsorption.

Finally, note that to allow the grafting between the configurations (B) and Down-DFT (blue MEP), we placed at the limit of the validity domain of the SM method: a stress of 2.50 Å was applied to make the system going out of its potential well and to search for a given state. This clearly indicates that strong constraints can be introduced when searching for a driven or pre-guided configuration with usual methods such as the NEB method.

For a sake of completeness, we applied other methods to look for the Down-DFT configuration. A drag method, keeping fixed only one atom of the phosphate group in the direction toward the surface, was carried out as usually applied to avoid introducing too much stress in the system, and performed on all the phosphate group atoms, one after the other. None of those tests has resulted in Down-DFT configuration, all of them have resulted in Down-QMSM, which will be described in the next section. Using a drag method, it is nevertheless possible to obtain the Down-DFT configuration: both the x, y and z coordinates of the P atom were kept fixed during the constrained relaxation and using the SM to drive and impose the direction of the P atom move toward the target Al-dft atom. If such a highly-constrained drag method is carried out without using SM as a guide, the final obtained configuration is not Al-dft atom. Those calculations still validate that we have to apply a strong constraint to get the desired configuration. Moreover, we can note that SM allows the user to drive a

calculation toward a target configuration, since even a strong constraint can result in non desired state. Here, performing the drag method with a correct estimate of the deformation field imposed even by a small move, allows to find the expected configuration.

Finally, based on the difficulty of exploring the DFT target adsorbed configuration, we conclude that Down-DFT is not the most supposedly to be studied. It is thus necessary to prospect while letting the system evolve through free exploration and keeping a high degree of freedom.

3.2 Free exploration from a given configuration

The second exploration aims at using QMSM from configuration Up, and follows step by step the deformations proposed, with the sole constraint that the phosphate group must meet the surface. For such exploration, SM are used to determine the deformations that reduce the sum of the distances between all the phosphate group atoms (4 O atoms, 2 H atoms and 1 P atom) with the surface plane. This example illustrates a free exploration from a given relaxed state and let the system evolves only guided by its intrinsic flexibility.

Using SM and starting from configuration Up (Fig. 4 (A)), we determine all optimized stresses that allows to respect the reduction of the sum previously described (all deformations meeting this criterion are provided in Supp. Mat. S3†). As shown on Fig. 5, the stress that induces the most favorable deformation is applied on P atom and allows the O atom to bind to the surface with a non-anticipated Al atom (labelled Al-qmsm on Fig. 3 (b)), which is different from Al-dft.

On the first intermediate in Fig. 4 (E), we note that the H-bond between the phosphate group and the sugar still exists (d(HOP2-O4') = 1.85 Å). In such configuration (E), the system is stabilized by -1.09 eV. Finally, from this identified minimum, we pursue the exploration and apply again the deformations that would allow the system to evolve toward another grafting of the phosphate group on the surface, using the QMSM method. The deformation field induced by the stress imposed on HOP2 atom is shown in Fig. 5 (e) (the complete histogram is given in Supp. Mat. S4†). A new reaction intermediate is obtained, where the H-bond still exists, but is formed with an O atom present on the thymine and already grafted on the surface. This atom is labelled OTh on Fig. 3 (a) (d(HOP2-OTh) = 1.68 Å). At this stage, in (F) configuration (Fig. 4 (F)), the system still stabilizes by -0.17 eV (see Supp. Mat. S5† for the whole histogram). Again, the strain field is calculated to optimize the distance between the phosphate group and the surface. P atom displacement optimizes this criterion and the induced deformation field is shown in Fig. 4 (F). Then the new minimum (G) (Fig. 4 (G)) is obtained, which stabilizes the system by -0.30 eV. Finally, a final search for meaningful deformations is made from this new intermediate (see Supp. Mat. S6† for the whole histogram) where the displacement of the OP2 atom appears as the most promising, resulting in the system deformation provided in Fig. 5 (g). By applying this last deformation, we manage to graft the phosphate group on the surface with two anchoring points, where the H atom finally leaves the phosphate group to bind to the surface. This final configuration is still stabilized with an energy gain of -0.21 eV.

Finally, we use the NEB method to calculate the activation barriers between each of the identified intermediates. The orange MEP on Fig. 6, even if the final configuration is less stable by 0.25 eV than Down-DFT configuration, shows that no activation barriers from (E) to (H) (the largest one is 0.02 eV from (F) to (G)) exists along this second investigated diffusion.

Thanks to QMSM, we find key reaction intermediates that preserve important interactions like H-bond, while accounting for the intrinsic deformations occurring within the molecule. The phosphate group HOP2 atom, initially forming an H bond with the O4' sugar atom, plays a central role in the overall diffusion of the phosphate group. The hydrogen bond remains present along the migration, jumping from interaction to interaction with the oxygen atoms present on the dTMP, which finally allows to pass from minimum to minimum along the diffusion of the phosphate group without crossing any barrier (Fig. 6 (orange MEP)).

Table 1 reports the values of typical angles (γ and χ in Fig. 3) of the dTMP during the grafting on the surface comparing driven *vs.* free exploration. Along driven exploration, γ varies up to 6 ° whereas along free exploration the variation is about 27 °, same for χ angle where the largest amplitude variation is reached during the free exploration (16 ° *vs.* 12 °). A free exploration allows to introduce a greater flexibility than the driven exploration.

QMSM coupling as applied in this section demonstrates that a very broad criterion, *i.e.* moving a group of atoms and not only forcing one atom, can result in a more realistic atomistic mechanism that takes into account the important interactions and deformations of a complex system along the diffusion.

Table 1 Geometric information (γ and χ angles) of the dTMP on reached configuration when grafted on the alumina surface as described in Fig. 4

Path	Config.	γ (°)	χ (°)
	Alone	-77.02	-122.61
	A	-179.26	70.80
Driven	B	-179.28	68.94
exploration	C	-173.09	58.87
	E	168.81	55.25
Free	F	159.65	54.28
exploration	G	146.64	59.31
	H	152.92	59.31

4 Diffusion in crystalline materials

Diffusion in crystalline materials[31] finds many applications in industrial domains as micro-electronics[32,33], energy[34,35], nuclear[36,37], corrosion[38] for instance. For such applications, it is necessary to link the atomic scale description to the macroscopic scale features for the design and/or optimization toward tailored final properties. For example, it may be necessary to perfectly place an atom at a given location in the material during the manufacturing process or to predict its final configuration, but also to be able to predict the penetration of desired or undesired species that may or may not weaken the material. All these questions require an exhaustive knowledge of the mode of diffusion of the studied species with the hope of mastering it experimentally and technologically by controlling the conditions of manufacture or

use. Obtaining an exhaustive library of all possible events for these defects represents a real challenge for the materials science community, even for the simplest of these defects.

For this study, we focus on two very well-know point defects in silicon to which related behaviour of silicon self-interstitials can alter the functioning of the microelectronic devices: the mono-interstitial (I_1) defect shown as an additional atom placed in the silicon crystalline lattice and the tetra-interstitial (I_4) cluster, where four additional interstitial atoms form a stable agglomerate of atoms. Many theoretical studies have been carried out on these defects to make a link between atomic configuration and global material properties[39–42]. Many explorations with different energetic models have been achieved[43–46]. The more precise results fitting with experiment are obtained with ab initio calculation[47–50]. In this part, we apply QMSM method, to both demonstrate that the QMSM method can reach an exhaustive screening with limited calculation time by exploiting the symmetry of so well-organized crystalline materials and identify some relevant mechanisms of migration.

4.1 Diffusion of mono-interstitial atom I_1 in silicon bulk

The mono-interstitial I_1 in crystalline silicon has two stable configurations with neutral charge[49]. The more stable configuration is the *Dumbbell* configuration (Fig. 7), characterized by two atoms sharing the site of one. The other possible state is the *Hexagonal* configuration (less stable by 0.04 eV compared to Dumbbell) (Fig. 8) where the interstitial atom is positioned between six silicon atoms on the diamond lattice. Both configurations present a great number of symmetry planes and axes. When it comes to the question of how these point defects diffuse, due to this high degree of symmetry and rotational invariant, multiple possibilities of diffusion direction and final configurations appear, while paths to reach them must be solved in details.

4.1.1 Defining an energy criterion to guide the exploration

In the previous application (Section 3), we have used a geometric criterion, chosen to monitor the propensity of the system to undergo an assumed structural rearrangement. Here, we choose to guide the evolution of the system following an energy criterion, in order to identify the deformations associated with a low energetic cost, but without prior knowledge of the configuration to be reached. In this study, the energy criterion is based on the Keating formalism[51] where the strain energy of the system ($E_{Keating}$ (see Eq. 1)) is calculated to evaluate the local strain, *i.e.* the deviation of the deformed structure from a reference configuration, as already used for studying the Si/SiO_2 interface[52].

$$E_{Keating} = \frac{1}{2}\sum_i k_b(b_i - b_0)^2 + \frac{1}{2}\sum_{ij} k_\theta(cos\theta_{ij} - cos\theta_0)^2 \qquad (1)$$

where k_b is the bond stretching constant (eV/Å), k_θ the angle bending constant (eV), b_i and b_0 are respectively the bond length of the i^{th} bond and its equilibrium distance (Å), θ_{ij} and θ_0 are the angle between the i^{th} and the j^{th} bonds and the equilibrium angle. Parameters are extracted from Table I [KT(s-h) parameters] in Ref.[53]. The strain energy is taken to be zero for the reference structure when starting the QMSM exploration (either Dumbbell or Hexagonal configurations) and positive for any deviation from it. The total strain energy is obtained by summing over all bonds shorter than 2.50 Å. This value was set considering the equilibrium Si-Si bond length, including elastic distortions that appear during atomic diffusion.

4.1.2 QMSM results from exploration on mono-interstitial system

QMSM coupling, according to Fig. 1, is used to investigate the diffusion of the I_1 defect both from the Dumbbell and the Hexagonal configurations. QMSM screening finds perfectly all the symmetries of the system, for which all deformations are obtained with respect to rotation, translation and permutation symmetry operations. This is illustrated on Fig. 9 showing the screening from the D atom of the Dumbbell configuration (see Fig. 7). The projections of the 3D hemisphere on plane (001) as obtained with the screening, are plotted respectively in the direction [001] on Fig. 9 (a), and in the direction [00-1] on Fig. 9 (b). The two projections are similar and exhibit larger strain energies (red areas) notably in the direction of the neighboring atoms, whereas the smaller ones are depicted by green shapes, revealing low-strain deformation energy and appropriate directions. The dark blue area appears towards the area left empty by the atom of the crystal that participates in the Dumbbell configuration, that has become D'.

In the following, we state that all symmetries are respected, so studying one so-defined hemisphere is sufficient. Considering the symmetries of the configuration (see Fig. 7), the QMSM exploration is performed on the interstitial atoms D and the two first neighbor atoms D1 and D2, as labelled in Fig. 7 (a) belonging to the same hemisphere. We impose a stress in the range of [1.00 Å - 2.00 Å]. Only the events that allow the defect to migrate are presented. Only the exploration on the atom D for a stress limit lower than 2.00 Å allows to leave the initial configuration. The configurations reached are described in Fig. 7 (b) where the initial Dumbbell configuration is shown in yellow (named D yellow) and summarized in Tab. 2. The energy barriers are calculated using NEB calculations, all DFT parameters are described in the Computational Details section. QMSM identifies several configurations as possible final states: two Hexagonal and two Dumbbell configurations. Finally, refining the search using NEB calculations between identified states with QMSM, the I_1 defect in Dumbbell configuration can only migrate toward an Hexagonal configuration: when determining the MEP from Dumbbell to Dumbbell, an intermediate position is found, which is an Hexagonal configuration. The associated barrier from Dumbbell to Hexagonal is large of 0.29 eV.

The same exploration is performed from the Hexagonal configuration from H, H1, H2, H3 and H4 as labelled in Fig. 8 (a), which are the closest atoms to Hexagonal assuming the symmetry of the configuration. As previously, the exploration is performed on these atoms with a stress in the range of [1.00 Å - 2.00 Å] and only the events that allow the defect to migrate are detailed.

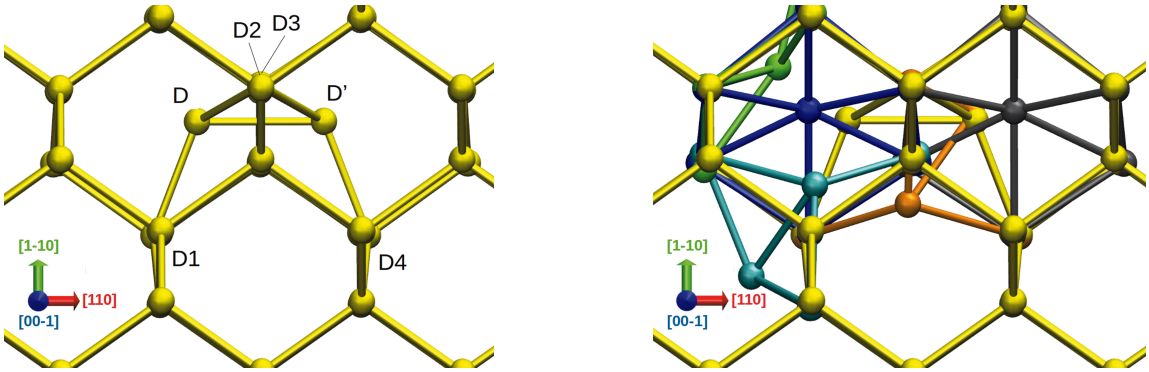


Fig. 7 (a) Dumbbell configuration of a I_1 Si atom in crystalline Si in plane (001), named D yellow. D and D' form the Dumbbell. D, D', D1 to D4 atoms are explored with QMSM. (b) Configurations reached from QMSM exploration. For the sake of clarity, each configuration is shown independently in Supp. Mat. S7.

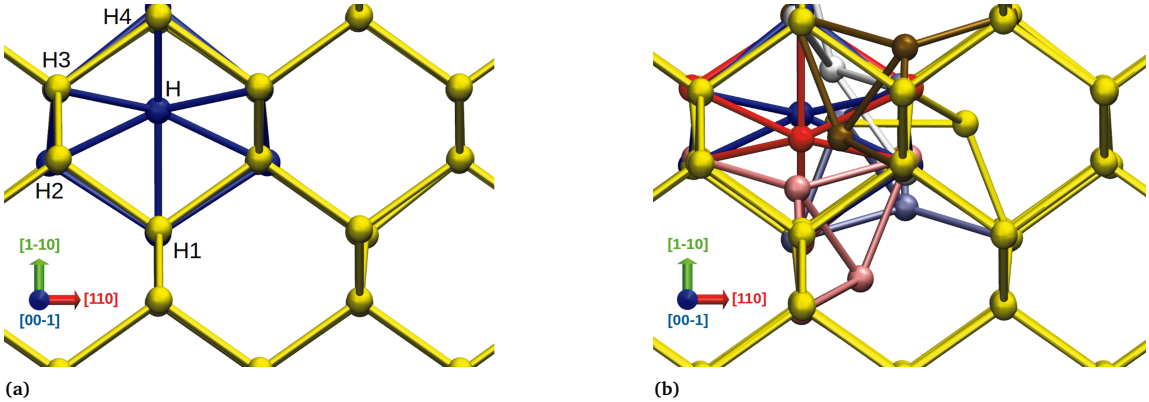


Fig. 8 (a) Hexagonal configuration of a I_1 Si atom in crystalline Si in plane (001), named H blue. H is the interstitial atom. H, H1 to H4 atoms are explored with QMSM. (b) Configurations reached from QMSM exploration. For the sake of clarity, each configuration is shown independently in Supp. Mat. S7.

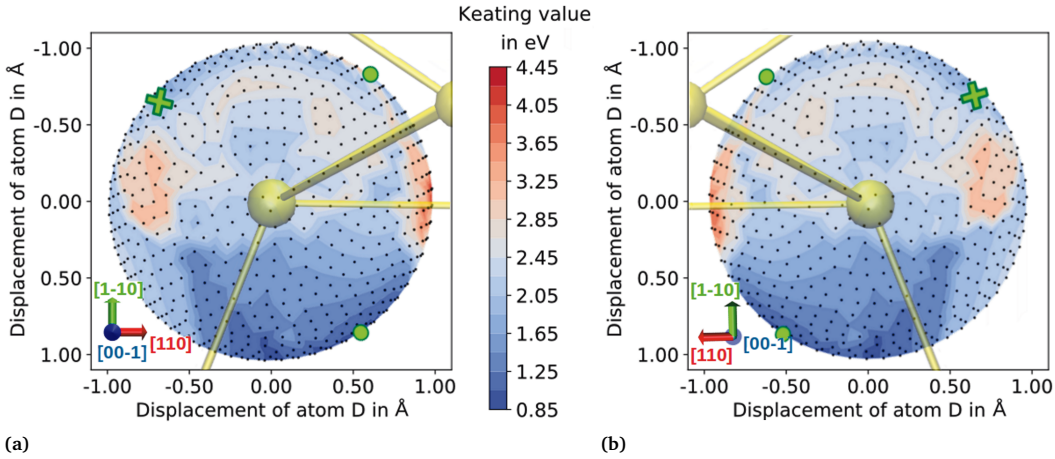


Fig. 9 QMSM screening around D atom of the Dumbbell configuration. (a) and (b) are the projections on plane (001) of the sphere respectively in the directions [001] and [00-1]. Abscissa and ordinate are the displacement of atom D in Å. Color scheme follows $E_{Keating}$ value in eV, ranging from blue to red areas respectively for low and high values. Green shapes correspond to deformations which minimize the $E_{Keating}$ value: green dots are directions in the symmetry (001) plane, green crosses are directions outside the symmetry plane.

Table 2 List of configurations identified by QMSM from Dumbbell configuration Figure 7 (a). H and D letters refer respectively to relaxed Hexagonal and Dumbbell configurations with associated colors as shown in Figure 7 (b) 'direct' indicates that no intermediate exists, in such case the corresponding activation barrier is provided.

Moved atom	Reached config.	Magnitude	Intermediate
D	H green	1.50 Å	H blue
D	H blue	1.50 Å	direct (0.29 eV)
D	D cyan	1.50 Å	H blue
D	D orange	1.75 Å	H black

The reached configurations are shown in Fig. 8 (b), where the initial Hexagonal configuration is the blue one, and summarized in Tab. 3. QMSM exploration identifies several final configurations revealing three main trends for the diffusion of an Hexagonal atom: i) an Hexagonal configuration, symmetric to the initial one regarding the plane created by the six crystalline atoms bonded with the interstitial atom. Such move requires an activation barrier of 0.03 eV (H red in Fig. 8 (b)). ii) an Hexagonal configuration in one of the first adjacent Hexagonal plane formed by six atoms. The diffusion from H configuration to adjacent H configuration requires an activation barrier of 0.19 eV (H orange in Fig. 8 (b)). iii) a Dumbbell configuration where an atom in crystalline position goes in interstitial position and forms a Dumbbell configuration, with an associated activation barrier of 0.25 eV (D yellow in Fig. 8 (b)). In this last case, several orientations of the Dumbbell are reachable because of the symmetry.

Table 3 List of configurations identified by QMSM from Hexagonal configuration. H and D letters refer respectively to relaxed Hexagonal and Dumbbell configurations with associated colors as shown in Figure 8 (b). 'direct' indicates that no intermediate exists, in such case the corresponding activation barrier is provided.

Moved atom	Reached config.	Magnitude	Intermediate
H	H red	1.00 Å	direct (0.03 eV)
H	D yellow	1.75 Å	direct (0.25 eV)
H	H white	1.50 Å	direct (0.19 eV)
H1	D purple	1.50 Å	direct (0.25 eV)
H2	D pink	1.50 Å	H red
H4	D brown	1.50 Å	H red

In a crystalline system, QMSM finds all minima around a given configuration. We identify the configurations of interest towards which the system tends to evolve and with the help of the knowledge of the symmetry of the system. Knowing these configurations, further refinement using MEP characterisation methods such as NEB allows to identify intermediate states, even if the amplitude of the applied deformations is high in some cases.

The exploration in the studied domain from the Hexagonal configuration reveals that this configuration can evolve into several final configurations, either Hexagonal or Dumbbell, when deformed from different atoms. We note here that both configurations, Hexagonal (H red) or Dumbbell (D yellow), are reachable from the H atom but with different deformation amplitudes, respectively 1.00 Å and 1.75 Å. However, if the exploration is carried out from the atom of the silicon crystal that is the first neighbour of H, i.e. H1 the minimum deformation amplitude to reach a Dumbbell configuration is 1.50 Å *i.e.* less than the deformation applied on H to reach the same configuration.

This observation clearly indicates that it is not only necessary to study the dynamics of a Hexagonal-type defect from the single atom in the interstitial position (H), but that it is also necessary to study the possibility of neighbouring atoms to make the system evolve toward another configuration by their ability to deform the initial atomic configuration themselves. This will be further discussed at the end of this section.

These results are consistent with Ref. [48] where the diffusion of self-interstitial in silicon occurs via successive jumps between the Hexagonal and Dumbbell sites, in which the latter is named split-⟨110⟩ defects and consistent with other mechanisms described in literature[43,47]. Figure 10 summarizes a possible migration path from a Dumbbell to another Dumbbell located in the next silicon hexagon and rotated by 180°. The intermediate states of the migration and the MEP are also given (see Fig. 10 (a) and (b) respectively).

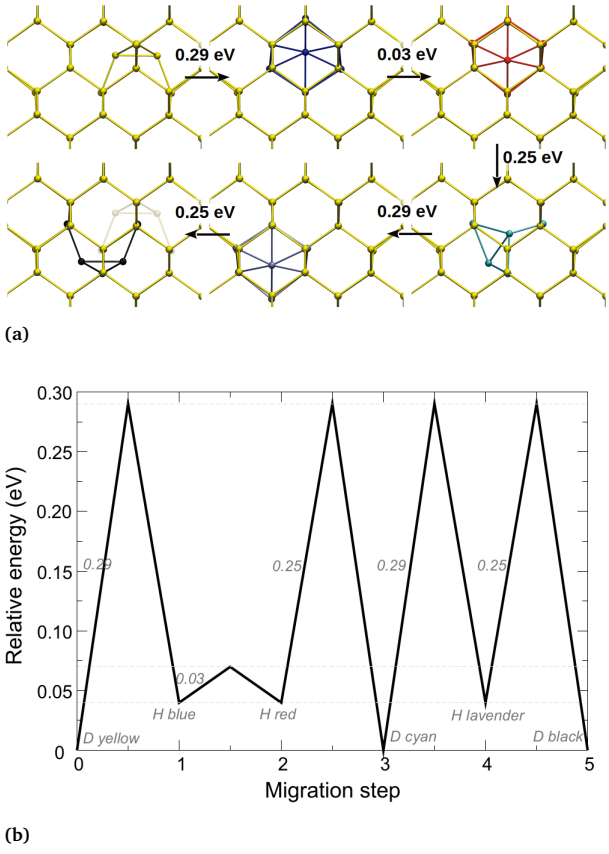


Fig. 10 (a) Reached configurations identified during the Minimum Energy Path (MEP). On the final configuration (in black at the bottom left), the initial state appears in shadow for a better understanding of the diffusion. (b) MEP of the diffusion of the mono-interstitial Si atom from between two Dumbbell configurations using Hexagonal configuration as intermediate states. Activation barriers are given in eV in italic grey font.

In this section, we show that the QMSM approach allows a quick identification of the favorable directions of diffusion based on the deformation of the crystal system. Thanks only to the

knowledge of a Hessian matrix, the SM exploration has a negligible computational cost. Moreover, the possibility for the user to propose criteria allows to reduce the time spent through a better selection of the diffusion directions to be retained and to limit the explored area. After having demonstrated the application of QMSM to a simple crystal system, we apply in the following paragraph, the QMSM approach to the diffusion of another point defect, more complex in terms of structure, namely the tetra-interstitial.

4.2 Diffusion of tetra-interstitial cluster I_4 in silicon bulk

We now focus on the identification of the atomic mechanisms allowing the migration of the tetra-interstitial (I_4) in its ground state as shown in Fig. 11 also named I_4^a [50,54]. Since we only study this configuration, we make the shortcut in the rest of this article to name the tetra-interstitial in its most stable state I_4. More precisely, we look for the atomic mechanisms that make I_4 going from an initial position in blue (I_4 blue) to the final position in red (I_4 red) in Fig. 11 (a) and (b) respectively. The I_4 structure forms two pentagons in two adjacent planes (both in (001) plane in the Fig. 11). When going from initial to final configurations of the I_4 defect, a total of twelve silicon atoms are involved in the migration process. This is explained in the following.

4.2.1 Defining a criterion for multi-atom displacements

QMSM is used to find the direction that is able to move a global configuration toward a known final configuration. The challenge is to understand the details of a complex diffusion involving the move of several atoms notably to identify which atom moves first and what are the possible intermediate states. For such goal, we implement a criterion based on the difference of positions between two configurations Δ_{Conf}. In the case of the tetra-interstitial, Δ_{Conf} is the difference between the deformed and the final configuration, which implies here to consider a total of twelve atoms. This criterion is defined in Eq. 2.

$$\Delta_{Conf} = \sum_i \sqrt{\sum_j (Cd_{i,j} - Cf_{i,j})^2} \quad (2)$$

where $Cd_{i,j}$ and $Cf_{i,j}$ are the j^{th} coordinates of atom i, respectively in the deformed configuration (after deformation with SM) and in the final configuration. The lower the criterion, the closer to the final configuration the deformed configuration is.

4.2.2 QMSM results from exploration on tetra-interstitial system

As in the previous section, QMSM is used to help in the identification of the MEP of the I_4 migration. QMSM exploration is applied from the initial configuration and from all intermediate configurations subsequently identified. Only directions selected for a stress of 1.50 Å that optimized the criterion of Eq. 2 are retained. From these identified directions, a drag method applied with DFT is performed by maintaining the three coordinates in x, y and z of the atom moved in the given direction. Once the transition point is reached, the system is relaxed. The configuration taken as reference is the Initial configuration. The configuration energies given in the following are therefore compared to this configuration.

Table 4 List of configurations identified by QMSM from I_4 blue configuration and from Intermediate states TIS-1, TIS-2 and TIS-3 as shown in Fig. 11. The configurational energy is compared to the total energy of I_4 blue initial considered as the reference system. For symmetric configurations, namely I_4 blue and TIS-2, we only provide information for one atom, the corresponding atom of the symmetric configuration is given in parentheses.

Starting config.	Moved atom	Reached config.
I_4 blue	T11 (T5)	TIS-1 (1.48 eV)
I_4 blue	T9 (T3)	TIS-1' (0.97 eV)
TIS-1	T5	TIS-2 (1.00 eV)
TIS-1	T9	TIS-2' (2.07 eV)
TIS-1	T3	TIS-2' (2.07 eV)
TIS-1	T1	TIS-2" (3.16 eV)
TIS-2	T9 (T3)	TIS-3 (1.48 eV)
TIS-3	T6	I_4 red
TIS-3	T3	I_4 red

The reached configurations are described in Fig. 11 (c) to (h) and summarized in Tab. 4. The activation barriers have been determined using NEB calculations. The complete migration path presented as a MEP (red curve), as well as the intermediate diffusion tests (black curves), are presented in Fig. 12. For the sake of clarity, we only discuss the displacements from the twelve atoms already introduced above. However, for some of the configurations obtained during the search for atomic mechanisms related to I_4 migration, other atoms seen as atoms in the crystalline position may induce structural modifications of the I_4 defect, as discussed in section 4.1.2. We do not discuss them hereafter.

From the Initial configuration, we select two directions of interest identified on atom T9 and T11 shown in Fig. 11 (c). Each direction is explored using the drag method and leads to two new configurations: from the T11 atom, the reached configuration is named TIS-1 shown in Fig. 11 (d) and from the T9 atom, the reached configuration is named TIS-1' shown in Supp. Mat. S8 (a)†. Between the two configurations obtained, TIS-1 is less stable by 1.48 eV compared to the Initial configuration. TIS-1' is more stable than TIS-1 with a configuration energy of 0.97 eV. Finally, the calculated activation barrier going from Initial to TIS-1 is large of 2.23 eV, whereas the activation barrier going from Initial to TIS-1' is large of 1.08 eV, with an associated back activation barrier toward Initial configuration of 0.11 eV.

Regarding the exploration from the TIS-1' configuration, we conclude that TIS-1' is metastable, tending to return to the Initial configuration, not consistent with the I_4 migration. This state is however of crucial interest in particular for the understanding of leakage currents observed in microelectronic silicon based-devices[55] whose origin is well identified as coming from the metastability of structural defects, out of the scope of the present study. We therefore do not pursue the exploration of the TIS-1' configuration.

From the exploration of configuration TIS-1, four directions are identified from four different atoms T5, T9, T3, T1 (see Fig. 11 (d)) leading to three different configurations. From T5 atom, TIS-2 intermediate configuration is obtained, featured by a configuration energy of 1.00 eV, the lowest in the built list

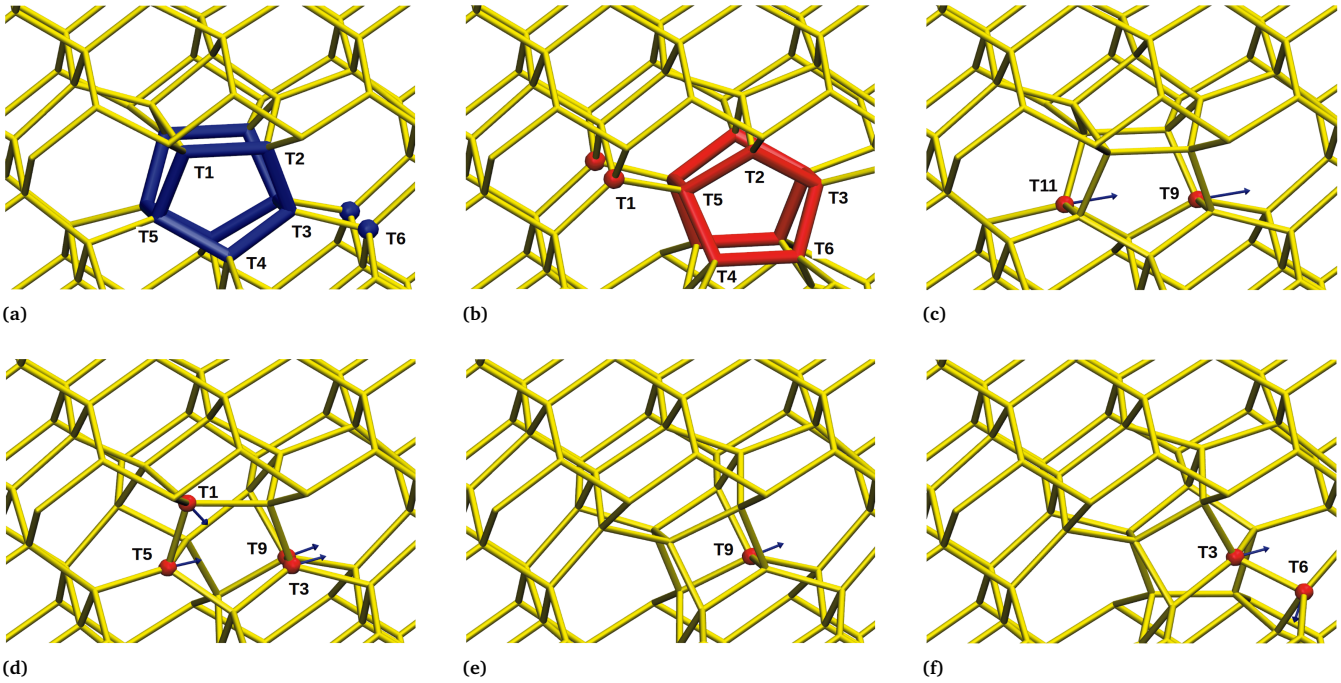


Fig. 11 (a) and (b) Ground states of the I_4 defect referred respectively as the Initial configuration (I_4 blue) and Final configuration (I_4 red) of the studied migration. I_4 ground state can be depicted as two adjacent pentagons. Atoms that take part in the diffusion are identified by the labels T1 to T6. For sake of clarity, only one pentagon has been labelled, but a total of 12 silicon atoms are involved in the considered migration. Additional labelling from T7 to T12 can be applied similarly on the other pentagon. (c) QMSM exploration from I_4 blue shown in (a). TIS-1, TIS-2 and TIS-3 configurations and associated directions resulting from the QMSM exploration are presented respectively in (d), (e) and (f). Corresponding minimum energy path is given in Fig. 12.

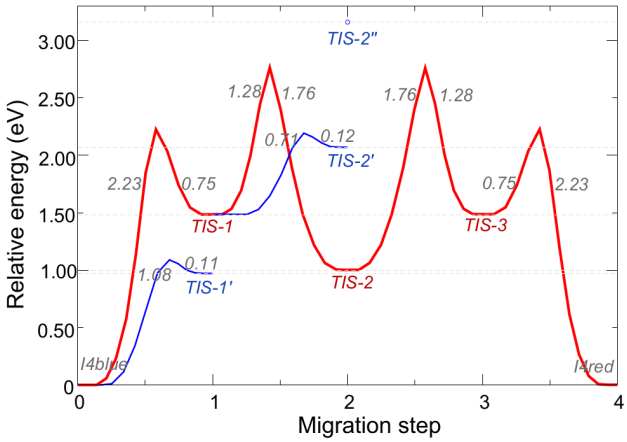


Fig. 12 Energy diagram of the configurations reached during the I_4 migration referring to Fig. 11. The whole MEP from I_4 blue to I_4 red is shown in red curve. Blue curves are MEP of tested diffusions outside from the MEP of interest. Activation energies are given in italic grey font and in eV.

shown in Fig. 11 (e). The activation barrier from TIS-1 to TIS-2 is large of 1.28 eV. From TIS-1, this diffusion stabilizes the system by 0.48 eV. Other explored directions from T9 and T3 result in TIS-2' intermediate shown in Supp. Mat. S8 (b)†. Further explored, TIS-2' does not result in meaningful configurations for I_4 migration. This state is actually revealed to be metastable, destabilizing the system (by 0.59 eV from TIS-1) that gets back to the TIS-1 configuration. The activation barrier from TIS-1 to TIS-2' is large of 0.71 eV and the corresponding back reaction exhibits an activation barrier of 0.12 eV. For details, an other configuration named TIS-2" and shown in Supp. Mat. S8 (c)† can be obtained, when moving T1 atom. TIS-2" destabilizes even more the system by more than 2.00 eV from TIS-1, with a configuration energy of 3.16 eV.

TIS-2 is a central intermediate of the migration of I_4. TIS-2 shown in Fig. 11 (e) is a highly symmetric configuration where initial pentagons form lozenge-like patterns. In such TIS-2 configuration, T1 (T7) and T6 (T12) atoms and T5 (T11) and T3 (T9) atoms are symmetric. Applying the QMSM exploration on TIS-2, only two directions are identified from T9 and T3 atoms, symmetric of each other. Only one atom is thus studied: the exploration from T9 results in an intermediate configuration named TIS-3 and shown in Fig. 11 (f). TIS-3 is the same atomic configuration as TIS-1. Path from TIS-2 to TIS-3 is similar to the back reaction from TIS-2 to TIS-1. Such move requires an activation barrier of 1.76 eV. Likewise, applying the QMSM exploration on TIS-3 identified two directions from T3 and T6 atoms (see Fig. 11 (f)).

Both lead to the final position of I_4 by the back reaction from Initial configuration TIS-1. The activation barrier is large of 0.75 eV. In this last observation, as a critical discussion, we can notice that the NEB does not inform about the exact mechanism of diffusion. The QMSM approach gives information on the atom that initiates the last movement, in this case two atoms can play this role.

Finally the I_4 migration exhibits a symmetric reaction path featured by three intermediate states, where the migration occurs by moving one atom after the other. Intermediate configuration TIS-2, highly symmetric and stable regarding the other configurations, has a central role in the diffusion of I_4. The migration mechanism of I_4 preferably keeps the cluster atoms in a compact state rather than dissociating the cluster atoms. The I_4 defect cluster is a very stable defect. The mechanisms of structural modification identified in this work are consistent with experimental observations, indicating that the I_4 migration is a really tight process. The first barrier obtained between the initial configuration and TIS-1 is large to overcome, more difficult than the one associated with a local modification of the structure observed in the metastable configuration TIS-1'.

Beyond the study of the migration of a defect in silicon, this work shows the need to use the QMSM approach to guide the user through the choices to be made in the case of complex and/or large initial configurations. Neither a random choice, nor an exhaustive study can be performed manually. QMSM is useful in two ways. First, it allows identifying which atom has to be considered in a specific configuration. But QMSM goes even further by proposing an exact direction for the movement of that atom. We note here that the exploration is significantly reduced to a few atoms. In this example reduced to 12 atoms of interest, which is yet a high number when it comes to testing their move in a 3D space, only eight atoms are useful, for a total of four configurations considered. Moreover, from a given configuration, the range of possibilities is reduced to a single case.

Importantly, we have presented here two different ways of using the SM to guide DFT calculations, based on two types of information extracted from the QMSM selection process: i) using the SM strain field to deform a structure that is further energy minimized; ii) or using the direction of the optimized stress provided by the SM analysis in a subsequent drag method. The first approach is the most direct to implement since it only requires one DFT calculation. However, it is also the most guided since the deformation is applied to the whole system, which in turn is oriented toward a new configuration. As a consequence, depending on the energy landscape, it is possible to apply a deformation that brings the system far enough to find a new configuration while missing intermediate states. On the contrary, the second approach is based on the step by step application of a single atomic stress, while letting the rest of the system free to evolve. This approach is then more demanding in computational resources, but is less prone to miss possible configurations. The example of the Dumbbell exploration illustrates well this consideration. In the end, both approaches can be used. Applying the SM strain field is more efficient to extract the system out of the potential well, but using the stress direction as a guide with a drag method is recommended when it comes to elucidating the diffusion path. In both cases, the results can be refined using NEB calculations to determine the activation barrier between the identified states.

5 Discussion

On the basis of these examples, thanks to an exploration guided by the knowledge of the intrinsic deformation of a given system provided by SM, we have illustrated the capability to explore potential energy surfaces from a minimum, whether the final configuration is known or not, whether the reaction path is known or not, thus allowing the identification of important reaction intermediates.

The low cost of the SM approach in the exploration phase (screen, score steps) when used with DFT is an important added value. Once calculated for a relaxed atomic configuration, the SM constitute an immediately accessible data bank with almost no computational cost. In the provided examples, a quick exploration is performed in the 3D space by using a sphere of equally distributed points as described in section 2. This sphere, constructed for each atom and allowing to define a 3N exploration direction vector (see Fig. 5 (a)), makes it possible to quickly screen all the directions and thus reduce the number of evaluations for the escape of the minimum energy configuration. Regardless of the reaction or displacement considered, it is possible for the user to experiment with one or more exploration criteria, to apply it to all atoms of the system or a region of interest only. Without any additional calculation, it is also possible for the user to explore (screen) several directions from the same configuration on a single atom or on several atoms, if the scores are difficult to separate or if several scores (the second, the third etc.) seem relevant. This is illustrated in the case of the monointerstitial (Fig. 9) and tetrainterstitial diffusion (Fig. 11).

Moreover, each time a new reaction intermediate is identified, it is possible to evaluate and quickly redefine the direction of exploration by resetting an appropriate variable. Indeed, depending on the scores obtained for the metastable states during the reaction, we can propose a more suitable direction of displacement, or the identification of a better atom or set of possible candidates to pursue the diffusion or the more global mechanism studied. This is illustrated throughout the DNA grafting reaction on the alumina surface (follow the blue arrows in Fig.5 (b) to (g)). This is also obtained in the case of the migration of I_4 into silicon described in section 4.2.2 and shown in Fig. 11. We only present the migration of I_4 from the blue to red configuration, as the aim of the paper was not to study the migration of I_4 into silicon. However, it should be noted that along the migration path of I_4, several intermediates were identified (see the MEP in Fig. 12 showing the relative energies of the TIS-1', TIS-2', TIS-2" configurations (structures not shown)) and were also explored, but are not detailed in the paper as they did not lead to the red configuration.

This paper is a new definition of SM when used with DFT. By definition, electronic structure relaxation codes implementing DFT do not offer the possibility of systematic exploration of the PES, but use algorithms to characterize the minimum energy paths, as used in this paper. Few methods exist to escape the

energy minimum and explore the PES from this local minimum without knowledge of the final configuration. Recently, such a development has been made with the ART methodology by the authors [56], which allows to randomly explore the local minimum with the great advantage of not knowing in advance the final configuration to be reached. Despite its advantages in terms of computational cost, this coupling could be further improved by guiding the escape from the local minimum thanks to the selection of an optimized exit direction.

Beyond this first attempt, this paper reveals the potential of the SM method to guide PES exploration methods, including other techniques than those based on DFT. The implementation of the QMSM method, and more generally of the SM method to reduce the trials to escape the local minimum of complex systems, will be considered in the near future, in particular with ab initio or non ab initio Molecular Dynamics codes. Indeed, it is easy to think that with the SM, the definition of the reaction coordinates, *i.e.* the collective variables, can be defined in a direction identified as relevant, based on considerations of deformation of the system studied under a given stimulus. A set of directions can also be considered, which can be of great interest in metadynamics for example. It is also important to mention again this advantage, pointing out that these collective displacement variables can be easily redetermined by the SM during the reaction as soon as a new local minimum is identified.

6 Conclusions

In this article, we have presented the QMSM approach that makes use of the Static Modes to explore all atomic stresses and associated strain fields in a systematic way. This QMSM screening step results in the selection of the most relevant deformations regarding a property or behavior of interest. The results are then used as a starting point to guide subsequent DFT calculations in search of new configurations.

As illustrated when grafting a dTMP on alumina surface, QMSM is as well adapted to the search of a given minimum as to carry out an exploration without prior knowledge of the final configuration. In the first case, the driven exploration shows that, even if the investigated mechanism is not the most favorable, QMSM can force the exploration to find a path between two known minima, by proposing not intuitive intermediate steps. In the second case, we have shown that a free exploration, associated with a less constrained criterion, allows to fully account for the intrinsic flexibility of the system, letting it evolve through important interactions and more realistic mechanisms toward a not presupposed configuration.

The second application that we have treated, which concerns the study of defect diffusion in silicon, demonstrates that the proposed methodology is also efficient when applied to crystalline systems, since it respects the symmetry and can take advantage of the latter to find all possible configurations around an initial state. In this part, we have also shown how the deformation energy can be used as a relevant criterion to select the SM that are the most favorable to guide the exploration and allow the system to get out of the initial state. Finally, in the case of a more complex system, like the tetra-interstitial system, our approach allows to reduce the number of atoms to be explored and used in ab initio calculations.

Through these applications, we have proposed different ways to use the Static Modes to guide the user through the choices to be made when neither a random nor an exhaustive exploration is adapted. The QMSM process then results in identifying precisely which atom has to be considered and constrained, and in what direction. The range of possibilities is then drastically reduced and can be tested with a very acceptable computational cost. Importantly, the screening, scoring and selection steps can be adapted to satisfy a criterion that is chosen regarding the system of interest. Therefore, the QMSM coupling can be used for general applications. We believe that this approach represents a step forward in making the exploration of complex systems more reachable, and thus participates in the design of new, predictive and multi-level modeling strategies.

Indeed, beyond the choice made here of using SM to guide DFT calculations, we discuss the possibility of using the same selection process to propose relevant reaction coordinates, i.e, collective variables, in other methodologies like metadynamics, and more widely, biased molecular dynamics methods.

Computational details

DNA model on Al_2O_3 surface

We used the Density Functional Theory plus Dispersion (DFT-D) method based on the Generalized Gradient Approximation (GGA), employing method with Becke-Johnson damping (DFT-D3) [57,58]. We used the Projector Augmented Wave method (PAW [59]), as implemented in VASP [60], to keep frozen the core electrons, and a cut-off energy of 500 eV. A dipole correction was introduced, oriented along the normal vector of the surface plane. Geometry optimization was performed with 10^{-6}eV per cell and 10^{-4}eV/Å per cell criteria for energy and force, respectively. A 2 x 2 x 1 Monkhorst-Pack grid was used for the calculations. The corundum (0001) Al_2O_3 surface was modeled by a periodic slab. The unit cell was repeated as (4 x 2 x 2), and a vacuum space of 25 Å was added along the normal vector. Finally the supercell is 19.057 x 16.503 x 34 Å^3 and it is composed by 96 Al, 160 O and 37 H. To simulate the bulk, O and Al atoms of the bottom layers are kept fixed. Drag method calculations are performed moving the fixed atom by 0.2 Å between each step. Climbing-Image Nudged Elastic Band method [61] calculations are performed using eight images between the initial and final state for the determination of the activation barriers.

Interstitial species in silicon bulk

We use the density functional theory with a Perdew-Zunger (LDA) exchange-correlation functional as implemented in Quantum Espresso [62]. Norm-conserving pseudopotential have been used. The occupation smearing used is Methfessel-Paxton first-order spreading [63] with a value of the gaussian spreading for Brillouin zone of 0.004 Ry, the cut-off energy is set to 20.0 Ry. Geometry optimization was performed with 10^{-6} eV per cell and 10^{-5}

eV/Å per cell criteria for energy and force, respectively. We use a mesh of 1 x 1 x 1 grid with the integration of Monkhorst Pack. The cubic supercell is 30.61 x 30.61 x 30.61 $Å^3$ and is composed of 217 and 220 silicon atoms, respectively for monointerstitial and tetra-interstitial defects. Drag method calculations are performed moving the fixed atom by 0.15 Å between each step. Climbing-Image Nudged Elastic Band method [61] calculations are performed using fourteen images between the initial and final state for the determination of the activation barriers.

Author contributions

AH and MB: Conceptualization and supervision, methodology, analysis, writing, review, editing - GL: Conceptualization and supervision, methodology - LF: methodology, calculations.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

The simulations were performed using HPC resources from CALMIP (Grant P18039). We thank Antoine Jay, from LAAS-CNRS, for helpful discussion.

Notes and references

1 G. R. Schleder, A. C. M. Padilha, C. Mera Acosta, M. Costa and A. Fazzio, J. Phys.: Mater., 2019, **2**, 032001.
2 J. Kästner, Wiley Interdiscip. Rev.: Comput. Mol. Sci., 2011, **1**, 932–942.
3 G. Bussi and A. Laio, Nat. Rev. Phys., 2020, **2**, 200–212.
4 M. Trochet, A. Sauvé-Lacoursière and N. Mousseau, J. Chem. Phys., 2017, **147**, 152712.
5 A. L. Dewyer, A. J. Argüelles and P. M. Zimmerman, Wiley Interdiscip. Rev.: Comput. Mol. Sci., 2018, **8**, e1354.
6 S. Maeda and Y. Harabuchi, J. Chem. Theory Comput., 2019, **15**, 2111–2115.
7 J. P. Unsleber and M. Reiher, Annu. Rev. Phys. Chem, 2020, **71**, 121–142.
8 M. Brut, A. Estève, G. Landa, G. Renvez and M. Djafari Rouhani, Eur. Phys. J. E, 2009, **28**, 17–25.
9 G. Renvez, A. Estève, M. Brut, G. Landa, M. D. Rouhani and A. Dkhissi, Phys. Chem. Chem. Phys., 2011, **13**, 14611–14616.
10 M. Brut, A. Estève, G. Landa, G. Renvez and M. D. Rouhani, J. Phys. Chem. B, 2011, **115**, 1616–1622.
11 M. Brut, A. Estève, G. Landa and M. Djafari Rouhani, J. Phys. Chem. B, 2014, **118**, 2821–2830.
12 M. Brut, A. Trapaidze, A. Estève, A. Bancaud, D. Estève, G. Landa and M. D. Rouhani, Appl. Phys. Lett., 2012, **100**, 163702.
13 D. Mercier, J. C. Rouchaud and M. G. Barthés-Labrousse, Appl. Surf. Sci., 2008, **254**, 6495–6503.
14 E. Ruckenstein and Z. F. Li, Adv. Colloid Interface Sci., 2005, **113**, 43–63.
15 S. Bonakala, A. D. Pathak, A. Deyko, C. Christova, I. Rudra and G. Verbist, ACS Appl. Mater. Interfaces, 2020, **12**, 18101–18109.
16 A. Mahapatra, D. Prochowicz, M. Mahdi Tavakoli, S. Trivedi, P. Kumar and P. Yadav, J. Mater. Chem. A, 2020, **8**, 27–54.
17 X. Lin, D. Cui, X. Luo, C. Zhang, Q. Han, Y. Wang and L. Han, Energy Environ. Sci., 2020, **13**, 3823–3847.
18 D. Hetemi, V. Noël and J. Pinson, Biosensors, 2020, **10**, 4.
19 L.-R. Lee, V. K. Karapala, Y.-L. Lin, H.-C. He and J.-T. Chen, J. Phys. Chem. C, 2020, **124**, 11870–11876.
20 B. Kasprzyk-Hordern, Adv. Colloid Interface Sci., 2004, **110**, 19–48.
21 P. Jain, B. Patidar and J. Bhawsar, J. Bio- Tribo-Corros., 2020, **6**, 43.
22 V. Maurice and P. Marcus, Prog. Mater. Sci., 2018, **95**, 132–171.
23 P. Claus, Top. Catal., 1998, **5**, 51–62.
24 H. Chouirfa, H. Bouloussa, V. Migonney and C. Falentin-Daudré, Acta Biomater., 2019, **83**, 37–54.
25 D. Loffreda, Surf. Sci., 2006, **600**, 2103–2112.
26 D. Costa, C.-M. Pradier, F. Tielens and L. Savio, Surf. Sci. Reports, 2015, **70**, 449–553.
27 O. Borck and E. Schröder, J. Phys.: Condens. Matter, 2005, **18**, 1–12.
28 M. Poberžnik, D. Costa, A. Hemeryck and A. Kokalj, J. Phys. Chem. C, 2018, **122**, 9417–9431.
29 F. Tielens, M. Gierada, J. Handzlik and M. Calatayud, Catal. Today, 2020, **354**, 3–18.
30 T. Calais, B. Playe, J.-M. Ducéré, J.-F. Veyan, S. Rupich, A. Hemeryck, M. Djafari Rouhani, C. Rossi, Y. J. Chabal and A. Estève, J. Phys. Chem. C, 2015, **119**, 23527–23543.
31 J. Zhong, L. Chen and L. Zhang, J. Mater. Sci., 2020, **55**, 10303–10338.
32 S. Ke, D. Li and S. Chen, J. Phys. D: Appl. Phys., 2020, **53**, 323001.
33 S. Y. Chen, Z. W. Wu, K. X. Liu, X. J. Li, N. Luo and G. X. Lu, J. Appl. Phys., 2013, **113**, 044901.
34 F. Sturm, M. Trempa, S. Schwanke, K. Schuck, C. Kranert, C. Reimann and J. Friedrich, J. Cryst. Growth, 2020, **540**, 125636.
35 J. Collins, J. P. de Souza, M. Hopstaken, J. A. Ott, S. W. Bedell and D. K. Sadana, iScience, 2020, **23**, 101586.
36 A. Lushchik, E. Feldbach, E. A. Kotomin, I. Kudryavtseva, V. N. Kuzovkov, A. I. Popov, V. Seeman and E. Shablonin, Sci. Rep., 2020, **10**, 7810.
37 L. Messina, T. Schuler, M. Nastar, M.-C. Marinica and P. Olsson, Acta Mater., 2020, **191**, 166–185.
38 N. Djamel and A. Samira, J. Mol. Liq., 2021, **323**, 114642.
39 V. S. Proshchenko, P. P. Dholabhai, T. C. Sterling and S. Neogi, Phys. Rev. B, 2019, **99**, 014207.
40 I. Martin-Bragado, I. Avci, N. Zographos, M. Jaraiz and P. Castrillo, Solid-State Electron., 2008, **52**, 1430–1436.
41 A. Jay, M. Raine, N. Richard, N. Mousseau, V. Goiffon, A. Hémeryck and P. Magnan, IEEE Trans. Nucl. Sci., 2017, **64**, 141–148.

42 A. Platonenko, F. S. Gentile, J. Maul, F. Pascale, E. A. Kotomin and R. Dovesi, Mater. Today Commun., 2019, **21**, 100616.

43 L. A. Marqués, L. Pelaz, P. Castrillo and J. Barbolla, Phys. Rev. B, 2005, **71**, 085204.

44 M. Trochet, L. K. Béland, J.-F. Joly, P. Brommer and N. Mousseau, Phys. Rev. B, 2015, **91**, 224106.

45 Y. A. Du, T. J. Lenosky, R. G. Hennig, S. Goedecker and J. W. Wilkins, Phys. Status Solidi B, 2011, **248**, 2050–2055.

46 L. A. Marqués, M. Aboy, M. Ruiz, I. Santos, P. López and L. Pelaz, Mater. Sci. Semicond. Process., 2016, **42**, 235–238.

47 W.-K. Leung, R. J. Needs, G. Rajagopal, S. Itoh and S. Ihara, Phys. Rev. Lett., 1999, **83**, 2351–2354.

48 R. J. Needs, J. Phys.: Condens. Matter, 1999, **11**, 10437–10450.

49 P. Rinke, A. Janotti, M. Scheffler and C. G. Van de Walle, Phys. Rev. Lett., 2009, **102**, 026402.

50 Y. A. Du, R. G. Hennig, T. J. Lenosky and J. W. Wilkins, Eur. Phys. J. B, 2007, **57**, 229–234.

51 P. N. Keating, Phys. Rev., 1966, **145**, 637–645.

52 N. Salles, N. Richard, N. Mousseau and A. Hemeryck, J. Chem. Phys., 2017, **147**, 054701.

53 S. Lee, R. J. Bondi and G. S. Hwang, Phys. Rev. B, 2011, **84**, 045202.

54 N. Arai, S. Takeda and M. Kohyama, Phys. Rev. Lett., 1997, **78**, 4265–4268.

55 C. Durnez, V. Goiffon, C. Virmontois, P. Magnan and L. Rubaldo, IEEE Trans. Electron Devices, 2020, **67**, 4940–4946.

56 A. Jay, C. Huet, N. Salles, M. Gunde, L. Martin-Samos, N. Richard, G. Landa, V. Goiffon, S. De Gironcoli, A. Hémeryck and N. Mousseau, 2020, **16**, 6726–6734.

57 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, J. Chem. Phys., 2010, **132**, 154104.

58 S. Grimme, S. Ehrlich and L. Goerigk, J. Comput. Chem., 2011, **32**, 1456–1465.

59 G. Kresse and J. Furthmüller, Comput. Mater. Sci., 1996, **6**, 15–50.

60 G. Kresse and D. Joubert, Phys. Rev. B, 1999, **59**, 1758–1775.

61 G. Henkelman, B. P. Uberuaga and H. Jónsson, J. Chem. Phys., 2000, **113**, 9901–9904.

62 P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru and S. Baroni, J. Chem. Phys., 2020, **152**, 154105.

63 M. Methfessel and A. T. Paxton, Phys. Rev. B, 1989, **40**, 3616–3621.