

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Uncertainty Quantification and Reduction of Molecular Dynamics Models

---

Xiaowang Zhou and Stephen M. Foiles

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.68507>

---

## Abstract

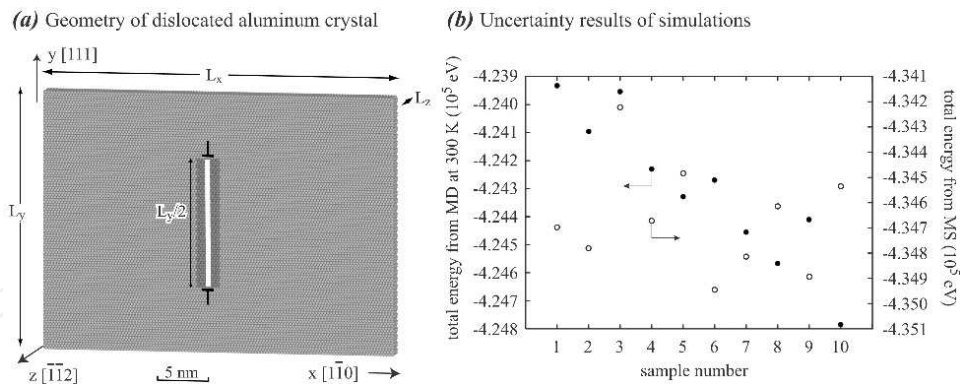
Molecular dynamics (MD) is an important method underlying the modern field of Computational Materials Science. Without requiring prior knowledge as inputs, MD simulations have been used to study a variety of material problems. However, results of molecular dynamics simulations are often associated with errors as compared with experimental observations. These errors come from a variety of sources, including inaccuracy of interatomic potentials, short length and time scales, idealized problem description and statistical uncertainties of MD simulations themselves. This chapter specifically devotes to the statistical uncertainties of MD simulations. In particular, methods to quantify and reduce such statistical uncertainties are demonstrated using a variety of exemplar cases, including calculations of finite temperature static properties such as lattice constants, cohesive energies, elastic constants, dislocation energies, thermal conductivities, surface segregation and calculations of kinetic properties such as diffusion parameters. We also demonstrate that when the statistical uncertainties are reduced to near zero, MD can be used to validate and improve widely used theories.

**Keywords:** molecular dynamics, molecular statics, uncertainty quantification, model calibration, materials science, thermodynamics, kinetics

---

## 1. Introduction

In atomistic simulations, a material is represented by the positions of an assembly of atoms whose energy is represented through a model of the interatomic forces. Molecular dynamics (MD) simulations follow the motion of this collection of atoms. From these simulations, one can extract information about the thermodynamics and kinetics of materials and key material defects. As an example of an MD material simulation, **Figure 1(a)** shows an aluminium crystal



**Figure 1.** Observation of MD uncertainties. (a) An aluminium crystal containing an edge dislocation dipole and (b) the total energies of the dislocated aluminium crystal obtained from MD and MS simulations of 10 different samples.

whose  $[1\bar{1}0]$ ,  $[111]$ , and  $[\bar{1}\bar{1}2]$  crystallographic orientations are aligned respectively with the  $x$ -,  $y$ - and  $z$ - coordinate directions. The initial atom coordinates can be assigned according to structure, orientation and lattice constant of the crystal. To make the system interesting, **Figure 1(a)** also contains two edge dislocations created by removing a  $(1\bar{1}0)$  plane as indicated by the white vertical line. The width of the removed region, therefore, equals exactly the Burgers magnitude  $b = |\langle 110 \rangle a/2|$ . To close the gap of the missing plane, surrounding atoms as indicated by the dark region are shifted towards the gap. The system can also possess a temperature. This is achieved by assigning velocities to all atoms under the Boltzmann distribution condition. Normally, periodic boundary conditions are used to remove free surfaces and infinitely extend the system. This means that the system shown in **Figure 1(a)** is periodically repeated in the  $x$ -,  $y$ - and  $z$ - coordinate directions, with the periodic lengths equal to the corresponding system dimensions  $L_x$ ,  $L_y$  and  $L_z$ . Based on an interatomic potential model that can be used to calculate system energy and interatomic forces [1], an MD simulation essentially solves atom positions as a function of time from Newton's equations of motion [2, 3].

The simplest MD simulations conserve energy and do not change system sizes  $L_x$ ,  $L_y$  and  $L_z$ . With such NVE (meaning that the number of atoms, system volume and system energy are constant) simulations, constant target temperature and pressure usually cannot be maintained. By using Nose-Hoover dragging forces [4] to increase or decrease atom kinetic energies depending on if the temperature is lower or higher than the desired value, MD simulations can be performed at a constant temperature. By using the Parrinello-Rahman algorithm [5] to allow the periodic lengths  $L_x$ ,  $L_y$  and  $L_z$  to increase or decrease depending on if the pressure is higher or lower than the desired value, MD simulations can also be performed at a constant pressure.

Once an interatomic potential is given, the MD methods described above enable many material problems to be computationally studied without any prior knowledge of these problems. For example, MD reveals phonon vibration spectrum and thermal transport properties even when applied to defect-free systems. When systems contain point defects, MD simulates the diffusion of these defects. When systems contain dislocations, such as **Figure 1(a)**, MD computes dislocation core structures and core energies. When external forces/loads are applied to the system, MD explores a variety of other problems including deformation, fracture and

structure evolution. When adatoms are continuously added to a surface, MD shows the structure evolution during vapour deposition synthesis processes. Due to the broad applicability and high predictability of MD simulations, the problem of the uncertainty margin of MD results is becoming increasingly important.

In principle, results of molecular dynamics simulations necessarily contain errors as compared with experimental observations. These errors come from a variety of sources, including inaccuracy of interatomic potentials, short length and time scales, idealized problem description and statistical uncertainties of MD simulations themselves. This chapter focuses on quantification and reduction of one important model uncertainty: statistical uncertainty of molecular dynamics simulations.

## 2. An overview perspective of uncertainty quantification methods

The ultimate goal of evaluating and reducing the statistical uncertainty of MD simulations is to minimize differences between predictions and experimental observations. To establish a useful context, we first briefly describe quantification methods for other uncertainties during multiscale simulations of materials.

Uncertainties are commonly divided into two types: aleatoric uncertainty arising from randomness and epistemic uncertainty arising from lack of knowledge. Examples of the aleatoric uncertainty include head or tail when flipping a coin or a high precision length measured with a coarse scale ruler. Typically, the aleatoric uncertainty can be described by a probability distribution function. Increasing data can result in more accurate characterization of this distribution, but cannot reduce its variance. Examples of the epistemic uncertainty include prediction from an inaccurate (or incorrect) model, or the length measured by a low-quality ruler. Usually, the epistemic uncertainty cannot be described by a probability distribution. This uncertainty, however, can be reduced when additional data or knowledge are incorporated (e.g., when the model is improved or the error of the ruler is calibrated). Note that sometimes the epistemic uncertainty can be treated as the aleatoric uncertainty. For example, due to the thermal expansion, rulers are usually associated with an epistemic error on a given day. This epistemic uncertainty may become an aleatoric uncertainty if the measurements are made throughout the entire year.

There are many issues that influence the comparison of MD results with experimental observations. The most commonly discussed approximation is the accuracy (epistemic uncertainty) of the interatomic potential. Ideally, this represents the true energy of the arrangement of atoms. In practice, a computationally convenient and physically motivated functional form of the potential is assumed and parameterized to match either fundamental electronic structure calculations or experimental data [1]. Only recently have systematic evaluations of these errors begun to be performed [6, 7]. As one practical approach, Moore et al. [7] performed a parameter sensibility study where the parameter of an interatomic potential is varied one at a time and its effects on properties (e.g., lattice constant, elastic constants, cohesive energy and enthalpy of mixing) are determined using MD simulations. Such a study reveals the relative

importance of each of the potential parameters. However, it does not provide information on the accuracy of potential.

In principle, we can always image the existence of an ideal potential that will give the exact solution to the problem of our interest, provided that this potential is fitted to the right values of a list of properties  $\{k_1, k_2, \dots, k_n, u_1, u_2, \dots, u_m\}$ , where  $k_1, k_2, \dots, k_n$  are the list of properties that are known to be important (e.g., lattice constant, elastic constants, cohesive energy, etc.), and  $u_1, u_2, \dots, u_m$  are a sufficient list of important properties that will make the potential accurate but are unknown to us as what these properties are due to the lack of knowledge. In practices, however, we will never achieve such an ideal potential because not only we do not know  $u_i$  ( $i = 1, 2, \dots, m$ ), but also no potential can be fitted exactly to the target values of all  $k_i$  ( $i = 1, 2, \dots, n$ ). Based on this recognition, a relevant approach to quantify the epistemic uncertainty of the potential is to create an ensemble of potentials that predict a distribution of properties  $\{k_1, k_2, \dots, k_n\}$  centring around the true experimental values and quantify the effects of this distribution on the target properties computed with MD simulations. This approach may still not yield a satisfactory quantification of the epistemic uncertainty of the potential currently. However, the quantified epistemic uncertainty will continuously improve as more and more  $u_i$  properties are understood and become  $k_i$  with improved knowledge.

There are additional issues associated with MD simulations. For the study of complex defects, issues can arise from the boundary conditions imposed on the simulations and from the structural idealizations often imposed. For example, in a recent study of faceting of grain boundaries in Fe, there were qualitative differences between the MD-predicted facet length and facet junction geometries and experimental observations [8]. The source of the disagreement was the idealized geometry used in the MD simulations. The simulations assumed an ideal coincident site lattice misorientation between the crystal lattices while the experiment deviated slightly from this ideal misorientation. This deviation introduced interfacial dislocations that fundamentally changed the faceting behaviour. The use of improved geometries, often at the computational cost of using larger systems, can be used to estimate the related epistemic uncertainty. Likewise, the time scales of MD simulations (on the order of nanoseconds) raise issues with processes that occur on longer time scales. For example, in simulations of multi-component systems, diffusive processes of substitutional impurities often occur on time scales beyond direct MD simulations, and simulations of mechanical deformation can be strongly influenced by the high strain-rates required by MD simulation times. Increasing simulation time can provide an estimate of the trends of the related epistemic uncertainty.

To study material problems at engineering scales, multiscale approaches linking models of different scales are needed. Beyond the specific uncertainties associated with MD simulations, there are also initial studies of the broader question of how those uncertainties propagate through a material modelling hierarchy [9–11]. To study how an aleatoric uncertainty of the interatomic potential propagates through the MD to a continuum model, we can perform many MD simulations using different interatomic potentials sampled from the aleatoric uncertainty distribution. Results of each MD simulation are used as inputs to perform a separate continuum simulation of the final material properties. Many continuum simulations then give an aleatoric uncertainty distribution. To yield a highly converged continuous distribution of



the final results, thousands or more MD simulations are needed. This is often computationally impractical.

Assume that a continuum scale model requires a list of properties  $P_{i, MD}$  ( $i = 1, 2, \dots, N$ ) from MD calculations as inputs. If these properties are independent (e.g., thermal conductivities obtained at different temperatures), then the direct Monte Carlo sampling [12] can be used to propagate uncertainties efficiently. First, numerous MD simulations are performed to determine distribution of each  $P_{i, MD}$ . Because only distribution of one property is concerned, the number of MD simulations needed to yield a smooth distribution of that property is significantly reduced. Knowledge of distribution of each of the  $P_{i, MD}$  properties can then be used to sample as many  $\{P_1, P_2, \dots, P_N\}$  sets [12] as one may desire. These data sets can be used in continuum simulations to yield a smooth distribution of the final results.

Experimentally, no samples can have exactly the same microstructure in terms of size and population of grains, shape and volume fraction of phases, defect densities, chemical composition and purity. As a result, experimental measurements of mechanical properties of materials always involve uncertainties. Because microstructures obtained from the same processing satisfy a certain distribution, such uncertainties are aleatoric. On the other hand, some properties such as diffusivities are difficult to measure. As a result, there are considerable disagreements for the diffusivity data reported by different groups [13]. Such uncertainties can be considered as epistemic. Note that experimental uncertainties are often the problem of interest, but they are different from model uncertainties. It is possible to use multiscale modelling to predict the experimental uncertainties. For example, MD simulations can be used to determine the cohesive zone laws [14, 15] of different grain boundaries. These cohesive zone laws can be incorporated in continuum models to simulation intergranular fracture. Through a continuum simulation of the intergranular fracture from a large number of realizations of initial grain structures, the experimental uncertainties due to the variation of grain microstructures can be calculated. Because experimental uncertainties are superimposed on model uncertainties, it is required that model uncertainties be reduced (or at least quantified) before experimental uncertainties can be confidently studied. The quantification and reduction of the statistical uncertainty of molecular dynamics simulations are therefore important.

### 3. Statistical uncertainty of molecular dynamics methods

Due to thermal noises, MD simulations are always associated with a statistical uncertainty. To examine this problem, an MD simulation of the computational system shown in **Figure 1(a)** is performed for a period of 20 ps at a temperature of 300 K using a previously developed Al-Cu interatomic potential [16]. After the first 10 ps is ignored to enable a preliminary equilibration, the total system energy is calculated every 1 ps for the remaining 10 ps. The total energies for these 10 snapshot samples are shown in **Figure 1(b)** using the filled circles. It can be seen that the total energies for the 10 samples are not exactly the same, but rather span a range of nearly 900 eV. Two types of uncertainties can be identified here. First, there is a general decreasing trend with sample number (corresponding to time). This systematic error arises from a continued

equilibration with increasing simulation time. Second, there are some occasional fluctuations of the results. This statistical error arises from thermal noises.

Molecular statics (MS) is another frequently applied computational method [2] to study materials. Rather than solving Newton's equation of motion, MS determines equilibrium atom positions by minimizing the total potential energy of the system at the 0 K temperature (i.e., there is no kinetic energy of atoms). To examine if MS simulations have the uncertainty issue when studying dislocations, 10 MS simulations are performed on the configuration of **Figure 1(a)** using different random number seeds. The 10 total system potential energies obtained from the 10 MS simulations are included in **Figure 1(b)** using unfilled circles. Interestingly, MS simulations, which do not involve thermal noises, also involve large uncertainties. In fact, differences among the 10 samples are comparable with the MD simulations ( $\sim 800$  eV or above). This MS error, however, appears to be entirely statistical.

The uncertainty discussed above pertains to total energy of the system. The system considered in **Figure 1(a)** contains 129,600 atoms. As a result, the relative error shown in **Figure 1(b)** is less than  $900/129,600 = 0.007$  eV/atom. It is important to note that the MS errors revealed in **Figure 1(b)** are larger than one would normally see in literature. This is because literature simulations are usually applied to either defect-free systems or much smaller system dimensions. When defects relax (e.g., a perfect dislocation dissociates into two partials bounding a stacking fault as in the present case), many local energy minimums occur and therefore MS results become uncertain because there are really no robust methods available today to identify the global minimum energy configuration. Furthermore, while current MS methods can achieve high accuracies for relative properties (e.g., energy per atom), it is unrealistic to achieve small global errors for large systems (unless accuracies of relative properties can be infinitely improved when system sizes are increased). Global errors are important to many applications. In **Figure 1(a)**, for example, the dislocation line energy is defined as the total system energy difference between dislocated and perfect crystals, divided by total dislocation length  $2L_z$  along the  $z$  direction. When  $L_z$  is not too big, say,  $\sim 25$  Å as in the present case, a total energy error of 900 eV will result in meaningless dislocation line energy calculations considering that the line energies of dislocations are usually less than 5 eV/Å [17]. In the following, we will discuss methods to quantify and reduce the statistical uncertainty margin of MD simulations as revealed here.

#### 4. Methods for quantifying molecular dynamics statistical uncertainty

Experimentally measured properties are average behaviour of systems over the time scale of the measurement, which is usually much longer than the MD time scales. To reflect experimental properties, it is appropriate to calculate time-averaged properties during MD simulations. Two different approaches can be used to perform statistical uncertainty quantification for time-averaged MD simulations based on fundamental principles of statistics [18].

The first approach is based entirely on the statistical nature of MD results. Assume that an MD simulation is performed for a total period of  $t_{tot}$ . We can divide  $t_{tot}$  into  $N$  segments with the end point of each segment being  $t_i = i\Delta t$  ( $i = 1, 2, \dots, N$ ) where  $\Delta t = t_{tot}/N$ . Any time-averaged

property can be calculated for each of the time intervals  $\Delta t_i = t_i - t_{i-1} = \Delta t$ , and as a result, each MD simulation will produce  $N$  values of the property  $P$ . If we denote each estimate of  $P$  to be  $P_i$  ( $i = 1, 2, \dots, N$ ), the best estimate of the property can be calculated as

$$\bar{P} = \frac{\sum_{i=1}^N P_i}{N} \quad (1)$$

The uncertainty of the samples  $P_i$  can be quantified by the sample standard deviation defined as

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (P_i - \bar{P})^2}{N - 1}} \quad (2)$$

The best estimate  $\bar{P}$  is also associated with an uncertainty  $\bar{\sigma}$ .  $\bar{\sigma}$  is reduced from  $\sigma$  through the well-known relationship [18]

$$\bar{\sigma} = \frac{\sigma}{\sqrt{N}} \quad (3)$$

Eqs. (1)–(3) are effective in determining the variation of the calculated properties. They do not give direct indication of how physical the results are. In many applications, properties  $P, Q, R, \dots$  are often related through some well-justified physical functions, say,  $F(P, Q, R, \dots) = 0$ . The second approach is based on the deviation of the calculated properties from these functions. In particular, a deviation parameter can be defined as  $\xi$

$$\xi = \sqrt{\frac{\sum_{i=1}^N F(P_i, Q_i, R_i, \dots)^2}{N}} \quad (4)$$

As a first example to calculate  $\xi$ , elastic constants of single cubic crystals satisfy  $C_{11} = C_{22} = C_{33}$ ,  $C_{33} = C_{44} = C_{55}$ ,  $C_{12} = C_{13} = C_{21} = C_{23} = C_{31} = C_{32}$  and  $C_{ij}$  ( $j = 4, 5, 6, i = 1, 2, \dots, j-1$ ) = 0. Accordingly, we can define four deviation parameters as

$$\xi_1 = \sqrt{\frac{\sum_{i=1}^3 [C_{ii} - (C_{11} + C_{22} + C_{33})/3]^2}{3}} \quad (5)$$

$$\xi_2 = \sqrt{\frac{\sum_{j=1}^3 \sum_{\substack{i=1, \\ i \neq j}}^3 [C_{ij} - (C_{12} + C_{13} + C_{21} + C_{23} + C_{31} + C_{32})/6]^2}{6}} \quad (6)$$



$$\xi_3 = \sqrt{\frac{\sum_{i=4}^6 [C_{ii} - (C_{44} + C_{55} + C_{66})/3]^2}{3}} \quad (7)$$

$$\xi_4 = \sqrt{\frac{\sum_{j=4}^6 \sum_{i=1}^{j-1} (C_{ij}^2 + C_{ji}^2)}{24}} \quad (8)$$

If we want to determine how physical our overall results are, the best MD estimates of  $C_{ij}$  can be used in Eqs. (5)–(8) to calculate  $\xi_1 - \xi_4$ . This simply means that  $C_{ij}$  are averaged over the entire simulation time  $t_{tot}$  rather than the short time interval  $\Delta t$ . We can also use  $C_{ij}$  obtained within different time intervals (multiple of  $\Delta t$ ) to examine time convergence of the calculated properties towards the true physical values.

As another example to calculate  $\xi$ , diffusivity  $D$  is related to pre-exponential factor  $D_0$  and activation energy barrier  $Q$  through the Arrhenius equation,  $D = D_0 \exp(-\frac{Q}{kT})$  or  $\ln D_0 - \frac{Q}{kT} - \ln D = 0$ , where  $k$  and  $T$  are respectively Boltzmann constant and temperature. If MD can be used to calculate diffusivities  $D_i$  at different temperatures  $T_i$  ( $i = 1, 2, \dots, N$ ), then we can fit the Arrhenius equation to get  $D_0$  and  $Q$ . We can then define a deviation error parameter for the calculated diffusivities from true values as

$$\xi = \sqrt{\frac{\sum_{i=1}^N \left( \ln D_0 - \frac{Q}{kT_i} - \ln D_i \right)^2}{N}} \quad (9)$$

Note that although the error parameter  $\xi$  can validate models, it does not directly measure the uncertainty margin of a property. However,  $\xi$  is related to the direct uncertainty margin  $\sigma$  (or  $\bar{\sigma}$ ) because  $\xi \rightarrow 0$  necessarily leads to  $\sigma \rightarrow 0$  (or  $\bar{\sigma} \rightarrow 0$ ). In the following, we demonstrate specific examples on how to quantify  $\sigma$  (or  $\bar{\sigma}$ ) and  $\xi$  in MD simulations.

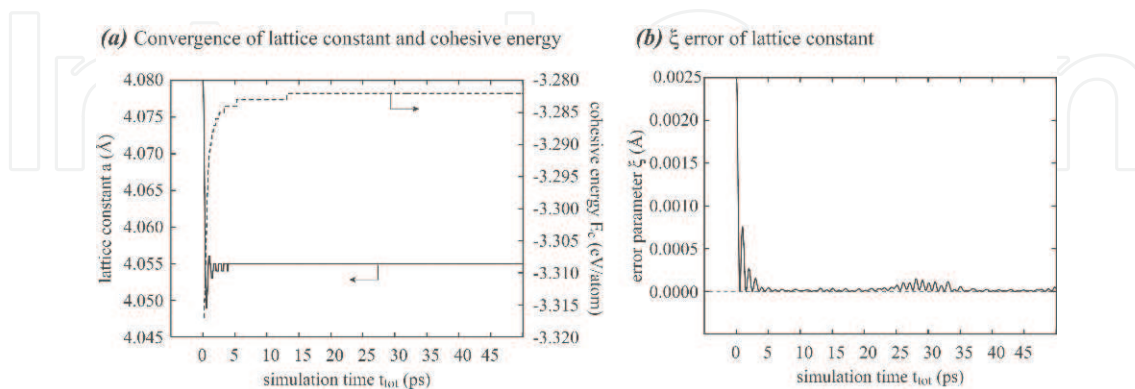
## 5. Lattice constant and cohesive energy

We now quantify the uncertainty margins of the finite temperature lattice constant and cohesive energy of aluminium calculated using MD simulations based on a literature interatomic potential [16]. The periodic computational system includes  $5 \times 5 \times 5$  unit cells of a face-centred-cubic (fcc) crystal. The initial lattice is intentionally strained in the  $x$ -,  $y$ - and  $z$ - directions by 0.01,  $-0.01$  and 0.02, respectively, and all atoms are randomly disturbed from their lattice sites subjecting to a maximum displacement of  $0.05 \text{ \AA}$ . A zero pressure NPT (number of atoms, pressure, and temperature are constant) MD simulation is then performed at 300 K

using a time step size of 0.004 ps. Since the lattice constants ( $a_x$ ,  $a_y$  and  $a_z$ ) in the three coordinate directions are not the same initially, their geometric mean  $a = \sqrt[3]{a_x a_y a_z}$  is used as the overall lattice constant. Here, geometric mean is used instead of arithmetic mean to conserve volume. The short-term average lattice constant and cohesive energy (per atom) are calculated every 10 time steps (i.e.,  $\Delta t = 0.04$  ps). The best estimates (refer to running averages here) of these properties are calculated using Eq. (1) as a function of simulation time  $t_{tot}$ . The results of these best estimates are shown in **Figure 2(a)**. **Figure 2(a)** indicates that despite the initial disturbed crystal that biases the average calculations towards a non-equilibrium structure at short time, the finite temperature lattice constants and cohesive energy calculated from MD approaches convergence rapidly. After 15 ps simulation, both lattice constant and cohesive energy essentially become constant, and as a result, there is no significant uncertainty associated with this simulation.

Note that we do not explicitly show the standard deviation defined by Eq. (3). However, the information is implicitly revealed in **Figure 2(a)**, because the standard deviation must approach zero when the calculated properties become constant. On the other hand, cubic crystal lattice constants satisfy a relation  $a_x = a_y = a_z$ . This allows us to define a deviation parameter  $\xi = \sqrt{\frac{(a_x - \sqrt[3]{a_x a_y a_z})^2 + (a_y - \sqrt[3]{a_x a_y a_z})^2 + (a_z - \sqrt[3]{a_x a_y a_z})^2}{3}}$  to measure how physical the results are.  $\xi$  is calculated as a function of  $t_{tot}$ , and the results are shown in **Figure 2(b)**. Considering the small scale in the vertical axis, the non-cubic deviation is very small. This further confirms that the calculated values have extremely small uncertainty margin.

This example indicates that the uncertainty margin of time-averaged MD simulations can be easily reduced to a negligible level when calculating simple properties, such as lattice constant and cohesive energy. This is because these quantities are relative properties (i.e., per unit cell for lattice constant and per atom for cohesive energy), do not involve defects (i.e., no large number of local energy minimums) and can be obtained from small systems. More challenging cases will be presented below.



**Figure 2.** Effect of simulation time on uncertainty of MD calculation of lattice constant and cohesive energy of an fcc aluminium crystal. (a) Lattice constant and cohesive energy and (b) deviation of lattice constant from the cubic relations.

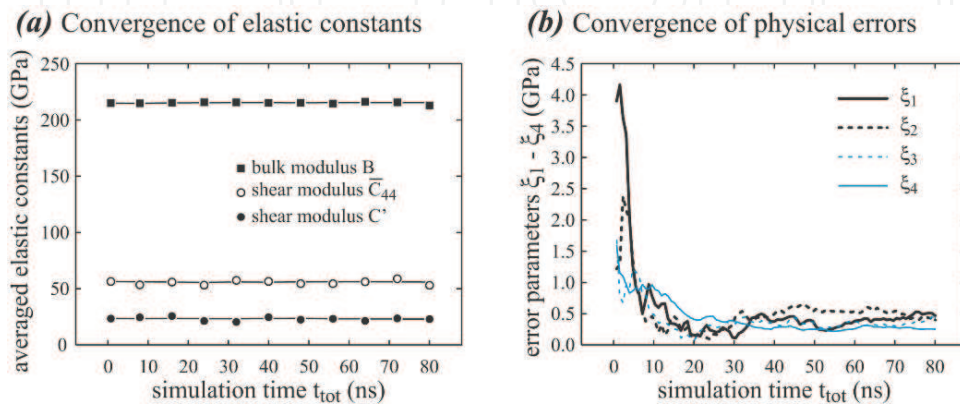
## 6. Elastic constants

Compared with lattice constant and cohesive energy, calculations of finite temperature elastic constants encounter a bigger uncertainty problem. This is because elastic constants are defined by  $C_{ij} = \partial\sigma_i/\partial\varepsilon_j$ , where  $\sigma_i$  and  $\varepsilon_j$  are the stress and strain components in the Voigt notation. Within the finite difference method, elastic constants are calculated as  $C_{ij} = \delta\sigma_i/\delta\varepsilon_j$ , where  $\delta\sigma_i$  is a small change of stress in responding to a small imposed strain  $\delta\varepsilon_j$ . Accurate calculations can only be achieved when the uncertainty margin of the  $\delta\sigma_i$  calculation is significantly smaller than a very small  $\delta\varepsilon_j$  value. We now explore this problem using fcc palladium as an example. The simulations employ the literature interatomic potential [19].

First, the equilibrium finite temperature palladium lattice constant that accounts for thermal expansion is calculated using the approach described above. This equilibrium lattice constant is then used to create an fcc palladium crystal containing  $4 \times 4 \times 4$  unit cells. Positive and negative small strains of the  $j$ th component  $\pm \delta\varepsilon_j = 10^{-4}$  ( $j = 1, 2, \dots, 6$ ) are separately applied to the system. MD simulations using an NVT (number of atoms, volume and temperature are constant) ensemble are performed for 100 ns to relax both the positively and negatively strained systems. An NVT ensemble is needed to maintain the imposed strain. After discarding the first 20 ns, time-averaged stresses  $\sigma_i$  ( $i = 1, 2, \dots, 6$ ) are calculated for the remaining  $t_{\text{tot}} = 80$  ns. The MD elastic constants  $C_{ij}$  are then calculated using a finite-difference scheme

$$C_{ij} = \frac{\sigma_i(\delta\varepsilon_j) - \sigma_i(-\delta\varepsilon_j)}{2\delta\varepsilon_j} \quad (10)$$

By repeating the same process for all  $i, j = 1, 2, \dots, 6$ , we determine all elastic constants. These elastic constants are converted to average values based on the cubic relations, i.e., the bulk modulus  $B = (C_{11} + C_{22} + C_{33} + 2C_{12} + 2C_{13} + 2C_{23})/9$ , shear moduli  $C' = (C_{11} + C_{22} + C_{33} - C_{12} - C_{13} - C_{23})/6$  and  $\bar{C}_{44} = (C_{44} + C_{55} + C_{66})/3$ . If we divide the 80 ns into 80 segments,  $B$ ,  $C'$  and  $\bar{C}_{44}$  are calculated for each segment, and the results are shown in **Figure 3(a)**, where data points are values for some selected segments and lines represent running averages. It can be seen that the uncertainty margin of the averaged elastic constants is very small especially for the running averages, which are virtually constant in the scale of the figure. Note that the data shown in



**Figure 3.** Effect of simulation time on uncertainty of MD calculations of finite temperature elastic constants of an fcc palladium crystal. (a) Cubic-averaged elastic constants and (b) deviation of individual elastic constants from the cubic relations.

**Figure 3(a)** have been averaged based on the cubic relations. Individual elastic constants  $C_{ij}$  may deviate from these relations. The four error parameters  $\xi_1 - \xi_4$  for individual elastic constants to deviate from the cubic relations are calculated using Eqs. (5)–(8). If individual elastic constants are calculated as running averages over the entire  $t_{tot}$ , then results for  $\xi_1 - \xi_4$  are shown in **Figure 3(b)** as a function of  $t_{tot}$ . **Figure 3(b)** further reveals that at average time below 20 ns, the calculated elastic constants might have relatively large uncertainties as they have not fully converged to physical values. However, satisfactorily converged results can be achieved when the average time exceeds 20 ns or above.

## 7. Dislocation energy

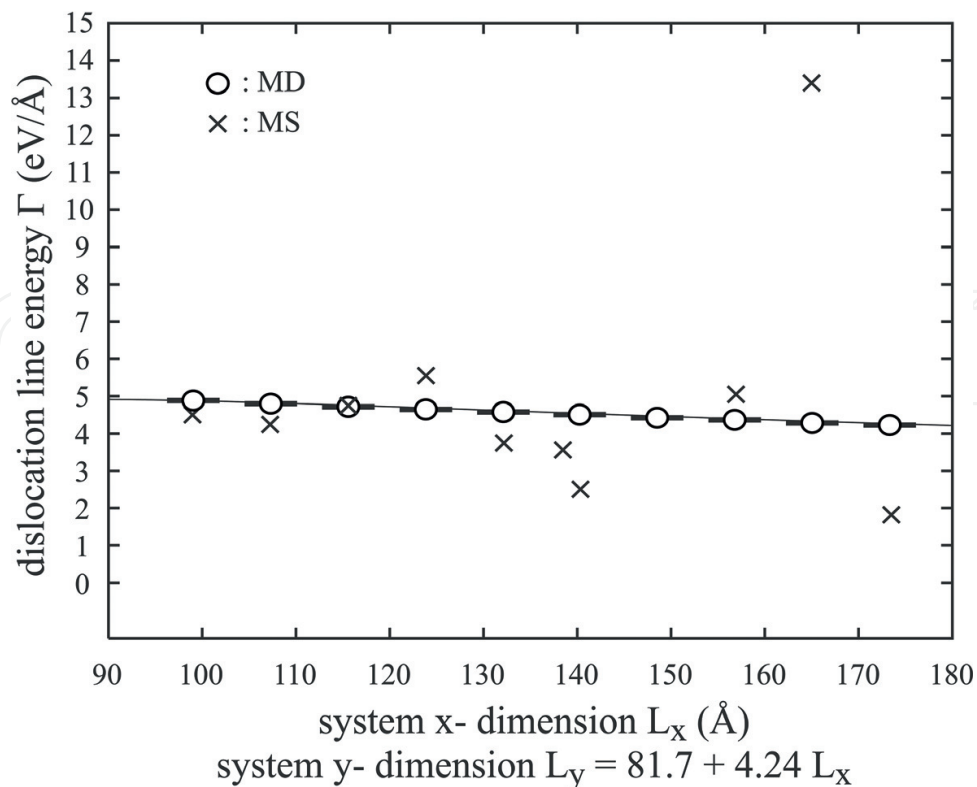
Dislocation relaxation causes a large number of local energy minimums, the long elastic field of dislocations requires the use of large systems and dislocation energies are related to total system energies rather than per-atom energy. All of these contribute to large uncertainties as can be seen in **Figure 1(b)**. As a result, reducing uncertainty margin during MD calculations of dislocation energies becomes extremely important. Here, we illustrate this by calculating core energies of edge type of misfit dislocation in zinc-blende CdS [20] using the literature interatomic potential [21]. We also calculated dislocation energies for aluminium using exactly the same geometry as shown in **Figure 1(a)**, and the same results were obtained [22].

The crystals used for the calculations contain  $n_x$  (101) planes in  $x$ -,  $n_y$  (010) planes in  $y$ - and  $(\bar{1}01)$   $n_z$  planes in  $z$ -. At a fixed  $n_z = 6$  ( $L_z \sim 25$  Å), 10 system dimensions of  $n_x \times n_y = 24 \times 86, 26 \times 92, 28 \times 98, 30 \times 104, 32 \times 110, 34 \times 116, 36 \times 122, 38 \times 128, 40 \times 134$  and  $42 \times 140$  are studied. Under these dimensions, the lengths  $L_y$  and  $L_x$  roughly satisfy the relation  $L_y = 81.7$  (Å) +  $4.24 L_x$ , and the smallest ( $n_x \times n_y = 24 \times 86$ ) and the largest ( $n_x \times n_y = 42 \times 140$ ) systems correspond to  $L_x \times L_y \sim 100 \times 500$  Å<sup>2</sup> and  $\sim 170 \times 820$  Å<sup>2</sup>, respectively. Similar to **Figure 1(a)**, a dislocation dipole is created by removing a (101) plane (a thickness of the Burgers magnitude  $b$ ) of height  $d = 40$  (010) planes ( $\sim 230$  Å).

MD simulations are performed at 300 K for 4 ns to equilibrate the systems, and another 16 (=  $t_{tot}$ ) ns to calculate time-averaged energies of both perfect crystals and crystals containing the dislocation dipoles. Let  $E_p$  and  $N_p$  represent the energy and total number of atoms in the perfect crystal, and  $E_d$  and  $N_d$  represent the energy and total number of atoms in the dislocated crystal. Since atoms are equivalent in the perfect crystal, the total energy of a perfect crystal containing the same number of atoms as in the dislocated crystal can be obtained by scaling  $E_p$  with the ratio  $N_d/N_p$ . Hence, the dislocation line energy is calculated as

$$\Gamma = \frac{E_d - \frac{N_d}{N_p} E_p}{2L_z} \quad (11)$$

where  $2L_z$  is the total dislocation length. Based on a time segment of  $\Delta t = 16$  ps to calculate sample  $\Gamma_i$ , both best estimate dislocation energies  $\bar{\Gamma}$  (simplified as  $\Gamma$ ) and their standard deviations  $\bar{\sigma}$  are calculated using Eqs. (1)–(3). The results of  $\Gamma$  and  $\bar{\sigma}$  are shown in **Figure 4** as unfilled circles and error bars, respectively. In **Figure 4**, the thin light line is obtained from a continuum model [20], and the crosses represent data from MS simulations.



**Figure 4.** CdS misfit dislocation energy as a function of system dimensions  $L_x$  and  $L_y$  under the constraint  $L_y = 81.7$  (Å) +  $4.24 L_x$ . Note that error bars of MD data are essentially horizontal lines, indicating negligible uncertainty margins.

**Figure 4** indicates that despite the challenge for convergence during short-time MD simulations as seen in **Figure 1(b)**, the uncertainty margin of time-averaged MD results of dislocation energies can be reduced to a negligible level if the average time is increased to 16 ns or above. As a consequence of the high convergence, all the MD data points fall right on top of the continuum line. This means that if constructed from the continuum function, the error parameter  $\xi$  would also be near zero. Contrarily, the MS results only approximately agree with the continuum line at small system dimensions (however, our smallest dimension of  $L_x \sim 100$  Å would correspond to 80,000 atoms, which is big according to the literature MS standard) and become meaningless for large systems. The uncertain margin of MS simulations can also be quantified and reduced by performing ensemble averages of a large number of MS simulations with initial configurations taken from various snapshots of an MD simulation (so that they are at different thermally activated states). This is left as an exercise for readers as we only address MD simulations in this chapter.

## 8. Diffusion parameters

For alloyed systems, or systems involving defects, the number of possible atomic diffusion mechanisms can be tremendous. In such cases, diffusivities can be most effectively calculated from the mean square displacement of the diffusing species obtained from MD simulations. Diffusivities at different temperatures can be further used to derive pre-exponential factor and



activation energy of diffusion through an Arrhenius fit. The only challenge of this approach is that it is usually associated with large statistical errors. We now explore this issue using hydrogen diffusion in aluminium as an example. We use the literature Al-H potential [13] in the calculations.

Aluminium fcc crystals containing 8 {100} planes in each of the three <100> coordinate directions are used for simulations. The initial crystals are created based on the room temperature experimental lattice constant  $a = 4.05 \text{ \AA}$  [23]. The system dimension is therefore around  $32 \text{ \AA}$ , corresponding to 2048 aluminium atoms. For comparison, we also calculate the theoretical lattice constants at finite temperatures following the approach described above, and find  $a = 4.05 \text{ \AA}$  at 300 K and  $a = 4.06 \text{ \AA}$  at 700 K. Bulk crystals are simulated using periodic boundary conditions, and a single hydrogen atom is introduced in the computational cell.

First, a warm-up MD simulation is performed for more than 0.1 ns to equilibrate the system at the target temperature  $T$ . Following this, an MD simulation is performed for a total period of  $t_{tot}$ . If the time step size is  $dt$ , the total number of simulated steps  $n = t_{tot}/dt$ . The time-dependent hydrogen location,  $\mathbf{r}(t)$ , is recorded on a time interval of  $\Delta t$ , i.e., at times  $t = i\Delta t$ ,  $i = 1, 2, \dots, m$  ( $m = t_{tot}/\Delta t$ ), where  $\Delta t$  is a multiple of  $dt$ . These locations allow calculations of the relative hydrogen displacement per time interval  $\Delta t$ . For example, if the starting and ending times of the  $\Delta t$  interval are  $(i - 1)\Delta t$  and  $i\Delta t$ , respectively ( $i = 1, 2, \dots, m$ ), the displacement can be calculated as  $\Delta \mathbf{r}_i(\Delta t) = \mathbf{r}(i\Delta t) - \mathbf{r}(i\Delta t - \Delta t)$ . Once  $\Delta \mathbf{r}$  per  $\Delta t$  is known, the relative displacement per larger time intervals of  $k\Delta t$ , measured between  $(i - 1)\Delta t$  and  $(i - 1 + k)\Delta t$ , can be simply obtained as  $\Delta \mathbf{r}_i(k\Delta t) = \sum_{j=i}^{i-1+k} \Delta \mathbf{r}_j(\Delta t)$ , where  $i = 1, 2, \dots, m + 1 - k$ . This means that we can calculate  $m + 1 - k$  values of  $\Delta \mathbf{r}_i(k\Delta t)$ . Clearly, the number of  $\Delta \mathbf{r}_i(k\Delta t)$  values is large when  $k \ll m$ . Under this condition, a highly converged mean square displacement can be obtained from

$$\overline{[\Delta \mathbf{r}(k\Delta t)]^2} = \frac{\sum_{i=1}^{m+1-k} |\Delta \mathbf{r}_i(k\Delta t)|^2}{m + 1 - k} \quad (12)$$

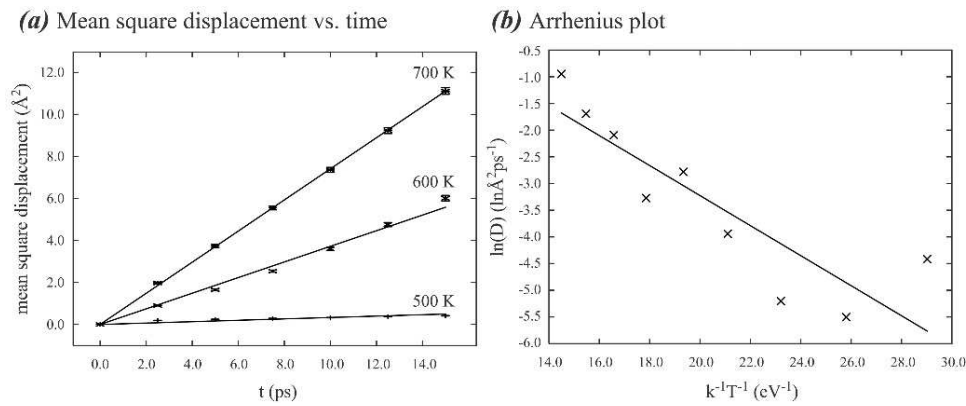
This mean square displacement is a linear function of time  $t$ . In particular,  $\overline{[\Delta \mathbf{r}(k\Delta t)]^2} = 6Dt$ , where  $D$  is diffusivity [24]. Fitting the MD data to  $6Dt$  in a small time range  $t \ll t_{tot}$  (i.e.,  $k \ll m$ ) allows us to determine diffusivity  $D$ . Eq. (4) can be used to estimate the uncertainty of this linear fit.

The MD procedures described above can be used to calculate diffusivities at different temperatures. The results can be fitted to the Arrhenius equation to get the pre-exponential factor  $D_0$  and activation energy  $Q$  of hydrogen diffusion in aluminium [13]. Eq. (4) can also be used to estimate error of this Arrhenius fit.

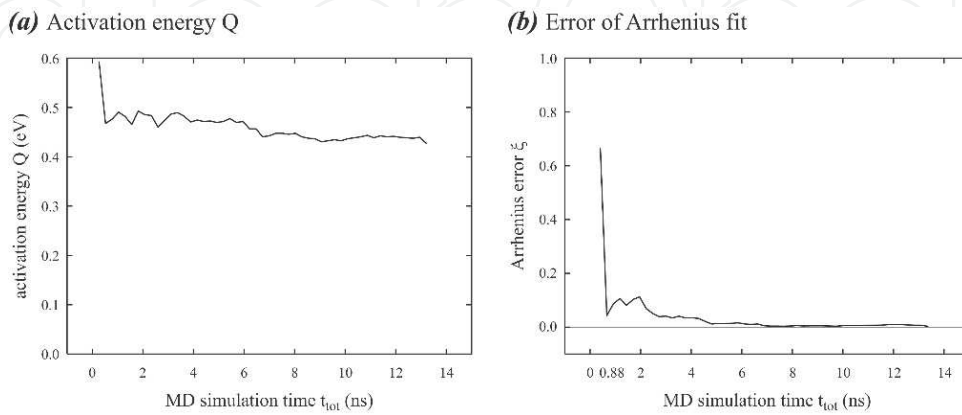
Based on an NVT ensemble, MD simulations are performed to calculate the mean square displacement of the hydrogen atom at various temperatures between 400 and 800 K using  $t_{tot} = 0.88 \text{ ns}$ ,  $dt = 0.001 \text{ ps}$  and  $\Delta t = 0.0088 \text{ ps}$ . The mean square displacements for a small time range ( $t < 15 \text{ ps}$ ) are fitted to  $6Dt$ . A small time range is used to increase the terms in the average sum. For example, for  $t = 15 \text{ ps}$ , the total number of terms in Eq. (12) equals  $N = m + 1 - k = t_{tot}/\Delta t + 1 - t/\Delta t > 98296$ . The MD mean square displacement results and the fitted  $6Dt$  lines

are shown in **Figure 5(a)** for three chosen temperatures 500, 600 and 700 K. The diffusivities derived from the mean square displacement are fitted to the Arrhenius equation, and the results are shown in **Figure 5(b)**. It can be seen that although the mean square displacement appears to satisfy well the linear function of time, the statistical error for the Arrhenius fit is significant.

To examine convergence of the results with respect to simulation time  $t_{tot}$ , extensive MD simulations at a variety of temperatures are performed using  $dt = 0.001$  ps and  $\Delta t = 4.4$  ps. Arrhenius fits are performed at different total MD simulation time  $t_{tot}$ , and the error parameter  $\xi$  for the Arrhenius fits is calculated using Eq. (9). The results for the fitted activation energy and the associated error parameter as a function of  $t_{tot}$  are shown respectively in **Figure 6(a)** and **(b)**. It can be seen that the activation energy approaches a constant value after the simulation time reaches 10 ns and above. Correspondingly, the Arrhenius error reduces to near zero at  $t \geq 10$  ns. To verify that highly converged results are indeed obtained at  $t_{tot} = 13.2$  ns, the corresponding mean square displacement as a function of time at selected temperatures and the Arrhenius plot are shown respectively in **Figure 7(a)** and **(b)**. It can be seen that ideally linear plots are obtained for both mean square displacement and Arrhenius fit, indicating

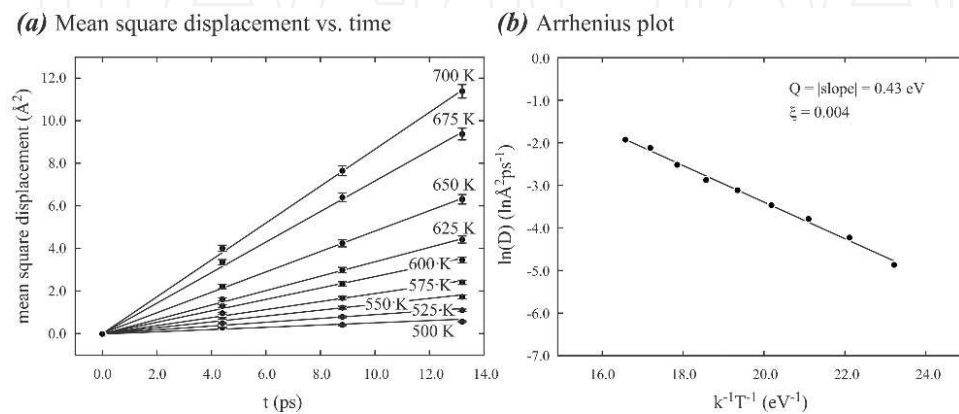


**Figure 5.** Uncertainty examination of hydrogen diffusion parameters in aluminium calculated from MD simulations at  $t_{tot} = 0.88$  ns,  $dt = 0.001$  ps and  $\Delta t = 0.0088$  ps. (a) Mean square displacement and (b) Arrhenius plot.



**Figure 6.** Convergence of diffusion calculation as a function of simulation time. (a) Activation energy and (b) Arrhenius error.

negligible errors. Interestingly, the activation energy determined from the slope of the Arrhenius fit,  $Q = 0.43$  eV, is in excellent agreement with MS calculation at 0 K,  $Q = 0.41$  eV [13]. Note that MS can only be applied for a single atomic diffusion mechanism as in the present case. MD simulations can be applied to alloyed and defected systems that may involve thousands or more different atomic jump paths.



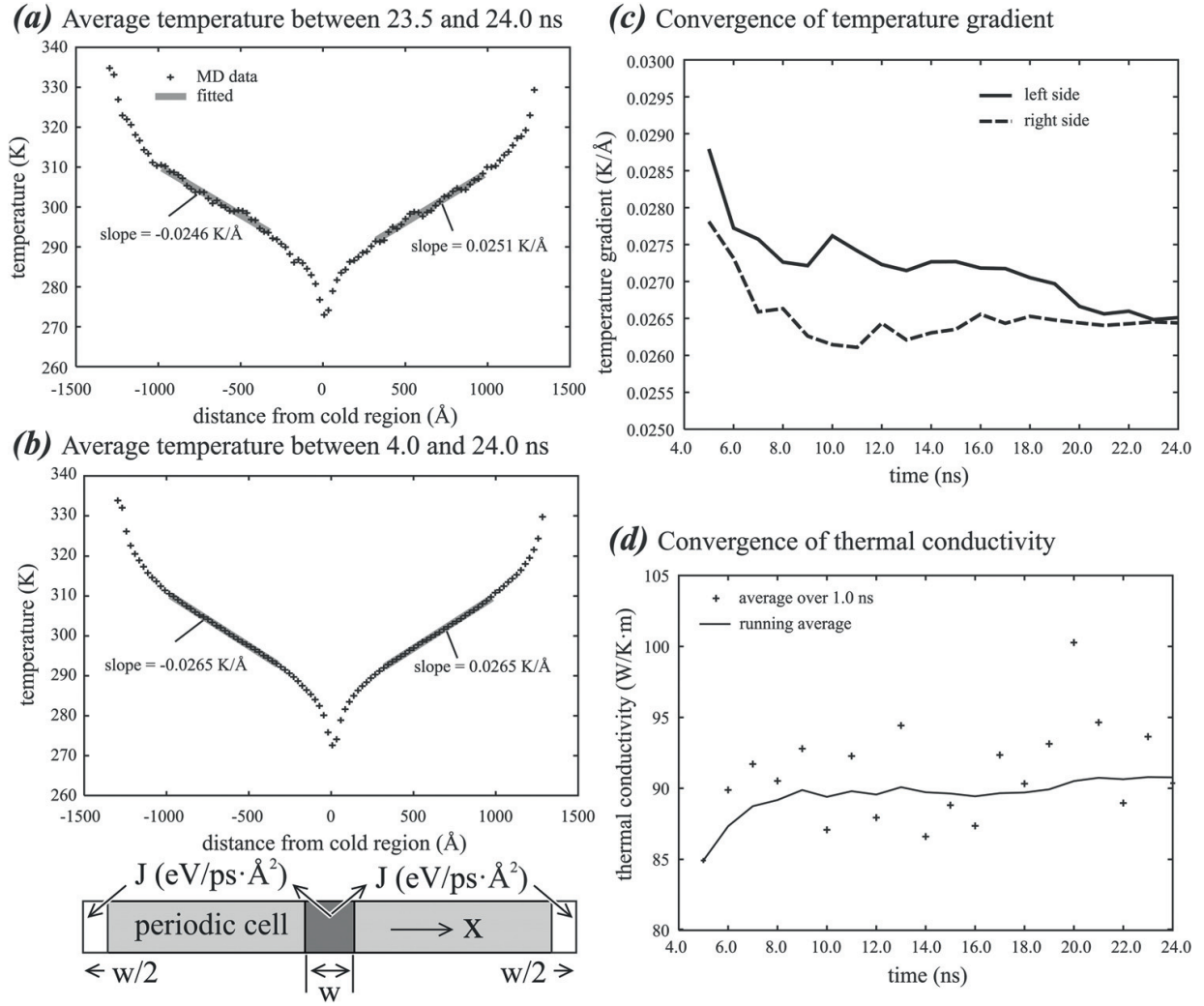
**Figure 7.** Statistical uncertainty examination of hydrogen diffusion parameters in aluminium calculated from MD simulations at  $t_{tot} = 13.2$  ns,  $dt = 0.001$  ps and  $\Delta t = 4.4$  ps. (a) Mean square displacement and (b) Arrhenius plot.

## 9. Thermal conductivity

Another good example to examine uncertainty is thermal conductivity calculations which are usually associated with large statistical errors. Here, we explore calculations of thermal conductivities for a bulk GaN crystal using a ‘direct method’ [12]. The geometry of such a method is illustrated in the left bottom legend of **Figure 8**. Assuming that heat flux is along the  $x$ -axis, two regions of width  $w$  are created in the cell. One region is in the middle, and the other region is on the two ends (due to the periodic boundary condition, the two regions of width  $w/2$  shown in the legend are in fact one region). Through velocity rescaling, a constant heat flux of  $J$  (say, in unit eV/ps·Å²) is continuously removed from the middle region and an exactly the same amount of heat flux is continuously added to the end region. When a steady state is reached, this creates a temperature gradient  $\partial T/\partial x$  from the cold (middle) to the hot (end) regions. This temperature gradient can then be used to calculate thermal conductivity  $\kappa$  through the Fourier’s law

$$\kappa = \frac{-J}{\partial T/\partial x} \quad (13)$$

Our calculations use the GaN literature potential developed by Bere and Serra [25, 26]. A wurtzite GaN crystal with 500 (0001) planes in the  $x$ -direction, 6 ( $\bar{1}100$ ) planes in the  $y$ -direction and 10 (11 $\bar{2}0$ ) planes in the  $z$ -direction is used. The crystal is uniformly divided into



**Figure 8.** Results of 300 K thermal conductivity calculations. (a) Temperature averaged over a short period of 0.5 ns; (b) temperature averaged over a long period of 20 ns; (c) convergence of temperature gradient and (d) convergence of thermal conductivity.

100 bins along the  $x$ -axis so that position-dependent temperature can be calculated as the time-averaged temperature for each of the bins. MD simulation is performed for 24 ns using an NVE ensemble at an initial temperature of 300 K, a heat flux of  $J = 0.0015 \text{ W/ps} \cdot \text{Å}^2$ , a heat source/sink width of  $w = 60 \text{ Å}$  and a time step size of  $dt = 0.001 \text{ ps}$ . After the first 4 ns is discarded to enable the temperature gradient to reach a steady state, time-averaged temperatures are calculated for the remaining simulations.

To examine the convergence of the temperature gradient calculations, **Figure 8(a)** and **(b)** compares the temperature profiles obtained from a short average time of 0.5 ns (average between 23.5 and 24 ns) and a long average time of 20 ns (average between 4 and 24 ns). It can be seen that extremely scattered data are obtained at the short average time. A related phenomenon is that the temperature gradients obtained from the left and the right sides of the cold region do not closely match, indicating non-convergence. Contrarily, the data averaged over the longer time are extremely smooth, and the temperature gradients obtained from both

sides of the cold region are the same up to the fourth decimal point. This suggests that the statistical margin of the temperature gradients is greatly reduced by increasing the average time. To quantify this, we show the running averages of the left and the right temperature gradients in **Figure 8(c)**. **Figure 8(c)** confirms that although the two temperature gradients differ significantly at short times, they approach the same plateau at  $t \rightarrow 24$  ns.

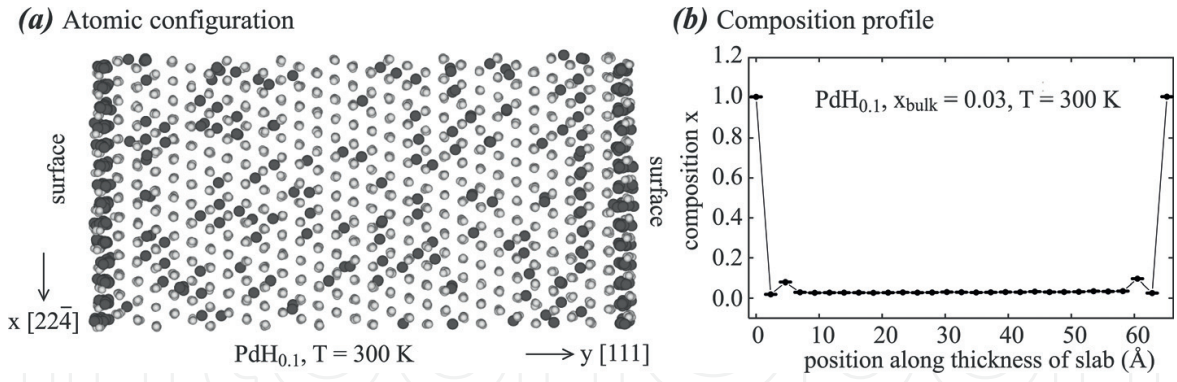
To understand the uncertainty margin of the final thermal conductivity, we divide the 20 ns simulation average time into 20 segments and calculate the thermal conductivities  $\kappa_i$  for each of the segments  $i = 1, 2, \dots, 20$ . We also calculate the running average of these conductivities. The results are shown in **Figure 8(d)**. It can be seen that  $\kappa_i$  is associated with a significant uncertainty margin  $\sigma$ , which can be calculated using Eq. (2). However, the running average quickly approaches a saturated value of  $\sim 91$  W/K·m. Note that the running average at time  $t = 20$  ns is exactly the average measurement of the 20  $\kappa_i$  as defined by Eq. (1). The uncertainty margin of this average is  $\bar{\sigma} = \sigma/\sqrt{20}$ .

## 10. Composition profile

Population of chemical species in a material often needs to be studied. For instance, hydrogen segregation to a crack tip causes hydrogen embrittlement. Hydrogen segregation to a surface impacts the adsorption/desorption performance of solid state hydrogen storage materials. Calculation of composition profiles is a good approach to quantify these segregation effects. However, due to the discrete nature of crystals, a snapshot composition profile is not smooth and is hence associated with a significant uncertainty margin. Here, we demonstrate the calculation of uncertainty margin of a composition profile obtained from MD simulations. We use the hydrogen segregation on (111) palladium surface as an example. The literature Pd-H potential [19] is used.

The fcc palladium crystal contains 5040 Pd atoms with 21  $(11\bar{2})$  planes in the  $x$ - direction, 20 (111) planes in the  $y$ - direction and 12  $(1\bar{1}0)$  planes in the  $z$ - direction. Based on an NPT ensemble to relax stresses, an MD simulation is performed at a temperature of  $T = 300$  K and a hydrogen composition of  $x = 0.1$  (i.e., the chemical formula for the system is  $\text{PdH}_{0.1}$ ). The corresponding numbers of H atoms are randomly introduced into the octahedral interstitial sites so that the initial composition is nominally uniform. To simulate the (111) surfaces, periodic boundary conditions are used in the  $x$ - and  $z$ -directions and a free boundary condition is used in the  $y$ - direction. To ensure a full equilibration between the surfaces and bulk, we first perform a pre-conditioning MD simulation that involves 1.5 ns annealing at 600 K, another 1.5 ns to cool the system from 600 K to the target temperature  $T = 300$  K, and a final 1.5 ns annealing at the target temperature. With the pre-conditioned sample, a second MD simulation is performed for 33 ns at the target temperature, where 100 snapshots of atom positions are recorded on a time interval of 330 ps. Hydrogen composition is calculated for each atomic layer and is averaged over the 100 snapshots. One snapshot of atomic configuration and the averaged composition profile normal to the surface are shown respectively in **Figure 9(a)** and **(b)**. In **Figure 9(b)**, the error bars represent the standard deviation calculated using Eq. (3).





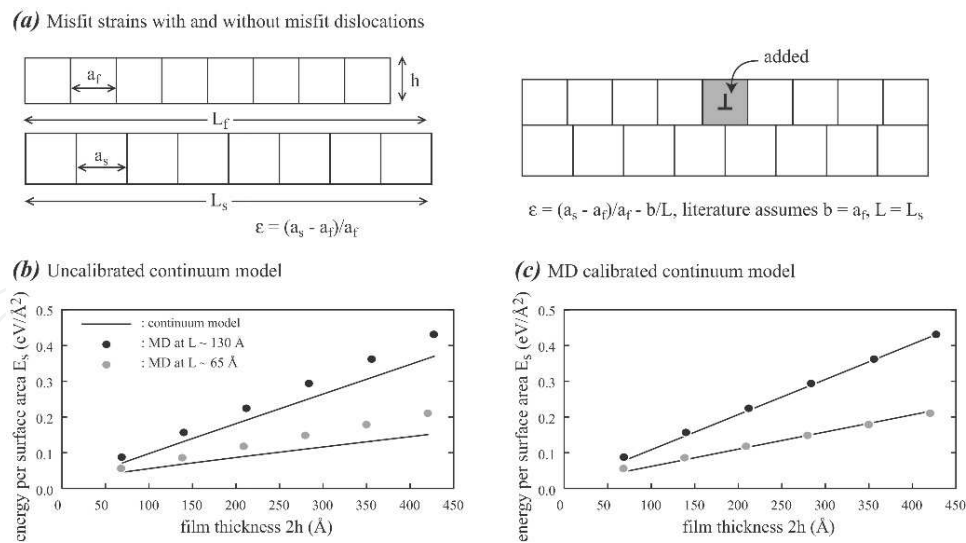
**Figure 9.** Hydrogen segregation on (111) surfaces. (a) A snapshot configuration and (b) averaged composition profile.

Considering that the initial composition is nominally uniform, **Figure 9(a)** shows visually a strong hydrogen surface segregation effect. This is confirmed in **Figure 9(b)** where the surface composition reaches the saturation value of 1.0 as compared to the bulk value of  $x_{bulk} \sim 0.03$ . Interestingly, all data points shown in **Figure 9(b)** have negligible error bars. Also it is important to note that the composition profile is highly symmetric with respect to the two end surfaces and the composition is ideally constant in the bulk region. These further confirm that our high temperature pre-annealing and the ensemble-average of many snapshots have successfully equilibrated the system and reduced the statistical error, resulting in highly converged composition profile.

## 11. Calibration of continuum models

When the uncertainty margin is reduced to near zero, MD simulations are well suited to validate and calibrate other models. Here, we apply MD to calibrate a continuum misfit dislocation theory. As shown in **Figure 10(a)**, consider that a film is grown on a substrate surface. If the film lattice constant  $a_f$  is smaller than the substrate lattice constant  $a_s$ , then the film must be stretched by a strain of  $\varepsilon = (a_s - a_f)/a_f$  in order to grow on the substrate. This creates a large strain energy. However, if misfit dislocations are formed in the film (i.e., adding a lattice unit in the film), then the strain for the film to match the substrate is reduced to  $\varepsilon = (a_s - a_f)/a_f - b/L$ , where  $b$  is the Burgers magnitude of the dislocation and  $L$  is the dislocation spacing. While formation of dislocations reduces lattice mismatch strain energy, it causes additional dislocation energy. The continuum misfit dislocation models express the total system energy as a function of dislocation density so that by minimizing the total energy, the equilibrium dislocation density can be predicted. This concept has been used to develop a variety of continuum misfit dislocation models [27–30].

In previous application of the continuum misfit dislocation models, the dislocation Burgers magnitude  $b$  is usually taken from the film lattice constant, and the dislocation spacing  $L$  is usually taken from the substrate [31–33]. Referring to **Figure 10(a)**, these mean that  $b = a_f$  and  $L = L_s$ . Despite that these choices appear to be natural, they have not been justified. Questions arise on why the Burgers magnitude should be defined by the film lattice constant because if a



**Figure 10.** Calibration of a continuum misfit dislocation theory. (a) Misfit strain with and without misfit dislocation; (b) comparison of MD data with an uncalibrated continuum model and (c) comparison of MD data with a calibrated continuum model.

misfit dislocation can be viewed as an extra plane in the film, it can be equally viewed as a missing plane in the substrate. To understand this issue, we have performed MD analyses of a CdS film on a CdTe substrate [20]. Our MD energy (normalized by interfacial area) is compared with the original continuum model [30] in **Figure 10(b)**. It can be seen that MD results do not perfectly match the continuum prediction. Because the uncertainty margin of MD simulations on dislocation energy and strain energy (or equivalently elastic constant) calculations has been reduced to near zero as shown in **Figures 3 and 4**, the results should match perfectly if the MD and the continuum models are consistent. We find, however, that if an approximation of the dislocation interaction energy array assumed in the original continuum model is corrected, and if the Burgers magnitude is taken from substrate rather than film (i.e.,  $b = a_s$ ), and the dislocation spacing is taken from the film rather than substrate (i.e.,  $L = L_f$ ), then MD results can match the continuum prediction perfectly as shown in **Figure 10(c)**.

The Burger magnitude must be defined from substrate, whereas the dislocation spacing must be defined from the film can be analytically derived. Assume that in a dislocation-free system,  $n_f$  planes of film with plane spacing  $a_f$  are matched with  $n_s$  (assumes that  $n_s = n_f$ ) planes of substrate with plane spacing  $a_s$ . If the substrate is much thicker than the film, it can be assumed to be rigid. Then the film is subject to a mismatch strain of  $(n_f a_s - n_f a_f)/(n_f a_f) = (a_s - a_f)/a_f$ . If we consider a scenario where a half plane is inserted to the film, then the film is subject to a residual strain of  $[n_f a_s - (n_f + 1)a_f]/[(n_f + 1)a_f] = (a_s - a_f)/a_f - a_s/L_f$ . Obviously,  $L_f = (n_f + 1)a_f$  is exactly the length of unstrained film. Alternatively, if we consider a scenario where a half plane is removed from the substrate, then the film is subject to a residual strain of  $[(n_f - 1)a_s - n_f a_f]/(n_f a_f) = (a_s - a_f)/a_f - a_s/L_f$ . Interestingly,  $L_f = n_f a_f$  is again exactly the length of unstrained film. Hence, when the substrate is fixed, a dislocation always causes a consistent residual strain of  $(a_s - a_f)/a_f - a_s/L_f$  whether the dislocation is viewed as inserting a half plane in the film or removing a half plane from the substrate. By comparing the residual strain shown in **Figure 10 (a)**, we see that the magnitude of the Burgers vector  $b$  should be the substrate value  $a_s$  rather

than the film value  $a_f$  and the dislocation spacing  $L$  should be the film value  $L_f$  rather than the substrate value  $L_s$ . This can also be understood in a different way. According to the definition of strain  $\varepsilon = \Delta L/L_0$  where  $\Delta L$  is change of sample length and  $L_0$  is the length of unstrained sample, it is clear that the unstrained film length  $L_f$  should be used in place of  $L_0$  because in our case, the substrate is fixed and only the film is strained. On the other hand, our fixed substrate represents an infinite substrate thickness so that the thickness weighted average plane spacing equals the substrate plane spacing. As a result, the substrate spacing  $b_s$  (or  $a_s$ ) should be taken as the Burgers vector. This example shows how an MD model with reduced uncertainty margin can reveal errors of widely used theories.

## 12. Conclusions

A brief overview is given for uncertainty quantification methods of multiscale models. We demonstrate that rigorous quantification of all model uncertainties is still challenging. However, robust methods are already available today to reliably quantify and reduce the statistical uncertainties of molecular dynamics (MD) simulations. In particular, by averaging over time, the statistical uncertainties of MD calculated properties can always be reduced to near zero as long as the MD simulation is sufficiently long. Counterintuitively, the statistical uncertainties of time-averaged MD simulations are significantly smaller than those of molecular statics simulations especially for large systems with many local energy minimums. For instance, the dislocation energies calculated from time-averaged MD simulations match exactly the continuum predictions, whereas the dislocation energies calculated from MS diverge at large system dimensions. It is also demonstrated that the statistical uncertainties in long MD diffusion simulations can be reduced to such a low level that ideally linear Arrhenius behaviour of diffusion is captured. This means that MD simulations can be used to study diffusion for any complex systems containing any number of diffusion paths. This is extremely important considering that the past MS method to calculate diffusion energy barrier is usually only applicable to single, known atomic jump paths. When zero statistical uncertainty margin is achieved, MD simulations have been successfully used to validate and improve a widely-used misfit dislocation theory. Most importantly, zero statistical error means that MD simulations do not introduce additional errors beyond those inherent in the interatomic potential and simplified systems. Such MD simulations, therefore, isolate out other uncertainties, facilitating their quantifications. All these show that when statistical uncertainties are quantified and reduced, MD simulations can impact material research that would be otherwise impossible.

## Acknowledgements

The Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the US Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. The authors gratefully acknowledge research support from the U.S. Department of

Energy, Office of Energy Efficiency, and Renewable Energy, Fuel Cell Technologies Office, under Contract Number DE-AC04-94AL85000.

## Author details

Xiaowang Zhou<sup>1\*</sup> and Stephen M. Foiles<sup>2</sup>

\*Address all correspondence to: [xzhou@sandia.gov](mailto:xzhou@sandia.gov)

1 Sandia National Laboratories, Livermore, California, USA

2 Sandia National Laboratories, Albuquerque, New Mexico, USA

## References

- [1] Finnis M. Interatomic Forces in Condensed Matter, Oxford Series on Materials Modelling. Oxford: Oxford University Press; 2003.
- [2] Plimpton S. Fast parallel algorithms for short-range molecular-dynamics. *Journal of Computational Physics*. 1995;**117**:1–19.
- [3] LAMMPS download site: [lammmps.sandia.gov](http://lammmps.sandia.gov)
- [4] Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A (Atomic, Molecular, and Optical Physics)*. 1985;**31**:1695–1697.
- [5] Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*. 1981;**52**:7182–7190.
- [6] Cailliez F, Pernot P. Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation. *Journal of Chemical Physics*. 2011;**134**:054124.
- [7] Moore AP, Deo C, Baskes MI, Okuniewski MA, McDowell DL. Understanding the uncertainty of interatomic potentials' parameters and formalism. *Computational Materials Science*. 2017;**126**:308–320.
- [8] Medlin DL, Hattar K, Zimmerman JA, Abdeljawad F, Foiles SM. Defect character at grain boundary facet junctions: Analysis of an asymmetric  $\Sigma = 5$  grain boundary in Fe. *Acta Materialia*. 2017;**124**:383–396.
- [9] Koslowski M, Strachan A. Uncertainty propagation in a multiscale model of nanocrystalline plasticity. *Reliability Engineering & System Safety*. 2011;**96**:1161–1170.
- [10] Chernatynskiy A, Phillpot SR, LeSar R. Uncertainty quantification in multiscale simulation of materials: A prospective. *Annual Review of Materials Research*. 2013;**43**: 157–182.



- [11] Dienstfrey A, Phelan Jr FR, Christensen S, Strachan A, Santosa F, Boisvert R. Uncertainty quantification in materials modelling. *The Journal of the Minerals, Metals and Materials Society*. 2014;**66**:1342–1344.
- [12] Zhou XW, Aubry S, Jones RE, Greenstein A, Schelling PK. Towards more accurate molecular dynamics calculation of thermal conductivity: Case study of GaN bulk crystals. *Physical Review B (Condensed Matter)*. 2009;**79**:115201.
- [13] Zhou XW, El Gabaly F, Stavila V, Allendorf MD. Molecular dynamics simulations of hydrogen diffusion in aluminum. *Journal of Physical Chemistry C*. 2016;**120**:7500–7509.
- [14] Zhou XW, Moody NR, Jones RE, Zimmerman JA, Reedy ED. Molecular-dynamics-based cohesive zone law for brittle interfacial fracture under mixed loading conditions: Effects of elastic constant mismatch. *Acta Materialia*. 2009;**57**:4671–4686.
- [15] Lloyd JT, Zimmerman JA, Jones RE, Zhou XW, McDowell DL. Finite element analysis of an atomistically derived cohesive model for brittle fracture. *Modelling and Simulation in Materials Science and Engineering*. 2011;**19**:065007.
- [16] Zhou XW, Ward DK, Foster ME. An analytical bond-order potential for the aluminium copper binary system. *Journal of Alloys and Compounds*. 2016;**680**:752–767.
- [17] Hirth JP, Lothe J. *Theory of Dislocations*. New York: McGraw-Hill; 1968.
- [18] Mandel J. *Statistical Analysis of Experimental Data*. Weinheim: Wiley; 1964.
- [19] Zhou XW, Zimmerman JA. An embedded-atom method interatomic potential for Pd-H alloys. *Journal of Materials Research*. 2008;**23**:704–718.
- [20] Zhou XW, Ward DK, Zimmerman JA, Cruz-Campa JL, Zubia D, Martin JE, van Swol F. An atomistically validated continuum model for strain relaxation and misfit dislocation formation. *Journal of the Mechanics and Physics of Solids*. 2016;**91**:265–277.
- [21] Zhou XW, Ward DK, Martin JE, van Swol FB, Cruz-Campa, JL, and Zubia D. Stillinger-Weber potential for the II-VI elements Zn-Cd-Hg-S-Se-Te. *Physical Review B (Condensed Matter)*. 2013;**88**:085309.
- [22] Zhou XW, Sills RB, Ward DK, Karnesky RA. Atomistic calculations of dislocation core energy in aluminium. *Physical Review B (Condensed Matter)*. 2017;**95**:054112.
- [23] Donnay JDH, Ondik HM, editors. *Crystal Data, Determinative Tables*. 3rd ed. Vol. 2 (Inorganic Compounds). USA: U. S. Department of Commerce, National Bureau of Standards, and Joint Committee on Powder Diffraction Standards; 1973.
- [24] Reif F. *Fundamentals of Statistical and Thermal Physics*. New York: McGraw-Hill; 1965.
- [25] Bere A, Serra A. Atomic structure of dislocation cores in GaN. *Physical Review B (Condensed Matter)*. 2002;**65**:205323.
- [26] Bere A, Serra A. On the atomic structures, mobility and interactions of extended defects in GaN: Dislocations, tilt and twin boundaries. *Philosophical Magazine*. 2006;**86**:2159–2192.



- [27] Jain SC, Gosling TJ, Willis JR, Totterdell DHJ, Bullough R. A new study of critical layer thickness, stability and strain relaxation in pseudomorphic  $\text{Ge}_x\text{Si}_{1-x}$  strained epilayers. *Philosophical Magazine A: Physics of Condensed Matter Structure Defects and Mechanical Properties*. 1992;**65**:1151–1167.
- [28] Jain SC, Harker AH, Cowley RA. Misfit strain and misfit dislocations in lattice mismatched epitaxial layers and other systems. *Philosophical Magazine A: Physics of Condensed Matter Structure Defects and Mechanical Properties*. 1997;**75**:1461–1515.
- [29] Willis JR, Jain SC, Bullough R. The energy of an array of dislocations—Implications for strain relaxation in semiconductor heterostructures. *Philosophical Magazine A: Physics of Condensed Matter Structure Defects and Mechanical Properties*. 1990;**62**:115–129.
- [30] Nix WD. Mechanical-properties of thin films. *Metallurgical Transactions A: Physical Metallurgy and Materials Science*. 1989;**20**:2217–2245.
- [31] Maroudas D, Zepeda-Ruiz LA, Weinberg, WH. Interfacial stability and misfit dislocation formation in InAs/GaAs(110) heteroepitaxy. *Surface Science*. 1998;**411**:L865–L871.
- [32] Payne AP, Nix WD, Lairson BM, Clemens BM. Strain relaxation in ultrathin films - a modified theory of misfit-dislocation energetics. *Physical Review B (Condensed Matter)*. 1993;**47**:13730–13736.
- [33] Pizzagalli L, Cicero G, Catellani A. Theoretical investigations of a highly mismatched interface: SiC/Si(001). *Physical Review B (Condensed Matter)*. 2003;**68**:195302.

