Finding reaction pathways and transition states: r-ARTn and d- ARTn as an efficient and versatile alternative to string approaches

Antoine Jay, Christophe Huet, Nicolas Salles, Miha Gunde, Layla Martin-Samos, Nicolas Richard, Georges Landa, Vincent Goiffon, Stefano de Gironcoli, Anne Hémeryck, et al.

JCTC Journal of Chemical Theory and Computation

Structure Prediction

Finding reaction pathways and transition states: r-ARTn and d-ARTn as an efficient and versatile alternative to string approaches

Antoine Jay, Christophe Huet, Nicolas Salles, Miha Gunde, Layla Martin-Samos, Nicolas Richard, Georges Landa, Vincent Goiffon, Stefano de Gironcoli, Anne Hemeryck, and Normand Mousseau

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.0c00541 • Publication Date (Web): 14 Aug 2020

Finding reaction pathways and transition states: r-ARTn and d-ARTn as an efficient and versatile alternative to string approaches

Antoine Jay,[*,†] Christophe Huet,[‡] Nicolas Salles,[¶] Miha Gunde,[†] Layla Martin-Samos,[¶] Nicolas Richard,[§] Georges Landa,[†] Vincent Goiffon,[‡] Stefano De Gironcoli,[‖,¶] Anne Hémeryck,[†] and Normand Mousseau[⊥]

†*LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France*

‡*ISAE-SUPAERO, Universit de Toulouse, F-31055 Toulouse, France*

¶*CNR-IOM, Democritos National Simulation Center, Istituto Officina dei Materiali, c/o SISSA, via Bonomea 265, IT-34136 Trieste, Italy*

§*CEA, DAM, DIF, Bruyères-le-Châtel, F-91297 Arpajon, France*

‖*SISSA, via Bonomea 265, IT-34136 Trieste, Italy*

⊥*Département de Physique and Regroupement québécois sur les matriaux de pointe, Université de Montréal, Département de Physique, C.P. 6128, succursale Centre-ville H3C 3J7 Montréal Canada Montréal, Canada*

E-mail: antoine.jay@laas.fr

Abstract

Finding transition states and diffusion pathways is essential to understand the evolution of materials and chemical reactions. Such characterization is hampered by the heavy computation costs associated with exploring energy landscapes at *ab-initio* accuracy. Here, we revisit the activation-relaxation technique (ARTn) to considerably

reduce its costs when used with density functional theory (DFT) and propose three adapted versions of the algorithm to efficiently *(i)* explore the energy landscape of complex materials with the knowledge of a single minimum (ARTn); *(ii)* identify a transition state when two minima or a guess transition state are given (refining ART or r-ART) and *(iii)* reconstruct complex pathways between two given states (directed ART or d-ART). We show the application of these three variants on benchmark examples and on various complex defects in silicon. For the later, the presented improvements to ART leads to much more precise transition states while being two to six times faster than the commonly used string methods such as the Climbing Image Nudged Elastic Band method (CI-NEB).

Introduction

In the non-diffusive limit, the thermodynamical and dynamical properties of systems are determined by a discrete set of local minima and connected transition states[1] in the configurational space. Such set defines the Potential Energy Surface (PES) of the system.

Because of the evanescent nature of transition states and the presence of complex reaction mechanisms underlying the transitions, the exploration of the potential energy surfaces of molecules and materials remains a major challenge for modern computational chemistry and solid state physics. Several approaches addressing this question have been proposed and improved. These range from the earlier constrained-relaxation perpendicular to a guessed jump direction,[2] to the more sophisticated transition path sampling[3] that might provide free energy pathways connecting nearby known states.

Among all these methods, the Activation Relaxation Technique (ARTn),[4,5] which is the focus of the present work, was originally designed to solve the challenging problem of exploring PESs around initial ground states by means of a semi-automatic and unbiased approach. ARTn has been successfully applied on a wide range of systems such as impurities,[6] molecules,[7] amorphous and glassy materials,[8–11] and protein aggregation and flexibility.[12,13]

Coupling ARTn with a topological analysis tool (NAUTY[14]), kinetic ART (k-ART) offers the most efficient off-lattice kinetic Monte Carlo method for following the long-time kinetics of complex systems such as defects in crystal,[15–18] disordered materials[19] and cascades.[20,21] Despite its success when coupled to force fields, its use within first-principle energy and force engines has remained marginal.[7,22,23]

In order to maintain a reasonable balance between human workload and computational cost, the *ab initio* community has mostly adopted string methods such as the Nudged Elastic Band (NEB[24]). NEB method and its variants are implemented in most of the user-community softwares, such as ABINIT,[25] QUANTUM ESPRESSO,[26] VASP[27] or BIGDFT.[28] In spite of their popularity, string approaches present a certain number of drawbacks that significantly reduce their efficiency or even preclude their use in systems with complex PES, in particular when massive minimum-energy pathway (MEP) searches are needed.

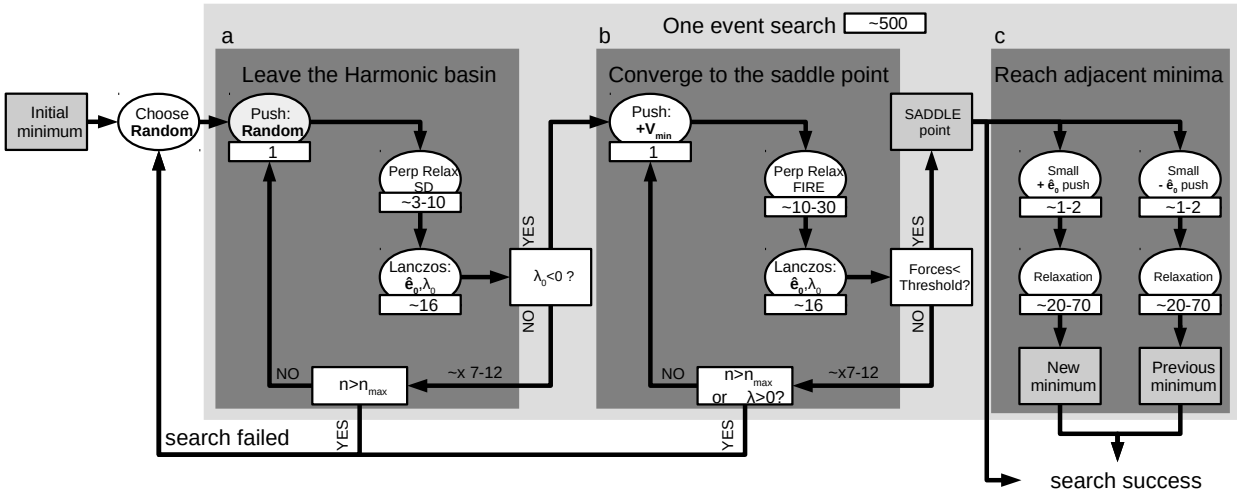


Figure 1: ARTn workflow. The three main stages are pictured in the dark gray boxes. Stage (a): Leave the harmonic basin; stage (b): converge to the saddle point; and stage (c): reach the adjacent minima. In stage (c), each relaxation from the saddle point is optional. They enable respectively to obtain the final structure and to check that the starting configuration is directly linked to the saddle point. Average iteration numbers and number of force calls per step are also indicated.

To address these drawbacks, we have revisited the implementation of ARTn, introducing

a number of algorithmic and implementation improvements that make it more suitable for applications in the *ab initio* context. In particular, the success rate in finding MEPs has been significantly enhanced by using mixing functions, reaching 87 % for the system studied. Two variants of the ARTn algorithm are also presented: the refining ART (r-ART), that converges directly to a saddle point from a guessed configuration, and the directed ART (d-ART), that is able to automatically reconstruct a complex path connecting several minima and saddle points. Benchmarks with NEB are provided in the following sections using an improved and tight coupling with PWscf, the Density Functional Theory (DFT)-based energy and force engine of Quantum ESPRESSO v6.5.[26] This coupling benefits from an improved code integration that exploits the electronic density of previous energy and force evaluations. These improvements and variants greatly enhance the applicability of ARTn for massive and intensive MEP explorations at *ab initio* accuracy, and offer an efficient and versatile alternative to standard string approaches.

Increasing success rate during exploration

Previous ARTn workflow

The standard workflow of ARTn is divided into three main stages described in Figure 1:

Leave the harmonic basin The initial configuration is iteratively pushed in a random direction until being out of its harmonic basin. This deformation is generally local, centered on the atoms of interest and their nearest neighbors. The user can set these atoms and their radius around which the initial deformation is imposed. In complex materials, searches can be launched around each atom of the system without topological preferences. If constrains are not imposed, all atoms are submitted to the deformation. After each increment along this random direction, the whole system is relaxed in the hyperplane perpendicular to this pushing direction. After this relaxation, the lowest eigenvalue (λ_0) of the Hessian matrix is

calculated using the Lanczos algorithm.[29] The iterations continue until this lowest eigenvalue falls below a preset negative threshold, indicating that the system has left the harmonic basin and moved beyond the inflection line. Note that since the system starts from a local minimum, the first steps are inevitably still in the harmonic basin so that evaluations of λ_0 can start only after a few iterations, to reduce computational efforts.

Converge to the saddle point Once above the inflection line, the activation part of ARTn starts. The configuration is pushed from the inflection to the saddle point, following the direction defined by the eigenvector $\hat{e}_0$ corresponding to λ_0. The magnitude of the applied displacement is decreased as the system approaches the saddle point.[30,31] Each push is followed by an atomic relaxation in the hyperplane perpendicular to the pushing direction, as in (a). These actions are performed iteratively until the total force falls below a given threshold defining the saddle point.

Reach the adjacent minima From the saddle point, the configuration is further pushed along the eigenvector both away from the initial minimum ($+\hat{e}_0$) and towards it ($-\hat{e}_0$). The two configurations obtained from the two pushes are then fully relaxed and compared with the initial minimum. This later step ensures that the saddle point is directly connected to the initial minimum without an intermediate minimum. The triplet of states: initial minimum – saddle point – final minimum corresponds to a fully connected event.

A video showing the exploration of a 2D PES is given as an illustration in supplementary materials.

Mixing function

Extensive saddle point searches on ARTn's original implementation show that, for most systems, up to 75 % of the trial pushes for reaching a saddle point lead to a perpendicular relaxation into the initial configuration, *i.e.*, the system falls back into the initial harmonic basin. In many cases, this failure occurs in the first few steps after the system has crossed

the inflection line. This suggests that the transition from the last random push in stage (a) to the first push along the eigenvector in stage (b) is too sharp. A significantly higher success rate can be achieved by softening the transition between these two vectors (see Figure 2).

To do so, we added an intermediate stage between stage (a) and (b) that starts after the first occurrence of a negative eigenvalue during stage (a). The goal of this new stage is to smoothly rotate the push vector from the initial random direction $\hat{r}_{rand}$, to the one of the Hessian eigenvector $\hat{e}_0$ corresponding to λ_0, according to:

$$\hat{r}_e = (1 - \eta)\hat{r}_{rand} + \eta\hat{e}_0 \tag{1}$$

where η grows linearly from 0 to 1 in a given number of steps, and its slope is a user-defined parameter. In the 4I example given below, only three steps are enough to significantly increase the success rate: $\eta = i/3$ with i going from 1 to 3.

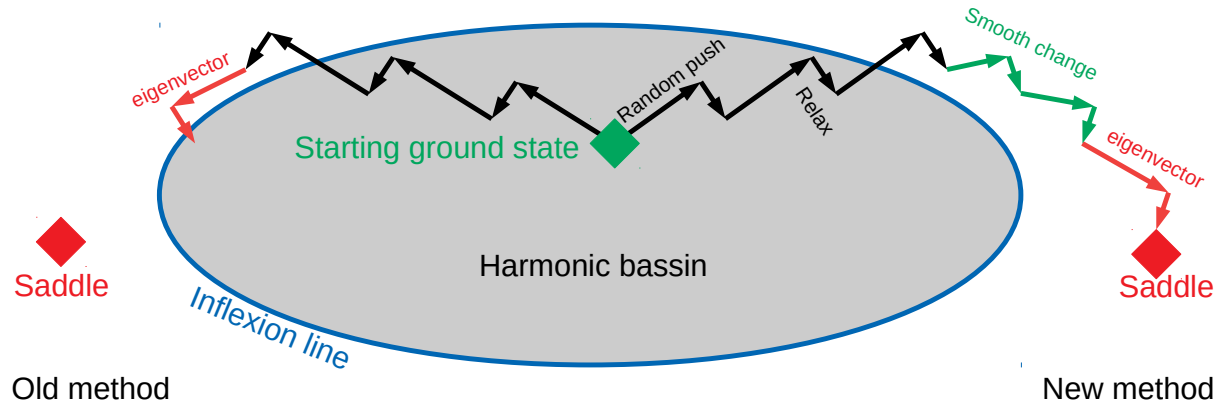


Figure 2: Leaving the harmonic basin with and without the smooth switching function. An abrupt change between pushing along the random direction and pushing along the eigenvector direction, with their respective perpendicular relaxations, might bring the algorithm back to the starting basin. A smooth switch helps to bring the system further away from the inflection line, which increases the success rate in finding basins.

Table 1: Intermediate path used by ARTn-DFT to find the saddle point from the ground state 4I-Si structure given in Figure 3 toward the metastable structure indicated by a square. ΔE (resp. Δr): difference of energy (resp. atomic positions) between obtained structure and reference structure. Relax: number of perpendicular relaxations done after each push. Forces: total sum of the forces on each atom (tot.), its components projected ($\parallel$) and orthogonal ($\perp$) to the pushing direction, and its total number (#) of evaluations since the beginning of the search. λ_0: lowest eigenvalue calculated by Lanczos. *nc* stands for not calculated.

ΔE	Δr	Relax	Forces (eV/Å)				λ_0
eV	Å	#	tot.	$\parallel$	$\perp$	#	eV/Å^2
harmonic basin: push in random direction: $\hat{r}_{rand}$							
2.39	1.46	6	5.46	-4.90	2.39	9	*nc*
2.67	1.67	6	4.93	-4.86	0.83	18	*nc*
3.11	1.86	6	5.15	-5.09	0.75	27	*nc*
3.59	2.04	6	5.37	-5.35	0.48	68	0.35
4.11	2.21	6	5.61	-5.61	0.29	94	0.29
4.66	2.38	6	5.85	-5.85	0.26	118	0.22
5.24	2.55	6	6.07	-6.06	0.21	144	0.12
5.85	2.71	6	6.32	-6.30	0.52	169	-0.07
reorientation towards the activation direction: $\hat{r}_e$							
4.39	2.58	6	6.27	-3.37	5.29	194	-1.43
2.87	2.43	6	3.91	-1.18	3.73	219	-1.63
2.14	2.31	6	2.54	-0.35	2.52	244	-1.07
inflection passed: push along eigenvector: $\hat{e}_0$							
2.14	2.46	5	1.84	-0.26	1.82	266	-0.73
2.09	2.58	10	0.89	-0.14	0.88	293	-1.11
1.98	2.63	15	0.63	-0.01	0.63	325	-1.26
1.83	2.63	20	0.27	0.00	0.27	362	-1.15
1.79	2.64	25	0.04	0.00	0.04	404	-0.99

Quadri-interstitial

To illustrate the impact of these improvements on ARTn, a portion of the self quadri-interstitial PES in silicon is explored. This complex defect is known to possess a very stable ground state and to exhibit a large variety of metastable minima.[32–34] Until now, because of its complexity, the PES had only been extensively explored with empirical force fields.[16] We use a 216 silicon model and add four silicon interstitial atoms, addressed as 4I-Si. A first-principle ARTn PES exploration from the 4I-Si ground state is launched using five parallel searches (for a total of $5 \times 36 = 180$ cores) and set to stop once twenty pathways are found. The initial deformation (stage (a)) is applied on the four interstitial atoms and their first neighbors, and a 0.05 eV/Å threshold for the total force is used to define convergence at the saddle point.

Table 1 summarizes the evolution of the energy difference, forces and Hessian lowest eigenvalue for one typical event search (all explorations exhibit a qualitatively similar behavior). In the initial stage (harmonic basin sub-block in Table 1), the system is still in the harmonic well. By definition, the lowest eigenvalue of the Hessian matrix (λ_0) is then always positive. As the system is pushed away from its local initial minimum, the energy difference (ΔE in Table 1), the displacement measured from the minimum (Δr in Table 1) and the total forces increase rapidly. In this stage, the main part of the forces is parallel to the pushing direction. The system is, however, still in the initial basin as these increases are distributed over many atoms. As the system reaches one of the edges of the harmonic basin (first negative Hessian eigenvalue, λ_0 in Table 1), the intermediate stage of the smooth switching starts and the pushing direction is slowly becoming aligned with the eigenvector direction, using the switching function described previously. Here, this reorientation is performed over three steps. After this intermediate stage, the system has left the basin and is above the inflection line. It is now pushed along $\hat{e}_0$ and the forces parallel to this vector decrease at each push as the system moves towards the saddle point. Because the system has to recover for the initial deformation, the main part of the forces is orthogonal to $\hat{e}_0$ as

the system rapidly reaches the ridge between the two minima. After five activation steps and 404 force evaluations, the total force falls below the threshold.

Table 2: Comparison of the computational costs and efficiency between the previous (Prev. ARTn) and current ARTn (New ARTn) implementations for a blind exploration of the 4I-Si PES. Both algorithms are stopped after the successful identification of 20 paths (20 saddle points and their respective final metastable minimum).

	Forces		Event	
Method	Tot call	call / event	failed	success rate
Prev. ARTn	17 887	894	41	33%
New ARTn	7 956	398	3	87%

For the full simulation, twenty-three search attempts were required to find the twenty

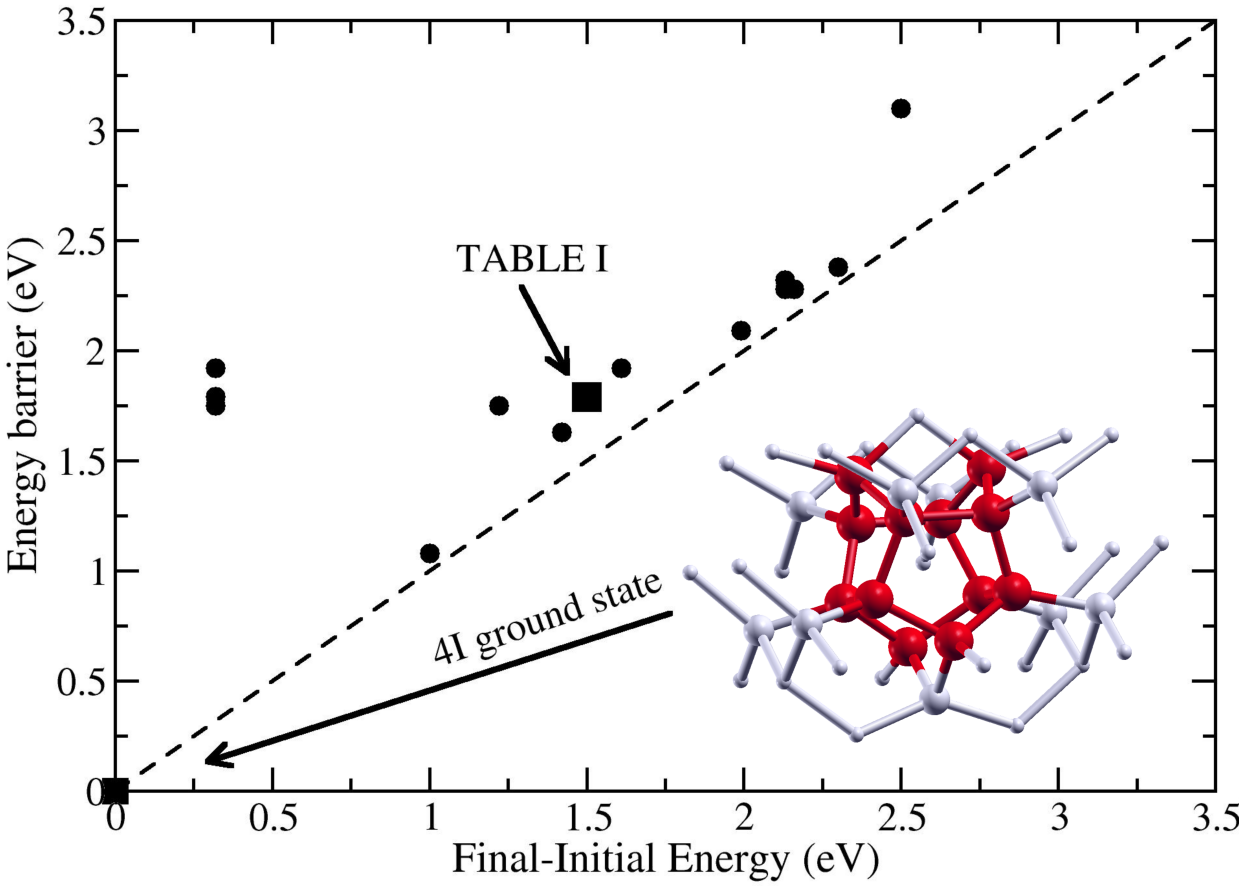


Figure 3: Saddle point energy and relative final energy of the metastable structures obtained from an exploration of the PES around the ground state of the 4I-Si. The 4I-Si ground state energy is taken as the reference. The ground state atomic structure is represented by silicon atoms out of their crystalline site (red balls) and their on-site neighbors (white balls).

paths, which represents a rate of 87 % of success for the improved ARTn. Success rate comparisons with the previous ARTn version are provided in Table 2. Figure 3 summarizes the barrier heights and final relative energies for the twenty paths connecting the ground state configuration to twenty new excited states (metastable minima). As can be seen, ARTn with the added smooth turn reduces by more than 55 % the computational costs of event searches while preserving the overall quality of exploration.

r-ART and d-ART

ARTn standard implementation is designed to explore the energy landscape starting from a single known energy minimum (see Figures **??** and **??**). However, the activation part alone, *i.e.* stage (b) in Figure 1, can be exploited separately to quickly converge to a saddle point starting from a guessed configuration outside the harmonic basin.

Refining ART (r-ART) implements this new feature in a revised algorithm that enables to rapidly find a saddle point from a given initial guessed value above the inflection. Only this guessed value is needed. It can be either provided directly by the user (Figure **??**), or generated by r-ART as the middle configuration (or any other fraction) between two input minima (Figure **??**). The initial eigendirection can also be either provided by the user, for example in highly symmetric cases, or directly calculated by r-ART using its Lanczos implementation. Next, the system follows the stage b) of the algorithm to converge to the saddle point, and stage c) to ensure the pathway connectivity.

As r-ART, *directed ART* (d-ART) is also built on the activation part of ARTn, but it searches for a fully connected path between a starting and arrival minima that can be separated by more than one saddle point. After finding the minima around the first saddle point (Figure **??**), d-ART will compare these minima with the input ones. If at least one of the new minima is different from the starting and arrival configurations, d-ART will continue its search until a fully connected path between the starting and arrival minima is found, as

schematized in Figure **??**. It enables, therefore, for an automatic search of complex pathways and saddle points, with several intermediate metastable minima and saddle points.

Performances against string approaches

String approaches, like NEB, rely on the use of a chain of configurations, often addressed as intermediate images, that connects two minima with an equal inter images distance controlled by a spring force. Therefore, a reasonable guess for the two minima needs to be provided as pre-requirement. In addition, an initial guess for the intermediate configurations between the two minima also needs to be provided. For this guess, most of the user-community codes implement a simple linear interpolation (Figure **??**). The MEP is found by minimizing the whole chain of configurations perpendicularly to the path tangent (Figure **??**), *i.e.* the atomic positions of each of the intermediate images need to be relaxed. With such approach, the number of relaxing steps is directly related to the quality of the initial path guess, that can be optimized, for example, by using more reasonable interatomic distances than a simple interpolation, with methods such as the Image Dependent Pair Potential algorithm (IDPP[35]). However, even with optimization methods, the initial MEP guess induces a bias in the saddle point search, especially in complex energy landscapes, by sometimes missing the actual MEP of the atomic system as illustrated in Figure **??**. In some cases, this problem can be addressed by the addition of an intermediate skillfully chosen image (Figure **??**).

The NEB stopping condition relies on the convergence threshold of the total forces perpendicular to the chain. As such, the accuracy on the saddle point depends on the resolution of the chain tangent vector, which can only be improved by increasing the number of intermediate images to reduce the intermediate images distance. The AutoNEB[36] enables to resolve the tangent more efficiently by automatically adding images only around the region of interest. Finally, only the image that converges to the saddle brings relevant information. The Climbing Image (CI-NEB[37]) enables to reduce the number of irrelevant images, by allowing the highest energy image to climb along the path without the inter images spring

force.

Despite all these improvements, string methods remain computationally heavy and lack in accuracy to converge to the saddle point. On the contrary, ARTn is faster as it does not need any images, and its convergence is improved at each step thanks to the iterative reevaluation of the eigenvector (Figures **??** and **??**). In the following, the performances of r-ART and d-ART in terms of number of force calls and precision are compared with the ones of string approaches. The number of force calls and precision only depends on ART algorithm (and its variants) and not on the specific software/approach that computes the forces, *i.e.* the performances in terms of force calls do not depends on the underlying energy and force engine.

r-ART applied to highly symmetric pathways: comparison with R-NEB and RMI-NEB

For simple symmetric pathways, two variations of NEB have been recently developed:[38] reflective NEB (R-NEB) and reflective-middle-image NEB (RMI-NEB). To reduce the number of force calls, they exploit the mirror symmetry that might be present in highly symmetric reaction pathways. The performances of r-ART are first compared to these two NEB variations on the migration of a mono-vacancy in silicon (supercell of 63 atoms), of a mono-vacancy in graphene, of a Li interstitial in graphene, and of a Li vacancy in Li_2O_2 which

Table 3: Comparison of the performances or r-ART (ART) against the ones of standard CI-NEB (std: 7 images), R-NEB (ref: 4 images + 3 symmetric), and RMI-NEB (one: 1 image) as provided in Table 2 of Ref. 38. Thanks to the symmetry, the tangent to the path (eigenvector) is known. The "precision" ΔE_B is the saddle point energy difference with respect to the one calculated with CI-NEB (E_B^{std}).

	Precision (meV)				Force calls			
	E_B^{std}	ΔE_B^{ref}	ΔE_B^{one}	ΔE_B^{ART}	N_{force}^{std}	N_{force}^{ref}	N_{force}^{one}	N_{force}^{ART}
Si	1057	0	0	0	343	196	43	51
C_6/V_c	804	0	61	0	434	248	44	50
C_6/ across	356	0	2	0	371	216	38	49
LiO_2	1705	0	0	0	215	129	43	51

are the benchmark examples of Ref. 38. For a relevant comparison, the same computational parameters have been used and the initial r-ART configuration is chosen as the middle point of the initial chain guess.

On these benchmark examples (see Table 3 for a summary), we find that r-ARTn is: *(i)* as accurate as R-NEB but at a lower computational cost, requiring two to three times fewer force calls, *(ii)* three to five times faster than CI-NEB, *(iii)* more accurate than RMI-NEB at a similar computational cost. It is important to recall that for RMI-NEB, R-NEB and CI-NEB, the number of force calls is proportional to the number of intermediate configurations that are included in the chain. In Ref. 38, as reported in Table 3, only seven intermediate images have been included due to the high symmetry of the path. As such, if a tighter criterion on the saddle point convergence is required, the NEB performances will drop proportionally to the number of intermediate images, while this is not the case for r-ART, that only depends on the force convergence threshold at the saddle point. These behaviors are illustrated in the next, more generic, test examples.

r-ART applied to silicon tetra-vacancy with non-trivial path: comparison with CI-NEB

In this section, we consider a more generic non-symmetric mechanism: the complex diffusion of the tetra-vacancy in silicon. Here, the transition state links the well known *chain* and *hexagonal non-symmetric* (hns) configurations[39] presented in Figure **??**, top panels.

The model system contains a total of 212 atoms and is addressed as 4V-Si. For r-ART, the total force threshold of $2\times\ 10^{-5}$ Ry/au at the saddle point is chosen to define the convergence. For the CI-NEB implemented in QUANTUM ESPRESSO v6.5,[26] the same threshold is used but it is applied on the forces orthogonal to the path, as no direct criterion can control the convergence to the saddle point along the path. The initial chain guess is a linear interpolation between the two minima. The r-ART starting configuration lies halfway between the two minima. r-ART calculations are compared with CI-NEB performed with

three different numbers of intermediate images.

As can be seen in Table 4, even after more than thousands force calls, CI-NEB is between one and two orders of magnitude less accurate than r-ART in resolving the position of the saddle point, depending on the number of intermediate images, with r-ART requiring four to seven times *fewer* force evaluations. Because of the large distance between the two minima, about 4.2 Å, and of the sinuosity of the path, a very large number of intermediate images should be used to achieve a saddle point force convergence similar to the one chosen for r-ART calculations, which results in a significant increase in computational costs.

The evolution of both CI-NEB (with 19 intermediate images) and r-ART saddle point search is shown in Figure **??**, where the energy (with respect to the total energy of one of the minima) is plotted as a function of the displacement vector projected on the total displacement. Lanczos, push and orthogonal relaxation steps are also highlighted in Figure **??**.

Table 4: In the case of 4V-Si, comparison of the convergence and of the cost to reach the saddle point between the CI-NEB implemented in QUANTUM ESPRESSO v6.5 and r-ART-DFT. The number of force calculations corresponds to the number of times an scf loop has been run by the DFT software. CPU time is defined as the product of the total search time by the number of CPUs. Total force is the rms sum of all the atomic forces at the saddle point. Maximum force is the maximum force on one atom among all at the saddle point. In r-ARTn*, 10 Ry of plane waves cutoff have been used until forces reach 10^{-2} Ry/au and 20 Ry afterward.

	force calc. (number)	CPU time (hcpu)	Tot. force (Ry/au)	Max. force (Ry/au)
CI-NEB 5 im	1127	1550	$1.1\ 10^{-3}$	$1.9\ 10^{-4}$
CI-NEB 13 im	1976	2565	$5.2\ 10^{-4}$	$1.0\ 10^{-4}$
CI-NEB 19 im	2071	2754	$1.9\ 10^{-4}$	$4.3\ 10^{-5}$
r-ARTn	301	309	$1.1\ 10^{-5}$	$2.1\ 10^{-6}$
r-ARTn*	353	121	$1.2\ 10^{-5}$	$2.3\ 10^{-6}$

Table 5: Same caption as in Table 4 but for the diffusion of the mono-vacancy that has one intermediate minimum and two saddle points. The number of forces includes the convergence to the saddle point and the two relaxations. The CI option is active for saddle 2.

			Saddle 1		Inter. min.		Saddle 2 (CI)	
	force calc. (number)	CPU time (hcpu)	Tot. force (Ry/au)	Max. force (Ry/au)	Tot. force (Ry/au)	Max. force (Ry/au)	Tot. force (Ry/au)	Max. force (Ry/au)
CI-NEB 5 im	273	371	$6.0\ 10^{-3}$	$3.0\ 10^{-3}$	$2.4\ 10^{-3}$	$1.8\ 10^{-4}$	$7.2\ 10^{-4}$	$3.0\ 10^{-4}$
CI-NEB 13 im	1080	1488	$6.0\ 10^{-4}$	$3.1\ 10^{-4}$	$6.1\ 10^{-4}$	$2.2\ 10^{-4}$	$4.7\ 10^{-4}$	$1.9\ 10^{-4}$
CI-NEB 19 im	1539	1998	$9.3\ 10^{-4}$	$3.6\ 10^{-4}$	$2.1\ 10^{-3}$	$5.1\ 10^{-4}$	$1.1\ 10^{-4}$	$3.4\ 10^{-5}$
d-ARTn 1rst part	542	524	$1.0\ 10^{-5}$	$1.4\ 10^{-6}$	$1.0\ 10^{-5}$	$2.1\ 10^{-6}$	-	-
d-ARTn 2nd part	284	291	-	-	$1.0\ 10^{-5}$	$2.1\ 10^{-6}$	$1.0\ 10^{-5}$	$2.3\ 10^{-6}$
d-ARTn (total)	826	815	$1.0\ 10^{-5}$	$1.4\ 10^{-6}$	$1.0\ 10^{-5}$	$2.1\ 10^{-6}$	$1.0\ 10^{-5}$	$2.3\ 10^{-6}$

d-ART used to reconstruct the multistep silicon mono-vacancy diffusion: comparison with CI-NEB

The silicon mono-vacancy (V-Si) migration with a model composed by 63 atoms was already used as benchmark example in Table 3 for comparison with Ref. 38. However, the converged pathway is more complex and can only be resolved with a model system with more than 200 atoms.[40] The MEP between the two ground states (D_{2d} symmetry) involves an intermediate metastable configuration with the C_s symmetry and two symmetric saddle points (top of Figure **??**). To demonstrate the application of d-ART, we consider, therefore, a mono-vacancy migration in a 215 atoms model, addressed as V-Si.

The d-ART performance is compared with that of CI-NEB as implemented in QUANTUM ESPRESSO v6.5[26] using 5, 13 and 19 intermediate images. Similarly to the previous example, even with 19 intermediate images, the total force in CI-NEB remains up to one order of magnitude higher than the one reached with d-ART at the saddle point (1.1×10^{-4} Ry/au for 19 images CI-NEB against 1.0×10^{-5} Ry/au for d-ART). Results on the saddle point 1 that do not have the climbing image and on the intermediate minimum are even worse: the total force remains two orders of magnitude higher than that obtained with d-ART. d-ART can find the fully connected path, including the two saddle points and the relaxation to the initial, final and intermediate minima after only 826 force evaluations. This represents 50% fewer force evaluations than CI-NEB with 19 intermediate images (1539), and four times less CPU time, due to the reduced number of DFT self-consistent cycles. Both saddle points are

resolved within 10^{-5} Ry/au total force.

Conclusion

The significant improvements in the implementation of ARTn has led to a considerable reduction in the computational costs needed to map the energy landscape of complex systems. This makes possible the use of ARTn with DFT calculations for a wide range of problems for which it is impossible to guess pathways from symmetry or physical arguments. For the examples presented here, with ARTn coupled to Quantum Espresso, the automatic searches of MEP show success rates at up to 87% in finding saddle points, resulting in a overall acceleration of 225 per cent.

In addition to these improvements, two important extensions of ARTn are proposed to address the needs of the DFT community to efficiently identify MEPs: refining ART (r-ART) and directed ART (d-ART). r-ART is designed to find a single transition state from a given or interpolated guess while d-ART can construct automatically a chain of states, including transition and intermediate states, between two given configurations.

The various comparisons made between these methods and NEB-based string approaches show that both r-ART and d-ART converge to saddle points with much higher precision than string methods for significantly less computational effort. They thus represent a valued alternative to the standard string approaches.

Together, ARTn, r-ART and d-ART provide a powerful tool that will greatly facilitate the characterization of energy landscapes in chemical reactions and complex materials.

Acknowledgement

The simulations were performed using HPC resources from CALMIP (Grant P1555). NM's work is supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

AJ, NS, MG, LMS, NR, GL, SDG, AH, NM are active members of the Multiscale And Multi-Model Approach for MaterialS In Applied Science consortium (MAMMASMIAS consortium), and acknowledge the efforts of the consortium in fostering scientific collaboration.

Supporting Information Available

A video showing the ARTn algorithm exploration of a two dimensions PES is available in supplementary material. This information is available free of charge via the Internet at http://pubs.acs.org

Software and data availability

ARTn-DFT is available freely upon request to any of the authors, it includes the standard exploration method, r-ART, d-ART, a complete documentation and many examples. The present version is only compatible with QE-6.5, also available freely at https://www.quantum-espresso.org. We are currently developing a plugin version for an easier use, hands on, maintenance and portability.

Computational details

DFT - ART coupling

Ab-initio approaches are computationally expensive because of the self-consistent (scf) determination of the ground-state electronic density, and have represented one of the major bottleneck for a broad adoption of ARTn by the DFT community. The number of scf cycles can, however, be significantly reduced if the starting Kohn-Sham potential is computed from the density of a previous energy and force calculation. This approach is already in use, in most of the user community DFT codes like ABINIT,[25] VASP[27] and Quantum ESPRESSO,[26]

for accelerating NEB calculations, atomic relaxations and molecular dynamics simulations. In the context of DFT codes based on pseudo-potentials (only valence electrons are treated explicitly), the corresponding core potential is updated on-the-fly at each step.

In the current ARTn-DFT, the implementation of the aforementioned computational optimisation in ARTn increases the speed of MEP searches by a factor of three (in all the test cases). Such a high efficiency is allowed by the iterative finite difference evaluation (to construct the tri-diagonal Lanczos matrix) of the lowest Hessian-matrix eigenvalue and eigenvector, which involves total displacements of 0.01 Å or less around a central configuration. In this case, the electronic density after an scf cycle is very similar to the density in the previous step.

Massive and automatic saddle point searches on a given system have also been enabled.

In addition, the coupling allows to fix atom positions simply by multiplying their force by a 0 or a 1 (T or F) that are commonly given in the DFT software inputs after the atomic positions.

This new version of ARTn also implements the use of two different sets of DFT parameters (cutoff, k-points,...) in the same run. This is particularly useful, for instance, to identify first rough saddle points and then resolve them accurately.

ARTn parameters

Stage (a): Leaving the harmonic basin. The size of the random displacement push is chosen to be 2 Å , distributed on all the atoms that are not a crystalline site for the first push, and reduced to 0.2 Å for the following pushes until reaching the inflection line (stage b)). 6 steps of perpendicular relaxations, after each push, have been imposed (see Table.1). The first evaluation of the lowest Hessian-matrix eigenvalue and corresponding eigenvector is chosen to occur after the third pushing step.

Mixing stage Three mixing steps are used to smooth the transition between stages (a) and (b).

Stage (b): Converging to the saddle point. The magnitude of the displacement push is proportional to the force when following the eigenvector as:[30,31]

$$dr = \min\left(\text{size}_{\max}, \frac{|f_{par}|}{\max(|\lambda_0|, 0.5)}\right), \tag{2}$$

where f_{par} is the force parallel to the eigenvector (eV/Å), λ_0 the negative eigenvalue (eV/Å^2) and $size_{max}$ the upper bound of the allowed displacement, set to 0.2 Å.

For the Quadri-interstitial test case, the number of perpendicular relaxations steps is increased from 5 to 25. For the other test cases, the perpendicular relaxation stops when the perpendicular forces are lower than the parallel ones, or when reaching the selected saddle-point total force convergence threshold.

Lanczos. For all stages, 16 iterations are used to converge the lowest eigenvalue and corresponding eigenvector.

DFT parameters

Unless stated differently, all DFT calculations use the Local Density Approximation (LDA),[41] the Brillouin zone is sampled at gamma only, the size of the plane wave basis is set to 20 Ry and the convergence threshold of the self-consistent energy to 10^{-10} Ry.

References

(1) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

(2) Bennett, C. H. Molecular Dynamics and Transition State Theory: The Simulation of Infrequent Events. *ACS Symp. Series.* **1977**, *46*, 63–97.

(3) Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **1978**, *68*, 2959.

(4) Barkema, G. T.; Mousseau, N. Event-Based Relaxation of Continuous Disordered Systems. *Phys. Rev. Let.* **1996**, *77*, 4358–4361.

(5) Malek, R.; Mousseau, N. Dynamics of Lennard-Jones clusters: A characterization of the activation-relaxation technique. *Phys. Rev. E* **2000**, *62*, 7723–7728.

(6) Marinica, M.-C.; Willaime, F.; Mousseau, N. Energy landscape of small clusters of self-interstitial dumbbells in iron. *Phys. Rev. B* **2011**, *83*, 094119.

(7) Machado-Charry, E.; Béland, L. K.; Caliste, D.; Genovese, L.; Deutsch, T.; Mousseau, N.; Pochet, P. Optimized energy landscape exploration using the ab initia based activation-relaxation technique. *J. Chem. Phys.* **2011**, *135*, 034102.

(8) Barkema, G. T.; Mousseau, N. Identification of Relaxation and Diffusion Mechanisms in Amorphous Silicon. *Phys. Rev. Let.* **1998**, *81*, 1865–1868.

(9) Kallel, H.; Mousseau, N.; Schiettekatte, F. Evolution of the Potential-Energy Surface of Amorphous Silicon. *Phys. Rev. Let.* **2010**, *105*, 045503.

(10) Rodney, D.; Schuh, C. Distribution of Thermally Activated Plastic Events in a Flowing Glass. *Phys. Rev. Let.* **2009**, *102*, 235503.

(11) Fan, Y.; Iwashita, T.; Egami, T. Energy landscape-driven non-equilibrium evolution of inherent structure in disordered material. *Nat. Comm.* **2017**, *8*, 15417.

(12) Wei, G.; Mousseau, N.; Derreumaux, P. Exploring the early steps of aggregation of amyloid-forming peptide KFFE. *J. Phys.: Cond. Matter.* **2004**, *16*, S5047.

(13) St-Pierre, J.-F.; Mousseau, N. Large loop conformation sampling using the activation relaxation technique, ART-nouveau method. *Proteins: Structure, Function, and Bioinformatics* **2012**, *80*, 1883–1894.

(14) McKay, B. D.; Piperno, A. Practical graph Isomorphism, II. *J. Symb. Comp.* **2014**, *60*, 94–112.

(15) Brommer, P.; Bland, L. K.; Joly, J.-F.; Mousseau, N. Understanding long-time vacancy aggregation in iron: A kinetic activation-relaxation technique study. *Phys. Rev. B* **2014**, *90*, 134109.

(16) Trochet, M.; Beland, L.; Joy, J.; Brommer, P.; Mousseau, N. Diffusion of point defects in crystalline silicon using the kinetic activation-relaxation technique method. *Phys. Rev. B* **2015**, *91*, 224106.

(17) Restrepo, O. A.; Becquart, C. S.; El-Mellouhi, F.; Bouhali, O.; Mousseau, N. Diffusion mechanisms of C in 100, 110 and 111 Fe surfaces studied using kinetic activation-relaxation technique. *Acta Mater.* **2017**, *136*, 303–314.

(18) Mahmoud, S.; Mousseau, N. Long-time point defect diffusion in ordered nickel-based binary alloys: How small kinetic differences can lead to completely long-time structural evolution. *Materialia* **2018**, *4*, 575–584.

(19) Joly, J.-F.; Bland, L. K.; Brommer, P.; Mousseau, N. Contribution of vacancies to relaxation in amorphous materials: A kinetic activation-relaxation technique study. *Phys. Rev. B* **2013**, *87*, 144204.

(20) Béland, L. K.; Anahory, Y.; Smeets, D.; Guihard, M.; Brommer, P.; Joly, J.-F.; Pothier, J.-C.; Lewis, L. J.; Mousseau, N.; Schiettekatte, F. Replenish and Relax: Explaining Logarithmic Annealing in Ion-Implanted c-Si. *Phys. Rev. Let.* **2013**, *111*, 105502.

(21) Jay, A.; Raine, M.; Richard, N.; Mousseau, N.; Goiffon, V.; Hmeryck, A.; Magnan, P. Simulation of Single Particle Displacement Damage in SiliconPart II: Generation and Long-Time Relaxation of Damage Structure. *IEEE Trans. Nucl. Sc.* **2017**, *64*, 141–148.

(22) Béland, L.; Mousseau, N. Long-time relaxation of ion-bombarded silicon studied with the kinetic activation-relaxation technique: Microscopic description of slow aging in a disordered system. *Phys. Rev. B* **2013**, *88*, 214201.

(23) Salles, N.; Richard, N.; Mousseau, N.; Hemeryck, A. Strain-driven diffusion process during silicon oxidation investigated by coupling density functional theory and activation relaxation technique. *J. Chem. Phys.* **2017**, *147*, 054701.

(24) Mills, G.; Jonsson, H.; Schenter, G. K. Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surf. Sci.* **1995**, *324*, 305–337.

(25) Gonze, X. The Abinit project: Impact, environment and recent developments. *Comp. Phys. Comm.* **2020**, *248*, 107042.

(26) Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Cond. Matter.* **2009**, *21*, 395502.

(27) Kresse, G.; Furthmüler, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169.

(28) Genovese, L.; Videau, B.; Ospici, M.; Deutsch, T.; Goedecker, S.; Méhaut, J. F. Daubechies wavelets for high performance electronic structure calculations: The BigDFT projectOndelettes de Daubechies pour le calculs de structure lecronique haute performance : Le code BigDFT. *Comp. Rendus Méca.* **2011**, *339*, 149–164.

(29) Lanczos, C. An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators. *J. Res. Nat. Bur. St.* **1950**, *45*, 255–282.

(30) Mousseau, N.; Béland, L. K.; Brommer, P.; Joly, J. F.; El-Mellouhi, F.; Machado-Charry, E.; Marinica, M. C.; Pochet, P. The Activation-Relaxation Technique: ART Nouveau and Kinetic ART. *J. At. Mol. Opt. Phys.* **2012**, *2012*, 925278.

(31) Cancs, E.; Legoll, F.; Marinica, M.-C.; Minoukadeh, K.; Willaime, F. Some improvements of the activation-relaxation technique method for finding transition pathways on potential energy surfaces. *J. Chem. Phys.* **2009**, *130*, 114711.

(32) Arai, N.; Takeda, S.; Kohyama, M. Self-Interstitial Clustering in Crystalline Silicon. *Phys. Rev. Let.* **1997**, *78*, 4265.

(33) Coomer, B. J.; Goss, J. P.; Jones, R.; Oberg, S.; Briddon, P. R. Interstitial aggregates and a new model for the I1/W optical centre in silicon. *Physica B* **1999**, *273-274*, 505–508.

(34) Du, Y.; Lenosky, T.; Hennig, R.; Goedecker, S.; Wilkins, J. Energy landscape of silicon tetra-interstitials using an optimized classical potential. *Phys. Stat. Sol. B* **2011**, *248*, 2050–2055.

(35) Smidstrup, S.; Pedersen, A.; Stokbro, K.; Jònsson, H. Improved initial guess for minimum energy path calculations. *J. Chem. Phys.* **2014**, *140*, 214106.

(36) Kolsberg, E. L.; Groves, M. N.; Hammer, B. An automated nudged elastic band method. *J. Chem. Phys.* **2016**, *145*, 094107.

(37) Henkelman, G.; Uberuaga, B. P.; Jonsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901.

(38) Mathiesen, N. R.; Jónsson, H.; Vegge, T.; Lastra, J. M. G. R-NEB: Accelerated Nudged Elastic Band Calculations by Use of Reflection Symmetry. *J. Chem. Theory Comput.* **2019**, *15*, 32155–3222.

(39) Makhov, D. V.; Lewis, L. Stable Fourfold Configurations for Small Vacancy Clusters in Silicon from ab initio Calculations. *Phys. Rev. Let.* **2004**, *92*, 255504.

(40) El-Mellouhi, F.; Mousseau, N.; Ordejn, P. Sampling the diffusion paths of a neutral vacancy in silicon with quantum mechanical calculations. *Phys. Rev. B* **2004**, *70*, 205205.

(41) Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **1981**, *23*, 5048.

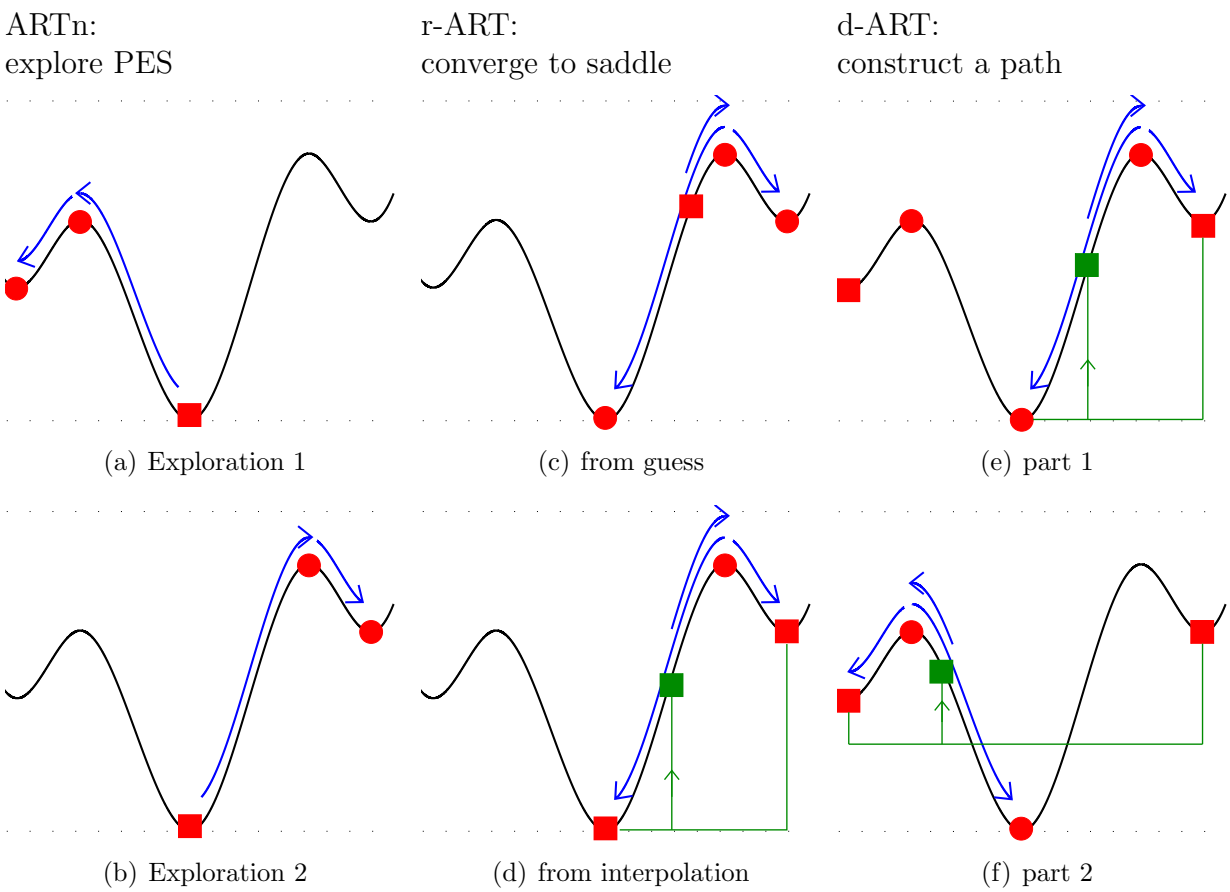
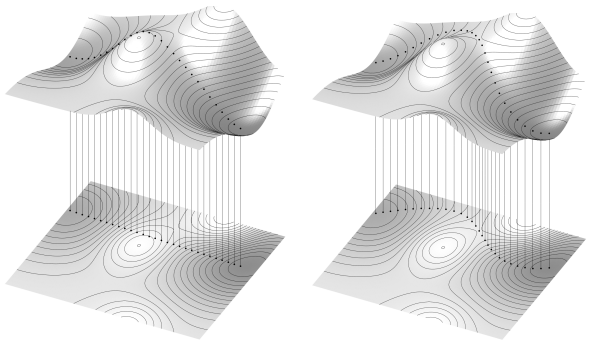
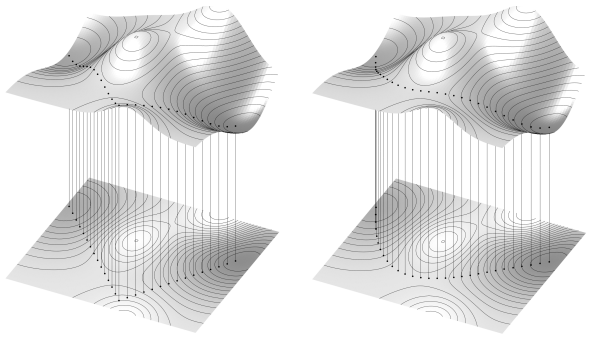


Figure 4: Schematic representation of searches performed by the three ARTn variants, namely whole ARTn, r-ART and d-ART. Starting configurations are represented by red squares, interpolated structures by green squares, and arrival configurations by circles. To guide the eye, the workflow is indicated by arrows.
Left panels (a) and (b) show two independent explorations from the same starting configuration.
Central panels describe how r-ART converges to a transition state, from a starting configuration near the saddle point (central panel (c)) or from an interpolation between two known configurations (central panel (d)).
Right panels show how d-ART constructs a multi-states diffusion pathway between two known minima. The system starts from an interpolated configuration and converges to a first saddle point, from which a new minimum configuration is found (right panel (e)). A new interpolated point is generated between the new configuration and the previous one, and the algorithm converges to a second saddle point (right panel (f)).



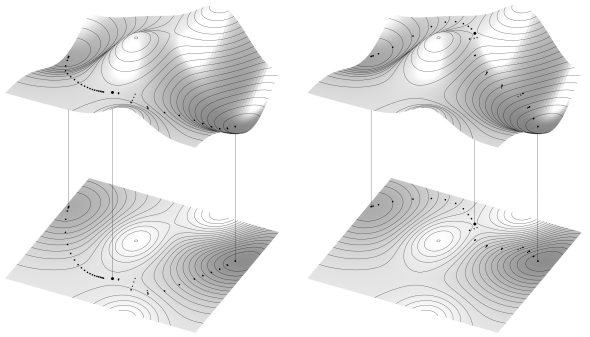
(a) Initial linear interpolation

(b) Not actual MEP



(c) Initial path using an additional intermediate point

(d) Correct MEP



(e) ARTn: search 1

(f) ARTn: search 2

Figure 5: Schematic representation of common biases introduced by the guessed chain of configurations used in string methods and ARTn PES exploration. (a) Standard linear interpolation. (b) Relaxation from this initial guess can result in a higher energy path. (c) Construction of an initial guess path with the introduction of an intermediate configuration close to the actual saddle point. (d) This approach results in the actual MEP. (e) and (f): ARTn exploration of the PES that results in the identification of two paths.

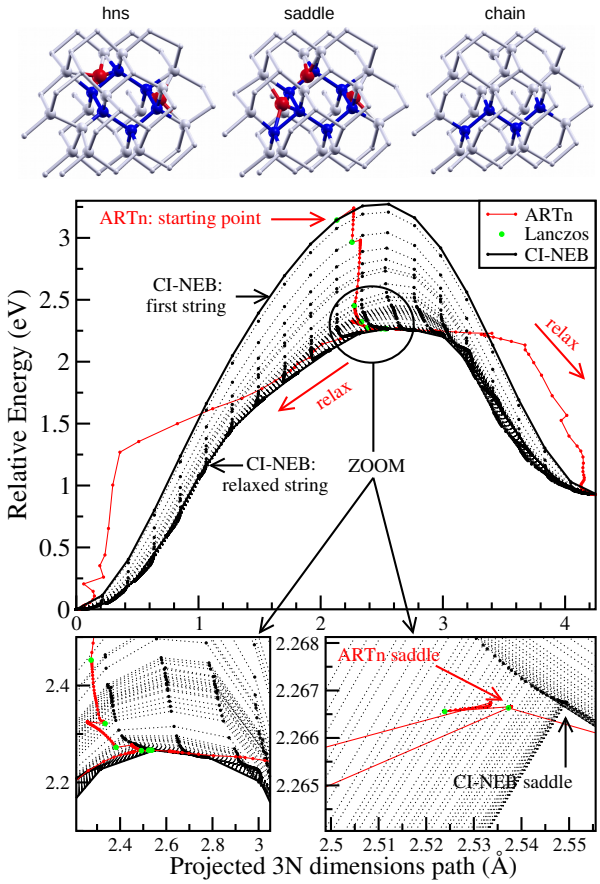


Figure 6: Top panel: Tetra-vacancy in a 212-atoms supercell switching between a *chain* and a *hns* configuration through the saddle point. Structures are represented by silicon atoms out of their crystalline site (red balls), their on-site neighbors (white balls), and their empty ideal sites (blue balls).
Bottom panels: Comparison of the energies of the structures visited by r-ART and CI-NEB and zoom around their saddle points. Big black points: force calculations by CI-NEB. Red points: force calculations by r-ART. Green points: Lanczos algorithm called by r-ART, which requires 16 force calculations for each. For clarity, some of the intermediate strings of the CI-NEB are removed.

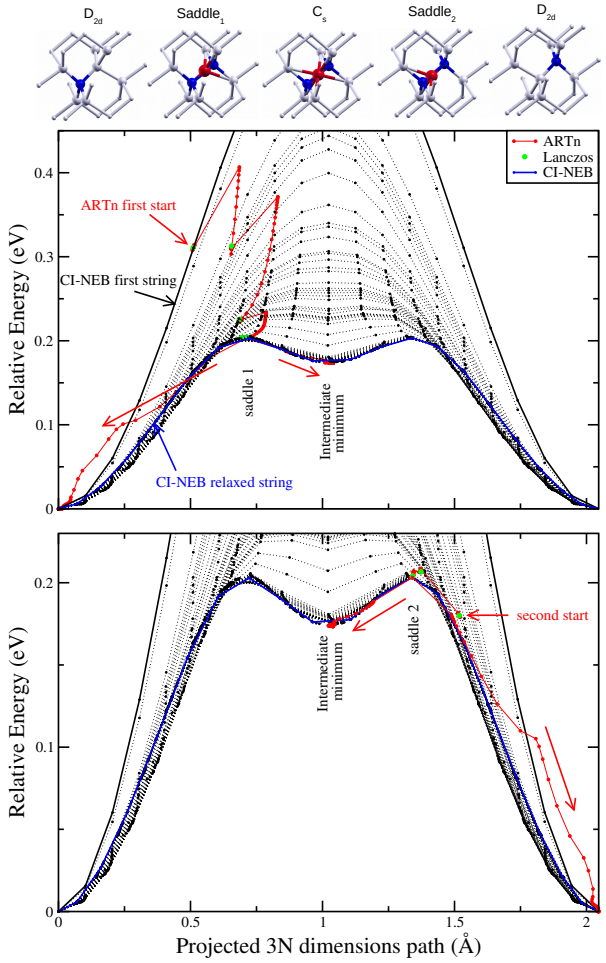


Figure 7: Top panel: Initial D_{2d}, intermediate C_s, and final D_{2d} structures obtained during the diffusion of the mono-vacancy (V-Si) and their intermediate saddle points. Structures are represented by silicon atoms out of their crystalline site (red balls), their on site neighbors (white balls), and their empty ideal sites (blue balls).
Bottom panel: Comparison of the energies of the structures visited by d-ART and by CI-NEB. Green points: Lanczos algorithm called by ARTn, which requires 16 force calculations for each.

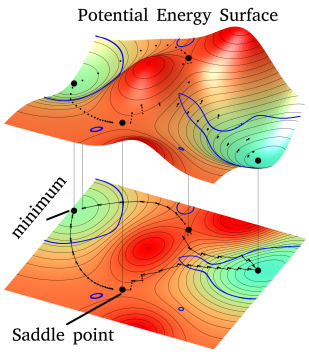


Figure 8: For Table of Contents Only