

RESEARCH ARTICLE | MARCH 07 2019

Improving collective variables: The case of crystallization

Yue-Yu Zhang ; Haiyang Niu; GiovanniMaria Piccini ; Dan Mendels ; Michele Parrinello



J. Chem. Phys. 150, 094509 (2019)

<https://doi.org/10.1063/1.5081040>



CrossMark



APL Energy

Latest Articles Online!

Read Now

Improving collective variables: The case of crystallization

Cite as: *J. Chem. Phys.* **150**, 094509 (2019); doi: [10.1063/1.5081040](https://doi.org/10.1063/1.5081040)

Submitted: 13 November 2018 • Accepted: 8 February 2019 •

Published Online: 7 March 2019



View Online



Export Citation



CrossMark

Yue-Yu Zhang,^{1,2} Haiyang Niu,^{1,2} GiovanniMaria Piccini,^{1,2} Dan Mendels,^{1,2} and Michele Parrinello^{1,2,a)}

AFFILIATIONS

¹Department of Chemistry and Applied Biosciences, ETH Zurich, c/o USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Ticino, Switzerland

²Institute of Computational Science, Università della Svizzera Italiana (USI), Via Giuseppe Buffi 13, CH-6900 Lugano, Ticino, Switzerland

^{a)}E-mail: parrinello@phys.chem.ethz.ch

ABSTRACT

Several enhanced sampling methods, such as umbrella sampling or metadynamics, rely on the identification of an appropriate set of collective variables. Recently two methods have been proposed to alleviate the task of determining efficient collective variables. One is based on linear discriminant analysis; the other is based on a variational approach to conformational dynamics and uses time-lagged independent component analysis. In this paper, we compare the performance of these two approaches in the study of the homogeneous crystallization of two simple metals. We focus on Na and Al and search for the most efficient collective variables that can be expressed as a linear combination of X-ray diffraction peak intensities. We find that the performances of the two methods are very similar. Wherever the different metastable states are well-separated, the method based on linear discriminant analysis, based on its harmonic version, is to be preferred because simpler to implement and less computationally demanding. The variational approach, however, has the potential to discover the existence of different metastable states.

© 2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5081040>

I. INTRODUCTION

Many chemical and physical phenomena are characterized by the occurrence of long lived metastable states. Under these circumstances, accurate sampling becomes computationally expensive or even prohibitive. In order to overcome this problem, many methods have been suggested.¹ A large fraction of these methods depends on the definition of collective variables (CVs). Typical examples are umbrella sampling,²⁻⁵ metadynamics,⁶⁻⁹ and variationally enhanced sampling.¹⁰ The efficiency of these simulations depends very much on the quality of the CV, and hence, the finding and improving of CVs is the object of intense investigations.¹¹

In our group, two methods have recently been developed, harmonic linear discriminant analysis (HLDA)¹² and variational approach to conformational dynamics (VAC).¹³ In HLDA, one constructs low dimensional CVs from the local fluctuations in the different metastable states. In contrast, in the VAC approach, one

starts with a biased simulation with non-optimal CVs and attempts to improve this initial guess using a variational principle approach that is based on the time-lagged independent component analysis (TICA).

In this work, we compare the performance of HLDA-generated CVs with those obtained using VAC. We focus on a typical application area of enhanced sampling methods, namely, homogeneous crystallization. In particular, we shall present results on two systems, Na and Al, already studied elsewhere.^{4,5,9,14}

It has recently been shown¹⁵ that the X-ray diffraction (XRD) intensities can be useful CVs. However, sometimes a single peak is not sufficient. Thus we shall search with HLDA and VAC for the best linear combinations of the diffraction peak intensities. We are not implying here that XRD intensities are necessarily the best possible CVs to study nucleation. However, they have been successfully used and provided a good example on how to improve from reasonable CVs, which is the situation commonly encountered in the practice.

II. METHODS

A. Well-tempered metadynamics

Here, we use as an enhanced sampling method well-tempered metadynamics (WTMetaD).^{16,17} We recall that WTMetaD is a procedure in which the dynamical evolution of the system is altered by the addition of an external bias potential $V(s)$ periodically updated as

$$V_n(s) = V_{n-1}(s) + G(s, s_n) \exp\left[-\frac{1}{\gamma-1}\beta V_{n-1}(s_n)\right], \quad (1)$$

where γ is the bias factor and $\beta = 1/k_B T$ is the inverse temperature. The modification of the potential in Eq. (1) results from the multiplication of a Gaussian kernel $G(s, s_n)$ centered at the current CV value s_n and scaled by $\exp\left[-\frac{1}{\gamma-1}\beta V_{n-1}(s_n)\right]$. This scaling factor decreases as $1/n$; thus, the change of the external bias potential becomes smaller as the metadynamics simulation progresses. The bias potential $V(s(\mathbf{R}))$ asymptotically takes the form

$$V(s) = -\left(1 - \frac{1}{\gamma}\right)F(s), \quad (2)$$

where $F(s)$ is the free energy associated with the CV s . The validity of Eq. (2) has recently been rigorously proven.¹⁸

B. Harmonic linear discriminant analysis

Very recently, a simple way of obtaining efficient CVs, called harmonic linear discriminant analysis (HLDA), has been proposed.¹² In its simplest version, one assumes that there are two separated states that can be identified by N_d descriptors $d_i(\mathbf{R})$ that are functions of the atomic coordinates \mathbf{R} and are arranged to form a vector $\mathbf{d}(\mathbf{R})$. We assume that the average values in the two basins, μ_A and μ_B , are different. The fluctuation of $\mathbf{d}(\mathbf{R})$ in the two states is given by the covariance matrices Σ_A and Σ_B .

The idea of HLDA is to find the direction along which the projected data of the two states are well-separated. This is obtained by maximizing the ratio of the between-class variation $S_B = (\mu_A - \mu_B)(\mu_A - \mu_B)^T$ and the within-class variance calculated by a harmonic average $S_W = \Sigma_A \Sigma_B / (\Sigma_A + \Sigma_B)$, which leads to the object function

$$J(W) = \frac{W^T S_B W}{W^T S_W W}, \quad (3)$$

where $J(W)$ is maximized by

$$W = \left(\frac{1}{\Sigma_A} + \frac{1}{\Sigma_B}\right)(\mu_A - \mu_B). \quad (4)$$

Thus the HLDA based CV takes the form

$$s^H(\mathbf{R}) = W^T \mathbf{d}(\mathbf{R}) = (\mu_A - \mu_B)^T \left(\frac{1}{\Sigma_A} + \frac{1}{\Sigma_B}\right) \mathbf{d}(\mathbf{R}). \quad (5)$$

This procedure has been inspired from linear discriminant analysis (LDA). However, it differs from the standard version of LDA in that the within-class variance is estimated from the harmonic average of the covariance matrices and not by the arithmetic average as done in LDA. The reasons for this choice have been illustrated in Ref. 12 and also in Ref. 19.

An extension of HLDA to the case in which several metastable states need to be considered is possible and following the nomenclature of the artificial intelligence community we refer to this as a multi-class situation and thereby we shall name the method as multi-class HLDA (MC-HLDA).¹⁹ We assume that there are M classes, the between-class variance matrix in this case is expressed as $S_B = \sum_{i=1}^M (\mu_i - \mu)(\mu_i - \mu)^T$, where $\mu = \frac{1}{M} \sum_{i=1}^M \mu_i$, and the within-class variance matrix is expressed as $S_W = \mathbf{I}/((\mathbf{1}/\Sigma_1) + (\mathbf{1}/\Sigma_2) + \dots + (\mathbf{1}/\Sigma_M))$. The projection matrix $K = [W_1|W_2|\dots|W_M]$ that maximizes the ratio $J(K) = \frac{K^T S_B K}{K^T S_W K}$ is the one whose columns are the eigenvectors corresponding to the generalized eigenvalue problem

$$(S_B - \lambda_i S_W) W_i = 0. \quad (6)$$

The optimized CVs are thus given by

$$s_i^V(\mathbf{R}) = W_i^T \mathbf{d}(\mathbf{R}). \quad (7)$$

Notice that, because S_B is of rank $(M - 1)$ or less, there will be at most $M - 1$ eigenvectors with non-zero eigenvalues λ_i .

C. Variational approach to conformational dynamics

Another way one can use to optimize CVs is based on a variational approach to conformational dynamics (VAC).¹³ It has been

TABLE I. The parameters of the WTMetaD simulation of Na and Al. The parameters are: target temperature of the thermostat (T), biased CVs (CVs), bias factor of the well-tempered distribution (γ), height of the Gaussians (ω), deviation of the Gaussians (σ), and time interval. Estimated melting temperature $T_m = 366$ K for Na²⁸ and $T_m = 926$ K for Al.²⁹

Number of trajectories	System	T (K)	T/ T_m	CVs	γ	ω (kJ/mol)	σ (a.u.)	Simulation time (ns)	Time interval (ps)
1	Na	375	1.025	$I^{(011)}$	30	3	2	350	1
2	Na	375	1.025	HLDA	30	3	0.012	600	1
3	Na	375	1.025	VAC	30	3	0.012	600	1
4	Al	850	0.918	$I^{(111)}$ HLDA	70	7.5	2	600	1
5	Al	850	0.918	Eig1 and Eig2 VAC	70	7.5	0.014	600	1
6	Al	850	0.918	Eig1 and Eig2	70	7.5	0.014	600	1

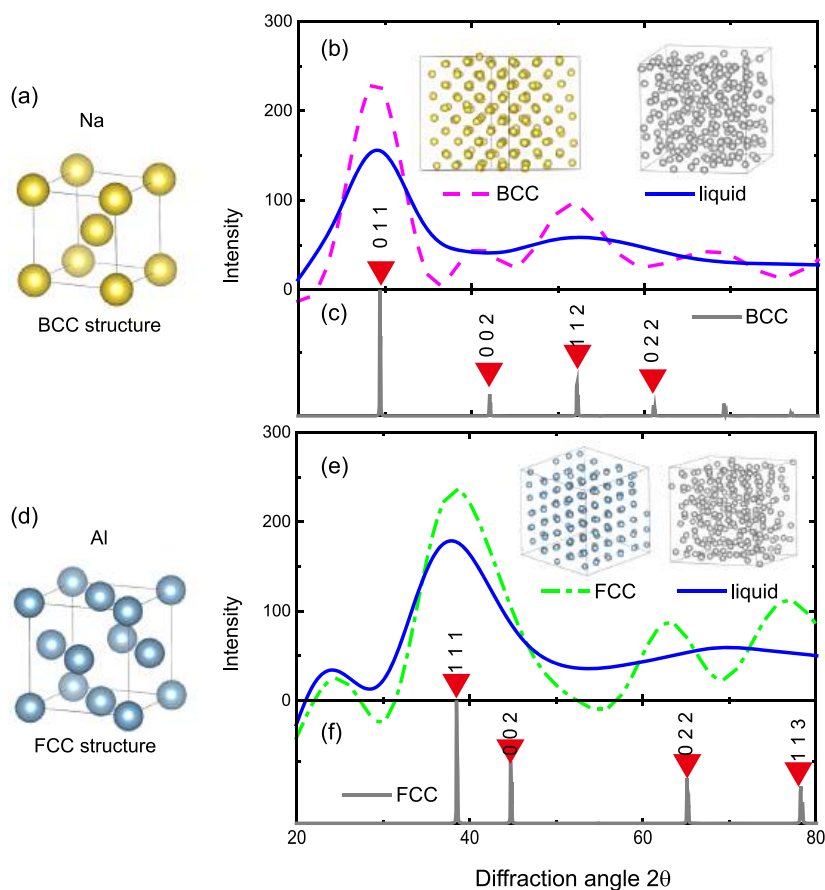


FIG. 1. (a) Na unit cell in the BCC phase. (b) Simulated XRD patterns of BCC and liquid Na at 375 K with 250 atoms. The cutoff radius R_c is set to 11 Å. Inset: two snapshots of the BCC (left) and liquid (right) phases. (c) Simulated XRD pattern for the ideal Na BCC structure with lattice constant 4.29 Å. The first four peaks are highlighted in red triangles. (d) Al unit cell in the FCC phase. (e) Simulated XRD patterns for FCC and liquid Al at 850 K with 256 atoms. The cutoff radius R_c is set to 9 Å. Inset: two snapshots of the FCC (left) and liquid (right) phases. (f) Simulated XRD pattern of the ideal Al FCC structure with lattice constant 4.05 Å.

shown that time-lagged independent component analysis (TICA) can provide an optimal solution of VAC. TICA is a well-known method for blind source separation problems in signal processing and recently has been applied in finding slow modes in molecular dynamics simulations.²⁰⁻²⁶

The first version of this method assumes that a sufficient number of trajectories in which the system underwent transitions from one well to another were already available.^{21,22} However, recently, this method has been extended and it is now possible to perform a VAC analysis starting from biased simulations by WTMetaD with

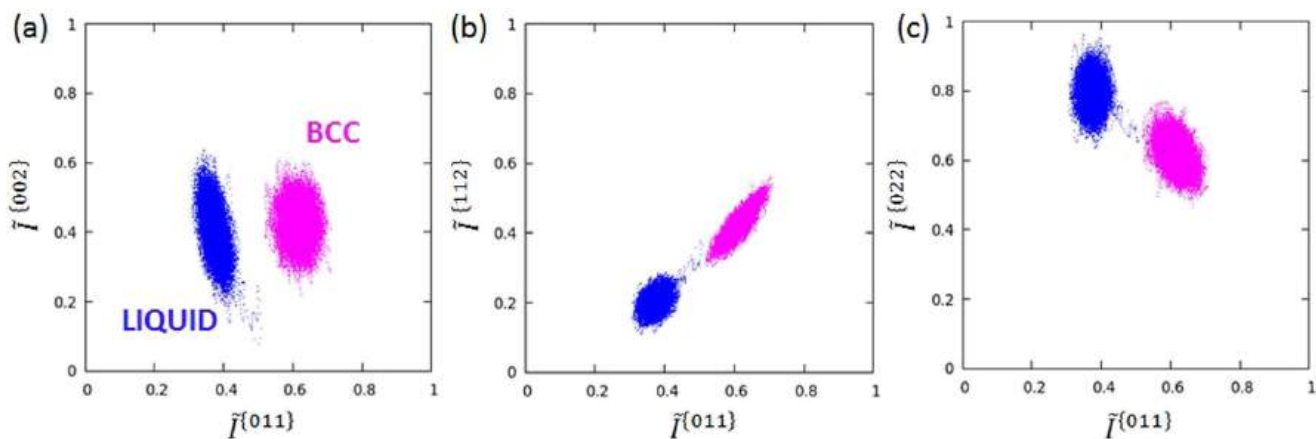


FIG. 2. Points sampled in the unbiased liquid and BCC basins are projected to different descriptors: (a) $\tilde{I}^{(011)} - \tilde{I}^{(002)}$, (b) $\tilde{I}^{(011)} - \tilde{I}^{(112)}$, and (c) $\tilde{I}^{(011)} - \tilde{I}^{(022)}$.

TABLE II. HLDA and VAC coefficients for the linear combination of XRD peak intensities for crystallization of Na.

	$I^{\{011\}}$	$I^{\{002\}}$	$I^{\{112\}}$	$I^{\{022\}}$
HLDA	0.857	0.201	0.464	0.096
VAC	0.923	0.011	-0.331	-0.198

non-optimal CVs.¹³ These CVs are then optimized. In the application of this principle, one expands the slow modes into a basis set of N_d descriptors $\mathbf{d}_i(\mathbf{R}_t)$ ($i = 1, 2, \dots, N_d$). The descriptors are normalized to one, and the average value is subtracted to get a quantity of zero mean. The best linear combination in these basis sets is obtained by maximizing the ratio $f(\mathbf{W}_1)$ with respect to a projection vector \mathbf{W}_1 ,

$$f(\mathbf{W}_1) = \frac{\mathbf{W}_1^T |\tilde{\mathbf{C}}(\tau)| \mathbf{W}_1}{\mathbf{W}_1^T |\tilde{\mathbf{C}}(0)| \mathbf{W}_1}, \quad (8)$$

where $\mathbf{C}_{ij}(\tau) = E[\mathbf{d}_i(\tilde{t})\mathbf{d}_j(\tilde{t} + \tau)]$ is the time lagged correlation between \mathbf{d}_i and \mathbf{d}_j , $\tilde{\mathbf{C}}(\tau) = [\mathbf{C}(\tau) + \mathbf{C}^T(\tau)]/2$ is the symmetrized correlation matrix, and $\tilde{\mathbf{C}}(0)$ is the correlation at lag time 0. In Refs. 13 and 24, it has been shown that Eq. (8) can be applied also to biased simulations provided that the time scale is modified as follows:

$$d\tilde{t} = e^{\beta(V(s,t)-c(t))} dt, \quad (9)$$

where $V(s, t)$ is the instantaneous value of bias in WTMetaD.

$$c(t) = -\frac{1}{\beta} \log \frac{\int ds e^{-\beta(F(s)+V(s,t))}}{\int ds e^{-\beta F(s)}} \quad (10)$$

is an energy offset, and $F(s)$ is the free energy of the system. The function $c(t)$ asymptotically tends to the reversible work done on the system by the external bias.

Maximizing the ratio in Eq. (8) leads to the generalized eigenvalue problem

$$\tilde{\mathbf{C}}(\tau) \cdot \mathbf{W} = \lambda(\tau) \tilde{\mathbf{C}}(0) \cdot \mathbf{W}. \quad (11)$$

The eigenvectors of Eq. (11) are ordered in descending $\lambda(\tau)$ values. The slowest decreasing modes are then chosen as CVs

$$s_i^V(\mathbf{R}) = \mathbf{W}_i^T \mathbf{d}(\mathbf{R}). \quad (12)$$

D. X-ray diffraction intensities

In a recent paper,¹⁵ it has been suggested that the intensities of the X-ray diffraction (XRD) peaks could act as good CVs since they are physically meaningful and can distinguish between different states.

The XRD intensities are computed by the Debye scattering function²⁷

$$I(Q) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N f_i(Q) f_j(Q) \frac{\sin(Q \cdot R_{ij})}{Q \cdot R_{ij}} W(R_{ij}), \quad (13)$$

where N is the total number of atoms, Q is the modulus of the scattering vector, $f_i(Q)$ and $f_j(Q)$ are the atomic scattering form factors, R_{ij} is the distance between atoms i and j , and $W(R_{ij}) = \sin(\pi R_{ij}/R_c)/(\pi R_{ij}/R_c)$ is a window function that handles the problem of the finite simulation box and R_c is the upper limit of R_{ij} .

III. COMPUTATIONAL DETAILS

A. Details of the molecular dynamics simulations

All simulations were performed in the isothermal-isobaric (NPT) ensemble. We simulated Na and Al using embedded atom models.^{28,29} Biased metadynamics simulations were performed

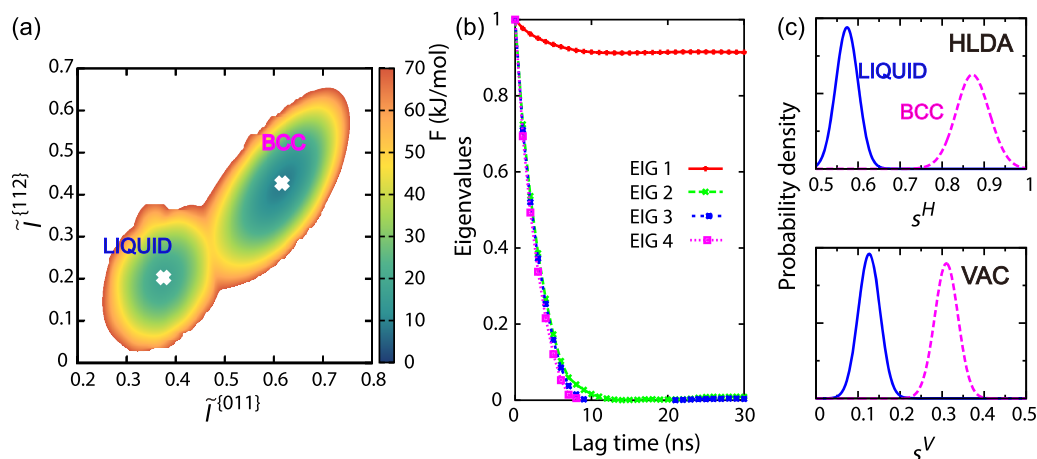


FIG. 3. (a) Free energy surface as a function of $\tilde{I}^{\{011\}}$ and $\tilde{I}^{\{112\}}$ descriptors for the liquid/solid phase transition in Na. The expectation value of each state is marked with a white cross. (b) Evolution of the eigenvalues in VAC as a function of different lag time τ . (c) Probability densities of the unbiased distributions projected onto the HLDA and VAC CVs, with the blue lines and magenta dashed lines corresponding to liquid and BCC phases, respectively.

using the LAMMPS molecular dynamics simulation code³⁰ patched with PLUMED 2 plugin.³¹ The integration of the equations of motion was carried out with a time step of 2 fs. We employed the stochastic velocity rescaling thermostat³² with a relaxation time of 0.1 ps. The target pressure of the isotropic barostat³³ was set to the atmospheric value, and a relaxation time of 10 ps was used.

We simulated the two systems at temperatures close to their melting points (375 K for Na and 850 K for Al). The total numbers of atoms were 250 in Na and 256 in Al.

B. Details of the WTMetad simulations

In order to determine the free energy surface $F(s)$ as a function of the CVs, we used WTMetad.¹⁶ The temporal length of the WTMetad runs ranges from 350 to 600 ns. The calculation details are summarized in Table I.

IV. RESULTS AND DISCUSSION

A. Choosing the descriptors

Both HLDA and VAC require defining an adequate set of descriptors. As mentioned in the Introduction, we shall use some of the XRD peak intensities. The cutoff R_c for Na and Al is 11 Å and 9 Å, respectively. Due to the small system sizes, the peaks resulting from Eq. (13) are broad, as shown in Fig. 1, but they can still distinguish well between solid and liquid.

In order to perform our analysis, we found that the intensities of the first four Bragg peaks were sufficient for our purpose. The four descriptors are labeled by the peak Miller indices $\{hkl\}$. They are $d_1 = I^{[011]}$, $d_2 = I^{[002]}$, $d_3 = I^{[112]}$, and $d_4 = I^{[022]}$ in Na, and $d_1 = I^{[111]}$, $d_2 = I^{[002]}$, $d_3 = I^{[022]}$, and $d_4 = I^{[113]}$ in Al.

In order to make a meaningful comparison of the HLDA and VAC coefficients from the original runs in which only $I^{[011]}$ for Na and $I^{[111]}$ for Al is biased, we rescaled the value $I^{[hkl]}$ to $\tilde{I}^{[hkl]}$, which covers the range $[0, 1]$, where 0 is the minimum value sampled and 1 is the maximum.

In the following, we shall measure the quality of the CVs not only by their ability to discriminate between states but also by the re-crossing frequency. The re-crossing frequency is the number of transitions between solid and liquid per unit time during the WTMetad. They could be measured from the CV trajectory (see Figs. 4 and 10). The rationale for this is that the re-crossing frequency between different basins is an important figure of merit. In fact, the slowest mode not included in the CV determines the re-crossing frequency.³⁴

B. The Na case

From the existing literature, it is known that the $F(s)$ of Na close to 375 K is dominated by 2 minima corresponding to the disordered liquid state and the ordered body centered cubic (BCC) state.⁹

We first applied HLDA to find the direction that well discriminate the two states. In each of them, we estimate the average values and the covariance matrix in 2-ns long unbiased runs, a time that is sufficient for these estimations to converge, and the points sampled are shown in Fig. 2. We then applied Eq. (5) to obtain the CV, and the result is shown in Table II. The descriptor $I^{[011]}$, which

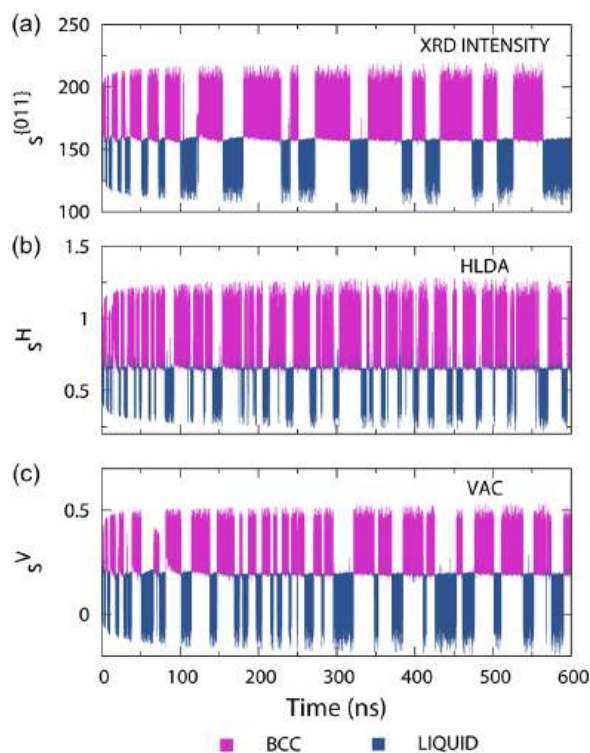


FIG. 4. Time evolution of the (a) XRD intensity CV $s^{[011]}$, (b) HLDA CV s^H , and (c) VAC CV s^V in metadynamics simulations for Na.

corresponds to the highest peak in the XRD pattern, dominates in the HLDA coefficients.

The VAC based calculation was started with a biased run in which only $I^{[011]}$ is used as CV. Following the procedure described in Ref. 13, we extract the slowest decaying modes in the basis of the selected descriptors. The eigenvalue of the VAC equation is displayed in Fig. 3(b) as a function of lag-time. Obviously one

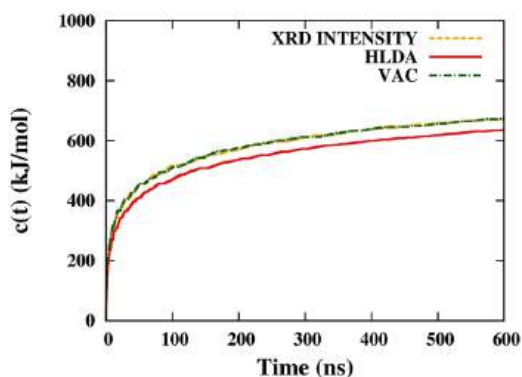


FIG. 5. The offset energy $c(t)$ [Eq. (10)] as a function of simulation time for Na. As to be expected from the theory of metadynamics, after an initial steep rise, the offset energy $c(t)$ reaches a slowly increase behavior, indicating that the asymptotic regime in which WTMetad is valid has been reached. In both systems, this regime is reached after about 200 ns.

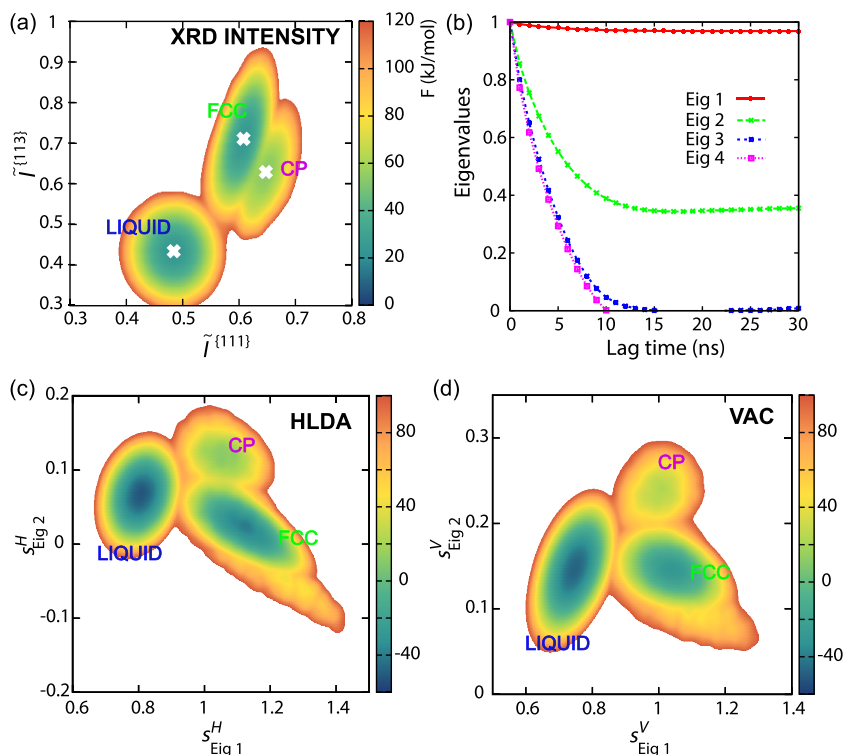


FIG. 6. (a) Free energy surface with respect to descriptors $\tilde{I}^{\{111\}}$ and $\tilde{I}^{\{113\}}$. The expectation value of each state is marked with a white cross. (b) Evolution of the eigenvalues in VAC as a function of lag time τ . [(c) and (d)] FES with respect to two-dimensional HLDA and VAC CVs. The liquid, FCC, and CP basins are well-separated in both CV spaces.

eigenvalue dominates the long time decay. The coefficients for the descriptors are shown in Table II. They are very similar to those of HLDA and also dominated by $I^{\{011\}}$.

The probability distributions related to these two new CVs were computed and shown in Fig. 3(c). Both of them can discriminate the two states, while the transition region is better separated by the direction of the HLDA CV s^H than that of the VAC CV s^V .

We now compare the performance of $I^{\{011\}}$, s^H , and s^V as a metadynamics CVs. Thus two additional runs based on s^H and s^V were performed. In Fig. 4, it is obvious that both s^H and s^V improve substantially the performance of the original simulation. As to be expected from the theory of metadynamics, after an initial step rise, the offset energy $c(t)$ reaches a slowly increasing behavior, indicating that the asymptotic regime in which WTMetad is valid has been

reached (Fig. 5). In both systems, this regime is reached after about 200 ns. If we count the number of recrossing in this regime, we find an average of 3.8 ($I^{\{011\}}$), 9.0 (HLDA), and 7.3 (VAC) re-crossings per 100 ns.

C. The Al case

The other case studied was the crystallization of Al. We first explored the free energy surface biasing $I^{\{111\}}$ at 850 K, close to the melting temperature. A visual inspection of the solid-like minimum reveals that the structure in the solid minimum is a combination of perfect face centered cubic (FCC) and stacking disorder close packed (CP) structures. Fortunately, we found that if we project the free energy surface (FES) onto $I^{\{111\}}$ and $I^{\{113\}}$ by reweighting,³⁵

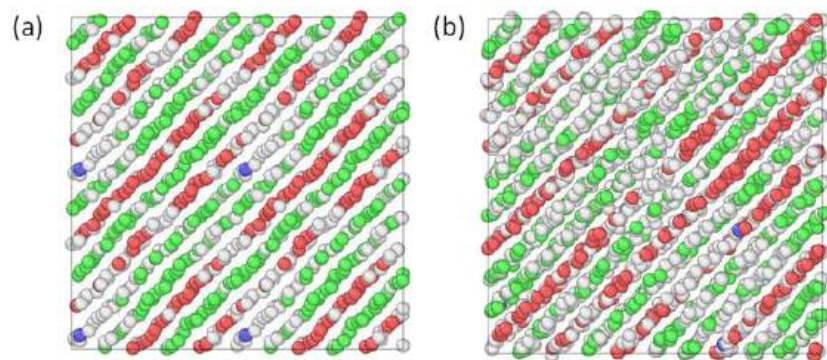


FIG. 7. Atomic model of the CP phase with 2048 atoms. (a) The initial configuration. (b) The configuration after a 2-ns unbiased simulation. The FCC-like atoms are colored in green, the HCP-like atoms are colored in red, and the atoms without a known structure are colored in gray.

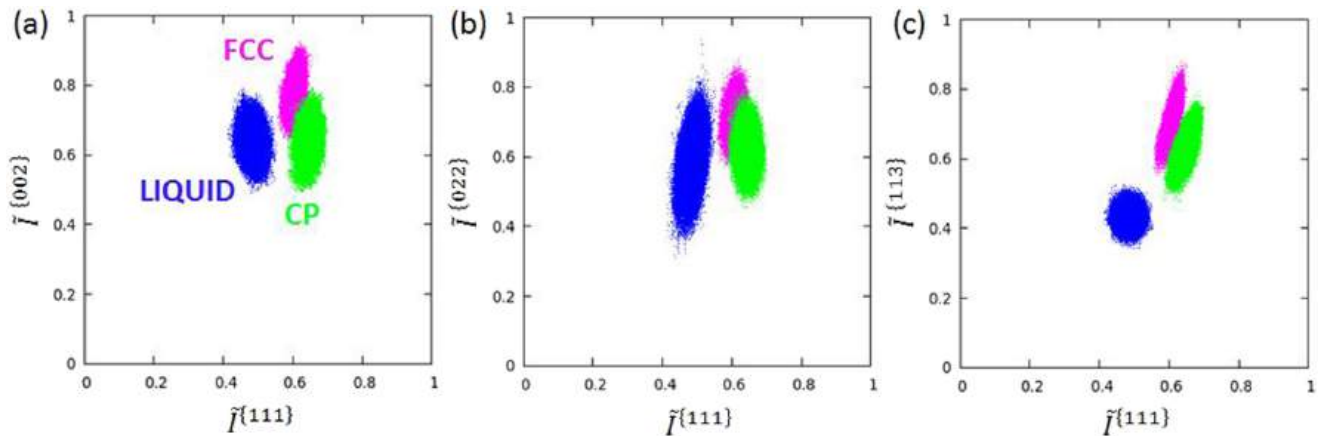


FIG. 8. Points sampled in the unbiased liquid, FCC, and CP basins are projected on different descriptors: (a) $\tilde{I}^{\{111\}} - \tilde{I}^{\{002\}}$, (b) $\tilde{I}^{\{111\}} - \tilde{I}^{\{022\}}$, and (c) $\tilde{I}^{\{111\}} - \tilde{I}^{\{113\}}$.

TABLE III. HLDA and VAC coefficients for the linear combination of XRD peak intensities for crystallization of Al.

		$I^{\{111\}}$	$I^{\{002\}}$	$I^{\{022\}}$	$I^{\{113\}}$
HLDA	Eig1	0.946	0.028	0.077	0.315
	Eig2	0.871	-0.044	-0.237	-0.429
VAC	Eig1	0.695	-0.023	0.217	0.685
	Eig2	0.932	-0.087	-0.220	-0.273

the solid minima are separated [see Fig. 6(a)]. The CP structure is stable even with a larger simulation cell with 2048 atoms, and the stacking faults remain throughout the 2-ns long unbiased simulation (Fig. 7). Such a phase of Al has indeed been stabilized in the experiments performed in cold rolled Al crystal³⁶ and nanocrystalline Al.^{37,38}

Having established somewhat empirically, that we are in the presence of a multi-class problem and can apply the multi-class generalization of HLDA.¹⁹ To prepare for the MC-HLDA

simulation, we performed three independent 2-ns-long runs start from the three states, and on this time scale no transition between these states was observed (Fig. 8). From this using Eq. (7), we obtain two CVs, s_{Eig1}^H and s_{Eig2}^H , and performed a biased run based on these CVs. The result shows not only that sampling is accelerated, but that the three classes are better separated [Fig. 6(c)].

We then constructed a VAC calculation starting from the biased trajectory that uses $I^{\{111\}}$ only as CV. Two eigenvalues dominate the long time decay [Fig. 6(b)], leading also to a 2-dimensional CV space, and the coefficients of the corresponding eigenvectors are listed in Table III. This is an important result since it shows that VAC can find out automatically that the Al case is in a multi-class scenario. This is to be contrasted with HLDA requiring that the classes have to be known beforehand.

Both the HLDA CV s^H and the VAC CV s^V spaces (Fig. 9) can discriminate better between the three basins than the XRD peak intensities (Fig. 8). While the coefficients of the descriptors appear to be different, the two FESs are very similar with the three minima well-separated, apart from a slight relative rotation [Figs. 6(c) and 6(d)].

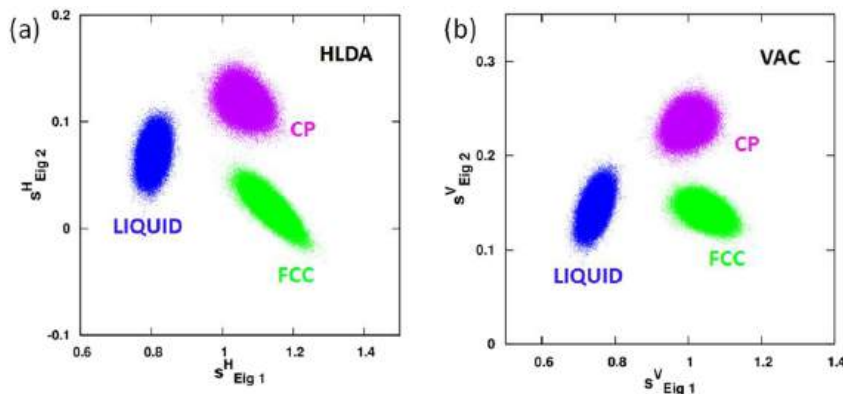


FIG. 9. Points sampled in the unbiased liquid, FCC, and CP basins are projected on (a) HLDA CVs and (b) VAC CVs.

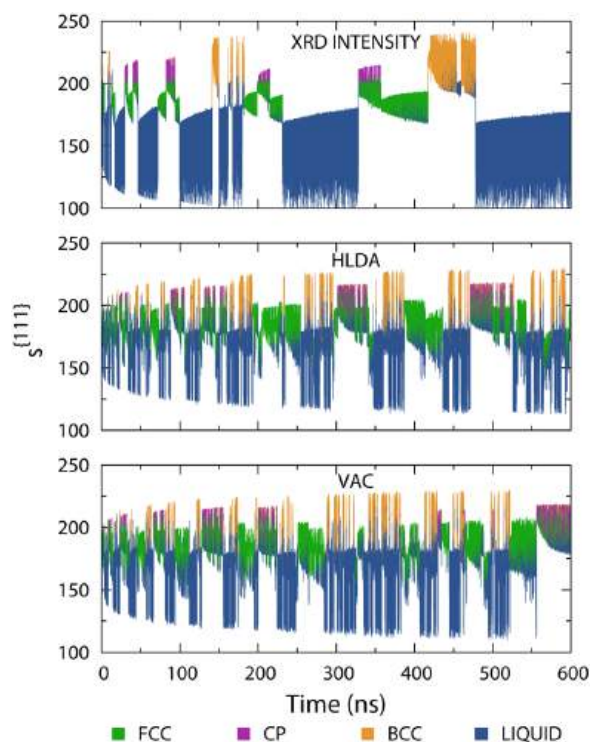


FIG. 10. Time evolution of XRD intensity $s^{(111)}$ in metadynamics simulations based on (a) $s^{(111)}$, (b) HLDA CV s^H , and (c) VAC CV s^V in Al.

To compare the relative performance of the $s^{(111)}$, s^H , and s^V , we performed two extra simulations using s^H and s^V , and the trajectories are shown in Fig. 10. Similarly to what done for Na, we measure how many transitions take place on average every 100 ns after the metadynamics has reached its asymptotic equilibrium (Fig. 11). If we count the number of re-crossings in this regime, we find an average of 0.8 ($s^{(111)}$), 3.0 (HLDA), and 3.3 (VAC) re-crossings per 100 ns.

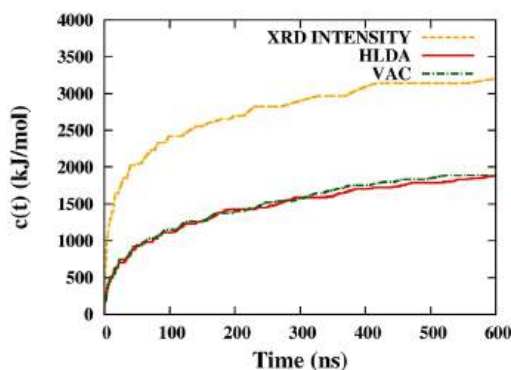


FIG. 11. The offset energy $c(t)$ [Eq. (10)] as a function of simulation time for Al.

V. CONCLUSION

In this paper, we tested the performance of HLDA and VAC in the case of homogeneous nucleation. We applied them to look for the most efficient collective variables that can be expressed as a linear combination of X-ray diffraction peak intensities. Both methods gave a good separation among the different states and substantially enhanced sampling. When the number of relevant classes is known in advance, HLDA is to be preferred to VAC because of its simplicity and the fact that it does not require a preliminary biased run. On the other hand, as we have seen in the case of Al, VAC can automatically find out the different classes that is not a minor advantage.

ACKNOWLEDGMENTS

The authors thank Faidon Brotzakakis, Yi Issac Yang, Valerio Rizzi, and Pablo M. Piaggi for useful discussions. Some of the calculations reported here were performed on the EULER cluster at ETH Zurich, and the others were on the Mönch cluster at the Swiss National Supercomputing Center (CSCS). This research was supported by the VARMET European Union (Grant No. ERC-2014-ADG-670227).

REFERENCES

- B. Peters, *Reaction Rate Theory and Rare Events* (Elsevier, 2003).
- J. S. Van Duijneveldt and D. Frenkel, "Computer simulation study of free energy barriers in crystal nucleation," *J. Chem. Phys.* **96**, 4655 (1992).
- R. M. Lynden-Bell, J. S. Van Duijneveldt, and D. Frenkel, "Free energy changes on freezing and melting ductile metals," *Mol. Phys.* **80**, 801 (1993).
- C. Desgranges and J. Delhommelea, "Molecular simulation of the crystallization of aluminum from the supercooled liquid," *J. Chem. Phys.* **127**, 144509 (2007).
- C. Desgranges and J. Delhommelea, "Molecular insight into the pathway to crystallization of aluminum," *J. Am. Chem. Soc.* **129**, 7012–7013 (2007).
- D. Quigley and P. M. Rodger, "A metadynamics-based approach to sampling crystallisation events," *Mol. Simul.* **35**, 613–623 (2009).
- F. Trudu, D. Donadio, and M. Parrinello, "Freezing of a Lennard-Jones fluid: From nucleation to spinodal regime," *Phys. Rev. Lett.* **97**, 105701 (2006).
- M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti, and M. Parrinello, "Molecular-dynamics simulations of urea nucleation from aqueous solution," *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6–E14 (2015).
- P. M. Piaggi, O. Valsson, and M. Parrinello, "Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations," *Phys. Rev. Lett.* **119**, 015701 (2017).
- O. Valsson and M. Parrinello, "Variational approach to enhanced sampling and free energy calculations," *Phys. Rev. Lett.* **113**, 090601 (2014).
- P. Tiwary and B. J. Berne, "Spectral gap optimization of order parameters for sampling complex molecular systems," *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2839–2844 (2016).
- D. Mendels, G. Piccini, and M. Parrinello, "Collective variables from local fluctuations," *J. Phys. Chem. Lett.* **9**, 2776–2781 (2018).
- J. McCarty and M. Parrinello, "A variational conformational dynamics approach to the selection of collective variables in metadynamics," *J. Chem. Phys.* **147**, 204109 (2017).
- D. W. Qi and R. A. Moorem, "Molecular dynamic simulations for crystallization of metallic liquids under different pressures," *J. Chem. Phys.* **99**, 8948 (1993).
- H. Niu, P. M. Piaggi, M. Invernizzi, and M. Parrinello, "Molecular dynamics simulations of liquid silica crystallization," *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5348–5352 (2018).
- A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," *Phys. Rev. Lett.* **100**, 020603 (2008).

- ¹⁷O. Valsson, P. Tiwary, and M. Parrinello, "Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint," *Annu. Rev. Phys. Chem.* **67**, 159 (2016).
- ¹⁸J. F. Dama, M. Parrinello, and G. A. Voth, "Well-tempered metadynamics converges asymptotically," *Phys. Rev. Lett.* **112**, 240602 (2014).
- ¹⁹G. Piccini, D. Mendels, and M. Parrinello, "Metadynamics with discriminants: A tool for understanding chemistry," *J. Chem. Theory Comput.* **14**, 5040 (2018).
- ²⁰L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Phys. Rev. Lett.* **72**, 3634 (1994).
- ²¹J. H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: Generation and validation," *J. Chem. Phys.* **134**, 174105 (2011).
- ²²G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, "Identification of slow molecular order parameters for Markov model construction," *J. Chem. Phys.* **139**, 015102 (2013).
- ²³M. M. Sultan and V. S. Pande, "tICA-Metadynamics: Accelerating metadynamics by using kinetically selected collective variables," *J. Chem. Theory Comput.* **13**, 2440 (2017).
- ²⁴Y. I. Yang and M. Parrinello, "Refining collective coordinates and improving free energy representation in variational enhanced sampling," *J. Chem. Theory Comput.* **14**, 2889 (2018).
- ²⁵R. T. McGibbon, B. E. Husic, and V. S. Pande, "Identification of simple reaction coordinates from complex dynamics," *J. Chem. Phys.* **146**, 044109 (2016).
- ²⁶A. Mitsutake, H. Iijima, and H. Takano, "Relaxation mode analysis of a peptide system: Comparison with principal component analysis," *J. Chem. Phys.* **135**, 164102 (2011).
- ²⁷B. Warren, *X-Ray Diffraction, Addison-Wesley Series in Metallurgy and Materials Engineering* (Dover Publications, 1969).
- ²⁸M. I. Mendelev, M. J. Kramer, C. A. Becker, and M. Asta, "Analysis of semi-empirical interatomic potentials appropriate for simulation of crystalline and liquid Al and Cu," *Philos. Mag.* **88**, 1723–1750 (2008).
- ²⁹S. R. Wilson, K. G. S. H. Gunawardana, and M. I. Mendelev, "Solid-liquid interface free energies of pure bcc metals and B2 phases," *J. Chem. Phys.* **142**, 134705 (2015).
- ³⁰S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *J. Comput. Phys.* **117**, 1 (1995).
- ³¹G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," *Comput. Phys. Commun.* **185**, 604 (2014).
- ³²G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *J. Chem. Phys.* **126**, 014101 (2007).
- ³³M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *J. Appl. Phys.* **52**, 7182 (1981).
- ³⁴M. Invernizzi and M. Parrinello, e-print [arXiv:1901.04455](https://arxiv.org/abs/1901.04455).
- ³⁵P. Tiwary and M. Parrinello, "A time-independent free energy estimator for metadynamics," *J. Phys. Chem. B* **119**, 736–742 (2015).
- ³⁶J. A. Wert, Q. Liu, and N. Hansen, "Dislocation boundary formation in a cold-rolled cube-oriented Al single crystal," *Acta Mater.* **45**, 2565–2576 (1997).
- ³⁷X. Z. Liao, F. Zhou, E. J. Lavernia, S. G. Srinivasan, M. I. Baskes, D. W. He, and Y. T. Zhu, "Deformation mechanism in nanocrystalline Al: Partial dislocation slip," *Appl. Phys. Lett.* **83**, 632 (2003).
- ³⁸X. Z. Liao, S. G. Srinivasan, Y. H. Zhao, M. I. Baskes, Y. T. Zhu, F. Zhou, E. J. Lavernia, and H. F. Xu, "Formation mechanism of wide stacking faults in nanocrystalline Al," *Appl. Phys. Lett.* **84**, 3564 (2004).