

# Monovacancy Properties From Atomistic Simulations Based on OpenKIM

Junhao Li,<sup>\*</sup> Matt Bierbaum, and James P. Sethna<sup>†</sup>

*Department of Physics and Laboratory of Atomic and Solid State Physics  
Cornell University, Ithaca, NY 14853, USA*

Noam Bernstein

*Center for Computational Materials Science  
Naval Research Laboratory,  
Washington, DC 20375, USA*

Ryan Elliott and Ellad Tadmor

*Department of Aerospace Engineering and Mechanics  
University of Minnesota, Minneapolis, MN 55455, USA*

(Dated: April 27, 2016)

In this paper, we describe how we obtain the monovacancy properties for all the elements and all the simple crystal structures predicted by the hundreds of interatomic potential models on OpenKIM, the knowledgebase of interatomic models. The properties we calculated include the vacancy formation energy, migration energy, relaxation volume, elastic dipole tensor, defect strain tensor, and saddle point elastic dipole and defect strain tensors. The structures we calculate include sc, fcc, bcc, diamond, and hcp. These results can provide use information for selecting interatomic models. We also examine how these properties depend on each other and other elemental properties.

## I. INTRODUCTION

A longtime goal of scientists' is to be able to calculate the properties of materials from their structures accurately and efficiently. There are several approaches to this problem, including the density functional theory (DFT) and the atomistic simulation. DFT has developed rapidly in the recent few decades and achieved decent accuracy for many small-scale systems. However, in many cases, the size, the complexity, or the timescale we want to simulate, demands a much more efficient method. And atomistic simulation with good interatomic potentials is one of the methods that can fulfill this demand.

Several different types of interatomic models have been developed over the past few decades, from pairwise models like Lennard-Jones, to those based on cluster functionals such as the embedded-atom method (EAM) [1, 2]. One problem that has hindered the development of atomistic simulation is that the models developed by various researchers are usually presented in different formats, making them difficult to reuse. The OpenKIM project [3, 4] solves this problem by providing a unified framework for model developers to submit, and for model users to perform atomistic simulations [5]. It also hosts a repository of reference data from other sources, such as DFT and experiments, for assessing the quality of these models.

In this study, we focus on vacancies, the simplest and the most common point defects. Finding interatomic models that can accurately describe vacancy effects is extremely beneficial for furthering our understanding and

application of materials. Vacancies mediate dislocation climb and creep [6] and the diffusion of substitutional defects [7]. They are also the dominant flow mechanism under external fields, such as gravity [8]. And in metrology, some scientists propose to use  $^{28}\text{Si}$  for a new definition of kilogram, and obtaining the equilibrium vacancy concentration is a crucial part of this work [9].

Defects in crystals not only have a free energy, describing the cost of formation, but they also have an environmental impact through the generation of long-range fields. Cracks have mode I, mode II, and mode III stress and strain fields; dislocations have strain fields determined by their orientation and Burger's vector, and so on. The long-range electrical impact of a point defect is described at lowest order by its charge, and to higher order by a charge multipole series; the long-range magnetic impact is partly the direct dipolar interaction between local moments and mostly the RKKY-type interactions mediated by conduction electrons. Neither of these are part of the current KIM potentials, although there are plans to incorporate charge states of atomic species in the next release. The long-range elastic environmental impact of our vacancies will thus be described at lowest order by the *elastic dipole* field – a strain field induced by the vacancy that dies  $1/R^3$  like the electrical dipole field.<sup>1</sup> We thus calculate these elastic dipole fields,

---

<sup>1</sup> There is a nomenclature problem in the field. The anisotropic part of the  $1/R^3$  elastic field around a defect in an isotropic medium has the  $L = 2$  symmetry of an electrical quadrupole, and is often described as the elastic quadrupole field. But calling the isotropic part a monopole field would be strange since it also decays as  $1/R^3$ . We believe neither choice is very satisfactory, but the alternative of naming it after Eshelby seems strange as well.

---

<sup>\*</sup> jl2922@cornell.edu

<sup>†</sup> sethna@lassp.cornell.edu

as well as the vacancy creation energy.

Each defect will have an environmental dependence as well. Vacancies in electric fields have a direct Coulomb force in insulators, and an important electromigration forces from the ‘electron wind’ in metals (which may be in the opposite direction from the electrical current); neither is accessible in the current KIM potentials. The pressure dependence of the vacancy creation energy, and the energy change due to external stresses, will be given by its elastic dipole field (discussed above). Our vacancy will diffuse due to pressure gradients and concentration gradients, with a rate given by the free energy barrier to motion (which we also calculate). In turn, the pressure or stress dependence of the free energy will be determined by the difference between the vacancy’s elastic dipole in the ground state and at the saddlepoint for the motion, which we also calculate. The entropic effects of vibrations at the ground state and saddlepoint, needed for the complete transition-state estimate for the prefactor of the diffusion constant [10], will in general depend upon whether the system described is a classical simulation or a quantum solid, and goes beyond the purview of our work here.

In order to assess how each interatomic potential model performs in vacancy-related simulations, we calculate the most important monovacancy properties for all the elements and all the simple crystal structures predicted by the interatomic potential models available on OpenKIM at this time. The properties we calculate include the vacancy formation energy (VFE), the vacancy migration energy (VME), the vacancy relaxation volume (VRV), the elastic dipole tensor (EDT), the defect strain tensor (DST), and the saddle point elastic dipole tensor (SPEDT) and defect strain tensor (SPDST). VFE and VME give us the local energy change. VRV and the four tensors describe the long-range environmental impact. The OpenKIM website properties page<sup>2</sup> provides a brief description to these quantities. Sec. II describes the method we used to obtain these quantities. The structures we calculate include sc, fcc, bcc, diamond, and hcp. The results can be compared with data from other sources, such as the density function theory (DFT), the quantum Monte Carlo (QMC), and the experiments, and provide use information when selecting interatomic models for vacancy-related simulations. They can also be used with continuum models to obtain an initial guess for further atomistic simulations [11]. In this paper, we compare our VFE and VME results for fcc and hcp structures to DFT calculations and experiments [12].

In addition, we explored how these properties relate to each other, and to other elemental properties, such as the cohesive energy and the bulk modulus. Both the vacancy formation energy and the vacancy migration energy show linear correlation with the cohesive energy,

with slopes of  $0.32 \pm 0.04$  and  $0.14 \pm 0.04$  respectively. The migration energy and the relaxation volume are correlated with the bulk modulus, the first with a positive slope of  $(0.57 \pm 0.03)\text{\AA}^3$ , and the second with a negative slope of  $(-4.2 \pm 0.7)\text{\AA}^6/\text{eV}$ . All of these results are consistent with physics intuition and previous literature [12]. These relationships can provide additional information in selecting models and screening structures.

The paper is organized as follows. In Sec. II we describe the method we use for calculating the monovacancy properties. In Sec. III we show the plots of our results, compare them with the reference data, and explore how they relate to each other and to other elemental properties. And Sec. IV concludes this paper.

## II. METHODOLOGY

In this section, we explain how we calculate the monovacancy properties, including the vacancy formation energy (VFE), the vacancy migration energy (VME), the vacancy relaxation volume (VRV), the elastic dipole tensor (EDT), the defect strain tensor (DST), and the saddle point elastic dipole tensor (SPEDT) and defect strain tensor (SPDST). For a brief description of these quantities, you can check the OpenKIM website properties page.

The calculation is based on OpenKIM [4] and the Atomic Simulation Environment (ASE) [13]. The code can be found in the OpenKIM repository with ID `TD_000000000000` (TODO). Here, we discuss the method we used for obtaining each of these quantities.

The free energy per unit volume of a crystal can be described by the following equation [14]

$$f = f_0 + n_d E_d + \frac{1}{2} n_d^2 E_{dd} + \frac{1}{2} C_{ijkl} \epsilon_{ij} \epsilon_{kl} - n_d \epsilon_{ij} G_{ij} \quad (1)$$

where  $n_d$  is the defect concentration,  $\epsilon_{ij}$  is the bulk strain,  $E_d$  is the vacancy formation energy,  $E_{dd}$  describes the energy from defect interaction,  $C_{ijkl}$  is the elastic stiffness tensor, and  $G_{ij}$  is the elastic dipole tensor.  $f_0$  is the chemical potential. If  $n$  is the number of atoms per unit volume and  $E_0$  is the cohesive energy per atom corresponding to a perfect crystal (what we defined as the reservoir), then  $f_0 = (n - n_d)E_0$ .

It is worthwhile to examine the physical meaning behind each terms. We assume the vacancies only interact with each other through the strain field and that the non-linear behavior of the elasticity is negligible. Then for a unit volume, the total cohesive energy can be described by

$$f = \int_V E_{\text{chemical}} + \frac{1}{2} C_{ijkl} \tilde{\epsilon}_{ij} \tilde{\epsilon}_{kl} dV \quad (2)$$

Here  $\tilde{\epsilon}$  is the strain at each  $dV$ , which consists of the bulk strain  $\epsilon$ , the strain caused by the nearest vacancy  $\tilde{\epsilon}_V$  and

<sup>2</sup> OpenKIM website properties page url: <https://openkim.org/properties>

other vacancies  $\tilde{\epsilon}_I$ .

$$\tilde{\epsilon}_{ij} = \epsilon_{ij} + \tilde{\epsilon}_{V,ij} + \tilde{\epsilon}_{I,ij} \quad (3)$$

Thus

$$\tilde{\epsilon}_{ij}\tilde{\epsilon}_{kl} = \epsilon_{ij}\epsilon_{kl} + \tilde{\epsilon}_{V,ij}\tilde{\epsilon}_{V,kl} + 2\epsilon_{ij}\tilde{\epsilon}_{V,kl} + I(\epsilon, \tilde{\epsilon}_V, \tilde{\epsilon}_I) \quad (4)$$

The first term is the energy due to the bulk strain. The second term together with the change in  $E_{\text{chemical}}$  give  $E_d$ . Since both of them peak at the vacant sites and drop rapidly when moving away,  $\partial f/\partial E_d \propto n_d$ . The third term gives how the formation energy change with respect to the bulk strain, and thus represents  $G$ . For the similar reason as above,  $\partial f/\partial G \propto n_d$ . And the last term, which includes all the remaining terms in the expansion, represents the interaction  $E_{dd}$ . To see how the total interaction energy changes, we first notice the strength of  $\tilde{\epsilon}_I$  is proportional  $n_d$  due to the superposition in linear elasticity. In addition, note that if the vacant sites form a Bravais lattice, which is the configuration we are using in our calculation, the peak of these  $\tilde{\epsilon}_I$  will also occur on these lattice points. Therefore the number of peaks is proportional to  $n_d$  as well. So the overall interaction energy to the lowest order approximation is related to  $f$  by  $\partial f/\partial E_{dd} \propto n_d^2$ .

From Eq. (1), the negative derivative of  $f$  with respect to  $n_d$  under zero strain gives vacancy formation energy

$$VFE = E_d = \left. \frac{\partial f}{\partial n_d} \right|_{\epsilon=0} + E_0 - n_d E_{dd} \quad (5)$$

Numerically, we use the supercell approach. For a supercell of volume  $V_0$  with a monovacancy,  $n_d = 1/V_0$ , Eq. (5) is equivalent to

$$(E_{V_0,d} - E_{V_0}) + E_0 = VFE + \frac{E_{dd}}{V_0} \stackrel{\text{def}}{=} g(V_0) \quad (6)$$

Here  $E_{V_0,d}$  and  $E_{V_0}$  are the free energy of the supercell with and without the defect respectively. And we define the expression in the middle as  $g(V_0)$ , thus  $VFE = g(\infty)$ . We obtain this  $g(\infty)$  through extrapolation. The uncertainty of this value is the root sum square of the uncertainty from the extrapolation and the uncertainty from the atomistic relaxation.

The negative derivative of  $f$  with respect to strain gives the stress

$$\sigma_{ij} = -\frac{\partial f}{\partial \epsilon_{ij}} = -C_{ijkl}\epsilon_{kl} + n_d G_{ij} - \frac{1}{2}n_d^2 \frac{\partial E_{dd}}{\partial \epsilon_{ij}} \quad (7)$$

And the derivative of  $\sigma_{ij}$  with respect to  $n_d$  with fixed strain  $\epsilon_0$  gives the elastic dipole tensor

$$G_{ij} = \left. \frac{\partial \sigma_{ij}}{\partial n_d} \right|_{\epsilon=\epsilon_0} + n_d \left. \frac{\partial E_{dd}}{\partial \epsilon_{ij}} \right|_{\epsilon=\epsilon_0} \quad (8)$$

Similarly, we obtain the value and the uncertainty of EDT through extrapolation.

The defect strain tensor  $\Lambda_{ij}$  is obtained by

$$\Lambda_{ij} = S_{ijkl}G_{kl} \quad (9)$$

where  $S_{ijkl}$  is the elastic compliance tensor, which is the inverse of the elastic stiffness tensor  $C_{ijkl}$ .

The relaxation volume, VRV, is the decrease in the volume of the material due to the monovacancy. There are two methods for calculating this quantity. One is through relaxing both the positions of the atoms and the size of the supercell. And the other is through its relationship with the defect strain tensor. For an orthorhombic supercell

$$VRV = V_0\epsilon_{ii} = \Lambda_{ii} \quad (10)$$

We verified that these two methods give consistent results. Since the second method does not involve additional relaxation, we use that one in our calculation. Note that the supercells used in our calculation for different structures are all orthorhombic cells. For sc, fcc, bcc, and diamond structures, we use the conventional cubic cell as the basis for constructing supercells, and for hcp structures, we use the one with lattice vectors  $(a, \sqrt{3}a, c)$  as the basis.

For vacancy migration, we use the nudged elastic band method to obtain the migration path and cubic spline interpolation to obtain the saddle point energy and configuration. We use a similar extrapolation method to obtain the vacancy migration energy (VME) and the saddle point elastic dipole and defect strain tensors. Since the configuration of the saddle point is still a monovacancy, the quantities shall scale in the same way with respect to the number of atoms.

In Fig. 1, we plot the normalized property values for fcc aluminum versus the number of atoms in a supercell, calculated with the embedded atom model (EAM) by Mishin and Farkas [15]. The linear relationship shown in this figure validates our extrapolation.

The calculated results can be found in the OpenKIM repository [4]. Appendix B provides scripts for retrieving these results.

A key quantity in these results that distinguishes different vacancies and migrations is the position of the vacancy. In the simplest case where symmetry is preserved, this is unambiguously the position of the center of the symmetry. However, in complex crystals or vacancies that break symmetries, there is no unique definition for the position of a vacancy. We propose the following definition, optimized for use in continuum calculations: The position of the vacancy (or any point defect) is the point, where, if we characterize the effect of the defect as a series of force multipoles at this point, the diagonal terms of the force quadruple are zero in the basis that diagonalizes the force dipole. Appendix A provides a more detailed explanation. This has been adopted as the KIM property definition for vacancy position.

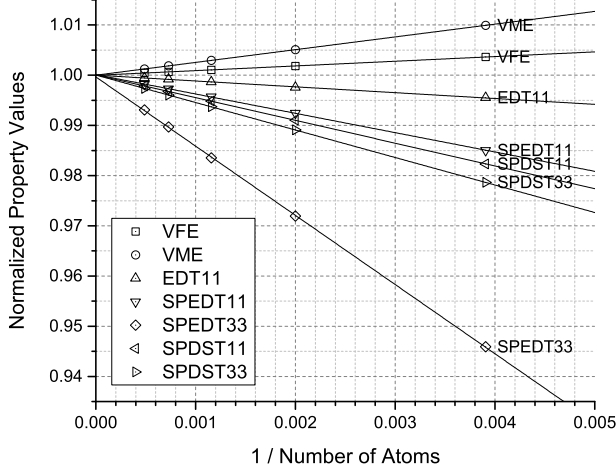


FIG. 1. Property values normalized by the extrapolated values versus the inverse of the number of atoms in a supercell. These values are calculated for fcc aluminum with the embedded atom model (EAM) from Mishin and Farkas [15]. The migration path is from (0.0, 0.0, 0.0) to (0.5, 0.5, 0.0) in a conventional fcc unit cell. VFE is the vacancy formation energy, VME is the vacancy migration energy, VRV is the vacancy relaxation volume, EDT is the elastic dipole tensor, DST is the defect strain tensor, and SP stands for the saddle point of the migration. The two numbers following the tensor are the row and column indices. Since VRV and DST11 are all proportional to EDT11, they will appear as the same curve after normalization, so we show EDT11 here only. EDT22 and EDT33 are the same as EDT11, and all the off-diagonal terms of EDT are zero. So are the terms in DST. The linear relationship shown in this figure validates our extrapolation. In order to avoid crowding the graph, the error bars are not shown here.

### III. RESULTS AND DISCUSSION

All the results calculated in this work are available in the OpenKIM repository [4] and can be either viewed directly on the website or downloaded through OpenKIM Query with the scripts provided in Appendix B.

In this section, we first compare these results with the reference data to see how these models perform in general. We then explore how these quantities related to each other and to other elemental properties.

#### A. Calculated Results Versus Reference Data

Fig. 2 compares our results and with the reference data from DFT calculations and experiments.

We can see from these figures that the EAM models, in general, produce similar VFE and VME as DFT calculations and experiments. While most pairwise models tend to produce larger values for VFE, especially for elements with Mendeleviev numbers ranging from 25 to 45,

which correspond to the transition metals and the post-transition metals.

To further explore how pairwise models come up with larger VFEs, we used fcc aluminum as an example and looked into the relationships between its cohesive energy  $E_0$ , unrelaxed vacancy formation energy  $E_u$ , and relaxed vacancy formation energy  $E_r$ . We found that, with a pairwise potential (Morse),  $E_0 = 2.70\text{eV} \approx E_u$ . While with an EAM potential (Mishin and Farkas),  $E_0 = 3.36\text{eV}$ ,  $E_u = 0.77\text{eV} \approx 0.2E_0 < E_0$ . The energy change due to relaxation,  $|E_u - E_r|$ , is much smaller than  $E_u$  in both cases, and thus can be ignored. We cursorily tested a few other transition metals and found similar relationships. Therefore, the overestimation of VFE is mostly due to the incorrect estimation of the total energy when the coordination number in the neighborhood of the vacancy changes.

Besides the EAM models, there are also several models in our database that, although only supports a few elements, do give accurate results. For example, the three-body potentials give decent results for diamond silicon.

In addition, several new models are also under active development and will be added into the OpenKIM repository soon, such as ... (TODO) The results for new models will be calculated automatically by the system once they are added.

#### B. Relationships Between Properties

The relationship between these vacancy properties and other elemental properties can provide additional information for choosing atomistic models and screening structures: one should be careful when the model we are going to use give results that are inconsistent with these relationships; and when the calculation of the vacancy properties are too expensive, we may use the information from other computationally less expensive properties to filter out some structures.

Previous literatures have shown that the vacancy formation and migration energies are strongly correlated with the cohesive energy and the bulk modulus. We verify these findings here and further explore other relationships, with EAM models and fcc structures.

Fig. 3 shows the relationship between the vacancy formation and migration energies and the cohesive energy. We can see there is a linear increasing trend between the vacancy-related energies and the cohesive energy. For the relationship between VFE and the cohesive energy, the slope is  $0.32 \pm 0.04$ , which agrees well with a previous DFT study, which gives a slope of 0.317 [12]. For the relationship between VME and the cohesive energy, the slope is  $0.14 \pm 0.04$ , which is also similar to the DFT result, 0.191.

Fig. 4 shows the relationship between the vacancy migration energy and the bulk modulus. We can see there is also a linearly increasing trend. The slope of the fitted line is  $(0.57 \pm 0.10) \text{ \AA}^3 = (3.5 \pm 0.6) \text{ eV/GPa}$ . This

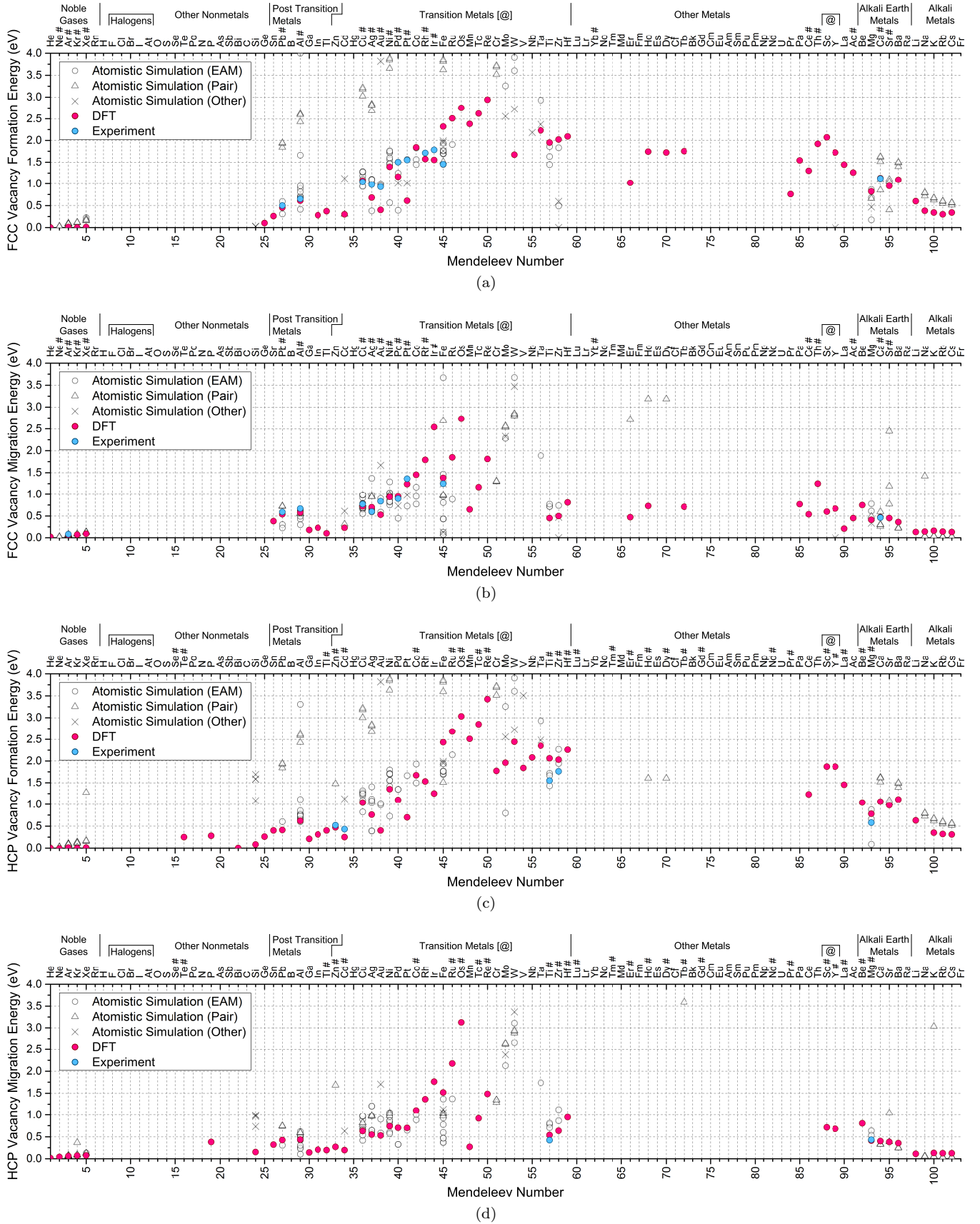


FIG. 2. Comparison with the reference data. The ‘#’ sign next to the symbol of each element indicates whether the structure presented in the figure is the element’s natural state. The error bars are smaller than the size of the symbols. See the OpenKIM model page for an interactive interface allowing investigation of the individual potentials for each symbol shown.

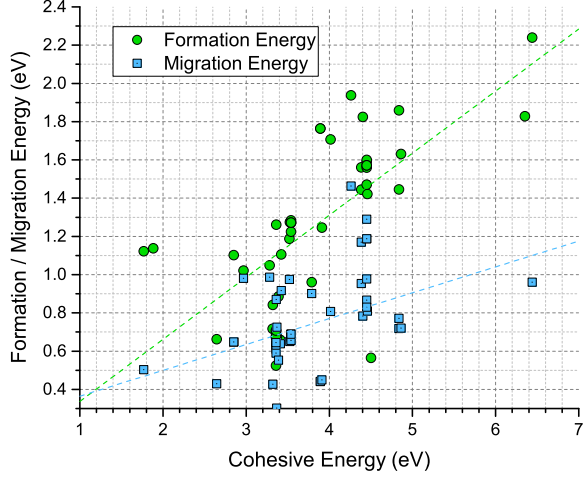


FIG. 3. Vacancy formation and migration energy versus cohesive energy. These are calculated with EAM models for fcc structures. As we can see, there is a linearly upward trend for both quantities. The fitted slope for the formation energy and the migration energy are  $0.32 \pm 0.04$  and  $0.14 \pm 0.04$ , respectively.

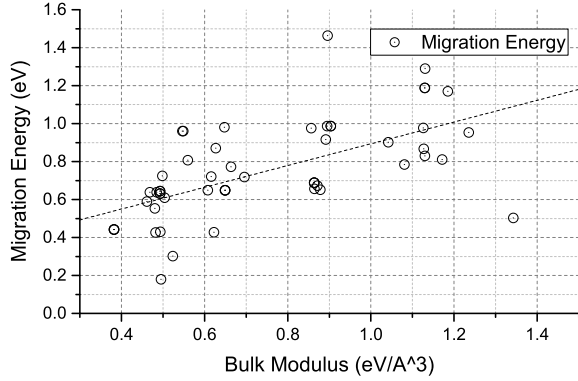


FIG. 4. Vacancy migration energy versus bulk modulus. These are calculated with EAM models for fcc structures. As we can see, there is a linearly upward trend. The fitted slope and intercept are  $(0.57 \pm 0.10) \text{ Å}^3$  and  $(0.32 \pm 0.08) \text{ eV}$ , respectively.

result is quantitatively a little bit smaller than the DFT prediction,  $5.5 \text{ eV/GPa}$ , but qualitatively still the same.

Fig. 5 shows the relationship between the vacancy relaxation volume and the bulk modulus. And we can see from the figure that there is a linear downward trend, which is consistent with the intuition that as the material becomes more stiff, the shrink due to the internal stress created by the vacancy decreases. The slope of the fitted line is  $(-4.2 \pm 0.7) \text{ Å}^6/\text{eV}$ .

We also found there is a strong correlation between the vacancy formation energy and the vacancy relaxation volume. Fig. 6 shows the relationship. The slope of the fitted line is  $(-0.14 \pm 0.02) \text{ eV/Å}^3$ . This downward trend is

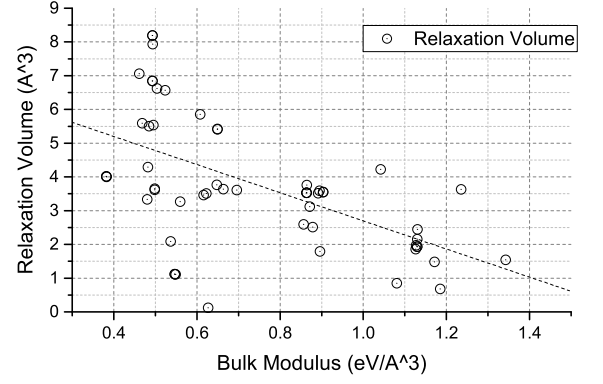


FIG. 5. Vacancy relaxation volume versus bulk modulus. These are calculated with EAM models for fcc structures. As we can see, there is a linearly downward trend. The fitted slope and intercept are  $(-4.2 \pm 0.7) \text{ Å}^6/\text{eV}$  and  $(6.8 \pm 0.6) \text{ Å}^3$ , respectively.

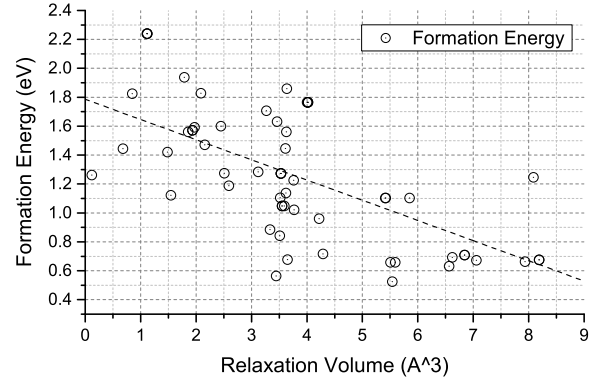


FIG. 6. Vacancy formation energy versus relaxation volume. These are calculated with EAM models for fcc structures. As we can see, there is also a linearly downward trend. The fitted slope and intercept are  $(-0.14 \pm 0.02) \text{ eV/Å}^3$  and  $(1.78 \pm 0.09) \text{ eV}$ , respectively.

also consistent with the intuition that the energy needed for the vacancy to pull the neighboring atoms towards itself decreases with the help of additional bulk strain.

The results above are from fcc structures and the models available on OpenKIM at this time. Appendix B provides Python scripts for retrieving the data from the OpenKIM repository, which can be used to obtain the relationships for all the structures with the latest data.

#### IV. CONCLUSION

In this study, we try to solve the problem of assessing the quality of interatomic models in terms of their ability to describe vacancy-related properties.

We developed tests using the OpenKIM API to perform atomistic calculations and obtain the most important monovacancy properties for all the simple structures

and all the interatomic models in the OpenKIM repository at this time.

Then, we compare the calculated data with the reference data from DFT and experiments. The results show that the EAM potentials, in general, agree with DFT calculations and experiments better than other kinds of potential models in the repository at this time. The pairwise potentials tend to overestimate the vacancy formation energy for transition metals and post-transition metals. One caveat to note is that neither the DFT results nor the experiments are assured to be accurate. It is always beneficial to have more accurate and more comprehensive reference data from other sources with which to compare our results.

And finally, we looked into how these properties related to each other and to other elemental properties. We got consistent results with physics intuition and previous studies, which further support the reliability of our results and these relationships. These relationships can provide additional information for choosing models and screening structures.

## Appendix A: Point Defect Position Definition

We model a point defect as a series of forces exerted on neighboring atoms. Let  $\mathbf{F}^v$  be the body force exerted by the defect on atom  $v$  situated at  $\mathbf{l}^v$ . Then the displacement field can be written as

$$u_i(\mathbf{x}) = \sum_{v=1}^N g_{ij}(\mathbf{x} - \mathbf{l}^v) F_j^v \quad (\text{A1})$$

$g_{ij}$  is the Green's function of a unit point force [16, 17].

Expanding  $u_i$  around an arbitrary point  $\mathbf{x}'$ , gives

$$u_i(\mathbf{x}) = g_{ij}(\mathbf{x} - \mathbf{x}') P_j^{(0)} + \frac{\partial g_{ij}(\mathbf{x} - \mathbf{x}')}{\partial x'_k} P_{kj}^{(1)} + \frac{\partial^2 g_{ij}(\mathbf{x} - \mathbf{x}')}{\partial x'_k \partial x'_l} P_{klj}^{(2)} + \dots \quad (\text{A2})$$

Here  $\mathbf{P}^{(k)}$  are the multipoles corresponding to  $\mathbf{x}'$ ,

$$P_j^{(0)} = \sum_{v=1}^N F_j^v = 0 \quad (\text{A3})$$

$$P_{kj}^{(1)} = \sum_{v=1}^N (l_k^v - x'_k) F_j^v = \sum_{v=1}^N l_k^v F_j^v \quad (\text{A4})$$

$$\begin{aligned} P_{klj}^{(2)} &= \sum_{v=1}^N (l_k^v - x'_k)(l_l^v - x'_l) F_j^v \\ &= \sum_{v=1}^N l_k^v l_l^v F_j^v - x'_k P_{lj}^{(1)} - x'_l P_{kj}^{(1)} \end{aligned} \quad (\text{A5})$$

$\mathbf{P}^{(1)}$  is the force dipole. It is independent of the  $\mathbf{x}'$  we choose. Also, it is symmetric, because the torque on the system

$$\tau_i = \sum_{v=1}^N l_j^v F_k^v - l_k^v F_j^v = P_{jk}^{(1)} - P_{kj}^{(1)} = 0 \quad (\text{A6})$$

Therefore,  $\mathbf{P}^{(1)}$  is diagonalizable.

$\mathbf{P}^{(2)}$  depends on the choice of  $\mathbf{x}'$ . Although we can perform multipole expansion with respect to any  $\mathbf{x}'$ , we hope our choice of  $\mathbf{x}'$  can give a direct impression of where the defect is situated. Therefore, we propose using the following criteria: choose  $\mathbf{x}'$  such that, in the basis where  $\mathbf{P}^{(1)}$  is a diagonal matrix,

$$P_{iii}^{(2)} = \sum_{v=1}^N (l_i^v - x'_i)^2 F_i^v = 0 \quad (\text{A7})$$

This gives an unambiguous definition for  $\mathbf{x}'$  and also makes  $\mathbf{x}'$  the center of the forces. And it is consistent with the case where the symmetry is preserved.

## Appendix B: Obtaining Results From OpenKIM

The results can be obtained either via the web interface directly <https://openkim.org/intro-tests/> or through OpenKIM Query [query.openkim.org](https://query.openkim.org). We describe the latter one in this section.

OpenKIM Query works for any programming languages as long as they support HTTP requests. We use Python here as an example. The backend of this query API is a MongoDB database, and in order to obtain the results we want, we need to construct our query with the correct MongoDB syntax. The website of MongoDB, [mongodb.org](https://www.mongodb.org), provides a detailed documentation of this.

Here is an example of using Python to obtain all the fcc aluminum lattice constants (TODO: replace by vacancy formation energy) calculated with various interatomic potentials:

```
import requests
res = requests.get('https://query.openkim.org/api?
    flat=on&query={"meta.type":"tr","property-id":
    "tag:staff@noreply.openkim.org,2014-04-15:
    property/structure-cubic-crystal-npt","meta.
    runner.kimcode":{"$regex":"^
    LatticeConstantCubicEnergy_fcc"},"species.
    source-value":"Al"}&limit=0&fields={"a.source-
    value":1,"meta.subject.kimcode":1}&database=
    data').json()
```

To obtain a plot of lattice constant versus bulk modulus for all the models and all the elements, we can do

```
import requests
import pandas as pd
import matplotlib.pyplot as plt

# Obtain data
```

```

res_lat = pd.read_json('https://query.openkim.org/
api?flat=on&query={"meta.type":"tr","property-
id":"tag:staff@noreply.openkim.org,2014-04-15:
property/structure-cubic-crystal-npt","meta.
runner.kimcode":{"$regex":"^
LatticeConstantCubicEnergy_fcc"}}&limit=0&
fields={"a.source-value":1,"species.source-
value":1,"meta.subject.kimcode":1}&database=
data')
res_bulk = pd.read_json('https://query.openkim.org/
api?flat=on&query={"meta.type":"tr","property-
id":"tag:staff@noreply.openkim.org,2014-04-15:
property/bulk-modulus-isothermal-cubic-crystal-
npt","meta.runner.kimcode":{"$regex":"^
ElasticConstantsCubic_fcc"}}&limit=0&fields={"
isothermal-bulk-modulus.source-value":1,"meta.
subject.kimcode":1,"species.source-value":1}&
database=data')

# Merge on species and model name
res_lat['species'] = res_lat['species.source-value']
    .apply(lambda x: x[0])
res_bulk['species'] = res_bulk['species.source-value']
    .apply(lambda x: x[0])

```

```

    ].apply(lambda x: x[0])
res = pd.merge(res_lat, res_bulk, how = 'inner', on
    = ['meta.subject.kimcode', 'species'])

# Remove outliers
bulk = res['isothermal-bulk-modulus.source-value']
res_cleaned = res[(bulk > 0) & (bulk < 1000)]

# Plot
res_cleaned.plot(x = 'a.source-value', y = '
    isothermal-bulk-modulus.source-value', style =
    'o')
plt.show()

```

We can also perform a linear regression with the data obtained above

```

from pandas.stats.api import ols
ols(y = res_cleaned['isothermal-bulk-modulus.source-
value'], x = res_cleaned['a.source-value'])

```

The code above has been tested in the Python 2.7 version of Anaconda 4.0.0.

- 
- [1] M. S. Daw, S. M. Foiles, and M. I. Baskes, *Materials Science Reports* **9**, 251 (1993).
  - [2] M. S. Daw and M. I. Baskes, *Physical Review B* **29**, 6443 (1984).
  - [3] M. Bierbaum, E. Tadmor, R. Elliott, T. Wennblom, A. Alemi, Y.-J. Chen, D. Karls, A. Ludvik, and J. Sethna, in *APS Meeting Abstracts*, Vol. 1 (2014) p. 26004.
  - [4] "OpenKIM: Knowledgebase of Interatomic Models," <https://openkim.org/>, [Online; accessed 11-February-2016].
  - [5] E. Tadmor, R. Elliott, J. Sethna, R. Miller, and C. Becker, *JOM Journal of the Minerals, Metals and Materials Society* **63**, 17 (2011).
  - [6] J. Weertman, *Journal of Applied Physics* **26**, 1213 (1955).
  - [7] P. M. Fahey, P. Griffin, and J. Plummer, *Reviews of modern physics* **61**, 289 (1989).
  - [8] J. Sethna, "Crystals, Rigidity, and Statistical Mechanics," <http://pages.physics.cornell.edu/~sethna/teaching/BasicTraining/BasicTraining.html> (2014), [Online; accessed 09-February-2016].
  - [9] B. Andreas, Y. Azuma, G. Bartl, P. Becker, H. Bettin, M. Borys, I. Busch, P. Fuchs, K. Fujii, H. Fujimoto, *et al.*, *Metrologia* **48**, S1 (2011).
  - [10] P. S. Burada, P. Hänggi, F. Marchesoni, G. Schmid, and P. Talkner, *ChemPhysChem* **10**, 45 (2009).
  - [11] S. I. Bozhevolnyi and V. S. Volkov, *Optics communications* **198**, 241 (2001).
  - [12] T. Angsten, T. Mayeshiba, H. Wu, and D. Morgan, *New Journal of Physics* **16**, 015018 (2014).
  - [13] S. R. Bahn and K. W. Jacobsen, *Computing in Science & Engineering* **4**, 56 (2002).
  - [14] D. A. Freedman, D. Roundy, and T. Arias, *Physical Review B* **80**, 064108 (2009).
  - [15] Y. Mishin, D. Farkas, M. Mehl, and D. Papaconstantopoulos, *Physical Review B* **59**, 3393 (1999).
  - [16] D. Seif and N. Ghoniem, .
  - [17] T. Ting and V.-G. Lee, *The Quarterly Journal of Mechanics and Applied Mathematics* **50**, 407 (1997).
  - [18] L. D. Landau and E. M. Lifshitz, *Course of theoretical physics* (Elsevier, 2013).
  - [19] M. Baskes, M. Daw, and S. Foiles, in *MRS Proceedings*, Vol. 141 (Cambridge Univ Press, 1988) p. 31.
  - [20] S. L. Frederiksen, K. W. Jacobsen, K. S. Brown, and J. P. Sethna, *Physical review letters* **93**, 165501 (2004).
  - [21] D. Mayer and D. Butler, *Ecological modelling* **68**, 21 (1993).
  - [22] M. De Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. K. Ande, S. van der Zwaag, J. J. Plata, *et al.*, *Scientific data* **2** (2015).
  - [23] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, *et al.*, *Apl Materials* **1**, 011002 (2013).
  - [24] A. Carlsson, *Solid State Physics* **43**, 1 (1990).
  - [25] B. Medasani, M. Haranczyk, A. Canning, and M. Asta, *Computational Materials Science* **101**, 96 (2015).
  - [26] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, *Physical review letters* **97**, 170201 (2006).
  - [27] T. E. Oliphant, *Computing in Science & Engineering* **9**, 10 (2007).
  - [28] J. A. Nelder and R. Mead, *The computer journal* **7**, 308 (1965).
  - [29] S. G. Louie, M. Schlüter, J. R. Chelikowsky, and M. L. Cohen, *Physical Review B* **13**, 1654 (1976).
  - [30] A. F. Voter, F. Montalenti, and T. C. Germann, *Annual Review of Materials Research* **32**, 321 (2002).
  - [31] T. Suzuki, I. Kosacki, and H. U. Anderson, *Solid State Ionics* **151**, 111 (2002).
  - [32] A. Kalabukhov, R. Gunnarsson, J. Börjesson, E. Olsson, T. Claesson, and D. Winkler, *Physical Review B* **75**, 121404 (2007).
  - [33] J.-E. Kluin and T. Hehenkamp, *Physical Review B* **44**,



- 11597 (1991).
- [34] M. Kass, C. R. Brooks, D. Falcon, and D. Basak, *Intermetallics* **10**, 951 (2002).
  - [35] Y. Kraftmakher, *Physics Reports* **299**, 79 (1998).
  - [36] P. Shewmon, *The Minerals, Metals & Materials Society, Diffusion in Solids. Second Edition.*(Retroactive Coverage)(United States), 1989, , 246 (1989).
  - [37] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, *Nature chemistry* **1**, 37 (2009).
  - [38] R. Bourassa and B. Lengeler, *Journal of Physics F: Metal Physics* **6**, 1405 (1976).
  - [39] B. McKee, W. Triftshäuser, and A. Stewart, *Physical Review Letters* **28**, 358 (1972).
  - [40] T. Korhonen, M. Puska, and R. Nieminen, *Physical Review B* **51**, 9526 (1995).
  - [41] W. Kohn, *Reviews of Modern Physics* **71**, 1253 (1999).
  - [42] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, *Reviews of Modern Physics* **73**, 33 (2001).
  - [43] J. Kotakoski, A. Krashennnikov, and K. Nordlund, *Physical Review B* **74**, 245420 (2006).
  - [44] C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, *Reviews of Modern Physics* **86**, 253 (2014).
  - [45] H. Zhou, S. Upreti, N. A. Chernova, G. Hautier, G. Ceder, and M. S. Whittingham, *Chemistry of Materials* **23**, 293 (2010).
  - [46] G. Ceder, *MRS bulletin* **35**, 693 (2010).