Computational Materials Science 84 (2014) 217–225

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Predicting vacancy migration energies in lattice-free environments using artificial neural networks

N. Castin [a,b,*], J.R. Fernández [a,c,d], R.C. Pasianot [a,c,d]

[a] Consejo Nacional de Investigaciones Científicas y técnicas (CONICET), Av. Rivadavia 1917, C1033AAJ Buenos Aires, Argentina
[b] Belgian Nuclear Research Centre, SCK•CEN, Nuclear Materials Science Institute, Boeretang 200, B-2400 Mol, Belgium
[c] Comisión Nacional de energía atómica (CNEA), CAC, Gcia. Materiales, Av. Gral. Paz 1499 San Martin, Buenos Aires, Argentina
[d] Instituto Sabato, UNSAM/CNEA, Avda. Gral. Paz 1499, 1650 San Martin, Argentina

ARTICLE INFO

Article history:
Received 19 August 2013
Received in revised form 7 November 2013
Accepted 5 December 2013

Keywords:
Kinetic Monte Carlo
Lattice-free
Artificial neural networks
Diffusion
Grain boundaries

ABSTRACT

We propose a methodology for predicting migration energies associated to the migration of single atoms towards vacant sites, using artificial neural networks. The novelty of the approach, which has already been proven efficient for bulk materials (e.g. bcc or fcc Fe-based alloys), is to allow for any structure, without restriction to a specific lattice. The proposed technique is designed in conjunction with a novel kind of lattice-free atomistic kinetic Monte Carlo model. The idea is to avoid as much as possible heavy atomistic simulations, e.g. static relaxation or general methods for finding transition paths. Such calculations, however, are applied once per Monte Carlo event, when a selected event is applied. The objective of this work is thus to propose a methodology for defining migration events at every step of the simulation, and at the same time assigning a frequency of occurrence to them (using artificial neural networks), in short computing times. We demonstrate the feasibility of this new concept by designing neural networks for predicting vacancy migration energies near grain boundaries in bcc FeCr alloys.

1. Introduction

Atomistic kinetic Monte Carlo (AKMC) methods are widespread tools to study diffusion-controlled microstructural and microchemical evolution in alloys during thermal ageing and under irradiation (see, e.g. Ref. [1] and references therein). In these models, short time-scale motion, such as thermal vibrations of the atoms around their equilibrium position, is disregarded: only significant transitions are considered, allowing for long time scales (seconds, hours or even years) to be simulated. For convenience, they are often formulated in a rigid-lattice world, where the atoms of the alloy are located on the positions corresponding to the crystallographic structure of interest. In that case, the evolution of the system is driven by the migration of point defects (vacancies and/or self-interstitials), whose position is exchanged with nearest neighbor atoms. The jump to occur is each time selected stochastically, based on the Monte Carlo method. Simulations conducted with these methods can potentially predict in detail, and with high accuracy, phenomena such as precipitation or segregation, which are known to affect the mechanical and chemical behaviors of materials ([1] and references therein), thereby determining their aging when in use. There are two main limitations to the application of AKMC tools: the first one is their high computing cost; the second one is the physical reliability of the method itself. Rigid-lattice AKMC models are indeed of straightforward implementation and use, but it is not easy to make them computationally faster; on the other hand most of the physics is contained in the energy barriers associated with the migration jumps of the point defects, E_m, which must embody both the thermodynamics and the kinetics of the system being studied. The accuracy with which these migration energies are determined represent therefore a key issue. First, AKMC simulations will be the more reliable, the most adequate is the cohesive model used as Hamiltonian (interatomic potential or *ab initio* methods). Secondly, the physics stemming out of the cohesive model will be the most accurately transferred into the simulation, the better the influence of the local atomic environment (LAE) on E_m is included in terms of chemistry and strain field. Generally, however, improving the reliability of E_m calculations entails a significant increase in the computational cost of the algorithm.

In previous works [1,2], we proposed a fast rigid-lattice AKMC algorithm to perform accurate simulations of thermal annealing processes in bulk materials. We could accurately simulate the kinetics of precipitation of either Cr [1] or Cu [2] in bcc Fe, obtaining results in excellent agreement with experimental data. The same method was also extended to more complex alloys (e.g. the FeCrW system [3]), and also to the modeling of the mobility and stability of vacancy clusters [3,4] in irradiated materials. The main

* Corresponding author at: Consejo Nacional de Investigaciones Científicas y técnicas (CONICET), Av. Rivadavia 1917, C1033AAJ Buenos Aires, Argentina. Tel.: +54 1167727254.
E-mail address: nicolas.m.b.castin@gmail.com (N. Castin).

0927-0256/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.commatsci.2013.12.016

idea of the model is to rely on a description as a rigid-lattice of the simulated system, and calculate the migration energies associated to the elementary vacancy jumps using the nudged elastic band (NEB) method [5], thereby making as little approximations as possible. Relevant interatomic potentials are used as cohesive energy functions. Artificial Neural Networks (ANN) [6] are designed to predict these migration energies, receiving as input vector a precise description of the LAE. Once trained on the basis of a finite database of relevant examples, the ANN is used in complete replacement of NEB, and the AKMC simulation proceeds very fast.

As such, this ANN-based AKMC algorithm is inapplicable to more complex systems where the rigid-lattice approximation would no longer be valid, such as at free surfaces, grain boundaries, or in the bulk of materials containing dislocations or nano-structural features such as nanovoids and dislocation loops. The limitation is twofold: (a) the impossibility to refer to a "perfect material configuration" at all times complicates or may even prohibit the definition, given the present state of the system, of all possible transitions to a nearby basin of the potential energy; (b) these transitions are *a priori* of different kinds from one event to another, which complicates the formulation of empirical laws, or the design of numerical procedures, to assign a frequency of occurrence to them.

Lattice-free AKMC models, designed to handle very general cases without being bound to restricting hypotheses, do exist. At every step, they explore the local curvature of the potential energy surface, and find transitions to nearby basins looking for saddle-points, as specified by the transition-state theory [7]. For example, in Ref. [8], Henkelmans and Jonsson proposed to use the Dimer method [9], which in theory fulfills the objective to search for all possible transitions. Other authors proposed different schemes, based on other search methods though having the same finality, e.g. the ART method [10] in Ref. [11]. In these schemes, the choice of the migration events, together with the calculation of their corresponding migration barriers, is made on the fly, allowing, given the initial state, to find all possible transition paths to other nearby local minima in the potential energy surface. The advantage is clearly the flexibility with respect to the simulated system. The main drawback, however, is the complexity of the method: any Dimer-like method requires appropriate parameterization to guarantee that most of possible transitions are found, which can be delicate for some systems. Also, as a consequence of the systematic repetition of search for saddle points, the required computing time is large, inherently limiting the practical application of the method to no more than a few thousands of events. This is certainly far insufficient to study long-term and slow processes such as precipitation or depletion of solutes at interfaces.

In this work, we develop the premises of a new kind of AKMC model that intends to be a compromise between the afore-discussed extremes. On the one hand, we aim at allowing for large-scale perturbations of a regular crystallographic structure. Here, we consider a simple example of interfaces: grain boundaries in bcc materials. On the other hand, the model is designed in such a way that heavy atomistic calculations are limited to the strict minimum, in two ways: (a) only static relaxation is necessary, and only once per MC event; (b) in the same line of ideas than our previous works, the migration energy associated to the MC events are very rapidly estimated using ANN's. As a first step towards this ambitious objective, our goal here is to demonstrate the viability of the model, by showing how these ANN's can be designed to predict the vacancy migration energy, in any position of the simulated material, i.e., close to the grain boundary and far from it.

The paper is organized as follows: in Section 2, we start, to fix the ideas, by describing in general terms the new lattice-free AKMC model. The goal is to discuss the hypotheses and approximations that are made, and most importantly to define the exact purpose of the ANN in the model. Next, in Section 3, we extensively describe how these ANN's are designed to accurately predict the migration energies in any structure, which is the bedrock of our model. Last, in Section 4, we show an application of the proposed methodology to a simple case of interfaces: grain boundaries in bcc FeCr alloys.

2. Lattice-free AKMC model based on artificial neural networks

Lattice-free AKMC models are aimed at replacing fully detailed atomistic simulations (molecular dynamics) by a larger-scale model based on transition-state theory [7]. The requirement is thus, at every step of the simulation, to:

1. Make a list of possible transitions from the present state, assumed to be an equilibrium position in the potential energy surface, towards nearby other stable positions.
2. Assign a frequency of occurrence Γ to all these possible transitions, calculated as a thermally activated event.

$$\Gamma = \Gamma_0 \cdot \exp\left(\frac{-E_m}{k_B T}\right) \tag{1}$$

Here, E_m is the migration energy and Γ_0 is an attempt frequency (which can be obtained using Vineyard's procedure [12]). Using general and non-approximate techniques, such as e.g. the Dimer [8], Monomer [13] or the activation-relaxation technique (ART) [10], these two steps are performed at the same time by exploring the local curvature of the potential energy surface around the present state. It is a complex operation, requiring a (very) long CPU time, because static relaxation and curvature calculations are involved.

As an alternative, we propose in this section a new concept of AKMC algorithm, where the above-listed operations are implemented in such a way to reduce the CPU time requirements to the extreme. This is explained in the two subsections below.

2.1. Generic procedure for defining transitions

To respect the transition-state theory, migration events in AKMC simulations should strictly correspond to the transition from the present state of the system towards a nearby metastable state. The migration path, from the initial state A towards the final state B should be, consequently, qualitatively similar to the one illustrated in Fig. 1(a): at the onset of the transition, the total energy of the system smoothly increases, until it reaches a local maximum. The latter is called saddle point, because it is characterized by a negative curvature in the direction of the transition, and a positive curvature in all others. Once the saddle point is crossed, the system falls into state B.

General methods for searching transitions (Dimer, Monomer, ART, etc.) directly search for saddle points from the initial state. Their contribution to the AKMC is thus twofold: (a) they perform themselves the identification of directions of migrations for the system towards the nearest saddle points, thereby establishing the list of events for the MC; (b) at the same time, they provide the value of the corresponding migration energy, necessary to calculate the events frequencies using Eq. (1).

Since we aim at avoiding these highly complex operations, we propose an alternative in two steps:

- Directions of migration are determined using an "event definition generic procedure" (EGP), where a complete exploration of the potential energy surface, searching for saddle points, is

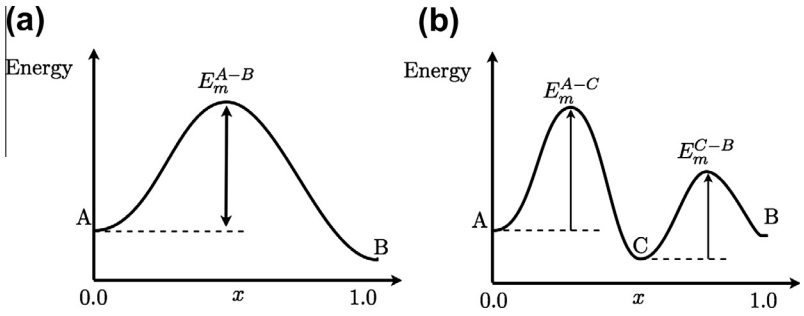

Fig. 1. Illustrative examples of minimum energy path obtained with NEB, given the initial state A and the final state B. The abscissa is a dimensionless coordinate x of advancement. (a) Ideal case, where the transition travels via a single saddle point. (b) Case where the transitions travels via intermediate metastable states, denoted as C.

not performed. Instead, events are defined relying on hypotheses regarding the types of transitions that are expected. The EGP we use in this work is described below.

- In the previous step, possible migration events were defined, and the final states where they lead to are known. The associated migration energies can thus be calculated using the NEB method. Similarly to our previous works [1–4], we assume Γ_0 constant in first approximation.

The reliability of our lattice-free model therefore relies on realistic hypotheses for the definition of the EGP. The latter must indeed be defined in an appropriate way, including all relevant transitions that might be encountered during the simulation. In this work, we consider the transport by single vacancies of solute atoms towards grain boundaries. A sensible definition of the EGP can thus be made as a generalization of rigid-lattice AKMC models, as depicted in Fig. 2(a). In bulk materials with a single vacancy (see e.g. [1] for bcc Fe-based alloys), i.e., far enough from the grain boundary, migration events are typically defined as the jump of that point defect to a first-nearest neighbor position. Other transitions (e.g. jumps to further-away neighboring positions), though possible, are not included because of the high energy barriers associated. Close enough to the interface, the lattice distortion cannot be neglected, and transition events are *a priori* no longer so systematically determined. If we assume, despite the lattice distortions, that the dominant transitions remain describable as the migration of the vacancy towards a neighboring position, the EGP can be defined as illustrated in Fig. 2(b). Not being in a region of the material that can be univocally compared with a perfect lattice, the exact position of the vacancy cannot be defined by searching for a missing atom. Since we consider grain boundaries, where the atomic positions retain an organized structure, we assume that the vacancy can be spotted by searching for a position where the local atomic density is the smallest. Mathematically speaking, it is equivalent to searching for a position that is the furthest away possible from all the atoms in the system. Several algorithms could be imagined to determine it. We here propose a relatively simple one based on the Voronoi polyhedra associated to every atom of the crystal (for this we use the "qvoronoi" program provided with the "qhull" package [14]), namely, the vertexes of these polyhedra correspond to positions that are furthest away from every nearby atom. Thus the vacancy position is defined as the vertex that is the furthest away from its closest atom.

Once the position of the vacancy is determined, migration events are defined as the jump towards this position of one of the nearest atoms, as depicted in Fig. 2(b). This is however insufficient to define the events completely, because it cannot be assumed that the final equilibrium position of the migrating atom is the position of the vacancy, nor that the surrounding atoms are not displaced in a significant way. For that reason, it is necessary to perform static relaxation for final states associated to all migration events. The procedure to determine each of them there-

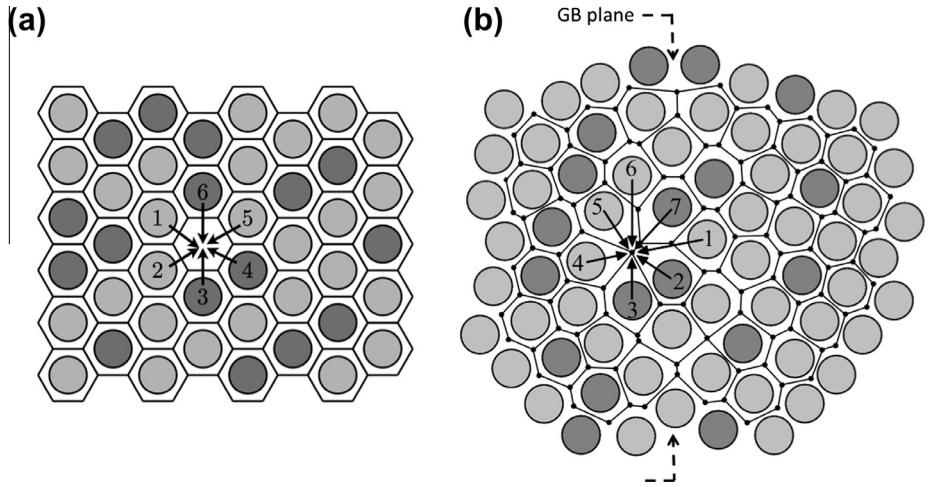

Fig. 2. Generic procedure to determine possible transitions (EGP) from the present state. Different colors denote different chemical species, e.g. Fe and Cr. (a) Far from the grain boundary, modeled as a rigid-lattice. The lines show the Wigner–Seitz cells delimiting the different atomic sites. Arrows show the migration events defined by the EGP, i.e., jumps towards the vacancy site of one of the closest neighbor atoms. (b) Close to the interface, no more modeled as a rigid-lattice. The lines show the Voronoi cells delimiting the volume associated to every atom. Migration events are defined as the jump towards a Voronoi cell vertex of any of the closest atoms to it. Only vertices with enough free volume are retained as possible destination.

fore proceeds in two steps, as depicted in Fig. 3. Note that this procedure perfectly transfers to bulk bcc, or, equivalently, far from the grain boundary. In that case, transitions defined by the EGP are identical to rigid-lattice AKMC: migration towards the vacancy of all the first-nearest neighbors atoms.

Migration events can always be defined using this procedure, and their associated migration energies can thus also always be defined. It is possible, however, to find limiting cases that break the hypotheses of the transition-state theory. Considering migration events not identified while searching for saddle points, migration paths as depicted in Fig. 1(b) can, in principle, be encountered: instead of directly migrating towards state B, the system passes via intermediate metastable states (denoted as C in the figure). Should this occur, it would mean that our EGP fails at identifying relevant transitions, and is not appropriate for the simulated system. However, as substantiated in Section 4, this was never the case.

2.2. Proposed KMC algorithm

Above, we have described how to define migration events and assign a frequency of occurrence to each of them. AKMC simulations can thus, in principle, be implemented. The associated CPU requirement, however, remains large, because static relaxation and the application of NEB are required as many times as possible migration events are defined. To fix the ideas, for bcc structures, the number of nearest neighbors to the vacancy is 8 if the latter is far from the GB, but can be as large as 12 in some positions close to it. Static relaxation and NEB must, consequently, be applied 10 times on average for every MC step. To remedy this inherent complexity of the method, we propose an alternative algorithm inspired from our previous works in bcc-Fe-based alloys [1–4]. The basic idea is to estimate with ANN the frequencies associated to all the events, thereby avoiding to calculate them directly.

Given a transition, its associated migration energy can be seen as a function of the initial and final states:

$$E_m = f(\bar{R}_i, \bar{R}_f) \tag{2}$$

Here, states $\bar{R}_i$ and $\bar{R}_f$ are given by the list of all atomic coordinates in the same reference frame. It thus looks, from this point of view, that the final state must be known to characterize completely the event. In reality, the latter is always correlated to the initial state, no matter how the transition was defined. If transition events are defined using a general lattice-free technique T (e.g. T = Dimer, Monomer, ART, etc.), the final state is obtained as the ending point of the transition path that is searched from the initial state: at every step of the simulation, intents to escape from the present state $\bar{R}_i$ are made by applying different perturbations, before exploring the local curvature of the local energy surface to find a saddle point. Given T, the migration energy associated to a transition that is found can, in very general terms, be written as:

$$E_m^{(e)} = f_T(\bar{R}_i, d\bar{R}^{(e)}) \tag{3}$$

Here, $d\bar{R}^{(e)}$ is the vector of the initial perturbation leading to event e, and $E_m^{(e)}$ is the associated migration energy. Function f_T is, *a priori*, very complex, and approximating it using e.g. an ANN may be too hard numerically, for at least two reasons: (a) given $\bar{R}_i$ and $d\bar{R}^{(e)}$, a small change to the latter may lead to a different final state. Also, that final state might be in completely opposite direction compared to the $d\bar{R}^{(e)}$ vector, which may be delicate to anticipate; (b) given $\bar{R}_i$, many different $d\bar{R}^{(e)}$, not necessarily similar to each other, may lead to the same final state. In our case, since migration events are defined using the EGP as described in the previous section, Eq. (3) can be written in a simpler way (T = EGP):

$$E_m^{(e)} = f_{\mathrm{EGP}}(\bar{R}_i, A^{(e)}, \bar{d}^{(e)}) \tag{4}$$

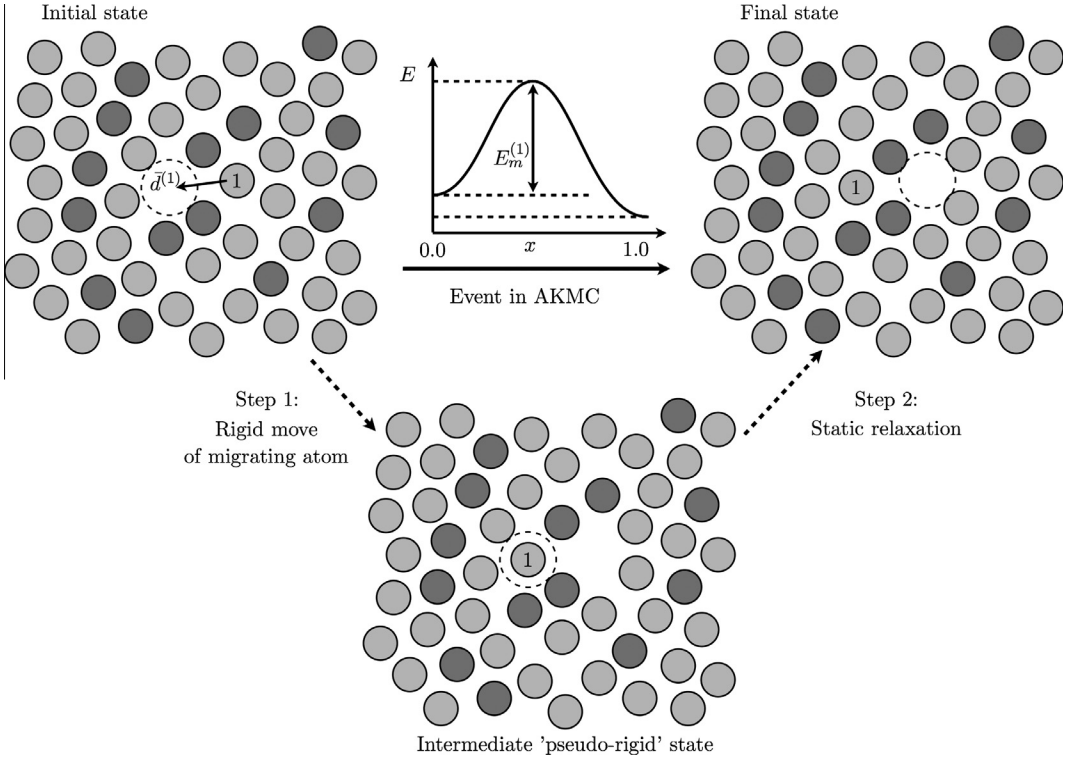

Fig. 3. Schematic representation for the generic procedure to determine possible transitions (EGP) from the present state. The depicted example corresponds to event number 1 in Fig. 2(b). In step 1, the migrating atom ($A^{(e)} = 1$ in Eq. (4)) is moved to its theoretical final position as determined by the EGP ($\bar{d}^{(e)} = \bar{d}^{(1)}$ in Eq. (4)), i.e., in the center of the vacant site (dashed circle), while the positions of the other atoms are unchanged. In step 2, this "pseudo-rigid" state is relaxed using static relaxation conjugate gradients, to find the true final state associated to the migration event.

Here, $A^{(e)}$ denotes which atom in the initial state migrates according to the EGP, and, $\bar{d}^{(e)}$ denotes the EGP direction of migration. Regarding the nature of the modeled system in this work, function f_{EGP} is of a manageable complexity, and can be accurately approximated using an ANN, as shown in the next section.

In summary, the proposed lattice-free AKMC model proceeds in four steps:

1. From the present state of the system, the procedure described earlier is followed to identify the position of the vacancy.
2. The EGP is applied to define a list of transition events. Only the first step in Fig. 3 is necessary at this point, i.e., only the migrating atoms $A^{(e)}$ and their respective migration directions, $\bar{d}^{(e)}$ must be determined. For each of them, the associated migration energy is estimated using a properly designed ANN, as described in the next section.
3. The migration events are put in competition using the MC algorithm (respectively to the migration frequencies using Eq. (1)), and one of them is selected.
4. The selected event is applied by executing the second step in Fig. 3, i.e., performing static relaxation to find the equilibrium position of all atoms. It is important to note that this step is, logically, not affected by ANN errors, since relaxation is performed using the potential directly. Physical relevance of the atomic configuration of the system is thus always guaranteed, at least as far as the used cohesive model is concerned. The ANN errors of predictions only affect the choice of migration events to apply, and thereby the time increment by the mean residence time algorithm.

The complexity of the proposed model is limited to the extreme, because static relaxation in step 4, which is undoubtedly by far the bottleneck, must be performed only once per MC event. Indeed, once trained, the required CPU time to perform ANN predictions in step 2 is many orders of magnitudes shorter. In non approximate lattice-free models, transition paths are systematically searched at all steps, implying that: (a) each search for a transition requires many operations, and are at least an order of magnitude more demanding in CPU time compared to static relaxation; (b) these searches must be performed a sufficient number of times to guarantee that all major transitions are included in the MC algorithm. The required number of searches can be seen as an accuracy parameter of the model, but more precisely depends on the kind of transitions that are expected to be found: this is settled by the user, choosing vectors $d\bar{R}^{(e)}$ in Eq. (3). For this work, since we expect of the order of 10 possible jumps corresponding to our EGP, it is reasonable to estimate that similar simulations with general lattice-free models would require 10 independent intents at the very least. As a conclusion, we can anticipate that our model is at least 2 orders of magnitude faster.

3. Design of the ANN to predict vacancy migration energies

Many aspects of the methodology do not fundamentally deviate from our previous works [1–4]. The main differences are twofold: (a) The definition of the input vector, describing the LAE around the migrating vacancy, cannot be the same since we are no longer dealing with a predefined lattice. In Section 3.1, we propose a new definition of this vector, no longer depending on a particular atomic structure, and incorporating in a natural way the specified variables in Eq. (4); (b) As explained later, the size of the input vector is larger, implying a number of numerical complications, in particular for training. For that reason, in Section 3.2, we analyze in detail the proposed ANN training scheme, and describe how its complexity can be managed using parallelized algorithms.

3.1. Choice of ANN input vector

In Eq. (4), function f_{EGP} receives as input variables a complete description of the initial state $\bar{R}_i$, as well as information regarding which atom $A^{(e)}$ migrates according to the EGP, and its theoretical direction of migration, $\bar{d}^{(e)}$. In our previous works [1–4], advantages of the rigid-lattice were used to formulate them as vector of integers, telling the chemical nature of the atoms sitting in the neighboring sites of the migrating atom. In a lattice-free context, where the list of relative coordinates for the neighbors is not identical (or at least nearly-identical) from one site to another, a similar description in cells can also be undertaken. The major drawback, however, is that in practice a quite large density of cells might be necessary to provide a sufficiently accurate description of the atomic environment, leading to a large vector of inputs for the ANN. In fact, we originally started this work formulating the ANN input vector in this way, but we found a number of limitations that shall not be discussed. We turned our attention to another formalism, inspired from works of other authors fitting ANN's to implement interatomic potentials [15,16], that respects a condition of rotational invariance of the LAE. We thus adopted a formulation based on series expansion of the atomic distribution in spherical harmonics, as extensively described in [17]. The migration energy is then formulated as a function of the local atomic densities in the initial state ρ_i:

$$E_m^{(e)} = f_{EGP}(\rho_i, A^{(e)}, \bar{d}^{(e)}) \tag{5}$$

The density is expanded as:

$$\rho(r,\theta,\varphi) = \sum_{n=1}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{nlm}\cdot R_n(r)\cdot Y_l^m(\theta,\varphi) \tag{6}$$

Here, (r, θ, φ) are the spherical coordinates, R_n are orthonormal radial functions, Y_l^m are the spherical harmonics functions, and the expansion coefficients C_{nlm} are calculated using:

$$C_{nlm} = \int\int\int_{r,\theta,\phi} \rho\cdot R_n(r)\cdot Y_l^m(\theta,\varphi)\cdot r^2 sin(\theta)\cdot d\theta\, d\varphi\, dr \tag{7}$$

As can be seen, the atomic density must be explicitly given. In Ref. [17], Bartók et al. suggest, for the sake of simplicity, to formulate ρ as a collection of Dirac delta's, each sitting at the position of every atom in the LAE. Contrary to the other works where these concepts are applied for pure materials [15–17], for our application, we should take the chemical nature of the atoms into account as well. One way to include this information is to multiply each Dirac delta by a constant coefficient S_a, whose value is chosen different for every atomic species: e.g. $S_a = 1$ if atom a is Fe, and $S_a = 2$ if atom a is Cr; this leads to:

$$C_{nlm} = \frac{1}{N_{at}}\sum_{a=1}^{N_{at}} S_a\cdot R_n(r_a)\cdot Y_l^m(\theta_a,\varphi_a) \tag{8}$$

For the radial function R_n, similarly to Bartók et al. in Ref. [17], we construct a set of orthonormal functions:

$$R_n(r) = \sum_{\alpha=1}^{N_{max}} W_{n\alpha}\,\phi_\alpha(r) \tag{9}$$

With:

$$\phi_\alpha(r) = \frac{e^{-\lambda r/R_c}}{N_\alpha} sin(\alpha\pi r/R_c) \tag{10}$$

$$N_\alpha = \sqrt{\frac{R_C(1-e^{-2\lambda})}{4\lambda(1+(\lambda/\alpha\pi)^2)}} \tag{11}$$

Here, N_α normalizes the functions, and R_c is a pre-determined cut-off radius. The sine part of the ϕ function, in Eq. (10), provides an

adequate basis for representing any radial function, providing that N_{max} is large enough. The decreasing exponential term was included to reduce the relative magnitude of the LAE close to the cut-off distance, to avoid numerical problems when atoms are very close to it. We use $\lambda = 3$. To complete the calculation of the radial functions, the W matrix is obtained as:

$$W = S^{-1/2} \text{ with } S_{\alpha\beta} = \int_0^{R_C} \phi_\alpha(r)\phi_\beta(r)dr \tag{12}$$

The C_{nlm} coefficients can be seen as a source of different (complex) numbers describing the LAE, presenting a number of qualities:

- Given values for n, l and m, the coefficient C_{nlm} is a single complex number, independently of the number of atoms in the LAE, and independently of the type of atomic structure.
- They are continuous, differentiable, and bounded functions. Being coefficients of a series expansion, each of them provides complementary information compared to the others.
- A number of coefficients as large as required can be defined and calculated. There are however numerical limitations with respect to the l parameter (see e.g. the GSL library user manual [18]). In practice, $l = 50$ can be seen as a (very) maximum value to take.

Regarding the formulation of function f_{EGP} in Eq. (4), a convenient definition of ANN input variables is to set the origin of the reference frame at the position of the migrating atom $A^{(e)}$, and align the z-axis with its direction of migration $\vec{d}^{(e)}$, as illustrated in Fig. 4. This way, the moduli $|C_{nlm}|$ are, by construction, invariant with respect to any rotation around that axis, and thereby constitute an adequate choice of input variables.

To conclude, the series expansion in Eq. (6) is performed using a limited number of terms: maximum values for the n and l parameters must thus be chosen, henceforth denoted as N_{max} and L_{max} respectively. Concerning the m parameters, because of parity relations, negative and positive values are related to each other. Including both is thus unnecessary, and we arbitrarily decided to keep the positive values. The total number N_{ANN} of input variables is thus given by:

$$N_{ANN} = N_{max}\left(\frac{L_{max}(L_{max}+1)}{2} + L_{max}\right) \tag{13}$$

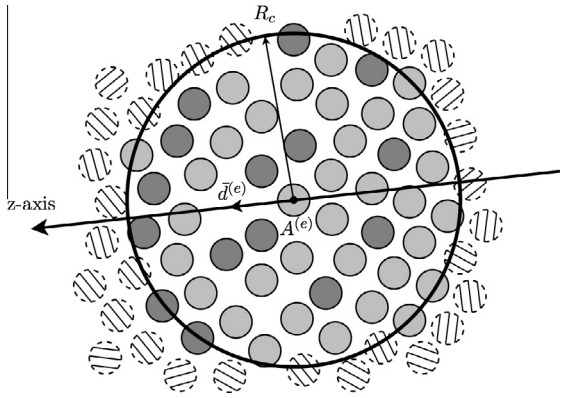

Fig. 4. Description of a migration event using the formalism in Eq. (5), as derived from the EGP. The origin of the reference frame is placed at the position of the migrating atom $A^{(e)}$, and the z-axis is aligned with its direction of migration $\vec{d}^{(e)}$. Only the atoms within a predefined cut-off radius R_c are taken into account to calculate the atomic density ρ_i.

3.2. Training algorithm

Artificial neural networks are, according to the universal approximation theorem, capable of approximating any continuous multivariate function to any desired degree of accuracy, provided that enough neurons are available [19]. This theorem, however, does not demonstrate that the "perfect" ANN can be obtained by training on the basis of a finite number of examples. In the absence of a devoted theoretical framework, ANN training is thus, in practice, formulated as a non-constrained and non-linear minimization problem, similarly to mean-square optimization. In this work, we use a training scheme essentially identical to our previous works [1–4]:

> It is a supervised training problem: The network synapses (whose number is henceforth denoted N_S) are optimized using a pre-calculated set of inputs with their corresponding desired outputs (training set). Progress is monitored during training following the evolution of the prediction error measured with another set of data (reference set) and training is stopped if the error ceases to decrease (early stopping).
> A local training scheme is used, implemented by an iterative gradient-based optimization algorithm. These iterations are called "epochs" in ANN jargon. Initial network synapses must thus be chosen. We initialize them at random between −1 and 1, our experience showing that it is not a sensitive parameter for this application.

Regarding the large number of input variables we expect in this application, and also the large number of training examples necessary (see next section), light training algorithms are, *a priori*, preferable choices: for example, stochastic-gradients-descent algorithms (see [20] and references therein), or batch algorithms such as simple steepest-descent [21], conjugate gradients [6] or the Resilient-Propagation [22]. These methods correspond to a complexity $O(TN_S)$, where T is the number of training examples. After intense trials, however, we concluded that none of these methods has an equal capacity at designing accurate ANN's as the Levenberg–Marquardt method [6]. The latter can be considered as a heavy algorithm: at every epoch, a system of N_S linear equation must be solved, resulting in a $O(TN_S^2)$ complexity. Consequently, ANN training with Levenberg–Marquardt is, for this application, delicate, mainly for two reasons: (a) the computing time to perform an epoch rapidly surpasses 1 h, using a properly parallelized software (house-made) with 8-cores or 12-cores modern personal computer; (b) The amount of memory necessary to store the A matrix of the linear system to solve, and the training set at the same time, may also surpass the available amount on these machines. Depending on the parallelization scheme, either the A matrix or the training set can be divided amongst the different processors, but not both at the same time. Too much processor communications would be required on the contrary case, and the parallelization would be inefficient.

Notwithstanding these numerical limitations, the Levenberg–Marquardt training algorithm can be used for this application, thanks to its particularities that alleviate the global complexity:

- In principle, given the training and reference set, the optimal ANN architecture should be determined, incrementing gradually the number of nodes in the network, either separately or using a constructive algorithm. This implies a substantial amount of experiments, multiplying the global complexity of training. The interested reader is referred to our previous works in Refs. [1,23] for more details. For this application, experience shows that these concerns are unnecessary, because the simplest architectures are almost always optimal. This is in fact

not surprising, since the ANN input variables are defined from series expansions of the LAE: further expansions (by introducing more nodes in the network) does not improve the accuracy of prediction. In this work, we elected the "cascade-correlation" architecture [24], using only one hidden node, as depicted in Fig. 5. This architecture provides, for a minimal N_s, the highest degrees of freedom in the output node.

- Using iterative methods, solving the linear system of equations at every training epoch results to be, in practice, faster than the calculation of the A matrix. We use a conjugate gradients method (see e.g. Ref. [25]) that can be parallelized in an efficient way (the bottleneck of each iteration being the multiplication of the A matrix by a vector). With a properly chosen stopping criterion, convergence is generally obtained in a number of iterations far below N_s.
- The number of training epochs before convergence is, also, conveniently small: 25–30 epochs at the most.

To conclude, ANN training with Levenberg–Marquardt results to be, for this application, manageable, despite its theoretical complexity, and relatively straightforward in practice because no parametric study (finding the optimal architecture) is necessary.

4. Application to grain boundaries in bcc FeCr alloys

In this section, we show the application of the ANN training technique described in Section 3, to grain boundary (GB) structures in bcc FeCr alloys. We consider five different $\langle 1\,1\,0\rangle$ symmetric tilt grain boundaries, taken from Ref. [26], namely: $\Sigma3\{1\,1\,1\}$, $\Sigma3\{1\,1\,2\}$, $\Sigma11\{1\,1\,3\}$, $\Sigma9\{1\,1\,4\}$ and $\Sigma9\{2\,2\,1\}$. A database of examples of migration events was generated in the following way:

- As cohesive energy, we used the two-band embedded-atom type interatomic potential by Bonny et al. [27].
  To favor diversity and avoid repetitive similarities in local atomic structures, a list of "original" atomic sites was extracted from ideal GB configurations (pure Fe compositions), as constructed with the method described in [26]. From these, we only retained sites for which the close neighbors are not equal or similar to each other. Also, we excluded sites too distant from the GB plane: the maximum distance was $2a_0$, i.e., a little beyond the cut-off of the potential ($1.86a_0$). The reason is to consider jumps in very general terms, thereby not excluding those that correspond, in bulk, to second-nearest neighbor positions. The presence of the interface might indeed lower their migration energies and make them competitive with first-nearest neighbors jumps. For convenience, we henceforth give all distances in units of the lattice parameter for bulk bcc Fe, as predicted by the potential: $a_0 = 2.8553$ Å.
- To generate each example of migration event, the following steps were followed:
  - One of the pre-determined "original" sites was selected, and a random concentration of Cr between 0% and 20% was introduced by substitution. Static relaxation was then applied, without volume optimization.
  - To include a vacancy, the atom corresponding to the site chosen in the previous step is removed, and static relaxation is applied again. The result is the initial state that corresponds to the migration event.
  - From this initial state, a migration event was selected at random amongst those defined using the EGP described in Section 2. The migrating atom and its direction of migration were thus first selected, and then the associated final state was determined using the procedure illustrated in Fig. 3.
  - Finally, given the initial and final states obtained in the two preceding steps, the migration energy was calculated using NEB and the interatomic potential.

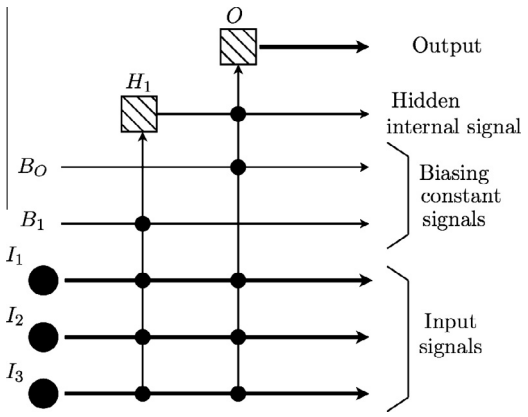

Fig. 5. "Cascade-correlation" architecture for the ANN, with three input signals (I_1, I_2, I_3), one hidden node (H_1) and one output node (O). B_1 and B_O are additional constant signals called nodes biases. The network is fully connected, i.e., every node receives as input all output signals coming from all preceding layers. Given the number I of inputs, the total number N_S of network synapses is given by $N_S = 2I + 3$.

A number of 31,000 migration events were calculated following this procedure, providing a total number of 62,000 examples: each calculation of migration path provides two transitions, from the initial to the final state, and *vice versa*. Migration path were automatically inspected, and no transition similar to the one depicted in Fig. 1(b) was found. This supports our hypothesis that the EGP, as defined in Section 2.1, is appropriate for this application. To complement the database, a number of 15,000 events far away from the grain boundary, i.e. in bulk bcc, were also added. In Fig 6(a), the distribution of distances d_{GB} between the migrating atoms (initial states) and the GB is shown. We see that limiting cases, i.e. either close to the GB or far away from it, are more represented, though without shadowing the intermediate cases. To complement the analysis, let us denote by δ_{LAE} the cumulative displacements, from the initial to the final state, of the atoms in the neighborhood of the migrating one:

$$\delta_{LAE} = \sum_{a \neq A^{(e)}} |\bar{R}_i^{(a)} - \bar{R}_f^{(a)}| \tag{14}$$

Here, $\bar{R}_i^{(a)}$ denotes the vector of coordinates of atom a in the initial state, and $\bar{R}_f^{(a)}$ its coordinates in the final state. We see in Fig. 6(b) that the proportion of cases with $\delta_{LAE} \leqslant 0.5a_0$ is large (72%), which is in good agreement with the hypotheses on which the EGP was defined in Section 2. The other cases, nevertheless, represent a minor but not negligible proportion of the database. This suggests that in some cases the total amount of atomic displacement induced by the migration event is larger (up to $3a_0$), and thereby may accumulate much relaxation during the simulation. This latter point was expected, and justifies the systematic use of static relaxation in the proposed AKMC.

Out of the complete database, the training set was defined selecting 65% of the events at random. Note that the two examples corresponding to the same transition (i.e., from initial to final and *vice versa*) were always included in the same set. In Fig. 7, we show the evolution of the residual prediction error on the reference set, using $N_{max} = 8$ and L_{max} varying from 8 to 36. Similar results are obtained using different values for N_{max}. We see that the residual error rapidly decreases at the beginning, and then the rate of progress saturates. One can anticipate that further reducing the residual error, by increasing N_{max} or L_{max}, is apparently possible, but it would most likely not be significant. It thus seems that "perfect"

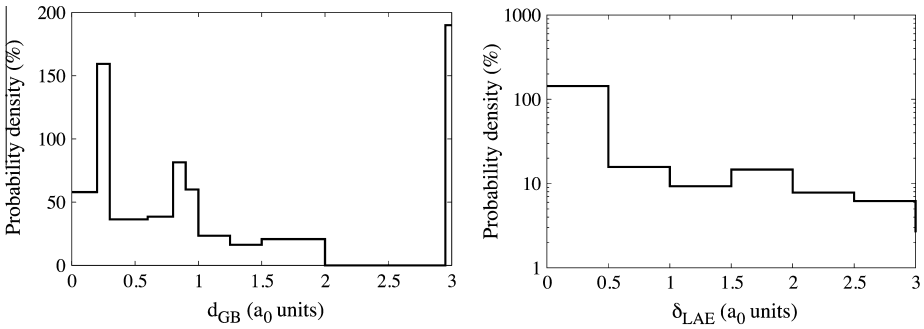

Fig. 6. Details of the composition of the database of migration events. (left) Distribution of distances d_{GB} between the migrating atom (initial state) and the grain boundary plane; (right) distribution of the cumulative atomic displacements in the LAE (δ_{LAE}).

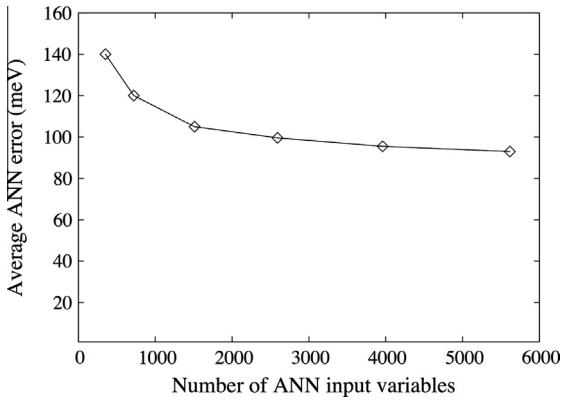

Fig. 7. Evolution of the ANN error of prediction with the number of input variables.

predictions cannot be achieved. This contrasts with the results obtained by Behler and Artrith et al. in Ref. [15–16], where ultra-high levels of accuracy were achieved fitting interatomic potentials on the basis of DFT-calculated energies. An essential difference with our work is that the energies predicted by the ANN are more directly and uniquely determined by the atomic structure. In our case, even tough the situation is in theory similar, the predicted migration energy is the consequence of a complete migration path, were much static relaxation is involved: getting highly accurate predictions is far more delicate.

In Fig. 7, the ANN input variables were calculated using $R_c = 2.75a_0$ in Eq. (10). This value was determined after an appropriate convergence study, similarly to the one extensively described in Ref. [1] for our work in bulk bcc FeCr (i.e., not close to an interface). In that earlier work, a similar value was determined, as necessary condition for giving enough information to the ANN regarding static relaxation and long-range chemical interactions. One might have expected that a larger R_c would be necessary in this work, because additional information about the local atomic structure (presence of the interface) is, *a priori*, necessary for the ANN. Our understanding is that this apparent lack of information is in reality compensated by the fact that the exact equilibrium positions of all atoms in the LAE are used in Eq. (8): information about the local structure, and the relative position and the type of the interface are deducible.

In Fig. 8, we compare directly the ANN predictions with the migration energies. Two separate figures are shown: On the left side, only events corresponding to $\delta_{LAE} \leqslant 0.5a_0$ were retained, because they represent the majority as discussed above; the other cases are shown on the right figure. We can see that the ANN predictions are, globally, very satisfactory, mainly regarding the high correlation with the corresponding migration energies. As already discussed, predictions do not perfectly compare with the migration energies. The magnitude of the error, however, is relatively acceptable compared to the large range of the migration energy (from nearly 0, up to 5 eV). Also, as can be seen in Fig. 9, the error evolves in a logical and anticipated way:

- The error increases with δ_{LAE}, duplicating its value from $\delta_{LAE} = 0.5a_0$ to $\delta_{LAE} = 3a_0$. It shows that the f_{EGP} function in Eqs. (4) and (5) is harder to approximate if δ_{LAE} increases. The accuracy for the largest δ_{LAE} remains, however, in an acceptable range, and could likely be ameliorated if more such examples were provided for training.
- The error decreases with d_{GB}, and converges smoothly to the value corresponding to bulk material: 33.4 meV as shown in Fig. 9. Identical accuracy was obtained in our previous work in bulk bcc FeCr (see Ref. [1]).

5. Conclusive remarks and future perspectives

We proposed a methodology to design artificial neural networks (ANN), for making accurate predictions of migration energies associated to vacancy jump events in lattice-free structures, e.g. near interfaces. Such a tool can be used as a base to define novel approaches in lattice-free AKMC models, which can accurately simulate the transport of solutes, at a computing cost largely reduced compared to non-approximate models. The here proposed ANN technique, as such, requires the definition of an adequate generic procedure for defining migration events (EGP). The major benefit is that heavy atomistic calculations, i.e., static relaxation, is only required once per MC events, and the local curvature of the potential energy surface must not systematically be explored. The drawback, however, is that hypotheses regarding the transition events to include in the simulation must be made, via the definition of the EGP.

In this paper, we showed an application of this methodology to simple examples of interfaces: grain boundaries in bcc FeCr alloys. A sensible EGP is in this case possible to define, assuming that the major transitions are describable in a similar way as in bulk materials, i.e., as migration of single vacancies towards a neighboring site. The quality of prediction of the ANN, after training, is clearly satisfactory, showing that this numerical approach has strong capacities at approximating the migration energies, in varying environments in terms of chemical composition and structural configuration. Naturally, our results are, strictly speaking, valid with respect to the database of examples that we generated, and used for designing the ANN. When used to perform AKMC simulations, it will be wise to assess the accuracy of the predictions for new atomic configurations, and retrain the ANN if necessary.

In a future work, we will conduct AKMC simulations of thermal annealing experiments in grain boundaries in FeCr alloys, in

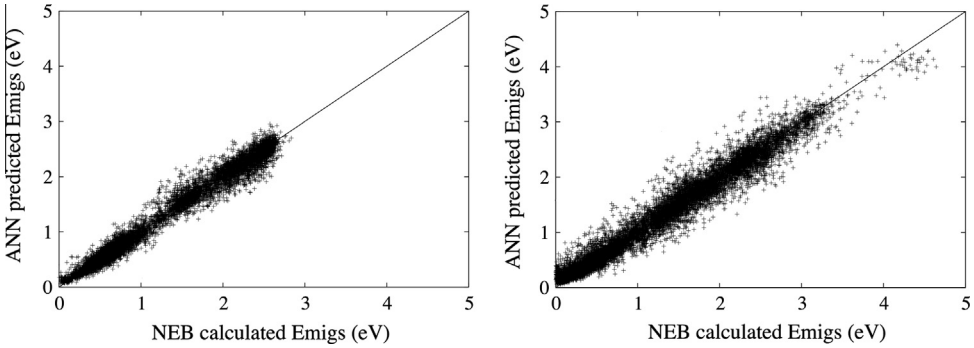

Fig. 8. Accuracy of ANN predictions, after training. The number of input variables is 5616 (N_{max} = 8 and L_{max} = 36). (left) Predictions for cases corresponding to $\delta_{LAE} \leqslant 0.5a_0$. The average prediction error is 64.4 meV. (right) Predictions for cases corresponding to $\delta_{LAE} > 0.5a_0$. The average prediction error is 123 meV.

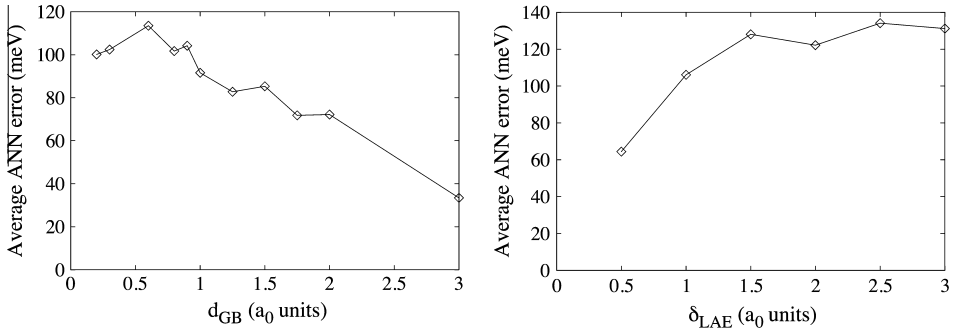

Fig. 9. Evolution of the average ANN error of prediction with: (left) the distance d_{GB} between the migrating atom (initial state) and the grain boundary plane. Note that the point with abscise $d_{GB} = 3a_0$ corresponds to bulk material; (right) the cumulative atomic displacement δ_{LAE} in the LAE.

similar conditions as the work with Metropolis Monte Carlo by Terentyev et al. in Ref. [26]. The capacity of our model to predict correct equilibrium conditions, as predicted by the interatomic potential, will thus be assessed.

To conclude, the method we presented in this paper can be generalized to avoid the definition of an EGP based on too restrictive hypotheses. As discussed in Section 2, ANN could be used in conjunction with general lattice-free techniques (e.g. Dimer, monomer, ART), by providing approximations to function f_T in Eq. (3), instead of f_{EGP} in Eq. (4). This would allow more degrees of freedom for the transitions to take place, and thereby eventually represent more accurately the kinetics as stemming from the cohesive model. As discussed, this cannot, most likely, be achieved in too complex cases, but nevertheless remains a promising starting point for proposing new lattice-free models. In any case, from the point of view of ANN training, and in particular the definition of its input variables, the method we proposed here can be generalized and adapted to new kinds of transitions, more complex than the migration of a single atom towards a given direction. The trick would reside in the new definition of axis z in Fig. 4. Finally, and perhaps more importantly, the model we proposed in this work can also, undoubtedly, be applied in a similar way to other kinds of interface than grain boundaries. The essential condition is, logically, that a relevant EGP can be defined, for which the knowledge of the main mechanisms of migration must first be acquired.

Acknowledgement

The authors acknowledge partial funding of this work by the PIP-CONICET 804/10 project.

References

[1] N. Castin, L. Malerba, J. Chem. Phys. 132 (2010) 074507.
[2] N. Castin, M.I. Pascuet, L. Malerba, J. Chem. Phys. 135 (2011) 064502.
[3] G. Bonny, N. Castin, J. Bullens, A. Backaev, T.C.P. Klaver, D. Terentyev, J. Phys.: Condens. Matter 25 (2013) 315401.
[4] N. Castin, M.I. Pascuet, L. Malerba, J. Nucl. Mater. 429 (2012) 315.
[5] G. Henkelman, H. Jónsson, J. Chem. Phys. 113 (2000) 9901.
[6] C.M. Bishop, Neural Networks for Pattern Recognition, Clarendon, Oxford, 1995.
[7] K.A. Fichthorn, W.H. Weinberg, J. Chem. Phys. 95 (1991) 1090.
[8] G. Henkelman, H. Jonsson, J. Chem. Phys. 115 (2001) 9657.
[9] G. Henkelman, H. Jonsson, J. Chem. Phys. 111 (1999) 7010.
[10] G.T. Barkema, N. Mousseau, Comput. Mater. Sci. 20 (2001) 285.
[11] L.K. Béland, P. Brommer, F. El-Mellouhi, J.-F. Joly, N. Mousseau, Phys. Rev. E 84 (2011) 046704.
[12] G.H. Vineyard, J. Phys. Chem. Solids 3 (1957) 121.
[13] V.P. Ramunni, M.A. Alurralde, R.C. Pasianot, Phys. Rev. B 74 (2006) 054113.
[14] C.B. Barber, D.P. Dobkin, H.T. Huhdanpaa, ACM Trans. Math. Softw. 22 (1996) 469.
[15] J. Behler, M. Parrinello, Phys. Rev. Lett. 98 (2007) 146401.
[16] N. Artrith, J. Behler, Phys. Rev. B 85 (2012) 045439.
[17] A.P. Bartók, R. Kondor, G. Csányi, arXiv:1209.3140 [physics.comp-ph].
[18] M. Galassi et al., GNU Scientific Library Reference Manual, third ed., ISBN 0954612078.
[19] K. Hornik, M. Stinchcombe, H. White, Neural Networ. 2 (1989) 359.
[20] T. Schaul, S. Zhang, Y. LeCun, No more Pesky Learning Rates, in: Proc. International Conference on Machine learning (ICML'13), 2013.
[21] S. Haykin, Neural Networks: A Comprehensive Foundation, MacMillian, New York, 1994.
[22] M. Riedmiller, H. Braun, A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm, in: Proc. of the IEEE Intl. Conf. on Neural Networks, 1993, pp. 586-591.
[23] N. Castin, R.P. Domingos, L. Malerba, Int. J. Comput. Intell. Syst. 1 (2008) 340.
[24] Scott E. Fahlman, C. Lebiere, The cascade-correlation learning architecture, in: Advances in Neural Information Processing Systems, vol. 2, 1990, pp. 524–532.
[25] M.R. Hestenes, E. Stiefel, J. Res. Nat. Bur. Stand. 49 (6) (1952).
[26] D. Terentyev, X. He, E. Zhurkin, A. Bakaev, J. Nuc. Mater. 408 (2011) 161.
[27] G. Bonny, R.C. Pasianot, D. Terentyev, L. Malerba, Philos. Mag. 91 (2011) 1724.