



**HAL**  
open science

# Sensitivity analysis and uncertainty propagation for SMA-TB potentials

M Trochet, Fabienne Berthier, Pascal Pernot

► **To cite this version:**

M Trochet, Fabienne Berthier, Pascal Pernot. Sensitivity analysis and uncertainty propagation for SMA-TB potentials. Computational Materials Science, 2022, 213, pp.111641. 10.1016/j.commatsci.2022.111641 . hal-03708211

**HAL Id: hal-03708211**

**<https://hal.science/hal-03708211>**

Submitted on 29 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# *Sensitivity analysis and uncertainty propagation for SMA-TB potentials.*

M. Trochet <sup>(a)</sup>, F. Berthier <sup>(a)</sup>, P. Pernot <sup>(b)</sup>

<sup>(a)</sup> Université Paris-Saclay, CNRS, Institut de chimie moléculaire et des matériaux d'Orsay, 91405, Orsay, France

<sup>(b)</sup> Université Paris-Saclay, CNRS, Institut de Chimie Physique, 91405, Orsay, France.

**Abstract:** Uncertainty propagation and sensitivity analysis have been applied to SMA-TB interatomic potentials that are often used in atomistic simulations to study metals and alloys. Usually, the parameters of potentials are estimated by a minimization method to best fit a set of experimental or *ab initio* computational data, such as bulk properties (*e.g.*, cohesive energy and elastic constants) without considering the uncertainty in these data. Our goal is to perform uncertainty propagation from the bulk properties to extract uncertainties and correlations of SMA-TB  $(\xi, A, q, p)$  potential parameters to generate valid parameter sets. Using a statistical framework, we estimate initial probability distributions that allow us to determine the uncertainties for each potential parameter of the SMA-TB. We show that many sets of potential parameters lead to bulk properties included in the initial probability distributions. Local and global Sobol' sensitivity analysis methods, which entails no additional computational cost, show how physical properties can be adjusted using interatomic potentials.

Keyword: Uncertainty propagation; SMA-TB potentials; Sensitivity analysis; transition metals

## 1. Introduction

Atomistic simulations are useful tools for studying the structure-property relationship. These simulations, which have a limited length and time scale, are an essential part of a multiscale approach in materials design. One of the challenges of multiscale modeling is the credibility of results. How reliable are the predictions obtained and what are the uncertainties? These questions, which are very general and affect many scientific fields, have led to the development of a new field of research on the quantification and propagation of uncertainties. Two kinds of uncertainty are identified, the *epistemic uncertainty* due to the lack of knowledge both about the physical system and the models, and the stochastic uncertainty of the simulations called *aleatory uncertainty* (random fluctuation due the temperature, number of events considered in Monte Carlo steps...) [1,2].

In this work we investigate the uncertainty parameters of interatomic potentials used to describe metals at atomic scale. Modeling bulk properties of transition metals with high accuracy plays a key role on more complex studies such as prediction of surfaces, kinetics and phase diagrams properties. Extensive *ab initio* calculations and experiments onto pure transition metals have been made during the last two-three decades to construct databases of multiple properties [3,4]. However, some properties are obtained from non-*ab initio* simulations, that are faster but less accurate, such as molecular dynamics (MD) or Monte Carlo (MC) simulations, that require to define interatomic potentials. The ability of these

simulations to correctly predict the properties of the studied materials depends on the interatomic potentials used and on their parameters. Many interatomic potentials have been developed in the last years (pair potentials, Lennard-Jones [5], Finnis Sinclair [6], EAM [7–10], SMA-TB [11–15], MEAM [16,17], REAXFF [18–20], Finnis Sinclair [6], ADP [21–23]...). These potentials depend on a more or less important number of parameters, from two parameters for Lennard-Jones potentials to twelve parameters for ADP potentials. The choice of one class of potentials rather than another is very subjective. This choice depends mostly on the objectives. The more complex the potentials, the more realistic they are, the greater the challenge of identifying the relationships between the results and the parameters. The simpler, less realistic potentials allow for easier rationalization of the results. Of course, the choice of a potential inevitably introduces a bias. All potentials have uncertainty related to the analytical form of the model, which affects predictions.

In this work, we consider N-body interatomic potentials issued from the second moment approximation of the tight-binding (SMA-TB) scheme derived by Ducastelle [11]. These interatomic potentials have been shown to be quite successful to study bulk, surface, and grain boundaries in metallic alloys [24–30]. They are the sum of an attractive and a repulsive term and are depending on four adjustable parameters ( $\xi, A, q, p$ ) which are determined by fitting experimental or *ab initio* values of cohesive energy, bulk modulus and shear elastic constants for pure metals, and mixing energies in the dilute limits for alloys. The parameters can be quite simply related to physical quantities in the framework of the 1st neighbor approximation. A binary alloy is described using twelve parameters and atomic radii. From such potentials it is possible to extract the thermodynamic driving forces of the alloys as a function of the parameters and/or the original data [29,31].

Once the interatomic potentials are calibrated, *i.e.* the best parameters are determined, they are used in MD or MC simulations to predict physical quantities not used for parameter fitting (surfaces energies, vacancy formation...). Most often, the authors do not consider the uncertainty of the physical properties used to calibrate the potentials: the parameters of the potential are estimated by an optimization method in order to reproduce as well as possible the reference data. The quantification and the propagation of uncertainty is rarely explored. Uncertainty due to the class of potential and parameter uncertainty are the major components of epistemic uncertainty [32].

Uncertainty quantification (UQ) is the characterization of uncertainty in input variables and models, while uncertainty propagation (UP) is the determination of uncertainty in outputs based on the uncertainty in input variables. UQ and UP are often combined under the general scope of uncertainty analysis (UA). Complementary to UA, sensitivity analysis (SA) aims to describe to which extent the model outputs are affected by changes in the model inputs. The more complex a model is, the more often some parameters have no influence on the results. It is then possible to propose a surrogate model simpler than the initial model, with fewer parameters. Performing both analyses (UA and SA) helps to reduce errors during the model calibration procedure, by prioritizing the input parameters, and increases confidence in the predictability of a given model, within the limits where SA and UA have been performed. Homma *et al.* [33] split SA in two classes: *local* and *global*. *Local* SA (LSA) is the study of the derivatives (Jacobian and sometimes Hessian) of the model outputs with respect to input parameters around the nominal value of the input parameters. *Global* SA (GSA) considers the output uncertainty over the entire variation domain of input parameters. There are several GSA methods and reviews comparing them, such as Iman and Helton [34], Saltelli and Homma [35] and Saltelli *et al.* [36,37]. We choose here to use Sobol' sensitivity analysis (a variance-based global SA method) on bulk properties to obtain insight on the relative importance of each SMA-TB parameters.

Establishing a parametric model is a two-steps procedure: (1) calibration and (2) prediction. In analytical form, let us define  $Y = F(X)$ , where  $Y = (y_1, y_2, \dots, y_l)$  is a set of  $l$  different quantities of interest (QoIs), estimated by a set of  $m$  models  $F = (f_1, f_2, \dots, f_m)$  used to describe the problem and  $X = (x_1, x_2, \dots, x_n)$  is the set of  $n$  adjustable parameters. The calibration step aims at the determination of  $X$  for a given model  $f_i$  that reproduce the  $l$  QoIs with the most accuracy (it is also known as an *inverse* problem). The prediction procedure uses the fitted models to determine a given QoI  $y_i$  from a model  $f_j$ , with the knowledge of its  $X$  adjustable parameters (this is known as the *direct* problem). In material science, the use of interatomic potentials is a good example of this two-steps procedure: we first define a model, here the SMA-TB model, then we perform the fitting procedure based on bulk properties and then the model is run to predict new physical properties.

In recent studies of interatomic potentials, work has been made to incorporate UA and SA during the calibration and prediction steps. UP of physical properties during the prediction procedure has been attempted, which provides insight on the predictability of the model. Moore *et al.* [38] performed a one-at-a-time (OAT) SA (which is a *local* SA method) of the MEAM parameters on ground states and thermal properties of uranium zirconium alloys, thus providing uncertainty when fitting and running MEAM interatomic potentials. Dhaliwal *et al.* [39] performed a *local* SA analysis on output QoIs, from which they observed that a 1% change in the nominal parameter set can lead to dramatic change in the physical properties of interest.

Our study proceeds in two stages:

1. In a first step, a calibration procedure with UP, we develop a statistical framework that draws random SMA-TB parameter sets within a predefined domain. Four specific bulk properties are evaluated analytically, for which UQ has been made based on available literature, and used to validate or discard SMA-TB parameter sets. Furthermore, Sobol' SA [33,40,41], which does not incur additional computational cost, has been performed to obtain further insight toward the relative importance of each SMA-TB parameters on these four bulk properties. From a large uncorrelated random sample (more than a million points), we obtained a selected small sample of SMA-TB parameter sets (about hundred points), where some SMA-TB parameters exhibit correlations. Thanks to UP, we can quantify the uncertainty of SMA-TB parameters, and estimate the joint and marginals distributions from the selected sets.
2. The second step is a prediction procedure, with UP using the estimated joint and marginals distributions with the correlation approximated by a gaussian copula [42,43].

For the SMA-TB model, we study four different approximations that can be seen as different models, only presenting two of them. Tarantola [44] states that the solution of an inverse problem should be seen as the collection of models (and their respective fitted parameters) that are acceptable given the uncertainty of observables. The procedure is iterative and discriminating: by providing new observables or by increasing precision measurement of existing ones, it is possible to discard models and/or sets of fitted parameters that were acceptable before. Thus, performing a prediction procedure for another physical property, such as the vacancy formation enthalpy, one can further reduce the accepted SMA-TB approximations and their respective accepted parameters. When no set of parameters allows to fit the uncertainties of a new observable then the potentials can be questioned.

Note that this method is akin to the Data Collaboration framework of Frenklach [45]. The existence of a solution at stage 1 reposes on the consistency between the calibration datasets, and between these data and model predictions. The absence of a solution is an interesting diagnostic in itself. An optimization-based framework could be seen as advantageous, as it will always provide a best parameter set, even in

the absence of data/data and data/model consistency. However, the estimation of parameters uncertainty in such conditions (*i.e.* for a non-valid statistical model) is still an open problem and often requires the development of auxiliary error models to account for inconsistencies [46–50]. We wish to avoid complex statistical scenarios for the simple models under consideration in this study.

The goal of this work is to provide a systematic method that can be applied to different classes of potentials to determine the relevant model inputs and estimate the error bars of atomic calculations.

The paper is organized as follows. In Section 2, we describe the SMA-TB potential (2.1) and derive the bulk physical properties in the first neighbor approximation (2.2). Section 3.1 contains a short summary of the ANOVA decomposition and the definition of Sobol' indices (3.1), a presentation of our statistical procedure to produce optimized SMA-TB parameters followed by a discussion of the application of the procedure to more complex potentials (3.2). Then, Section 4 aggregates results from the Sobol' analysis using the first order indices of the SMA-TB parameters (4.1), as well as the characterization of the uncertainty of these parameters for silver (4.2) and for each transition metal (4.3). We then present in Section 4.4 the prediction interval for the vacancy formation enthalpy from evaluation of 1000 samples drawn from the estimated correlated joint distributions. Finally, conclusions are presented in the last section.

## 2. Energetic model

### 2.1. SMA-TB interatomic potentials

SMA-TB interatomic potentials are derived from the tight binding model within the second moment approximation [11]. It has been extensively used in studies related to transition metals and related alloys. The interatomic interaction energy  $E_i$  of an atom  $i$  for a pure metal is defined as

$$E_i(\{r_{ij}\}) = A \sum_{j \neq i} e^{-p\left(\frac{r_{ij}}{r_0} - 1\right)} - \xi \sqrt{\sum_{j \neq i} e^{-2q\left(\frac{r_{ij}}{r_0} - 1\right)}} \quad (1)$$

where  $r_{ij}$  is the distance between atoms  $i$  and  $j$ ,  $r_0$  is the equilibrium interatomic distance between two atoms,  $\xi$  is the effective hopping integral,  $q$  describes the distance dependence of the hopping integral,  $A$  is associated with the strength of the repulsive energy contribution.  $p$ ,  $q$  are dimensionless parameters and  $A$  and  $\xi$  are in eV.

It can be noted that in the original paper [11], the expression of the energy does not depend on the parameter  $r_0$ . Expression (1) was derived later for practical reasons [51]. Under this formulation, the energy depends on  $r_0$  which has a special role since it is an observable property that can be measured with great accuracy.

This SMA-TB potential includes a sum over all others atoms  $j$  from the atom  $i$ . This sum can be split into a sum over the neighboring shells and a sum over all atoms within each shell. For an homogeneous system at 0K, we can assess that all atoms within a given shell  $k$  are located at the same distance from the  $i$  atom such as  $r_{ij} \rightarrow r_k$  and also  $d_k = \frac{r_k}{r_0}$ . We can therefore rewrite the expression of the energy of the atom  $i$  as :

$$E_i(\{d_k\}) = A \sum_k Z_k e^{-p(d_k - 1)} - \xi \sqrt{\sum_k Z_k e^{-2q(d_k - 1)}} \quad (2)$$

with  $Z_k$  the number of  $k^{\text{th}}$  neighbors.

In practice, for computational time reasons, the energies are not calculated by including all neighboring atoms. To avoid discontinuities in both energy and force, the exponentials are connected to zero with a fifth-order polynomial between the  $2^{\text{nd}}$  nearest neighbors and the  $4^{\text{th}}$  nearest neighbors (see Appendix A). The SMA-TB potentials thus defined have been successfully used to study bulk properties and superficial or intergranular segregation [26,52–54].

However, for a better understanding, we have also considered the energy calculations with  $l = 1, 2$  or  $3$  without connection polynomial. In what follows, we only discuss the approximation up to the first neighboring shell, called the 1NN model. The 1NN approximation is clearly unsuitable for accurate atomistic simulations, yet it allows an easy analytical derivation of the bulk physical properties ( $E_{coh}$ ,  $B$ ,  $C'$  and  $C_{44}$ ).

## 2.2. Bulk physical properties

We considered the following transition metals: Ag, Au, Cu, Ni, Pt and Pd. The parameters of the SMA-TB potential can be determined from bulk physical properties, energies of defect such as surfaces, interstitials or vacancies. In this work we have chosen quantities that characterize the homogeneous system: the cohesive energy of the metal  $E_{coh}$  and the elastic constants. In cubic crystals there are only 3 independent elastic constants, namely  $c_{11}$ ,  $c_{12}$  and  $c_{44}$ , due to symmetry. However, it is more common to measure the bulk modulus  $B$  and shear moduli  $c'$  and  $c_{44}$ . By convention, the bulk modulus  $B = \frac{1}{3}(c_{11} + 2c_{12})$  and the shear modulus  $c' = \frac{1}{2}(c_{11} - c_{12})$ . We present in Appendix B, the small deformation theory derivation of SMA-TB analytic expressions for elastic constants, and give in Appendix C the generalized expressions of  $E_{coh}$ ,  $B$ ,  $c'$  and  $c_{44}$  with respect to the SMA-TB parameters ( $\xi, A, q, p$ ) for any  $l$  value.

For any potential describing a single crystal, the equilibrium distance is determined by the minimum of the interatomic energy  $E_i(\{r_{ij}\})$ : the interatomic distance of equilibrium is  $r_0$  and the lowest energy is the cohesion energy  $E_{coh}$ . Within the 1NN approximation, the cohesive energy  $E_{coh} = E_i(d_1 = 1)$  with  $l = 1$  is

$$E_{coh} = AZ_1 - \xi\sqrt{Z_1} \quad (3)$$

At equilibrium, the forces cancel each other out, leading to

$$\left. \frac{\partial E_i(\{r_k\})}{\partial r_k} \right|_{r_k=r_0} = \left. \frac{\partial E_i}{\partial \varepsilon} \right|_{\varepsilon=0} = 0 \implies -\frac{1}{2}\xi\sqrt{Z_1}(-2q) + Z_1A(-p) = 0 \quad (4)$$

which can also be written as

$$\frac{Z_1A}{\xi\sqrt{Z_1}} = \frac{q}{p} \quad (5)$$

hence Ducastelle's definition  $x = q/p$  [11].

We can reformulate the cohesive energy in function of couples  $(\xi, x)$  or  $(A, x)$  instead of  $(A, \xi)$  such that

$$E_{coh} = -(1-x)\xi\sqrt{Z_1} = -\left(\frac{1-x}{x}\right)AZ_1 \quad (6)$$

For fcc lattices, using the small deformation approximation, the elastic constants are :

- for the bulk modulus

$$B = pq \frac{E_{coh}}{9 \left(\frac{V^{fcc}}{4}\right)} \quad (7)$$

- for the shear modulus  $c'$

$$c' = \frac{1}{6} \left(\frac{1-2x}{1-x}\right) pq \frac{E_{coh}}{4 \left(\frac{V^{fcc}}{4}\right)} \quad (8)$$

- and for the shear modulus  $c_{44}$

$$c_{44} = \frac{1}{3} \left(\frac{1-2x}{1-x}\right) pq \frac{E_{coh}}{4 \left(\frac{V^{fcc}}{4}\right)} \quad (9)$$

where  $V^{fcc} = (r_0\sqrt{2})^3$  is the lattice volume related to the equilibrium interatomic distance between two atoms.

It can be noted that each elastic constant is expressed as a function of the cohesive energy and the volume of the network. The values of  $E_{coh}$  and  $r_0$  are known with high accuracy. We can reformulate the elastic constants as:

$$c = f_c(p, q) g(E_{coh}, r_0) = f_c(p, q) g(A, \xi, r_0), \forall c \in [B, c', c_{44}] \quad (10)$$

with  $f_c(p, q)$  is a dimensionless function of  $p$  and  $q$  depending on the symmetry of the crystal and  $g$  is an energy per unit volume (equivalent to pressure). Once the element is chosen, the function  $g$  is known and the three elastic constants are determined by the two parameters  $p$  and  $q$ . For this reason, the 1NN approximation does not always succeed in fitting all the bulk properties of an element. Rosato *et al.* [51] also mention that the  $c_{44}/c'$  ratio is equal to 2, independently of the values of  $p$  and  $q$ . In the literature (experimental [11] and *ab initio* [3]) this ratio varies in fact between 1.5 and 3 for the fcc structure.

However, the 1NN approximation, gives qualitative information about the possible values of the SMA-TB parameters for transition metals. Considering the conversion ( $1eV.\text{\AA}^{-3} = 160.21GPa$ ) and a typical fcc lattice volume of the order of  $10^3\text{\AA}^3$ , the relationship between the cohesive energy and the bulk modulus indicates that  $B \propto 3E_{co}$  which leads to  $q \approx 3$  and  $p \approx 3q$ . According to equation 5, we then have  $\xi \gg A$ , which means that the cohesive energy is mainly defined by the attractive contribution  $E_{co} \approx -\xi\sqrt{Z_1}$ . This knowledge provides an excellent physical basis for a calibration procedure or for checking whether a fitted set of SMA-TB parameters is reasonable.

For reasons of readability, we only present here the formulas for the 1NN approximation. The extension to the 2NN and 3NN approximations is described in Appendix D. Finally, in Appendix E, we provide the specific expressions for the four thermodynamics properties for the SMA-TB potential connected with a fifth-order polynomial.

### 3. Statistical framework

#### 3.1. Sensitivity analysis

Sensitivity analysis (SA) is the study of how the model outputs respond to a change in model inputs [40,41]. This is essential for the modeling of complex systems such as bulk, surfaces and metal alloys. Saltelli *et al.* [40] and also Sobol' [55,56] earlier defined two main groups of SA methods: local SA and global SA. The local SA is based on the study of first partial derivative of each input parameter while keeping other parameters at their 'optimal' values. This local SA only explores importance of the input parameter toward the output around a narrow domain where the set of 'optimal' values of input parameters is thought to be located. It also ignores the interaction between parameters. The global SA explores the whole variation domain of input parameters and considers the global effect of an input parameter on the outputs. However, it is necessary to define an initial domain of input parameters. It is possible to define this parameter space from the units and the realistic order of magnitude of the parameters, or also from the set of parameters used in the literature.

One of the most used global SA method uses the decomposition of variance approach (ANOVA) and is named Variance-based SA ([40,57]). The *first-order* sensitivity index or *main* index of the input parameter  $X_i$  on the model output  $Y$  is defined as

$$S_i = \frac{V_i}{V(Y)} \quad (11)$$

where  $V(Y)$  is the total variance of the model output  $Y$  and the partial variance  $V_i$  is the variance of the conditional expectation  $V_i[E_i(Y \vee X_i)]$  ( $i$  denote all but the  $i^{th}$  input variable). It can be seen as the fraction of the model output variance that would disappear on average when  $X_i$  is fixed to a value within its variation range. The impact on the model output variance of the interaction between parameters  $X_i$  and  $X_j$  is defined as  $S_{ij} = V_{ij}/V(Y)$  with  $V_{ij} = V_{ij}[E_{ij}(Y \vee X_i, X_j)]$  and all higher order of sensitivity indices follow in a similar manner.

Another quantity has been proposed by Sobol' [56] called the "total effect index"  $S_{Ti}$  and can be defined as the sum of all sensitivity indices related to the input variables  $X_i$ . For example, in a system of three variables  $(x_1, x_2, x_3)$  the total effect indices are:

$$\begin{cases} S_{T1} = S_1 + S_{12} + S_{13} + S_{123} \\ S_{T2} = S_2 + S_{12} + S_{23} + S_{123} \\ S_{T3} = S_3 + S_{13} + S_{23} + S_{123} \end{cases} \quad (12)$$

When compared to the first-order indices,  $S_{Ti}$  provides information about interactions between the input parameters. One property of the total index is that  $\sum_i S_{Ti} \geq 1$  and it is equal to 1 only and only if there is no interactions between parameters. Therefore, should have  $S_{Ti} = S_i$  for a purely non-interacting model.

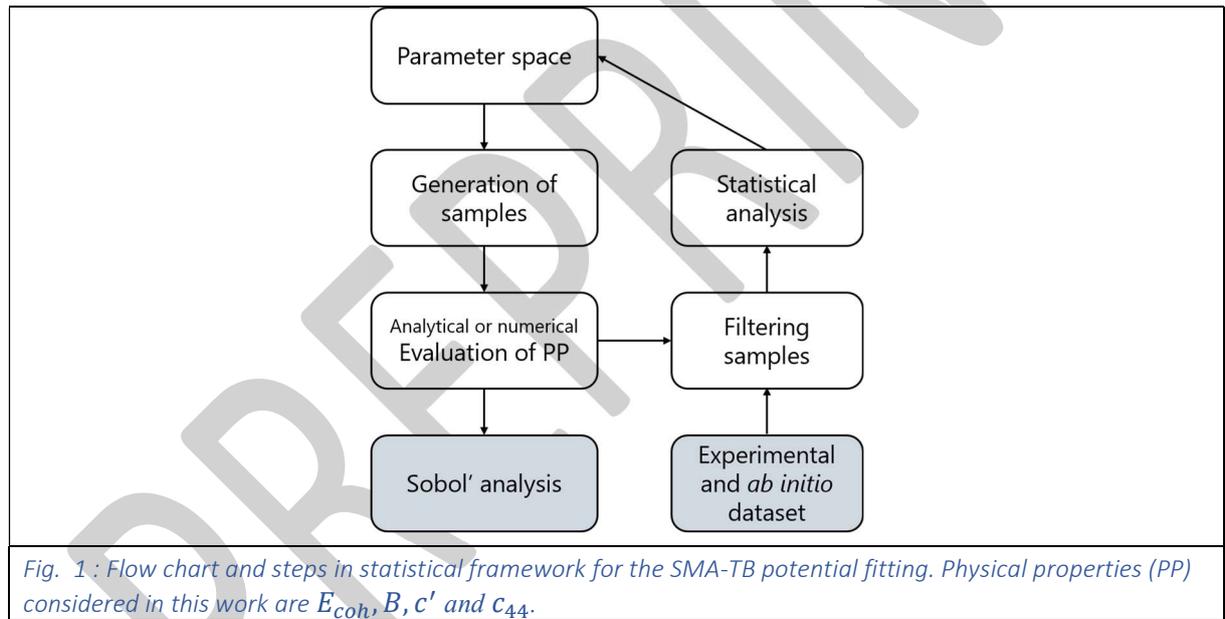
Let us recall, that main indices are a good indicator for inputs prioritization, and one should focus on parameters with the largest first-order indices. A null value of the total effect index indicates that this input is not contributing in any way to the output variance and therefore can be fixed within its initial predefined domain which is a simplification, of interest when the model is complex with many parameters. Less complexity means greater ease in analyzing all aspects of interest.

### 3.2. Calibration of input parameters

A standard way to identify the 'best fit' of the parameters of a given model is to minimize an error function that provides a number representing the difference between the data and the model prediction. This is usually the root mean square error (RMSE) or maximum likelihood [58]. However, the physical properties ( $E_{coh}, B, c', c_{44}$ ) used as reference data do not have fixed values, which excludes the determination of unique values for the parameters. Based on the literature, we found that the cohesive energy is known to an accuracy of 0.01 eV, which is very high quality, but the elastic constants can have a deviation of up to 40 GPa in some cases.

To introduce the uncertainty on the physical properties, we decided to reject any SMA-TB parameter set that does not predict a QoI within a predefined confidence interval. By using this kind of metric, we do not sort by 'best' to 'worst' the small final set of remaining samples that we obtain, since we do not have enough precision on the ( $E_{coh}, B, c', c_{44}$ ) to pretend to know which of the final ( $\xi, A, q, p$ ) values are better. Moreover, if the number of samples ( $\xi, A, q, p$ ) selected by these filters is high enough, we can measure the uncertainty on the SMA-TB parameters.

The main steps of the statistical framework are summarized in Fig. 1:



The selection/filtering procedure can be executed iteratively. We describe the main steps of the first iteration:

#### 1. Parameter space definition

This parameter space is chosen based on ( $\xi, A, q, p$ ) quadruples for all transition metals found in the literature and take into account possible uncertainties. Table 1 summarizes the *a priori* information for the ( $\xi, A, q, p$ ) quadruples for six studied transition metals [29].

#### 2. Generation of random samples

We then generate  $N_{tot}$  samples of SMA-TB parameters with quasi-random Sobol' sequences [59] using the Salib python package [60] to pave homogeneously the parameter space. The total number of samples is  $N_{tot} = (2D + 2)N_{sample}$ , where  $D = 4$  is the number of SMA-TB parameters and

$N_{sample}$  ranges from  $2^{12} = 4096$  up to  $2^{18} = 262144$  to check convergence of Sobol' indices and to generate enough samples that satisfy strict filtering, which we will explain later.

### 3. Evaluation of physical properties

We evaluate the four physical properties ( $E_{Co}$ ,  $B$ ,  $c'$ ,  $c_{44}$ ) for all parameters samples and for each of the six metals considered.

### 4. Sobol' analysis

The implementation of the Sobol' analysis does not require additional development. The sample of output variables obtained in step 3 are used for sensitivity analysis. We compute the first  $S_i$ , the second,  $S_{ij}$  and the total  $S_{Ti}$  Sobol' indices (hereafter referred to as  $S_i$ 's) of each output (Note:  $N_{sample} = 2^{12}$  is largely sufficient for the convergence of Sobol' indices).

### 5. Filtering of samples

We filter out parameter samples that do not predict output values within the predefined reference intervals established from the literature (experimental and *ab initio* dataset given in Table 2).

### 6. Statistical analysis

We compute mean, standard deviation, min, max of the distribution as well as the correlation between the SMA-TB parameters from remaining samples. This allow to perform the UP from the physical properties to SMA-TB parameter and characterize the UQ of SMA-TB parameter.

We iterate this procedure twice. For the second iteration, we create (step 2) a correlated joint probability distribution with the reduced parameter space and the correlation matrix obtained at the end of the first pass. In this case, the Sobol analysis is invalid because, the SALib package does not support the computation of Sobol' indices with dependent stochastic variables. Recent studies [57] and developments (like chaospy python package [61,62]) deal with stochastic dependent input variables by a new mathematical technique. However, it is still in an early development stage. In perspective, it would be an enrichment of our understanding of the sensitivity of parameters when the package will be able to handle the repartition of main effect, interaction and correlation of each parameter.

Extension of the proposed method to potentials with many parameters will face several problems. First, the identification of a consistent parameters set in a high-dimensional space by random sampling might prove difficult if a good solution is not known beforehand to limit the search space. Second, the absence of analytical expressions for the physical properties of interest would require to run an unrealistic number of simulations or to develop surrogate models, adding an unwanted uncertainty layer.

In Table 2, we gather experimental [63,64] and *ab initio* [3,65] results from the literature to obtain reference intervals for physical properties (last column). We also provide the resulting computed values with the SMA-TB parameters from Table 1. We can notice discrepancies between calculated values and reference intervals in Table 2. For the calculated values,  $c' = 29.45GPa$  and  $c_{44} = 96.896GPa$ , are not in the reference interval. For *Pd* and *Pt*, even the equilibrium interatomic distances are subject to discrepancies. However, we decide that  $r_0$  will be taken from [29] since our known input parameters set are computed based on these values of  $r_0$ . Note that when discrepancies appear between the calculated values and the min-max values obtained from the literature, we enlarge the reference interval used to filter the samples.

**Table 1:** ( $A, \xi, p, q$ ) quadruple for several transition metals found in literature [14] and corresponding physical properties from our work and ab initio [63] or experiment values [64] in parenthesis.

	$A$	$\xi$	$p$	$q$	$r_0$	$ E_{coh} $	$B$	$C'$	$C_{44}$
(unit)	eV	eV	-	-	Å	eV/at	GPa	GPa	GPa
Ni	0.1217	1.6396	10.7626	2.4350	2.490	4.44 (4.44)	186 (188)	29 (55)	97 (132)
Cu	0.1084	1.3434	10.3770	2.6335	2.560	3.50 (3.50)	141 (142)	21 (26)	69 (82)
Pd	0.1435	1.8857	11.7528	2.8749	2.722	4.98 (4.98)	209 (215)	37 (23)	107 (54)
Ag	0.1249	1.2672	10.3453	3.4236	2.890	2.95 (2.95)	108 (111)	14 (16)	41 (52)
Pt	0.2420	2.5060	11.1400	3.6800	2.760	5.87 (5.87)	280 (288)	39 (52)	108 (77)
Au	0.2134	1.8303	10.4201	4.1765	2.885	3.81 (3.81)	173 (165)	16 (15)	43 (42)

**Table 2:** Uncertainty of the physical properties for transition metals. The computed values are obtained from SMA-TB interatomic potentials with parameters from Table 1.  $r_0$  in the computed values column are those from [29]. The experimental values are taken from [63,64] and the *ab initio* values are from [65] and also, the Material Project database [3]. Some unreasonable old references values have been discarded. Bold computed values lie outside of the corresponding reference interval.

Elements	Properties	Computed values	Min ref. value	Max ref. value	Reference interval
Ni	$r_0$	2.490	2.475	2.492	-
	$ E_{coh} $	4.44	4.43	4.45	[4.43,4.45]
	$B$	186.10	178.40	192.57	[178,193]
	$c'$	29.45	39.235	55.205	[29,56]
	$c_{44}$	96.896	107.58	133.51	[96,134]
Cu	$r_0$	2.56	2.469	2.573	-
	$ E_{coh} $	3.50	3.49	3.51	[3.49,3.51]
	$B$	141.01	100.73	143.67	[100,144]
	$c'$	21.0	14.295	31.65	[14,32]
	$c_{44}$	68.8	50.01	84.72	[50,85]
Pd	$r_0$	2.7224	2.739	2.751	-
	$ E_{coh} $	4.98	4.97	4.99	[4.97,4.99]
	$B$	209.49	195.7	215.84	[195,216]
	$c'$	37.02	16.865	29.595	[16,38]
	$c_{44}$	107.59	55.67	90.32	[55,108]
Ag	$r_0$	2.890	2.874	2.892	-
	$ E_{coh} $	2.95	2.94	2.96	[2.94,2.96]
	$B$	108.38	100.23	109.99	[100,110]
	$c'$	14.12	14.37	19.05	[14,20]
	$c_{44}$	41.34	45.97	56.79	[41,57]
Pt	$r_0$	2.76	2.764	2.772	-
	$ E_{coh} $	5.87	5.86	5.88	[5.86,5.88]
	$B$	280.8	283.09	292.09	[280,293]
	$c'$	39.02	14.97	52.57	[14,53]
	$c_{44}$	107.03	68.28	81.5	[68,108]
Au	$r_0$	2.885	2.873	2.885	-
	$ E_{coh} $	3.81	3.80	3.82	[3.80,3.82]
	$B$	173.40	166.89	189.57	[166,190]
	$c'$	16	12.21	25.22	[12,26]
	$c_{44}$	43.1	42.22	65.41	[42,66]

## 4. Results

### 4.1. SMA-TB parameters sensitivity analysis

Following our statistical framework on the six metals, we present a global sensitivity analysis of each physical quantity, using uncorrelated input parameters. We consider successively a wide initial parameter space covering the values of all six metals and a tight one adapted to a single metal.

#### 1. Wide parameter space

We consider first a parameter space large enough to describe the six metals (Table 1). We create an uncorrelated joint probability distribution with the following (uniform) marginal distributions:

$$\begin{cases} \xi \in [1,3] \\ A \in [0.08, 0.3] \\ q \in [2, 5] \\ p \in [8, 14] \end{cases} \quad (13)$$

We show the first-order  $S_i$  and the total-order  $S_{Ti}$  Sobol' indices for each transition metal and for all physical properties in Fig. 2a.

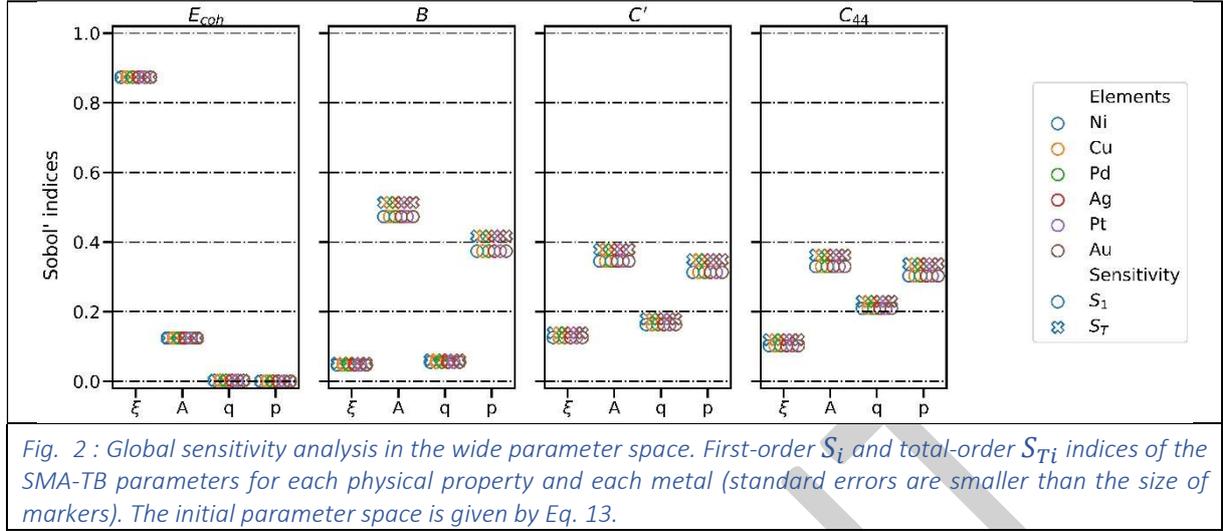
The first observation is that  $S_i$  does not depend on the metal. This is expected, since the only quantity that defines an element in the SMA-TB potential energy is  $r_0$ . The cohesive energy does not depend on  $r_0$  (see § 2), whereas the elastic constants do through the volume of the fcc lattice,  $V_{fcc}$ . A variation of this volumic term affects the mean value and variance of  $B$ ,  $c'$  or  $c_{44}$ , but does not change the Sobol' sensitivity indices, since the variation cancels out through the ratio of conditional variance and total variance.

The second observation is that the total-order sensitivity indices and first-order sensitivity indices have similar  $S_{Ti} \approx S_i$  values, for each metal and for each physical property, indicating no significant interaction between the parameters. We can thus say that the SMATB model itself is well stated on the basis of Hadamard's conditions [66,67] to the notion of an inverse problem.

The sensitivity indices indicate that  $\xi$  is the most important parameter, contributing to 88% of the variance of  $E_{coh}$ . It is followed by  $A$ , contributing 12%. These results agree with the physical meanings of  $\xi$  and  $A$ , since  $\xi$  is the effective jump integral that fundamentally defines the binding properties of the metal under consideration and  $A$  scales the repulsive energy contribution, which competes with the attractive contribution driven by  $\xi$  to stabilize the equilibrium interatomic distance and the cohesive energy.

For the elastic constants, the dominant parameters are  $A$  and  $p$ . The low sensitivity to  $\xi$  and  $q$  shows that the elastic properties are mainly driven by the repulsive contribution. This is especially true for  $B$ , for which the sensitivity to  $\xi$  and  $q$  is almost null. The sensitivity index of  $q$  is more significant for  $c'$  and  $C_{44}$ .

$A$  plays a leading role for all elastic constants ( $B, c', c_{44}$ ) with the help of  $p$ , where  $\xi$  almost defines entirely the cohesive energy  $E_{coh}$ . The parameter  $q$  is insignificant for  $E_{co}$  and  $B$ , while it is a little more important for  $c'$  and  $C_{44}$ . That means that  $q$  cannot be determined precisely in a calibration procedure. For each of these metals, the four physical properties are mainly defined by the triplet  $(A, p, \xi)$ , the value of  $q$  having a minor influence.



## 2. Tight parameter space

We consider a reduced parameter space centred around the parameter values of each metal given in Table 1, with 10% uncertainty. Thus, the reduced variation range of parameters for *Ag* is :

$$\begin{cases} \xi \in [1.14, 1.39] \\ A \in [0.11, 0.14] \\ q \in [3.08, 3.76] \\ p \in [9.31, 11.38] \end{cases} \quad (14)$$

The difference between the wide and tight parameter spaces is that, in the first case, the system has no *a priori* information about the metal under consideration, whereas, in second case, the equilibrium condition for a single metal is satisfied, since parameters values are close to their optimum. This is why the global  $S_i$ 's do not depend on the metal for the wide parameter space but do for the tight parameter space (Fig. 3).

Knowledge of the metal does not significantly affect  $S_i$ 's for  $E_{coh}$ . The Taylor expansion for  $r_{ij}$  in Eq. 1 around  $r_0$  shows that the parameters  $p, q$  do not contribute to  $E_{coh}$  with respect to  $A, \xi$ . On the contrary, the derivation of the elastic constants shows quadratic terms in  $p$  and  $q$ . Once the equilibrium condition is satisfied, we can assume that  $A$  and  $\xi$  are fixed.

The important information obtained by the global SA starting from a wide and then a tight initial parameter space is that the two parameters  $\xi$  and  $A$  are defining the metal through the value of  $r_0$  and the equilibrium condition. The unitless parameters  $q$  and  $p$  allow the physical properties to be reproduced. This common-sense result holds regardless of the approximation considered. This global SA may be useful for other potentials where the nature of the parameter is not so easily categorized.

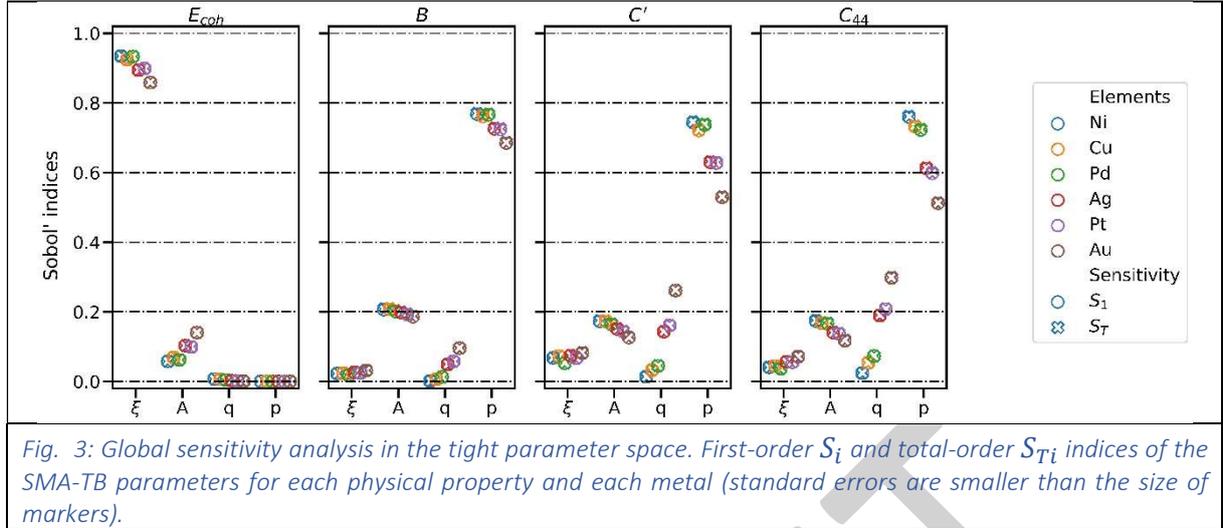


Fig. 3: Global sensitivity analysis in the tight parameter space. First-order  $S_i$  and total-order  $S_{T_i}$  indices of the SMA-TB parameters for each physical property and each metal (standard errors are smaller than the size of markers).

## 4.2. Generation of optimized parameters

In this section, we only present the results for  $Ag$ . Using the  $N_{tot}$  samples according to the wide parameter space (Eq. 13), we compute the four physical quantities analytically.

Fig. 4 displays the distributions of the physical properties obtained at the first iteration from the  $N_{tot}$  parameter samples. The shaded areas represent the reference intervals, collected from experimental or *ab initio* studies found in the literature for  $Ag$  (see Table 2). The wide initial parameter space leads to a large range of physical property values, some results having no physical meaning (e.g. negative elastic constants).

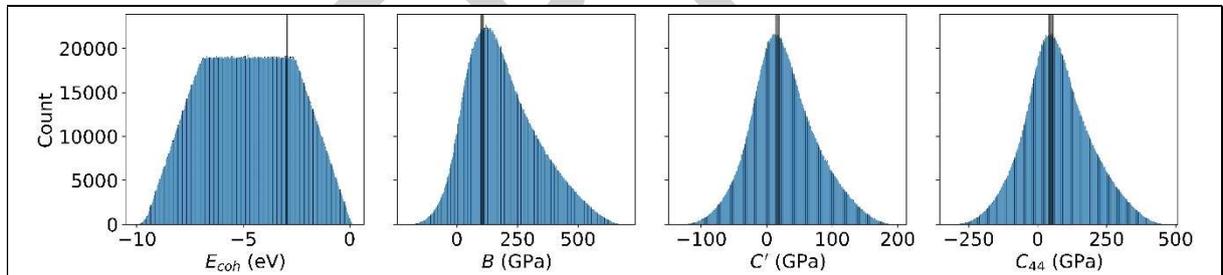


Fig. 4: Distributions of the bulk properties for  $Ag$  for the initial wide parameter space (Eq.13). Reference ranges in bulk properties found in the literature are indicated by the shaded regions.

The min-max values of the physical properties define the acceptable values. After applying these filters on the four physical properties, one is left with only 71 sets of parameters (see Fig. 5a). The filtering order does not affect this selection. Each time a filter is applied, the uncertainty in the physical properties restricts the parameter sets of the SMA-TB. By applying all the filters, all the knowledge of the physical properties is "inferred" to the parameters of the SMA-TB.

From the marginal distribution of the parameters selected by the first iteration, we perform a second iteration as explained in section 3.2. The distributions of the parameters remaining after the filtering in the second iteration (447 302 points) are shown in Fig. 5b.

The range of the remaining parameters is for  $Ag$

$$\begin{cases} \xi \in [1,10,1,501] \\ A \in [0.081,0.195] \\ q \in [2.018,3.599] \\ p \in [8.015,12.385] \end{cases} \quad (15)$$

The range of  $A$  and  $\xi$  values is reduced by about a factor two (Eq. 13).

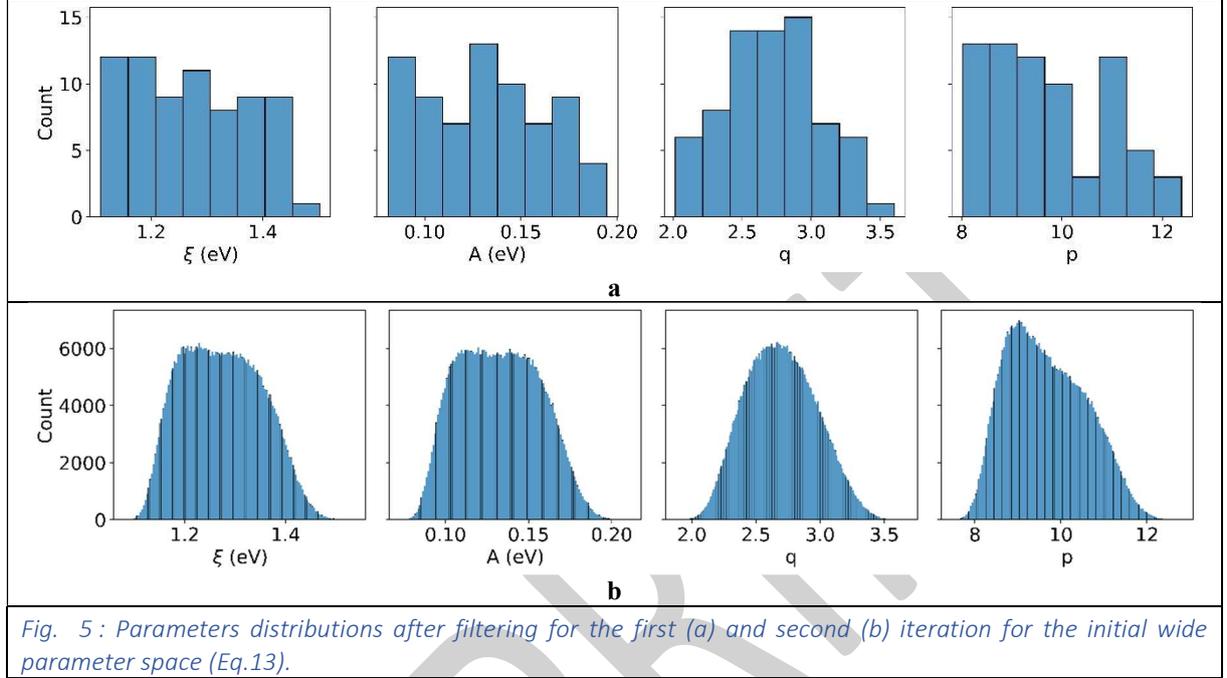


Fig. 5: Parameters distributions after filtering for the first (a) and second (b) iteration for the initial wide parameter space (Eq.13).

The Pearson correlation matrix of the SMA-TB parameters issued from the first iteration is

	$\xi$	$A$	$p$	$q$
$\xi$	1	0.99	-0.96	-0.48
$A$	0.99	1	-0.98	-0.57
$p$	-0.96	-0.98	1	0.68
$q$	-0.48	-0.57	0.68	1

(16)

The correlation matrices are approximately the same for both iterations. This matrix shows that there is a strong positive correlation between  $\xi$  and  $A$ , and negative between  $\xi$ ,  $A$  and  $p$ . The correlation with  $q$

is negative and moderate. All  $p$ -values are very low, meaning that the coefficients are statistically significant.

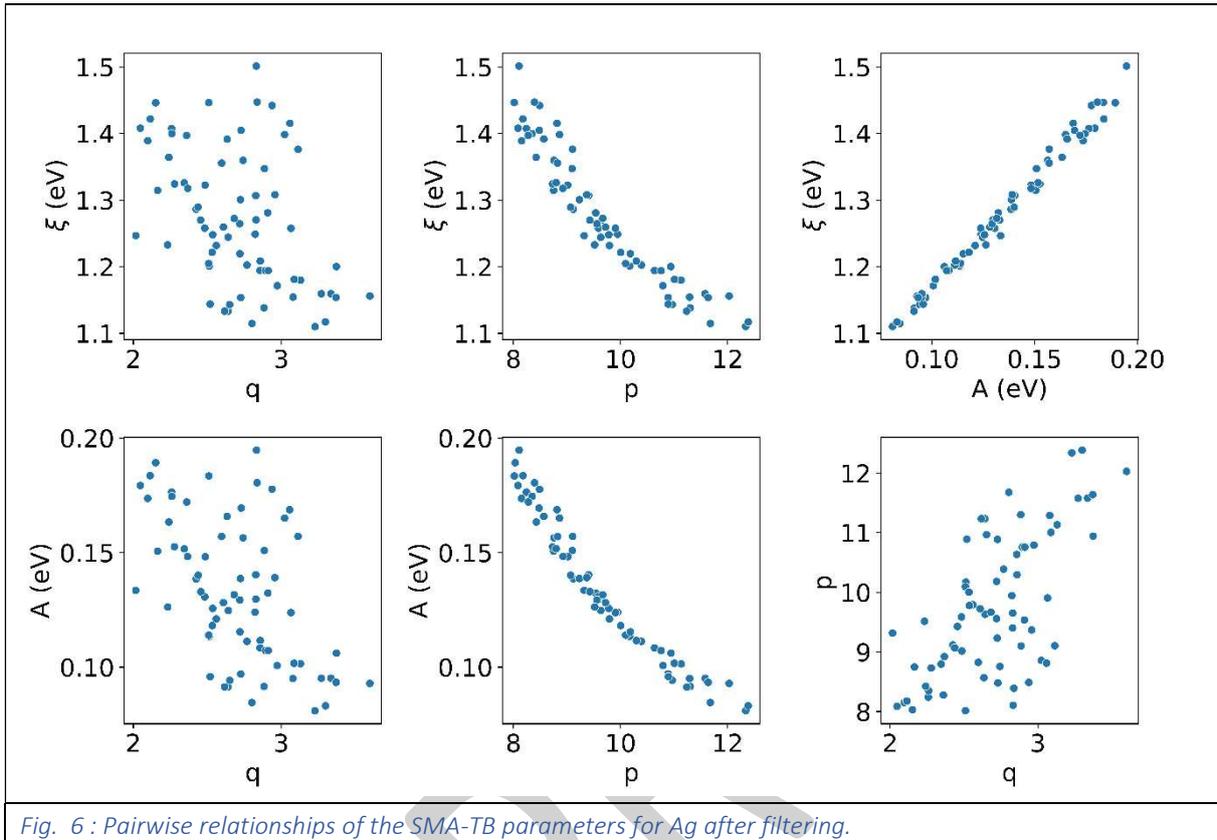


Fig. 6 : Pairwise relationships of the SMA-TB parameters for Ag after filtering.

The pairwise relationships for Ag with the remaining samples after filtering are shown in Fig. 6. For pairs of parameters with a high correlation coefficient, the points are not very scattered around the line. When the correlation coefficient is weak, the points are widely scattered. This figure highlights the strong correlation between  $A$  and  $\xi$  and between  $p$  and  $A$ ,  $\xi$ , and confirms Sobol' analysis that  $p$  is very important locally. One can see also that the relationship  $p \approx 3q$  is a very rough approximation.

Let us recall that the Ag physical properties compiled from the literature (see Table 2) are between  $-2.94$  and  $-2.96 eV$  for the cohesive energy;  $100.23$  and  $109.99 GPa$  for the Bulk modulus;  $14.37$  and  $19.05 GPa$  for  $c'$  and  $45.97$  and  $56.79 GPa$  for  $c_{44}$ . From Fig. 6 we can estimate the propagation of uncertainty from the physical properties through the SMA-TB parameters for Ag. We find that the mean of the  $\xi$  parameter is  $1.27 eV$  with a standard deviation of  $0.102 eV$ , or about 8% relative uncertainty (the minimum and maximum values are  $1.11$  and  $1.50 eV$ , respectively). For the  $A$  parameter, we find a mean of  $0.132 eV$  with a standard deviation of  $0.031 eV$ , *i.e.*, about 23% relative uncertainty (the minimum and maximum values are  $0.081$  and  $0.195 eV$ , respectively).  $\xi$  is better constrained than  $A$ .

### 4.3. Summary for the six transition metals.

We summarize in Table 3 the uncertainty obtained for SMA-TB parameters for each of the elements studied resulting from the second filtering iteration. Based on the uncertainty of physical properties of each of these transition metals, we narrowed down possible values of SMA-TB parameters sets. We compared our results with the different parameter sets from the literature [29]. Overall, their values are within our min-max interval for the remaining input parameter sets for all elements, which means that their parameters are within the uncertainty of the SMA-TB parameters that we found. We obtained

standard deviation that are at least 10 times lower than the average values obtained for each element and SMA-TB parameters. Also, when looking at statistics, we found that their parameters are within  $\pm 2\sigma$  for *Pt, Ni, Cu, Au* and within  $\pm 3\sigma$  for *Pd, Ag*.

<b>Table 3 : Statistics of the remaining input parameters after the 4 filters for each element after the first iteration starting from the wide parameters space (Eq. 13). Ref values are parameters values reported from Table 1.</b>							
Parameters	Stats	Ni	Cu	Pd	Ag	Pt	Au
	count	16	127	194	71	16	69
$\xi$	ref values	1.64	1.34	1.89	1.27	2.51	1.83
	mean	1.57	1.37	2.08	1.27	2.44	1.79
	std	0.07	0.10	0.19	0.10	0.12	0.18
	min	1.49	1.22	1.73	1.11	2.28	1.49
	max	1.72	1.63	2.44	1.50	2.66	2.12
<i>A</i>	ref values	0.12	0.11	0.14	0.12	0.24	0.21
	mean	0.10	0.12	0.20	0.13	0.22	0.20
	std	0.02	0.03	0.06	0.03	0.04	0.05
	min	0.08	0.08	0.10	0.08	0.17	0.12
	max	0.15	0.19	0.30	0.20	0.29	0.30
<i>q</i>	ref values	2.44	2.63	2.87	3.42	3.68	4.18
	mean	2.37	2.53	3.15	2.71	4.11	4.05
	std	0.20	0.35	0.49	0.35	0.28	0.31
	min	2.01	2.01	2.02	2.02	3.76	3.24
	max	2.68	3.26	4.20	3.60	4.60	4.66
<i>p</i>	ref values	10.76	10.38	11.75	10.35	11.14	10.42
	mean	11.78	9.90	10.57	9.76	12.26	10.90
	std	1.08	1.06	1.65	1.18	1.16	1.56
	min	9.72	8.09	8.11	8.02	10.52	8.53
	max	13.13	12.07	13.96	12.39	13.97	13.84

For each element, the correlation coefficients are approximately the same for  $\xi$  and *A*,  $\xi$  and *p*, *A* and *p* (see Table 4). In contrast, the correlations with *q* vary according the metal considered, they are weak for Ni and Cu, moderate for Pd and Ag, strong for Pt and Au.

<b>Table 4 : Pearson correlation coefficients issued from the local exploration for each transition metal.</b>						
Pearson coefficient	Ni	Cu	Pd	Ag	Pt	Au
$(\xi, A)$	0.99	0.99	0.99	0.99	1.0	1.0
$(\xi, p)$	-0.97	-0.91	-0.95	-0.95	-0.97	-0.98
$(\xi, q)$	-0.19	0.02	-0.35	-0.43	-0.80	-0.83
$(A, p)$	-0.99	-0.95	-0.97	-0.97	-0.97	-0.98
$(A, q)$	-0.33	-0.14	-0.45	-0.53	-0.82	-0.83
$(p, q)$	0.37	0.34	0.57	0.66	0.92	0.89

It is not useful to characterize precisely the nature of marginal distributions because we always restart the framework with a normal distribution obtained from the previous iteration. We use the Nataf transformation with the copula method [42], to include the correlation matrix derived from the parameter space resulting from the second iteration. Most of the samples generated in the third iteration are

validated by the filtering, which informs us that the use of Nataf copula transformation is a good approximation to generate a large amount of parameter sets.

As already mentioned, all the approximations considered (1NN, 2NN, 3NN) give similar results. For this reason, we do not present further figures for these approximations.

#### 4.4. Prediction procedure: application to the vacancy formation enthalpy

We generate samples from the correlated joint probability distribution with the reduced parameter space and the correlation matrix obtained at the end of the first iteration. We select 1000 samples after filtering of the second iteration to calculate the vacancy formation enthalpy. We use the Fire algorithm [68] which minimizes the potential energy at  $T = 0$  K and optimizes the interatomic distances around the vacancy. Results are shown in Fig. 7. For the cohesive energy, all predictions lie within the reference range by construction (see Fig. 7a).

We superimpose on the prediction distribution of the vacancy formation enthalpy (Fig. 7b) the reference range found in literature from *ab initio* calculation and the experimental value [69–71]. We find that on average the calculated values for Ag are too low, in agreement with [51]. However, a number of parameter sets give vacancy formation energies within the reference range.

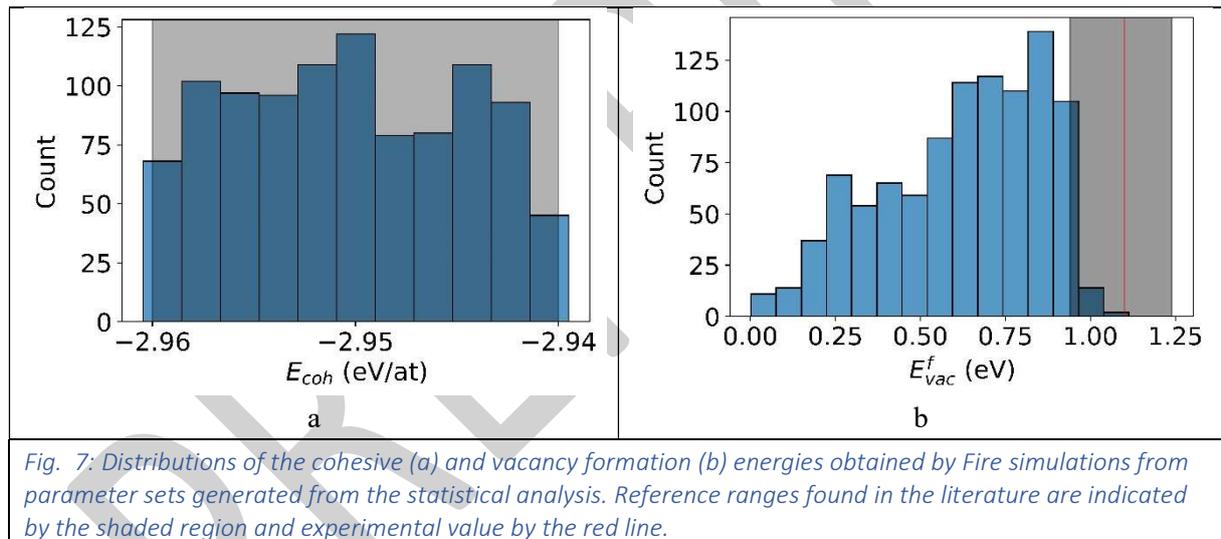


Fig. 7: Distributions of the cohesive (a) and vacancy formation (b) energies obtained by Fire simulations from parameter sets generated from the statistical analysis. Reference ranges found in the literature are indicated by the shaded region and experimental value by the red line.

It is then possible to make a new filtering on the vacancy formation energies to have a parameter space compatible with the reference data of  $(E_{coh}, B, c', c_{44}, E_{vac}^f)$ .

The parameter distributions derived from the filtering on the vacancy formation enthalpy, Fig. 8, shows that the width of the marginal distributions are smaller than the starting ones (Table 3). The mean values of the selected parameter sets are lower for  $\xi$  and  $A$ , higher for  $p$  and  $q$ . The correlation matrix changes only slightly.

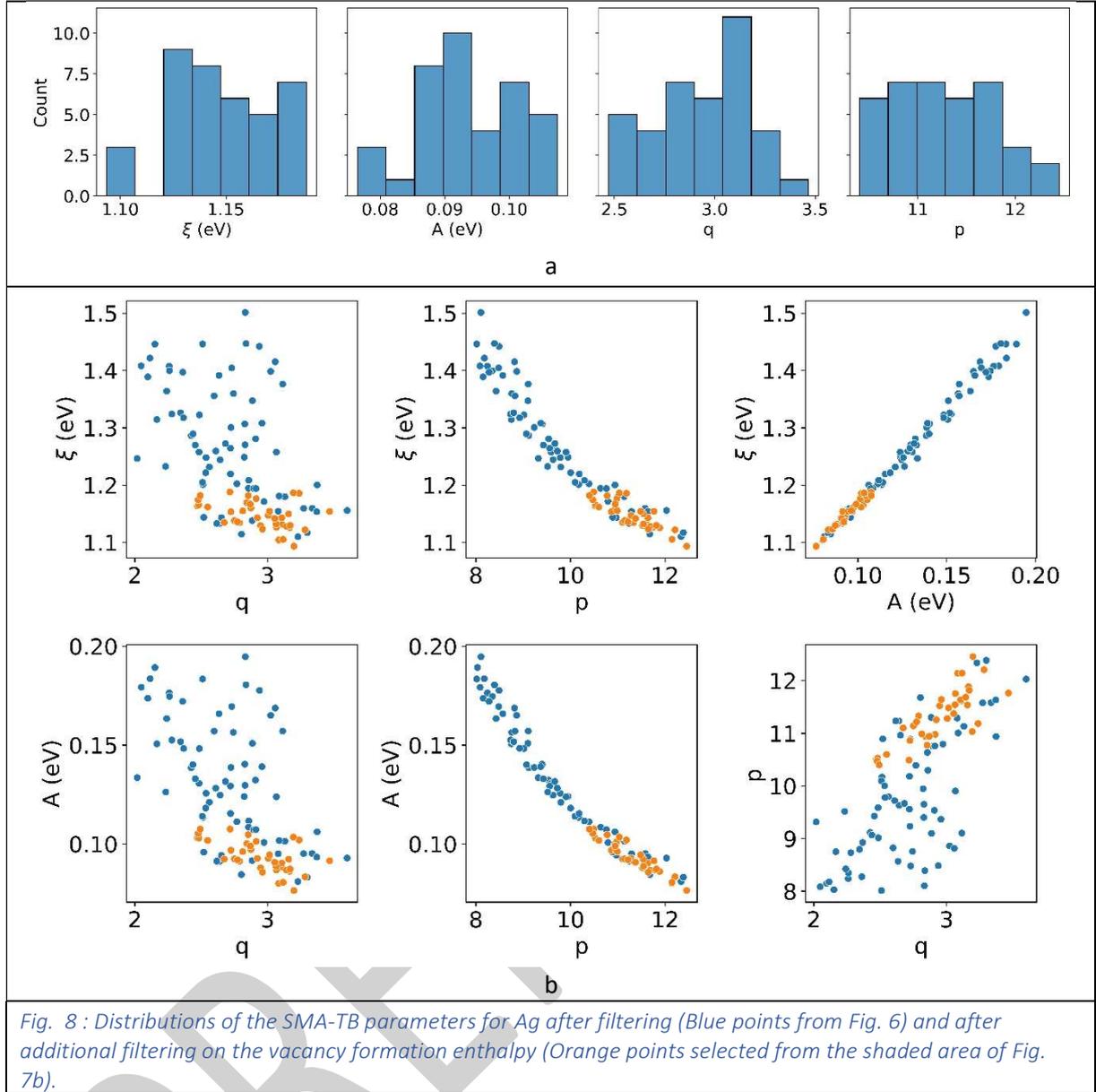


Fig. 8 : Distributions of the SMA-TB parameters for Ag after filtering (Blue points from Fig. 6) and after additional filtering on the vacancy formation enthalpy (Orange points selected from the shaded area of Fig. 7b).

We have shown that it is possible to find parameters that describe correctly both the bulk properties and the enthalpy of vacancy formation ( $E_{coh}, B, c', c_{44}, E_{vac}^f$ ). Regardless of the size of consistent parameters sets, they allow to recalibrate the parameter uncertainties. This procedure is incremental, it can be repeated as many times as necessary on many properties not included in the reference database (other elastic constants, interstitial formation energy, linear thermal expansion coefficient...). The calibration is refined at each step. Recalibrating the parameter space during the addition of a new physical property will inevitably reduce the parameter space. It can even remove complete formerly consistent subsets. This removal can be source of information from the model or the new property added.

With parameters estimated by a minimization method to best fit a set of experimental or ab initio computational data (without taking into account uncertainties), the potentials do not always reproduce exactly all the data. For example, the Pt and Pd parameters lead to a bad estimate of  $c_{44}$  (Table 1) while the other data are accurately reproduced. If the trade-off is considered unacceptable, it is difficult to know whether the disagreement stems from the estimation of the parameters on average values or from

a possible inadequacy of the potentials. In this work, since the potentials are calibrated, the proposed approach allows to reject the model when there are no parameter sets to account for an observable.

## 5. Conclusions

Uncertainty propagation and sensitivity analysis were performed on the application of SMA-TB interatomic potentials to predict bulk properties of metals. The sensitivity analysis provides a better understanding of the cause-and-effect relationship of interatomic potentials.

The sensitivity analysis shows how the SMA-TB parameters affect the cohesive energy and the elastic constants. The sensitivity of the SMA-TB parameters to physical properties depends on the domain of the chosen input parameters. First, a global analysis performed on the entire range of parameters for all metals (wide parameter space), shows that  $\xi$  is the most important parameter for the cohesive energy and  $A$  and  $p$  for the elastic constants. The sensitivity coefficients are identical for all the metals. Then, once the range of parameters is restricted for a specific metal (tight parameter space), the global sensitivity analysis shows that the most important parameters are still  $\xi$  and  $A$  for the cohesive energy but  $p$  and  $q$  for the elastic constants. In this case the sensitivity coefficients depend on the metal considered.

We defined a statistical framework to calibrate the SMA-TB potentials. This procedure is based on the generation of parameter sets, the analytical evaluation of physical properties, the characterization of the uncertainties of these physical properties from the literature and the selection of parameter sets leading to values of physical properties consistent with these reference ranges. Note that this method is akin to the Data Collaboration framework of Frenklach [45] and differs from the Bayesian setup by the absence of a weight function (the *likelihood*) on the selected parameters sets. This framework is aimed at being used as an iterative and discriminating tool, to select appropriate models and associated parameters based on the goal of the study. The prediction procedure credibility will be enhanced by an estimation of the uncertainty of QoIs. This is possible only with the help of the preliminary work done during the calibration phase, thanks to an estimation of parameters uncertainty and the study of correlations.

This approach allows us to estimate the parameters uncertainty. Statistical analysis of the selected parameter sets shows that  $A$ ,  $\xi$  and  $p$  are correlated. The characteristics of the parameter distributions are well defined, which allows us to create a correlated joint probability distribution from the reduced parameter space and the correlation matrix.

Finally, we have shown that the computation of properties out of the calibration set, such as the enthalpy of vacancy formation, can further help to reduce the dispersion of the parameters set.

The application of the developed statistical approach requires the ability to model the relationship between inputs and outputs. In this study, we have derived analytical expressions but numerical approximations can also be used (*e.g.*, ordinary or partial differential equations). It is also necessary to have *a priori* knowledge about the input parameters (from the literature or simplified models) to define the initial parameter space. Finally, the parameter set selection step requires sufficient data on the QoIs that we are modeling. To estimate the propagation of uncertainty from the outputs to the inputs, the uncertainty of the outputs must be known.

Finally, it would be worth applying the procedure described above to other classes of interatomic potentials. However, parallelization of the framework is highly recommended (maybe required) especially for models with more input parameters than the one presented here for a single element.

## Funding sources

This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (Labex charmmmat, ANR-11-LABX-0039-grant).

## Acknowledgments

The authors would like to thank C. Varvenne, B. Legrand, R. Tétot and J. Creuze for very fruitful discussions.

## Data Availability

The data needed to reproduce these results can be requested from [fabienne.berthier@universite-paris-saclay.fr](mailto:fabienne.berthier@universite-paris-saclay.fr)

PREPRINT

## 6. References

- [1] P. Acar, Recent progress of uncertainty quantification in small-scale materials science, *Prog. Mater. Sci.* 117 (2021). <https://doi.org/10.1016/j.pmatsci.2020.100723>.
- [2] A. Chernatynskiy, S.R. Phillpot, R. Lesar, Uncertainty quantification in multiscale simulation of materials: A prospective, *Annu. Rev. Mater. Res.* 43 (2013) 157–182. <https://doi.org/10.1146/annurev-matsci-071312-121708>.
- [3] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.* 1 (2013) 11002. <https://doi.org/10.1063/1.4812323>.
- [4] A. Vaitkus, A. Merkys, S. Gražulis, Validation of the Crystallography Open Database using the Crystallographic Information Framework, *Urn:Issn:1600-5767.* 54 (2021) 661–672. <https://doi.org/10.1107/S1600576720016532>.
- [5] K. Kanhaiya, S. Kim, W. Im, H. Heinz, Accurate simulation of surfaces and interfaces of ten FCC metals and steel using Lennard–Jones potentials, *Npj Comput. Mater.* 7 (2021) 1–15. <https://doi.org/10.1038/s41524-020-00478-1>.
- [6] X.D. Dai, Y. Kong, J.H. Li, B.X. Liu, Extended Finnis–Sinclair potential for bcc and fcc metals and alloys, *J. Phys. Condens. Matter.* 18 (2006) 4527–4542. <https://doi.org/10.1088/0953-8984/18/19/008>.
- [7] M.S. Daw, M.I. Baskes, Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals, *Phys. Rev. Lett.* 50 (1983) 1285–1288. <https://doi.org/10.1103/PhysRevLett.50.1285>.
- [8] M.S. Daw, M.I. Baskes, Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals, *Phys. Rev. B.* 29 (1984) 6443–6453. <https://doi.org/10.1103/PhysRevB.29.6443>.
- [9] P. Bacher, P. Wynblatt, S.M. Foiles, A Monte Carlo study of the structure and composition of (001) semicoherent interphase boundaries in CuAgAu alloys, *Acta Metall. Mater.* 39 (1991) 2681–2691. [https://doi.org/10.1016/0956-7151\(91\)90084-E](https://doi.org/10.1016/0956-7151(91)90084-E).
- [10] S.M. Foiles, J.B. Adams, Thermodynamic properties of fcc transition metals as calculated with the embedded-atom method, *Phys. Rev. B.* 40 (1989) 5909–5915. <https://doi.org/10.1103/PhysRevB.40.5909>.
- [11] F. Ducastelle, Modules élastiques des métaux de transition, *J. Phys.* 31 (1970) 1055–1062. <https://doi.org/10.1051/jphys:019700031011-120105500>.
- [12] G. Rossi, R. Ferrando, C. Mottet, Structure and chemical ordering in CoPt nanoalloys, *Faraday Discuss.* 138 (2008) 193–210. <https://doi.org/10.1039/B705415G>.
- [13] D. Cheng, S. Yuan, R. Ferrando, Structure, chemical ordering and thermal stability of Pt-Ni alloy nanoclusters, *J. Phys. Condens. Matter.* 25 (2013). <https://doi.org/10.1088/0953-8984/25/35/355008>.
- [14] S.A. Dokukin, S. V. Kolesnikov, A.M. Saletsky, A.L. Klavsyuk, Growth of the Pt/Cu(111) surface alloy: Self-learning kinetic Monte Carlo simulations, *J. Alloys Compd.* 763 (2018) 719–727. <https://doi.org/10.1016/j.jallcom.2018.05.335>.
- [15] J. Pirart, A. Front, D. Rapetti, C. Andreazza-Vignolle, P. Andreazza, C. Mottet, R. Ferrando, Reversed size-dependent stabilization of ordered nanophases, *Nat. Commun.* 10 (2019) 1–7. <https://doi.org/10.1038/s41467-019-09841-3>.

- [16] M.I. Baskes, Application of the Embedded-Atom Method to Covalent Materials: A Semiempirical Potential for Silicon, *Phys. Rev. Lett.* 59 (1987) 2666–2669. <https://doi.org/10.1103/PhysRevLett.59.2666>.
- [17] B. Jelinek, S. Groh, M.F. Horstemeyer, J. Houze, S.G. Kim, G.J. Wagner, A. Moitra, M.I. Baskes, Modified embedded atom method potential for Al, Si, Mg, Cu, and Fe alloys, *Phys. Rev. B - Condens. Matter Mater. Phys.* 85 (2012). <https://doi.org/10.1103/PhysRevB.85.245102>.
- [18] A.C.T. van Duin, S. Dasgupta, F. Lorant, W.A. Goddard Iii, W.A. Goddard, ReaxFF: A Reactive Force Field for Hydrocarbons, *J. Phys. Chem. A.* 105 (2001) 9396–9409. <https://doi.org/10.1021/jp004368u>.
- [19] M. Aryanpour, A.C.T. Van Duin, J.D. Kubicki, Development of a reactive force field for iron-oxyhydroxide systems, *J. Phys. Chem. A.* 114 (2010) 6298–6307. <https://doi.org/10.1021/jp101332k>.
- [20] K. Chenoweth, A.C.T. Van Duin, P. Persson, M.J. Cheng, J. Oxgaard, W.A. Goddard, Development and application of a ReaxFF reactive force field for oxidative dehydrogenation on vanadium oxide catalysts, *J. Phys. Chem. C.* 112 (2008) 14645–14654. <https://doi.org/10.1021/jp802134x>.
- [21] Y. Mishin, M.J. Mehl, D.A. Papaconstantopoulos, Phase stability in the Fe-Ni system: Investigation by first-principles calculations and atomistic simulations, *Acta Mater.* 53 (2005) 4029–4041. <https://doi.org/10.1016/j.actamat.2005.05.001>.
- [22] G.P. Purja Pun, K.A. Darling, L.J. Kecskes, Y. Mishin, Angular-dependent interatomic potential for the Cu-Ta system and its application to structural stability of nano-crystalline alloys, *Acta Mater.* 100 (2015) 377–391. <https://doi.org/10.1016/j.actamat.2015.08.052>.
- [23] G. Wang, Y. Xu, P. Qian, Y. Su, ADP potential for the Au-Rh system and its application in element segregation of nanoparticles, *Comput. Mater. Sci.* 186 (2021) 110002. <https://doi.org/10.1016/j.commatsci.2020.110002>.
- [24] F. Berthier, B. Legrand, G. Tréglia, How to compare superficial and intergranular segregation? A new analysis within the mixed SMA-TBIM approach, *Acta Mater.* 47 (1999) 2705–2715. [https://doi.org/10.1016/S1359-6454\(99\)00144-5](https://doi.org/10.1016/S1359-6454(99)00144-5).
- [25] F. Berthier, B. Legrand, G. Tréglia, New structures and atomistic analysis of the polymorphism for the  $\Sigma = 5$  (210) [001] tilt boundary, *Interface Sci.* 8 (2000) 55–69. <https://doi.org/10.1023/A:1008783220877>.
- [26] J. Creuze, F. Berthier, R. Tétot, B. Legrand, G. Tréglia, Intergranular segregation and vibrational effects: A local analysis, *Phys. Rev. B.* 61 (2000) 14470–14480. <https://doi.org/10.1103/PhysRevB.61.14470>.
- [27] F. Berthier, J. Creuze, R. Tétot, B. Legrand, Structural phase transition induced by interfacial segregation: a comparison between surface and grain boundary, *Appl. Surf. Sci.* 177 (2001) 243–251. [https://doi.org/10.1016/S0169-4332\(01\)00210-0](https://doi.org/10.1016/S0169-4332(01)00210-0).
- [28] R. Tétot, F. Berthier, J. Creuze, I. Meunier, G. Tréglia, B. Legrand, Cu-Ag (111) Polymorphism Induced by Segregation and Advacancies, *Phys. Rev. Lett.* 91 (2003) 176103. <https://doi.org/10.1103/PhysRevLett.91.176103>.
- [29] F. Berthier, J. Creuze, B. Legrand, Effective site-energy model: A thermodynamic approach applied to size-mismatched alloys, *Phys. Rev. B.* 95 (2017) 224102. <https://doi.org/10.1103/PhysRevB.95.224102>.
- [30] F. Berthier, J. Creuze, T. Gabard, B. Legrand, M.-C. Marinica, C. Mottet, Order-disorder or phase-separation transition: Analysis of the Au-Pd system by the effective site energy model,

- Phys. Rev. B. 99 (2019) 014108. <https://doi.org/10.1103/PhysRevB.99.014108>.
- [31] F. Berthier, J. Creuze, T. Gabard, B. Legrand, M.C. Marinica, C. Mottet, Order-disorder or phase-separation transition: Analysis of the Au-Pd system by the effective site energy model, *Phys. Rev. B.* 99 (2019) 1–11. <https://doi.org/10.1103/PhysRevB.99.014108>.
- [32] D.M. Yan Wang, *Uncertainty quantification in multiscale materials modeling*, Elsevier Series in Mechanics of Advanced Materials, n.d. <https://www.elsevier.com/books/title/author/9780081029428>.
- [33] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, *Reliab. Eng. Syst. Saf.* 52 (1996) 1–17. [https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/10.1016/0951-8320(96)00002-6).
- [34] R.L. Iman, J.C. Helton, An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models, *Risk Anal.* 8 (1988) 71–90. <https://doi.org/10.1111/j.1539-6924.1988.tb01155.x>.
- [35] A. Saltelli, T. Homma, Sensitivity analysis for model output, *Comput. Stat. Data Anal.* 13 (1992) 73–94. [https://doi.org/10.1016/0167-9473\(92\)90155-9](https://doi.org/10.1016/0167-9473(92)90155-9).
- [36] A. Saltelli, T.H. Andres, T. Homma, Sensitivity analysis of model output, *Comput. Stat. Data Anal.* 15 (1993) 211–238. [https://doi.org/10.1016/0167-9473\(93\)90193-W](https://doi.org/10.1016/0167-9473(93)90193-W).
- [37] A. Saltelli, T.H. Andres, T. Homma, Sensitivity analysis of model output. Performance of the iterated fractional factorial design method, *Comput. Stat. Data Anal.* 20 (1995) 387–407. [https://doi.org/10.1016/0167-9473\(95\)92843-M](https://doi.org/10.1016/0167-9473(95)92843-M).
- [38] A.P. Moore, C. Deo, M.I. Baskes, M.A. Okuniewski, D.L. McDowell, Understanding the uncertainty of interatomic potentials' parameters and formalism, *Comput. Mater. Sci.* 126 (2017) 308–320. <https://doi.org/10.1016/j.commatsci.2016.09.041>.
- [39] G. Dhaliwal, P.B. Nair, C.V. Singh, Uncertainty analysis and estimation of robust AIREBO parameters for graphene, *Carbon N. Y.* 142 (2019) 300–310. <https://doi.org/10.1016/J.CARBON.2018.10.020>.
- [40] A. Saltelli, S. Tarantola, K.P.-S. Chan, A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics.* 41 (1999) 39–56. <https://doi.org/10.1080/00401706.1999.10485594>.
- [41] A. Saltelli, P. Annoni, How to avoid a perfunctory sensitivity analysis, *Environ. Model. Softw.* 25 (2010) 1508–1517. <https://doi.org/10.1016/j.envsoft.2010.04.012>.
- [42] R. Lebrun, A. Dutfoy, A generalization of the Nataf transformation to distributions with elliptical copula, *Probabilistic Eng. Mech.* 24 (2009) 172–178. <https://doi.org/10.1016/j.probenmech.2008.05.001>.
- [43] R.B. Nelsen, *Copulas, Characterization, Correlation, and Counterexamples*, *Math. Mag.* 68 (1995) 193–198. <https://doi.org/10.1080/0025570X.1995.11996311>.
- [44] A. Tarantola, Popper, Bayes and the inverse problem, *Nat. Phys.* 2 (2006) 492–494. <https://doi.org/10.1038/nphys375>.
- [45] M. Frenklach, A. Packard, G. Garcia-Donato, R. Paulo, J. Sacks, Comparison of Statistical and Deterministic Frameworks of Uncertainty Quantification, *SIAM/ASA J. Uncertain. Quantif.* 4 (2016) 875–901. <https://doi.org/10.1137/15M1019131>.
- [46] M.C. Kennedy, A. O'hagan, *Bayesian calibration of computer models*, 2001.
- [47] J. Brynjarsdóttir, A. O'Hagan, Learning about physical parameters: the importance of model discrepancy, *Inverse Probl.* 30 (2014) 114007. <https://doi.org/10.1088/0266-5611/30/11/114007>.

- [48] P. Pernot, F. Cailliez, A critical review of statistical calibration/prediction models handling data inconsistency and model inadequacy, *AIChE J.* 63 (2017) 4642–4665. <https://doi.org/10.1002/aic.15781>.
- [49] K. Sargsyan, X. Huan, H.N. Najm, EMBEDDED MODEL ERROR REPRESENTATION FOR BAYESIAN MODEL CALIBRATION, *Int. J. Uncertain. Quantif.* 9 (2019) 365–394. <https://doi.org/10.1615/Int.J.UncertaintyQuantification.2019027384>.
- [50] F. Cailliez, P. Pernot, F. Rizzi, R. Jones, O. Knio, G. Arampatzis, P. Koumoutsakos, Bayesian calibration of force fields for molecular simulations, in: *Uncertain. Quantif. Multiscale Mater. Model.*, Elsevier, 2020: pp. 169–227. <https://doi.org/10.1016/B978-0-08-102941-1.00006-7>.
- [51] V. Rosato, M. Guillope, B. Legrand, Thermodynamical and structural properties of f.c.c. transition metals using a simple tight-binding model, *Philos. Mag. A.* 59 (1989) 321–336. <https://doi.org/10.1080/01418618908205062>.
- [52] I. Meunier, G. Tréglia, R. Tétot, J. Creuze, F. Berthier, B. Legrand, Misfit dislocation loops or incommensurate structure at an interface: Vibrational and anharmonic effects, *Phys. Rev. B.* 66 (2002) 125409. <https://doi.org/10.1103/PhysRevB.66.125409>.
- [53] J. Creuze, F. Berthier, R. Tétot, B. Legrand, Unexpected profiles of surface segregation vibrational entropies, *Surf. Sci.* 526 (2003) 121–132. [https://doi.org/10.1016/S0039-6028\(02\)02594-3](https://doi.org/10.1016/S0039-6028(02)02594-3).
- [54] E.B. Webb, G.S. Grest, D.R. Heine, J.J. Hoyt, Dissolutive wetting of Ag on Cu: A molecular dynamics simulation study, *Acta Mater.* 53 (2005) 3163–3177. <https://doi.org/10.1016/j.actamat.2005.03.021>.
- [55] I.M. Sobol, Sensitivity analysis for non-linear mathematical models, *Math. Model. Comput. Exp.* 1 (1993) 407–414.
- [56] I.M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.* 55 (2001) 271–280. [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6).
- [57] S. Kucherenko, S. Tarantola, P. Annoni, Estimation of global sensitivity indices for models with dependent variables, *Comput. Phys. Commun.* 183 (2012) 937–946. <https://doi.org/10.1016/j.cpc.2011.12.020>.
- [58] F.S.G. Richards, A Method of Maximum-Likelihood Estimation, *J. R. Stat. Soc. Ser. B.* 23 (1961) 469–475. <https://doi.org/10.1111/j.2517-6161.1961.tb00430.x>.
- [59] I.M. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys.* 7 (1967) 86–112. [https://doi.org/10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9).
- [60] J. Herman, W. Usher, SALib: An open-source Python library for Sensitivity Analysis, *J. Open Source Softw.* 2 (2017) 97. <https://doi.org/10.21105/joss.00097>.
- [61] J. Feinberg, H.P. Langtangen, Chaospy: An open source tool for designing methods of uncertainty quantification, *J. Comput. Sci.* 11 (2015) 46–57. <https://doi.org/10.1016/j.jocs.2015.08.008>.
- [62] J. Feinberg, V.G. Eck, H.P. Langtangen, Multivariate Polynomial Chaos Expansions with Dependent Variables, *SIAM J. Sci. Comput.* 40 (2018) A199–A223. <https://doi.org/10.1137/15M1020447>.
- [63] G. Simmons, H. Wang, Single Crystal Elastic Constants and Calculated Aggregate Properties. A Handbook, in: *Single Cryst. Elastic Constants Calc. Aggreg. Prop. A Handb.*, 1971.
- [64] C. Kittel, *Introduction to solid state physics*, 8th ed., John Wiley & Sons, New York, NY,

2004.

- [65] P. Janthon, S. Luo, S.M. Kozlov, F. Vin, J. Limtrakul, D.G. Truhlar, F. Illas, Bulk Properties of Transition Metals: A Challenge for the Design of Universal Density Functionals, (2014). <https://doi.org/10.1021/ct500532v>.
- [66] J. Hadamard, Sur les problèmes aux dérivées partielles et leur signification physique, *Princet. Univ. Bull.* (1902) 49–52.
- [67] S. Dobre, Analyses de sensibilité et d'identifiabilité globales. Application à l'estimation de paramètres photophysiques en thérapie photodynamique, Université Henri Poincaré - Nancy I, 2010.
- [68] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, P. Gumbsch, Structural relaxation made simple, *Phys. Rev. Lett.* 97 (2006) 1–4. <https://doi.org/10.1103/PhysRevLett.97.170201>.
- [69] W. Xing, P. Liu, X. Cheng, H. Niu, H. Ma, D. Li, Y. Li, X.-Q. Chen, Vacancy formation enthalpy of filled  $d$ -band noble metals by hybrid functionals, *Phys. Rev. B.* 90 (2014) 144105. <https://doi.org/10.1103/PhysRevB.90.144105>.
- [70] P.A. Korzhavyi, I.A. Abrikosov, B. Johansson, A. V. Ruban, H.L. Skriver, First-principles calculations of the vacancy formation energy in transition and noble metals, *Phys. Rev. B.* 59 (1999) 11693–11703. <https://doi.org/10.1103/PhysRevB.59.11693>.
- [71] Y.N. Starodubtsev, V.S. Tsepelev, K.M. Wu, Y.A. Kochetkova, N.P. Tsepeleva, Vacancy Formation Energy of Metals, *Key Eng. Mater.* 861 (2020) 46–51. <https://doi.org/10.4028/www.scientific.net/KEM.861.46>.

## Appendix A

### Polynomial function, $f$ and $g$ functions.

Recall the interaction energy with the polynomial fit function  $f$  and  $g$

$$E_i = -\xi \sqrt{Z_1 + Z_2 e^{-2q(d_2-1)} + Z_3 g(d_3 r_0)^2} + A (Z_1 + Z_2 e^{-p(d_2-1)} + Z_3 f(d_3 r_0)) \quad (A1)$$

The two functions  $f$  and  $g$  are polynomials of order 5

$$f(r_{ij}) = a_5 r_{ij}^5 + a_4 r_{ij}^4 + a_3 r_{ij}^3 + a_2 r_{ij}^2 + a_1 r_{ij} + a_0 \quad (A2)$$

Assuming that interactions, forces and curvatures are zero at distance  $4NN$

$$f(r_{4NN}) = 0; \left(\frac{\partial f}{\partial r}\right)_{r=r_{4NN}} = 0; \left(\frac{\partial^2 f}{\partial r^2}\right)_{r=r_{4NN}} = 0 \quad (A3)$$

$f$  is reduced to

$$f(r_{ij}) = a_5 (r_{ij} - r_{4NN})^5 + a_4 (r_{ij} - r_{4NN})^4 + a_3 (r_{ij} - r_{4NN})^3 \quad (A4)$$

The continuity at distance  $2NN$  of interactions, forces and curvatures is written

$$\hat{f}(r_{2NN}) = \hat{f} = e^{-p\left(\frac{r_{2NN}}{r_{1NN}}-1\right)}; \quad (A5)$$

$$\left(\frac{\partial \hat{f}}{\partial r}\right)_{r=r_{2NN}} = \hat{f}' = \left(-\frac{p}{r_{1NN}}\right) e^{-p\left(\frac{r_{2NN}}{r_{1NN}}-1\right)}; \quad (A6)$$

$$\left(\frac{\partial^2 \hat{f}}{\partial r^2}\right)_{r=r_{2NN}} = \hat{f}'' = \left(-\frac{p}{r_{1NN}}\right)^2 e^{-p\left(\frac{r_{2NN}}{r_{1NN}}-1\right)} \quad (A7)$$

which leads to the following system of equations

$$\begin{cases} \hat{f} = f(r_{2NN}) \\ \hat{f}' = \left(\frac{\partial f}{\partial r}\right)_{r=r_{2NN}} \\ \hat{f}'' = \left(\frac{\partial^2 f}{\partial r^2}\right)_{r=r_{2NN}} \end{cases} \Rightarrow \begin{cases} \hat{f} = a_5 (r_{2NN} - r_{4NN})^5 + a_4 (r_{2NN} - r_{4NN})^4 + a_3 (r_{2NN} - r_{4NN})^3 \\ \hat{f}' = 5a_5 (r_{2NN} - r_{4NN})^4 + 4a_4 (r_{2NN} - r_{4NN})^3 + 3a_3 (r_{2NN} - r_{4NN})^2 \\ \hat{f}'' = 20a_5 (r_{2NN} - r_{4NN})^3 + 12a_4 (r_{2NN} - r_{4NN})^2 + 6a_3 (r_{2NN} - r_{4NN})^1 \end{cases} \quad (A8)$$

Let  $x = (r_{2NN} - r_{4NN}) = r_0(\sqrt{2} - 2)$

$$\Rightarrow \begin{cases} \hat{f} = a_5 x^5 + a_4 x^4 + a_3 x^3 \\ \hat{f}' = 5a_5 x^4 + 4a_4 x^3 + 3a_3 x^2 \\ \hat{f}'' = 20a_5 x^3 + 12a_4 x^2 + 6a_3 x \end{cases} \quad (A9)$$

Knowing  $x; \hat{f}; \hat{f}'$  et  $\hat{f}''$  the coefficients  $a_5, a_4$  et  $a_3$  are given by

$$\Rightarrow \begin{cases} 2a_3 = \frac{\hat{f}''}{x} - 8\frac{\hat{f}'}{x^2} + 20\frac{\hat{f}}{x^3} \\ -a_4x = \frac{\hat{f}''}{x} - 7\frac{\hat{f}'}{x^2} + 15\frac{\hat{f}}{x^3} \\ 2a_5x^2 = \frac{\hat{f}''}{x} - 6\frac{\hat{f}'}{x^2} + 12\frac{\hat{f}}{x^3} \end{cases} \quad (\text{A10})$$

The attractive contribution has an equivalent form, it is obtained by replacing  $\hat{f}$  by  $\hat{g} = e^{-q\left(\frac{r_{2NN}}{r_{1NN}} - 1\right)}$ .

PREPRINT

## Appendix B

### Small deformation theory derivation of analytic expressions for elastic constants.

Within the small deformation theory, the total energy of a solid is given, using Voigt notation, by

$$E_{tot}(V_0, \{\varepsilon_k\}) = E_{tot}(V_0, \{0\}) + V_0 \left( \sum_{i=1}^6 \sigma_i \varepsilon_i + \frac{1}{2} \sum_{i,j=1}^6 c_{ij} \varepsilon_i \varepsilon_j \right) \quad (B1)$$

where  $E_{tot}(V_0, \{0\})$ ,  $V_0$ ,  $\sigma_i$  and  $\varepsilon_i$  are respectively the total energy, the volume in absence of deformations, the components of the stress and strain tensors;  $c_{ij}$  are the elastic constants. Remark, in Voigt notation,  $\varepsilon_1 = \varepsilon_{xx}, \varepsilon_2 = \varepsilon_{yy}, \varepsilon_3 = \varepsilon_{zz}, \varepsilon_4 = 2\varepsilon_{yz}, \varepsilon_5 = 2\varepsilon_{xz}$  and  $\varepsilon_6 = 2\varepsilon_{xy}$ . The elastic constants  $c_{ij}$  are obtained by

$$c_{ij} = \frac{1}{V_0} \frac{\partial^2 E_{tot}(V_0, \{\varepsilon_k\})}{\partial \varepsilon_i \partial \varepsilon_j} \Big|_{\{\varepsilon_k=0\}} = \frac{1}{V_0} \frac{\partial^2 E_{elastic}(V_0, \{\varepsilon_k\})}{\partial \varepsilon_i \partial \varepsilon_j} \Big|_{\{\varepsilon_k=0\}} \quad (B2)$$

with  $E_{elastic} = \frac{V_0}{2} \sum_{i,j=1}^6 c_{ij} \varepsilon_i \varepsilon_j$  the elastic energy.

Given the cubic symmetry, the elastic energy can be simplified

$$E_{elastic} = \frac{V_0}{2} \left( c_{11}(\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2) + 2c_{12}(\varepsilon_1 \varepsilon_2 + \varepsilon_1 \varepsilon_3 + \varepsilon_2 \varepsilon_3) + c_{44}(\varepsilon_4^2 + \varepsilon_5^2 + \varepsilon_6^2) \right) \quad (B3)$$

The bulk modulus is the measure of resistance to isotropic deformation, such as  $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_B$  and  $\varepsilon_4 = \varepsilon_5 = \varepsilon_6 = 0$ , the elastic energy associated to the bulk modulus is then

$$E_{elastic}^B = 3 \frac{V_0}{2} (c_{11} + 2c_{12}) \varepsilon_B^2 \quad (B4)$$

By convention the bulk modulus  $B = \frac{1}{3}(c_{11} + 2c_{12})$ , such that

$$E_{elastic}^B = 9 \frac{V_0}{2} B \varepsilon_B^2 \Rightarrow B = \frac{1}{9V_0} \frac{\partial^2 E_{elastic}^B(V_0, \{\varepsilon_k\})}{\partial \varepsilon_B^2} \Big|_{\{\varepsilon_k=0\}} \quad (B5)$$

The shear modulus  $c'$  is written by convention  $c' = \frac{1}{2}(c_{11} - c_{12})$ . It derives from the following deformation  $\varepsilon_1 = -\varepsilon_2 = \varepsilon_{c'}$ , the remaining components  $\varepsilon_k$  being null.

$$E_{elastic}^{c'} = 2 \frac{V_0}{2} (c_{11} - c_{12}) \varepsilon_{c'}^2 \quad (B6)$$

$$E_{elastic}^{c'} = 4 \frac{V_0}{2} c' \varepsilon_{c'}^2 \Rightarrow c' = \frac{1}{4V_0} \frac{\partial^2 E_{elastic}^{c'}(V_0, \{\varepsilon_k\})}{\partial \varepsilon_{c'}^2} \Big|_{\{\varepsilon_k=0\}} \quad (B7)$$

The deformation  $\varepsilon_{12} = \varepsilon_{c_{44}}$  then  $\varepsilon_4 = 2\varepsilon_{12} = 2\varepsilon_{c_{44}}$  leads to the shear modulus  $c_{44}$

$$E_{elastic}^{c_{44}} = \frac{V_0}{2} c_{44} (2\varepsilon_{c_{44}})^2 \quad (B8)$$

$$E_{elastic}^{c_{44}} = 4 \frac{V_0}{2} c_{44} \varepsilon_{c_{44}}^2 \Rightarrow c_{44} = \frac{1}{4V_0} \left. \frac{\partial^2 E_{elastic}^{c_{44}}(V_0, \{\varepsilon_k\})}{\partial \varepsilon_{c_{44}}^2} \right|_{\{\varepsilon_k=0\}} \quad (B9)$$

Note that since these deformations ( $\varepsilon_B, \varepsilon_{c'}, \varepsilon_{c_{44}}$ ) are not equivalent, it is essential to rewrite these expressions by introducing the deformed distance  $r'_{ij}$  with respect to the undeformed distance  $r_{ij}$  in Cartesian coordinate

$$\left\{ \begin{array}{l} r'_{ij} = \sqrt{x_{ij}^2(1 + \varepsilon_B)^2 + y_{ij}^2(1 + \varepsilon_B)^2 + z_{ij}^2(1 + \varepsilon_B)^2} = r_{ij}(1 + \varepsilon_B) \\ r'_{ij} = \sqrt{x_{ij}^2(1 + \varepsilon_{c'})^2 + y_{ij}^2(1 - \varepsilon_{c'})^2 + z_{ij}^2} \\ r'_{ij} = \sqrt{(x_{ij} + y_{ij}\varepsilon_{c_{44}})^2 + (y_{ij} + x_{ij}\varepsilon_{c_{44}})^2 + z_{ij}^2} \end{array} \right. \quad (B10)$$

Note: for (B10) relations,  $B$  is the standard bulk modulus.  $c'$  is most easily obtained by straining the crystal in the (100) direction while simultaneously compressing it in the (010) direction to conserve the volume, with lengths in the (001) direction remaining fixed.  $c_{44}$  is by straining the crystal in the (110) direction and fixing the volume by compressing in the (110) direction.

The total energy of a metal is written as the sum of the interaction energies and the elastic constant can be evaluated for a single atom as long as the total volume considered is reduced to the volume of the primitive cell. The elastic constant then simplifies to

$$E_{elastic} \Rightarrow N_{at} E_i; V_0 \Rightarrow N_{at} \left( \frac{V_{fcc}}{4} \right) \quad (B11)$$

As an example, the bulk modulus expression simplifies to

$$B = \frac{1}{9V_0} \left. \frac{\partial^2 E_{elastic}^B(V_0, \{\varepsilon_k\})}{\partial \varepsilon_B^2} \right|_{\{\varepsilon_k=0\}} = \frac{1}{9 \left( \frac{V_{fcc}}{4} \right)} \left. \frac{\partial^2 E_i^B(V_0, \{\varepsilon_k\})}{\partial \varepsilon_B^2} \right|_{\{\varepsilon_k=0\}} \quad (B12)$$

## Appendix C

### Elastic constants formulas

Elastic constants expressions with respect to the interatomic potential parameters for fcc metals are

$$B = \frac{1}{9 \left( \frac{V_{maille}^{fcc}}{4} \right)} \left[ -\xi \frac{1}{2} \left( -\frac{1}{2} \frac{(\sum_k^l Z_k d_k (-2q) e^{-2q(d_k-1)})^2}{(\sum_k^l Z_k e^{-2q(d_k-1)})^{\frac{3}{2}}} + \frac{\sum_k^l Z_k d_k^2 (-2q)^2 e^{-2q(d_k-1)}}{\sqrt{\sum_k^l Z_k e^{-2q(d_k-1)}}} \right) + A \left( \sum_k^l Z_k d_k^2 (-p)^2 e^{-p(d_k-1)} \right) \right] \quad (C1)$$

$$c' = \frac{1}{4 \left( \frac{V_{maille}^{fcc}}{4} \right)} \left[ -\xi \frac{1}{2} \left( \frac{\sum_k^3 ((-2q)^2 e^{-2q(d_k-1)} (v_k Z_k d_k^2) + (-2q) e^{-2q(d_k-1)} (w_k - v_k) Z_k d_k)}{\sqrt{\sum_k^3 Z_k e^{-2q(d_k-1)}}} \right) + A \left( \sum_k^3 (-p)^2 e^{-p(d_k-1)} (v_k Z_k d_k^2) + (-p) e^{-p(d_k-1)} (w_k - v_k) Z_k d_k \right) \right] \quad (C2)$$

$$c_{44} = \frac{1}{4 \left( \frac{V_{maille}^{fcc}}{4} \right)} \left[ -\xi \frac{1}{2} \left( \frac{\sum_k^3 ((-2q)^2 e^{-2q(d_k-1)} (4u_k Z_k d_k^2) + (-2q) e^{-2q(d_k-1)} (w_k - 4u_k) Z_k d_k)}{\sqrt{\sum_k^3 Z_k e^{-2q(d_k-1)}}} \right) + A \left( \sum_k^3 (-p)^2 e^{-p(d_k-1)} (4u_k Z_k d_k^2) + (-p) e^{-p(d_k-1)} (w_k - 4u_k) Z_k d_k \right) \right] \quad (C3)$$

Where  $u_k = \frac{1}{12}(1,0,1)$ ;  $v_k = \frac{1}{6}(1,4,1)$  and  $w_k = \frac{2}{3}(1,1,1)$  are due to symmetry of the fcc crystal.

## Appendix D

### Extension of the 1NN approximation

#### 1. Equilibrium condition

We first consider the equilibrium condition

$$\begin{aligned} \left. \frac{\partial E_l}{\partial \varepsilon} \right|_{\varepsilon=0} = 0 &\xrightarrow{l^{th} NN} -\xi \sqrt{Z_1} (-2q) \left( \frac{\sum_k^l \left( \frac{Z_k}{Z_1} (d_k) e^{-2q(d_k-1)} \right)}{2 \sqrt{\sum_k^l \left( \frac{Z_k}{Z_1} e^{-2q(d_k-1)} \right)}} \right) + Z_1 A (-p) \sum_k^l \left( \frac{Z_k}{Z_1} d_k e^{-p(d_k-1)} \right) = 0 \\ &\xrightarrow{l^{th} NN} \xi \sqrt{Z_1} q \left( \frac{\sum_k^l \left( \frac{Z_k}{Z_1} (d_k) e^{-2q(d_k-1)} \right)}{\sqrt{\sum_k^l \left( \frac{Z_k}{Z_1} e^{-2q(d_k-1)} \right)}} \right) = Z_1 A p \sum_k^l \left( \frac{Z_k}{Z_1} d_k e^{-p(d_k-1)} \right) \end{aligned} \quad (D1)$$

Sums on both side of the equation are only dependant on structures/symmetries of the crystal and on parameters  $q$  and  $p$  for the attractive and repulsive energies contributions, respectively. Let's us named these sums as

$$Q_{struc}^l(q) = \frac{\sum_k^l \left( \frac{Z_k}{Z_1} (d_k) e^{-2q(d_k-1)} \right)}{\sqrt{\sum_k^l \left( \frac{Z_k}{Z_1} e^{-2q(d_k-1)} \right)}} ; P_{struc}^l(p) = \sum_k^l \left( \frac{Z_k}{Z_1} d_k e^{-p(d_k-1)} \right) \quad (D2)$$

Such that the equilibrium condition for the  $l^{th}$  approximation is

$$\xi \sqrt{Z_1} q Q_{struc}^l(q) = Z_1 A p P_{struc}^l(p) \quad (D3)$$

We can now generalize the definition of  $x$  given by Ducastelle as  $x_l = q Q_{struc}^l(q) / p P_{struc}^l(p)$ , which leads to

$$\frac{Z_1 A}{\xi \sqrt{Z_1}} = x_l \quad (D4)$$

Note that  $x_1 = x = \frac{q}{p}$  since  $Q_{struc}^1(Z_1, d_1, q) = P_{struc}^1(Z_1, d_1, p) = 1$ .

#### 2. Cohesive energy

The cohesive energy is:

$$E_{co}^l = A Z_1 \beta_l^{E_{co}}(p) - \xi \sqrt{Z_1} \alpha_l^{E_{coh}}(q) \quad (D5)$$

with  $\alpha_l^{E_{coh}}(q) = \sqrt{\sum_k^l \frac{Z_k}{Z_1} e^{-2q(d_k-1)}}$  and  $\beta_l^{E_{co}}(p) = \sum_k^l \frac{Z_k}{Z_1} e^{-p(d_k-1)}$ .

Remark that  $E_{coh}$  depends only on  $\xi$  and  $A$  for  $l = 1$ , but depends on all four SMA-TB parameters for  $l > 1$ .

And it can be factorized in the same way as in the 1NN approximation, as

$$E_{coh}^l = -\xi\sqrt{Z_1} \left( \alpha_l^{E_{coh}}(q) - x_l \beta_l^{E_{coh}}(p) \right) = -AZ_1 \left( \frac{\alpha_l^{E_{co}}(q) - x_l \beta_l^{E_{co}}(p)}{x_l} \right) \quad (D6)$$

### 3. Bulk modulus

Expression of the bulk modulus  $B$  based on the  $l$ th approximation is

$$B^l = \frac{-\xi\sqrt{Z_1}\alpha_l^B(q) + AZ_1\beta_l^B(p)}{V_B^{eff}(r_0)} \quad (D7)$$

where  $\alpha_l^B(q) = q^2 \left( \frac{\left( \sum_k^l \frac{Z_k}{Z_1} d_k e^{-2q(d_k-1)} \right)^2}{\left( \sum_k^l \frac{Z_k}{Z_1} e^{-2q(d_k-1)} \right)^{\frac{3}{2}}} - 2 \frac{\sum_k^l \frac{Z_k}{Z_1} d_k^2 e^{-2q(d_k-1)}}{\sqrt{\sum_k^l \frac{Z_k}{Z_1} e^{-2q(d_k-1)}}} \right)$ ,  $\beta_l^B(p) = p^2 \left( \sum_k^l \frac{Z_k}{Z_1} d_k^2 e^{-p(d_k-1)} \right)$  are

functions of the geometry of the crystal and of parameters  $q, p$  and of  $V_B^{eff}(r_0) = 9V_{lat}(r_0)$  which is related to the volume.

Relationship between elastic constants and the cohesive energy can still be made, and we find for the bulk modulus

$$B^l = S_B^l(p, q) \frac{E_{coh}^l}{V_B^{eff}(r_0)} \quad (D8)$$

with  $S_B^l(p, q) = \frac{x_l \beta_l^B - \alpha_l^B}{\alpha_l^{E_{co}} - x_l \beta_l^{E_{co}}}$  which depends only on the parameters  $p$  and  $q$ , the structure and symmetry of the crystal.

### 4. Shear moduli

Each elastic constants  $c_{ij}$  can be written as the previous expression

$$c_{ij} = S_{c_{ij}}^l(p, q) \frac{E_{co}^l}{V_{c_{ij}}^{eff}(r_0)} \quad (D9)$$

In particular, the two shear moduli are given by

$$c'_l = S_{c'_l}^l(p, q) \frac{E_{co}^l}{V_{c'_l}^{eff}(r_0)} \quad (D10)$$

$$c_{44}^l = S_{c_{44}^l}^l(p, q) \frac{E_{co}^l}{V_{c_{44}^l}^{eff}(r_0)} \quad (D11)$$

## Appendix E

### Physical expression for the SMA-TB potential connected with a polynomial.

Finally, the three elastic constants within the SMATB potential connected with a fifth-order polynomial between the  $2^{nd}$  nearest neighbors and the  $4^{th}$  nearest neighbors have the following expressions

$$B = \frac{1}{9 \left( \frac{V_{fcc}}{4} \right)} \left[ \begin{array}{l} \frac{\xi \left( (-2q)Z_1 + (-2q)Z_2 d_2 e^{-2q(d_2-1)} + Z_3 2gg' \right)^2}{4 \left( Z_1 + Z_2 e^{-2q(d_2-1)} + Z_3 g^2 \right)^{\frac{3}{2}}} \\ - \frac{\xi \left( (-2q)^2 Z_1 + (-2q)^2 Z_2 d_2^2 e^{-2q(d_2-1)} + Z_3 \left( 2(g')^2 + 2gg'' \right) \right)}{2 \left( Z_1 + Z_2 e^{-2q(d_2-1)} + Z_3 g^2 \right)^{\frac{1}{2}}} \\ + A \left( (-p)^2 Z_1 + (-p)^2 Z_2 d_2^2 e^{-p(d_2-1)} + Z_3 f'' \right) \end{array} \right] \quad (E1)$$

$$c' = \frac{1}{4 \left( \frac{V_{fcc}}{4} \right)} \left( \frac{1}{6} \right) \left[ \begin{array}{l} -\xi \frac{1}{2} \left( \frac{\left( (2qZ_1)[2q-3] + (2q)^2 e^{-2q(d_2-1)} (4Z_2 d_2^2) \right)}{\sqrt{Z_1 + Z_2 e^{-2q(d_2-1)} + Z_3 g^2}} \right. \\ \left. + \frac{2Z_3 \left[ (g')^2 + gg'' + 3(gg') \right]}{\sqrt{Z_1 + Z_2 e^{-2q(d_2-1)} + Z_3 g^2}} \right) \\ + A \left( (pZ_1)[p-3] + p^2 e^{-p(d_2-1)} (4Z_2 d_2^2) \right) \\ \left. + Z_3 [f'' + 3f'] \right] \quad (E2)$$

$$c_{44} = \frac{1}{4 \left( \frac{V_{fcc}}{4} \right)} \left( \frac{1}{3} \right) \left[ \begin{array}{l} -\xi \frac{1}{2} \left( \frac{\left( (2qZ_1)[2q-1] + (-2q) e^{-2q(d_2-1)} (2Z_2 d_2) \right)}{\sqrt{Z_1 + Z_2 e^{-2q(d_2-1)} + Z_3 g^2}} \right) \\ \left. + \frac{2Z_3 \left( (gg' + ((g')^2 + gg'')) \right)}{\sqrt{Z_1 + Z_2 e^{-2q(d_2-1)} + Z_3 g^2}} \right) \\ + A \left( Z_1 p [p-1] - p e^{-p(d_2-1)} (2) Z_2 d_2 \right) \\ \left. + Z_3 [f' + f''] \right] \quad (E3)$$