



HAL
open science

A new Monte Carlo method for getting the density of states of atomic cluster systems

J.-M. Soudan, M. Basire, J.-M. Mestdagh, C. Angelié

► **To cite this version:**

J.-M. Soudan, M. Basire, J.-M. Mestdagh, C. Angelié. A new Monte Carlo method for getting the density of states of atomic cluster systems. *The Journal of Chemical Physics*, 2011, 135 (14), pp.144109. 10.1063/1.3647333 . hal-00655984

HAL Id: hal-00655984

<https://hal.science/hal-00655984v1>

Submitted on 16 Dec 2019

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

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

A new Monte Carlo method for getting the density of states of atomic cluster systems

J.-M. Soudan,^{a)} M. Basire, J.-M. Mestdagh, and C. Angelié

Laboratoire Francis Perrin, CNRS-URA 2453, CEA, IRAMIS, Service des Photons Atomes et Molécules, F-91191 Gif-sur-Yvette Cedex, France

A novel Monte Carlo flat histogram algorithm is proposed to get the classical density of states in terms of the potential energy, $g(E_p)$, for systems with continuous variables such as atomic clusters. It aims at avoiding the long iterative process of the Wang-Landau method and controlling carefully the convergence, but keeping the ability to overcome energy barriers. Our algorithm is based on a preliminary mapping in a series of points (called a σ -mapping), obtained by a two-parameter local probing of $g(E_p)$, and it converges in only two subsequent reweighting iterations on large intervals. The method is illustrated on the model system of a 432 atom cluster bound by a Rydberg type potential. Convergence properties are first examined in detail, particularly in the phase transition zone. We get $g(E_p)$ varying by a factor 10^{3700} over the energy range $[0.01 < E_p < 6000 \text{ eV}]$, covered by only eight overlapping intervals. Canonical quantities are derived, such as the internal energy $U(T)$ and the heat capacity $C_V(T)$. This reveals the solid to liquid phase transition, lying in our conditions at the triple point. This phase transition is further studied by computing a Lindemann-Berry index, the atomic cluster density $n(r)$, and the pressure, demonstrating the progressive surface melting at this triple point. Some limited results are also given for 1224 and 4044 atom clusters.

I. INTRODUCTION

The density of states (DOS) $g(E)$ of a system is a key function, allowing to compute all thermodynamic canonical quantities (free energy F , internal energy U , heat capacity C_V , etc.), at any temperature T by simple integration over the energy E . For example, the partition function, from which all other quantities can be derived, is (at the inverse temperature $\beta = (k_B T)^{-1}$)

$$Z(\beta) = \int_{-\infty}^{+\infty} dE e^{\ln[g(E)] - \beta E}. \quad (1)$$

$g(E)$ is also used in the Rice–Ramsperger–Kassel–Marcus statistical theory of reaction rates, in which the microcanonical dissociation rate is given by

$$K_D(E) = \frac{G(E - E_d)}{hg(E)}, \quad (2)$$

h being the Planck constant, $G(u)$ the integrated DOS of the transition state up to the energy u , and E_d the dissociation energy of the system. This formula has been applied to the unimolecular dissociation of molecules^{1,2} and clusters^{3,4} using Monte Carlo (MC) computations of g and G .

Given this key role, various Monte Carlo schemes have been developed in the last 20 years to determine $g(E)$, such as the multicanonical technique,⁵ the entropic sampling,⁶ the broad histogram method,⁷ and the Wang-Landau (WL) sampling.^{8,9} A common feature of these algorithms is to use the Metropolis algorithm with a bias $b(E)$, in which moves

from energy E_a to E_b are accepted with the probability

$$P(a \rightarrow b) = \min\left(1, \frac{b(E_a)}{b(E_b)}\right). \quad (3)$$

Consequently, in the standard Metropolis algorithm, the bin interval δE is populated according to an histogram $H(E)$, given by

$$H(E) \propto \frac{g(E)}{b(E)}. \quad (4)$$

A flat histogram is found when the bias matches the true density of states ($b(E) = g(E)$). An advantage of the flat histogram methods is that the MC point circulates in a large energy range, hence overcoming more easily the barriers in complex energy landscapes.

The WL algorithm has received a keen interest because it is self-adaptive. The initial bias $g^{\text{WL}}(E)$ is usually chosen constant and is modified simultaneously with the histogram of visits $H^{\text{WL}}(E)$ in terms of a modification factor f . $H^{\text{WL}}(E)$ converges to a flat histogram through, typically, 20–25 sequences of MC walks, each one corresponding to a value of the parameter f which starts around e^1 and is forced to tend towards 1, reducing consequently the noise. Then, $g^{\text{WL}}(E)$ converges towards the exact density of states. This algorithm has largely been applied to systems with discrete variables (spins), but also, to a lesser extent, to systems with continuous variables, such as Lennard-Jones (LJ) clusters of various sizes: LJ₂₅₀ (Ref. 10) (for which $g(E)$ was varied by a factor 10^{500}), LJ₃₁,¹¹ LJ₃₈,¹² and LJ₂₁₆.¹³ The classical density of states is usually considered, but the quantum one has also been tackled with the WL algorithm.^{14,15}

^{a)}Electronic mail: jean-maik.soudan@cea.fr.

These previous works about clusters refer to nanothermodynamics, a fascinating field in which surface effects and non-extensivity play an important role.^{16–19} Particularly, transition metal clusters are crucial ingredients for the catalytic growth of carbon nanotubes,²⁰ and their thermodynamic state is hardly accessible to experiment and controversial. To tackle theoretically such systems with continuous variables of a few hundreds of atoms or more, tightened by a realistic potential, is an important but very challenging program. The sequences of the WL method slow down when the parameter f approaches 1, and it happens that the convergence is not well controlled, particularly in a phase transition zone. Such convergence problems have been mentioned for LJ clusters.^{11,12} In Ref. 10, the WL sequence has been ended by a standard Metropolis walk in view of correcting the density of states obtained. It is also worth mentioning that Kim *et al.*²¹ proposed an adaptive mapping in terms of a temperature $T(E)$ to improve the convergence of the WL f -cascade.

The present work provides a novel approach to derive $g(E)$, particularly suited to systems with many degrees of freedom and with interacting potentials more involved than the LJ modelization. This new method converges in a well controlled way, still acting within the framework of MC flat histogram methods. Its originality is to get quickly a very good initial guess of the bias $b(E)$ through a new type of mapping function $\sigma(E_p)$ obtained at a few tens discrete points of the potential energy. As a result, $b(E)$ converges in only two reweighting iterations, by MC walks on a few overlapping ranges. This study focuses on the methodology and convergence of the algorithm. This is why we choose as an illustration the model system of a 432 atom cluster of 2 nm diameter, the interaction being given by the simple 6 parameter Rydberg type pair potential, for which billions of MC steps can be performed in one day. Some limited results are also given for clusters of 1224 and 4044 atoms.

This paper is organized as follows. In Sec. II, we explain the basis of our algorithm getting the density of states $g(E_p)$ of the potential energy in the configurational space, in terms of the mapping function $\sigma(E_p)$. In Sec. III, we illustrate briefly the method on a Debye (i.e., harmonic) cluster, and in detail on the Rydberg cluster, discussing particularly the convergence qualities. In Sec. IV, the canonical quantities $U(T)$ and $C_V(T)$ are first derived by integration, to put in evidence a phase transition. This one is further analyzed using the Lindemann-Berry index, the density $n(r)$ of the cluster, and finally a computation of the pressure as a function of the temperature. These analyses demonstrate clearly that this phase transition is a solid-liquid one in the presence of a gas, i.e., at the triple point. In Sec. V, the variation of the mapping function $\sigma(E_p)$ is explained, and a comparison of the performances with the WL algorithm is made. The conclusion of Sec. VI paves the road for treating a system with more realistic interactions than the Rydberg potential, such as the modified embedded atom method (MEAM) potential.^{22,23}

II. METHOD TO GET THE DENSITY OF STATES

In this section, we consider the configurational space of the system in which the potential energy $E_p(\vec{r}_1, \dots, \vec{r}_N)$ is

computed in terms of the positions \vec{r}_i of the N particles. The Metropolis algorithm is used with standard translational moves in this space, and a fixed bias $b(E_p)$ determined by the method explained below. It provides the density of states $g(E_p)$ associated with the potential energy, whereas $g_k(E_k)$ for the kinetic energy is known analytically. Given the separability of the total energy as $E = E_k + E_p$, canonical quantities at an inverse temperature β can be deduced by independent integrations over E_k and E_p , respectively, thermodynamic functions being additive at a given β (see Sec. IV). Thus, only computations relative to E_p and corresponding values ($g(E_p)$, $\sigma(E_p)$, ...) need to be conducted and are described in the following.

A. Relevance of $\sigma(E_p)$ to express the DOS $g(E_p)$

This section focuses on quantities $\ln[g(E_p)]$ (DOS) and $\ln[b(E_p)]$ (bias), which correspond to the configurational entropy of the system $S(E_p)/k_B$.^{7,18,19,21} They can be expressed as

$$\ln[g(E_p)] - \ln[g(E_{\text{ref}})] = \int_{E_{\text{ref}}}^{E_p} du \beta(u), \quad (5a)$$

$$\ln[b(E_p)] - \ln[b(E_{\text{ref}})] = \int_{E_{\text{ref}}}^{E_p} du \beta_b(u), \quad (5b)$$

where the function $\beta(E_p)$ (respectively $\beta_b(E_p)$) is the derivative of $\ln[g(E_p)]$ (respectively $\ln[b(E_p)]$) and appears as a configurational microcanonical inverse temperature. E_{ref} is an arbitrary reference energy. We found convenient and illuminating to express this function β (respectively β_b) under the form

$$\beta(E_p) = \frac{\sigma(E_p)}{E_p - E_z^0}, \quad (5c)$$

where E_z^0 is in principle the absolute minimum of the potential energy, and $\sigma(E_p)$ (respectively $\sigma_b(E_p)$) is a mapping function, which is an approximately extensive quantity with no dimension. In fact, Eq. (5c) can be used with an origin E_z different from the absolute minimum E_z^0 . It must be emphasized that the density of states $g(E_p)$ and the function $\beta(E_p)$ are quantities independent of the origin E_z , whereas $\sigma(E_p)$ depends on it. This dependence and the importance to determine E_z^0 is discussed hereafter in Sec. II B 1.

The justification of the mapping of Eq. (5c) results from the following model. For a separable potential with N_d degrees of freedom of the form

$$E_p = \sum_{i=1}^{N_d} |x_i|^{n_i}, \quad (6a)$$

the mapping function is expressed as

$$\sigma = \left(\sum_{i=1}^{N_d} \frac{1}{n_i} \right) - 1, \quad (6b)$$

and is therefore strictly constant. When the system is in the configurational vicinity of the absolute minimum of the potential energy basin, an harmonic expansion is valid and corresponds to the particular case $n_i = 2$. Then, when E_p tends

towards this bottom E_z^0 ,

$$\sigma(E_p) \rightarrow \frac{N_d}{2} - 1. \quad (7)$$

Note that the application of Eq. (7) for a cluster of 432 atoms with 6 frozen degrees of freedom gives $N_d = 1290$ and $\sigma(0) = 644$. This is the typical magnitude of σ for the system studied in Sec. III.

The previous separable model explains why the mapping function $\sigma(E_p)$ is more relevant than the parameter $\beta(E_p)$: (i) Equation (7) shows at low energies that $\sigma(E_p)$ is finite, provided E_z^0 has been determined, while $\beta(E_p)$ is diverging to $+\infty$. (ii) At higher energies, $\sigma(E_p)$ has almost everywhere a rather slow variation in terms of E_p (except in a phase transition zone). Equations (6a) and (6b) help to understand the variation of $\sigma(E_p)$ using a modelization of the mean potential perceived by each atom at the energy E_p . This discussion is further pursued in Sec. V A. (iii) The two previous properties allow to get a very good representation of $\sigma(E_p)$ by interpolation (or even extrapolation) between a few points (see first step of the algorithm). (iv) In a phase transition zone, $\sigma(E_p)$ exhibits a jump on a narrow energy range and displays, therefore, strikingly this phenomenon, while $\beta(E_p)$ shows only the so-called “back-bending” (an oscillating variation). This is analyzed in Sec. IV A. (v) As discussed in the preliminary step of the algorithm, far from being a limitation of our method, the determination of E_z^0 , by the dependency it introduces in $\sigma(E_p)$, is a powerful self-consistent way of checking, through Eq. (7), that no lower basin is reached during all the MC simulations.

It is to stress that all the previous equations and properties are equally valid for b and σ_b .

B. Algorithm

The method used to find the mapping function $\sigma(E_p)$ is implemented in several steps, which are described in detail in the following. They can be summarized as follows:

1. Preliminary step: determination of the minimum configurational energy E_z^0 .
2. First step: use of a local linear bias in a series of narrow potential energy ranges to get a first guess $\sigma_b^{(1)}(E_p)$ of the mapping function over an energy grid. Each value is deduced from the Gaussian form of the accumulating energy histogram and interpolated over the energy range to give $\sigma_b^{(1)}(E_p)$. The corresponding bias function $b^{(1)}(E_p)$ is computed from Eqs. (5b) and (5c).
3. Second step: standard Metropolis exploration on large intervals in potential energy, with $b^{(1)}(E_p)$ as an input for the Metropolis rule. A corrected mapping function $\sigma_b^{(2)}(E_p)$ is derived from the histograms $H^{(1)}$ and the mapping function $\sigma_b^{(1)}(E_p)$. Then, a new bias $b^{(2)}(E_p)$ is deduced.
4. To increase convergence and ergodicity, it can be necessary to iterate this last step using $b^{(2)}(E_p)$ in a new Metropolis exploration. Overlapping intervals can now be extended. $b^{(3)}(E_p)$ thus obtained converges now to the real $g(E_p)$.

1. Preliminary step

The Wales technique²⁴ is used to search the absolute minimum of $E_p(\vec{r}_1, \dots, \vec{r}_N)$. Obviously, it is not possible to guaranty that the true absolute minimum has been found, whatever the duration of the search. It is, thus, possible that during a longer MC trajectory, a deeper basin could be found and explored. This situation was in fact met in our first simulations: after a too short minimum search, we observed that, during the MC walk, the mapping function $\sigma(E_p)$ exhibited a sudden fall (see the first step of the method). This provides a very clear signature of a transition to a new and deeper basin. This fact can be explained by the following expression relating two different energy origins E_z :

$$\beta(E_p) = \frac{\sigma^{z0}(E_p)}{E_p - E_z^0} = \frac{\sigma^{z1}(E_p)}{E_p - E_z^1}. \quad (8)$$

$\beta(E_p)$ (Eqs. (5a) and (5c)) does not depend on the choice of the energy origin. The minimum found in the exploration, E_z^1 , is *a priori* larger than the absolute minimum E_z^0 . The two functions $\sigma^{z0}(E_p)$ and $\sigma^{z1}(E_p)$ corresponding to these values are related by Eq. (8). Then, if the MC point has explored ergodically, the true minimum basin, the left fraction expressing $\beta(E_p)$ tends to a finite value when E_p tends to E_z^1 , and consequently, $\sigma^{z1}(E_p)$ must tend to zero in the right fraction, so that the property of Eq. (7) is strongly violated. In this case, a new minimum search can be performed from a point extracted from the new basin explored during the MC walk. Hence, the value of the minimum is changed. If the sudden fall of $\sigma(E_p)$ mentioned above does not occur anymore during the remaining MC explorations, this is the proof that no lower basin has been reached. It is interesting to note that Eqs. (7) and (8) can be used to deduce E_z^0 with a very good accuracy from the bad value E_z^1 . Choosing a value E_p a little above E_z^1 and enforcing $\sigma^{z0}(E_p)$ to check Eq. (7) allows to estimate E_z^0 knowing $\sigma^{z1}(E_p)$.

It must be stressed again that our algorithm could be implemented without this preliminary search of the absolute minimum, but in this case the correct energy range of the potential energy would not be identified and the property of slow variation of $\sigma(E_p)$ would be lost.

2. First step: Linear bias

This step is the key trick of the method. The principle is to map the $\sigma(E_p)$ function at a series of points E_{p_i} , using a series of Metropolis explorations with a mapping function $\sigma_b^L(E_p)$, linear in terms of energy, from which a bias b^L is derived. Each exploration spans an interval beginning at E_α and of half-width Δ_α , the two-parameter linear bias being

$$\sigma_b^L(E_p) = \sigma_b^L(E_\alpha) + B_\alpha \cdot (E_p - E_\alpha). \quad (9)$$

Within the hypothesis of an ergodic exploration, the corresponding histograms $H(E_p)$ is accumulated in this interval according to Eq. (4). $H(E_p)$ is, therefore, extremum at the value E_{p_i} where the derivative of $\ln[g/b^L]$ is zero. Then, the following equality is satisfied:

$$\sigma(E_{p_i}) = \sigma_b^L(E_{p_i}). \quad (10)$$

A second-order expansion of $\ln [g/b^L]$ around this point gives the expression

$$H(E_p) \propto \exp[-A(E_p - E_{p_i})^2], \quad (11a)$$

with

$$A = \frac{B_\alpha - \sigma'(E_{p_i})}{2E_{p_i}}, \quad (11b)$$

determining the width of the function (σ' is the derivative of σ with respect to the potential energy, and the energy minimum has been fixed to zero). When the condition

$$B_\alpha > \sigma'(E_{p_i}) \quad (11c)$$

is satisfied, E_{p_i} is a maximum ($= E_{\max}$) and $H(E_p)$ has a Gaussian form (neglecting third-order terms). The maximum value will be found more accurately, if $H(E_p)$ explores a narrow energy zone, which means that the slope B_α must be chosen steep enough. At this point, Eq. (10) gives the desired value of $\sigma(E_{p_i})$. About 25 energy points are necessary for a good mapping over the $[0.01 - 6000 \text{ eV}]$ range of the system studied in the following. The values $\sigma(E_{p_i})$ are interpolated by spline functions to get a continuous representation of $\sigma(E_p)$. The slow variation of $\sigma(E_p)$ guaranties an accurate interpolation with a few points.

If the MC exploration were fully ergodic at this step, the determination of the mapping function $\sigma(E_p)$ would be achieved. In fact, we emphasize that the MC trajectory is limited to a narrow range of energy, and the system has a limited ability to hop from one energy basin to another. Then, full ergodicity is not reached and Eq. (10) gives only a first (but in fact very good) approximation $\sigma_b^{(1)}(E_p)$, which is used to compute the bias $b^{(1)}(E_p)$ introduced in the next step. The remarkable accuracy of this first step, in spite of the use of narrow energy ranges, is the key for the efficiency of our algorithm.

Note that choosing the particular value $\sigma_b^L(0) = 0$ leads to a canonical bias, with a slope corresponding to the inverse temperature, defined at the point E_{p_i} by

$$B_T(i) = (k_B T_i)^{-1} = \frac{\sigma_b^L(E_{p_i})}{E_{p_i}}. \quad (12)$$

The present two-parameter bias allows a much greater flexibility than the canonical one. This is necessary to get a good accuracy on the determination of $\sigma^{(1)}(E_p)$ over all the energy scale, and to map correctly the phase transition zone.

The conditions that the parameter B_α (and correlatively $\sigma_b^L(E_\alpha)$) must satisfy to optimize the accuracy are detailed in Appendix A. The latter sums up the main conclusions. The first condition that B_α must meet is Eq. (11c) which provides a lower limit to it. It is constraining only in the phase transition zone where σ' is high. Then, the half-width Δ_α must be chosen to be a fraction of the energy at the middle of the interval to monitor a sufficiently well-localized zone. Moreover, the Gaussian must be well drawn, centered around the center of this interval. This gives for B_α a lower limit much smaller than B_T (Eq. (A2c)), which leaves a large possibility of choice. Finally, the determination of σ_p using Eq. (10) is made by calculating the energy E_{\max} where the histogram is

maximum. This is performed by fitting the region of this maximum by a third-order polynomial decomposition of $\sigma_b^L(E_p)$. A third-order expansion around E_{\max} is necessary to take into account a slight asymmetry of the histogram, compared to a Gaussian. This asymmetry contribution can modify the value of σ by several units in some zones and cannot be neglected (Eqs. (A5b) and (A5c)). We have also performed an analysis of the MC fluctuations giving the statistical accuracy $\delta\sigma$ (Eq. (A10)). It is interesting to stress that what is important for the flatness of the histograms obtained at the following step is the variation of σ over an energy range and not its absolute value. This is analyzed in Sec. II B 3.

3. Second step: Metropolis exploration

At this step, a standard Metropolis exploration is done with the previous bias $b^{(1)}(E_p)$. In contrast to the previous step, it is now performed on a series of large potential energy overlapping ranges $[E_\alpha^{\text{inf}}, E_\alpha^{\text{sup}}]$. These intervals, indexed by α , have an expansion $R_\alpha = E_\alpha^{\text{sup}}/E_\alpha^{\text{inf}}$. Hence, we now take benefit of the circulation of the MC point in these large intervals to overcome the energy barriers and approach ergodicity. The extension of each range is nevertheless first limited by the need to make a sufficient number of round trips between the extremities of the interval, in view of obtaining good statistics. This restriction is related to the need of a sufficiently high Metropolis acceptance rate, which limits the translational moves. It is stressed that this limitation concerns any flat histogram method using independent intervals. Indeed, only replica exchange methods between several intervals, or concerted movements speeding the diffusion on the energy scale, can extend the energy spanned by the MC point, thus allowing double funnel structures of the energy landscape to be overcome (see Conclusion). It must be noticed that the different intervals overlap, so that they can give slightly different mapping functions $\sigma_b^{(2)}(E_p)$ on the common part of two overlapping energy ranges. This is the consequence of the statistical fluctuations $\delta\sigma$ and of a non-ergodic exploration in, at least, one interval. In this case, an interpolation was made between the two mapping functions from one extremity of the common range to the other.

In the frame of the ergodic hypothesis, the (in principle) exact value of σ is deduced from the histogram $H^{(1)}$ and from the mapping function $\sigma_b^{(1)}(E_p)$ by a reweighting derived from Eq. (4)

$$\sigma = \sigma_b^{(1)} + E_p \cdot \underbrace{\frac{d \ln H^{(1)}}{d E_p}}_{\sigma_H^{(1)}} + \sigma_{cd}, \quad (13)$$

σ_{cd} being a correction due to the discretization of energy into bins δE , given in Appendix B. The parameter δE is chosen in such a way that σ_{cd} is generally negligible, which is the case if $\delta E \lesssim 2E/\sigma$ (Eq. (B6)). It is important to note that a systematic discrepancy $\Delta\sigma = \langle \sigma_H^{(1)} \rangle$ over an interval of expansion R_α gives a relative variation,

$$H_{\text{sup}}/H_{\text{inf}} \simeq R_\alpha^{\Delta\sigma}. \quad (14)$$

Hence, the mapping $\sigma_b^{(1)}(E_p)$ obtained from the first step is useful, only if the error $\Delta\sigma$ does not exceed a few units. This

is actually the case, as shown in the following. The minimization of the H variation over the interval is, however, a second limitation for its extension R_α . The application of Eq. (13) gives a new and improved approximation $\sigma_b^{(2)}(E_p)$ of $\sigma(E_p)$, and then a new bias $b^{(2)}(E_p)$. At this step, it is not known yet if a perfect ergodicity has been reached, so if $\sigma_b^{(2)}(E_p)$ has fully converged towards $\sigma(E_p)$.

Then, a second iteration can be performed, identical to what has just been explained, but with the new estimate $b^{(2)}$ of the bias. With this new bias, the ranges R_α can be extended to improve the ergodicity and diminish the number of interval junctions. The extension of the range remains nevertheless limited by the need of a sufficient number of round trips. This last iteration was found to be necessary at low energies, in the solid zone, where the ergodicity was difficult to obtain. On the contrary, in the fluid zone this iteration serves only to check that $b^{(2)}$ gives a flat histogram.

4. Final remarks

We conclude this section remarking that the previous σ mapping is suited to a system with continuous variables, such as the position coordinates described here in a classical picture. Its adaptation to, for example, quantified spin systems remains to be explored but would need another type of slowly varying parameter.

It is also to notice that the linear bias $\sigma_b^L(E_p)$ is akin to the temperature-linear bias used in a recent work,²⁵ in the context of the replica exchange method. In this work a temperature $T_\alpha(E)$ depending linearly on E is used in a set of intervals α to compute a bias $w_\alpha(E)$. For a sufficiently high slope, a localized histogram with a Gaussian shape is built. The set of non-canonical ensembles α is used in a replica exchange method. At the end, the entropy $S(E)$ is calculated via the weighted histogram analysis method. Contrary to that, our σ -linear mapping provides directly a first estimate of $\ln[g(E)]$ (or $S(E)$).

III. ILLUSTRATION OF THE METHOD AND DISCUSSION OF THE $\sigma(E_p)$ CONVERGENCE

A. Computational details

The calculations were performed with a PC Dell T5400 with eight processors of 2.86 GHz. The initial crystal is cut spherically with a radius of 10 Å, in bulk bcc at $T = 0$ K, corresponding to 432 atoms. The crystal is placed in a spherical cavity of radius $R_W = 14$ Å, leaving 4 Å free around the crystal. An atom is fixed at the center to prevent translation and three other coordinates are frozen to prevent rotation. The same process is also used to cut spherical clusters of 1224 and 4044 atoms, with a single $R_W = 25$ Å sufficiently large to contain either of them.

B. A Debye crystal

As a preliminary test of the method, a Debye crystal has been studied, i.e., atoms bound to the first neighbours by a harmonic potential with a unique constant $K_{ij} = 1 \text{ eV}/\text{Å}^2$. In the range [50–100 eV], $\sigma(E_p) = 644.10$ has been obtained in 10^9 steps, in very good agreement with the analytical value for

an harmonic cluster having $N_d = 1290$ degrees of freedom: $\sigma = 644$ (Eq. (7)). However, in the range [4000–8000 eV], when atoms collide with the wall, $\sigma(E_p)$ decreased to 641.8. This is not surprising given the reduction of the accessible configurational space due to the wall (see Sec. V A).

C. A Rydberg crystal

The atomic interaction is described hereafter by a pair potential of the Rydberg type, analogous to the Rose function of the MEAM potential^{22,23}

$$\Phi = -E_c[1 + a^* + d(a^*)^3]e^{-a^*}, \quad (15a)$$

$$a^* = \alpha(r/r_e - 1). \quad (15b)$$

A cutoff function $f_c(x) = (1 - (1 - x)^2)^2$ is used for $0 < x < 1$, with $x = (r_c - r)/\delta r_c$. The six parameters are chosen to reproduce approximately the bulk data of iron around the equilibrium bcc structure. Nevertheless, the purpose of this work is not a quantitative study of this system, which cannot be well represented by a pair potential.²⁶ The parameters are: $E_c = 0.6371$ eV, $r_e = 2.64922$ Å, $\alpha = 5.170477$, $d = 0.05$, $r_c = 4.0$ Å, and $\delta r_c = 1.144422$ Å. With this potential, 2.4×10^9 MC steps can be performed for 432 atoms in one day.

The preliminary step of the method presented in Sec. II gave, after a long Wales search (about three days of computing time), an optimized minimum $E_z^0 = -1434.8245$ eV, with 13 other minima in a range of 0.6 eV. Note that an accuracy of 10^{-4} eV for E_z^0 is necessary, when exploring intervals beginning at 0.01 eV of the absolute minimum, in view of checking Eq. (7). In all the following MC explorations, we never encountered the fall of σ at low energies that can happen after a too short preliminary search, as it was mentioned in Sec. II. This provides high confidence that the absolute minimum has been found. In the following, if not otherwise stated, the origin of the energy is chosen at this absolute minimum (instead of being at the infinite separation of atoms) and, therefore, $E_z^0 = 0$.

Step 1 is illustrated in Fig. 1, which displays a Gaussian histogram obtained with a σ -linear bias near 350 eV. The full mapping $\sigma(E_p)$ is obtained with 25 points distributed from 1 to 6000 eV (not shown here). About 5×10^9 MC steps were necessary for a good convergence below 200 eV, but above this value, 5×10^8 were sufficient. It is important to stress that this preliminary mapping with 25 points has a computational cost corresponding to at most two intervals of step 2, so that it supplies a very good first guess of the σ curve at a remarkably low cost. In particular, it localizes strikingly and accurately the phase transition zone, which will be discussed in Sec. IV. Indeed, this curve shows a large slope of $\sigma(E_p)$ near 185 eV, the corresponding temperature being quantitatively determined at this stage within 1%.

The value of the energy at the maximum of the histogram presents fluctuations which determine the accuracy of the σ value. We find $\delta\sigma \simeq 1$ between 100 and 1000 eV, i.e., typically one-thousandth of the value of σ . This uncertainty is linked to the MC statistics and must not be mistaken with the

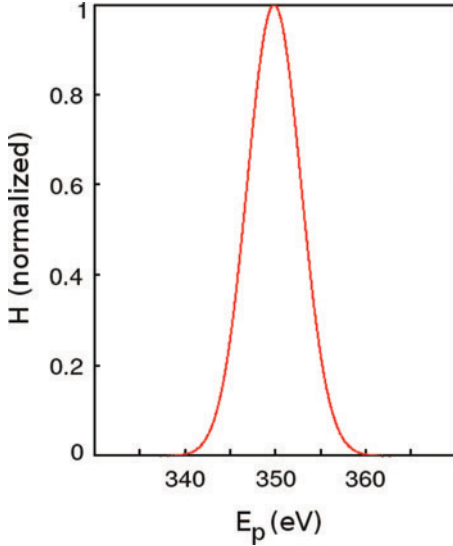


FIG. 1. Histogram obtained with a σ -linear bias ($\sigma_b^L(330) = 306$, $B = 40 \text{ eV}^{-1}$) giving $\sigma(349.93 \text{ eV}) = 1103.20$ for a 432 atom system.

error relative to the exact σ obtained from an actual ergodic exploration almost achieved in the following steps. $\delta\sigma$ is analyzed in Appendix A (Eq. (A10)) and its expression explains correctly the observed fluctuations.

Usually, the starting MC point of each Gaussian is prepared in the following way. The system evolves freely from a point obtained at lower energy towards a higher energy range of interest. The dynamical evolution of the Gaussian in this new interval can be studied, and it is observed that the energy of the maximum increases up to an asymptotical value. Moreover, a striking phenomenon appears in the phase transition zone just below the high slope region and is observed directly on the histogram dynamics. A single Gaussian is first formed as usual, but suddenly a new peak appears at slightly higher energy and grows to become a new Gaussian. This corresponds to a σ jump of several tens. This will be analyzed in more detail in Sec. IV B and interpreted as a transition of the MC point from a solid basin to a liquid one.

Step 2 is performed on the full energy range [0.01–6000 eV], covered by 12 overlapping intervals. The result is illustrated in Fig. 2(a) for a large energy range [230–1530 eV], typical of the fluid zone (line “first iteration”). The guess $b^{(1)}(E_p)$ was obtained in this range by 12 points coming from σ -linear bias explorations. The resulting histogram $H^{(1)}(E_p)$ after 50×10^9 steps varies between 4×10^{-2} and 1, implying an average correction $\Delta\sigma$ over the interval by +2.5, i.e., only a few thousandths of the σ mapping estimate. This correction is due to the lack of ergodicity of the previous step. With Eq. (13), a new estimate $b^{(2)}(E_p)$ of the density of states is computed.

A second iteration is performed with the bias $b^{(2)}(E_p)$ previously determined, using only eight overlapping intervals. The result is shown in Fig. 2(a) (line “second iteration”), demonstrating that the convergence to a flat histogram is achieved in this interval. This proves the necessity of large energy ranges to explore correctly the configurational space. This fact is also supported by the following observation: when performing step 2 just below the phase transition at 185 eV,

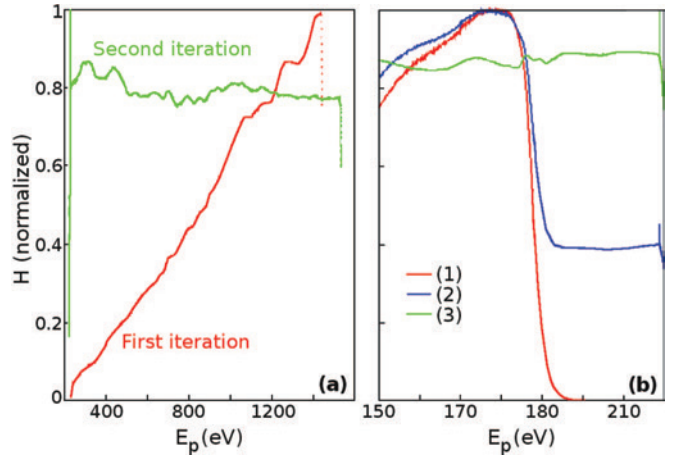


FIG. 2. (a) Histograms obtained with the bias $b^{(1)}$ (respectively $b^{(2)}$) after the first (respectively second) iteration, for the range [230–1530 eV] above the phase transition. (b) Histograms obtained in the phase transition zone [150–220 eV] after 7.5×10^9 MC steps (1), 11.7×10^9 MC steps (2), 50×10^9 MC steps (3). All histograms are obtained for a 432 atom system.

and extending the energy range towards low values (50 and next 25 eV), the new correction $\sigma_H^{(2)}$ reaches several units below 100 eV (+7 near 50 eV) at the end of this step. It is the evidence of a lack of ergodicity, which is not surprising for ranges remaining entirely in the solid zone. The circulation of the MC point near the fluid zone is necessary to approach ergodicity and to overcome energy barriers, as it was also shown in Refs. 11 and 12 in the context of the WL method.

Interestingly, the average number of steps to make a round trip between the extremities of an interval α of typically $R_\alpha \simeq 4$ is only 1×10^7 in the [14–55 eV] range. It raises up to 5×10^8 in the [50–160 eV] range just below the phase transition and to 3×10^8 in the [230–1530 eV] range just above. It falls down to 7.2×10^5 in the high energy range [1400–6000 eV], indicating the easiness of convergence when the phase is very fluid. We have also observed that the average number of MC steps necessary to make a single round trip varies approximately as the square of the interval extension in a given zone (see Appendix A, Eq. (A8)). This is the indication of a diffusion process over the energy scale.^{27,28} As shown in Appendix A, the uncertainty of the histograms is inversely proportional to the square of the number of round trips (Eq. (A8)). We conclude from these statements that it is necessary to achieve a compromise: sufficiently large intervals must be defined to approach ergodicity in the solid zone, but not too large to be able to make a few tens round trips, and thus to get a good statistical accuracy of about 10% on $g(E_p)$.

Another striking phenomenon is seen in the narrow [150–220 eV] range around the phase transition at 185 eV. This phenomenon is illustrated in Fig. 2(b) by successive snapshots of the histogram in this range. While the system wanders rapidly inside each half-range below or above 185 eV, 8×10^9 steps at least are necessary to switch from one half-range to the other, so that only three round trips are achieved during our 50×10^9 steps MC walk, although a flat histogram method is used. This is the evidence of a strong kinetic barrier between the two basins corresponding to the

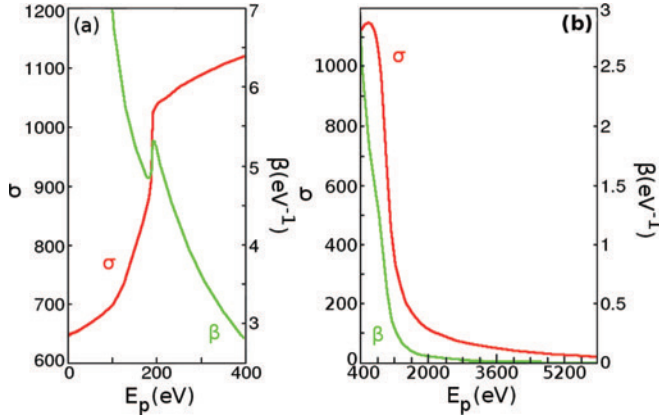


FIG. 3. Functions $\sigma(E_p)$ (red curve) and $\beta(E_p)$ (green curve, eV^{-1}) for a 432 atom system, with two different energy scales: (a) from 0 to 400 eV and (b) from 400 to 6000 eV.

solid and liquid phases. Collective movements are probably needed to overcome this barrier more easily (this point will be discussed in the Conclusion).

Figures 3(a) and 3(b) show the final mapping function $\sigma(E_p)$ between 0.01 and 6000 eV. It is equal to 644 at 0 eV, as it must be the case from Eq. (7) with $N_d = 1290$. A large slope, $d\sigma/dE_p = 24.5 \text{ eV}^{-1}$, is observed around $E_p^{ph} = 185 \text{ eV}$. The parameter $\beta(E_p)$, also shown on this figure, exhibits in this zone the corresponding triple intersection with the horizontal line (around $\beta = 5.10$), which is characteristic of a phase transition (see Sec. IV A).

Figure 4 shows the variation of $\text{Log}_{10}[g(E_p)]$ indicating that $g(E_p)$ spans 3700 orders of magnitude when the energy varies over the range [0.01 – 6000 eV]. Comparing with Fig. 3, it can be noticed that the phase transition is undetectable on the curve $\text{Log}_{10}[g(E_p)]$.

Finally, the curve $\sigma(E_p)$ was computed for the 1224 and 4044 atom clusters, from the minimum energy up to the liquid zone, in view of obtaining the melting temperature in terms of the cluster size. We also found for these two clusters that it was not possible to make a single round trip between the solid and the liquid zone for an interval overlapping the phase tran-

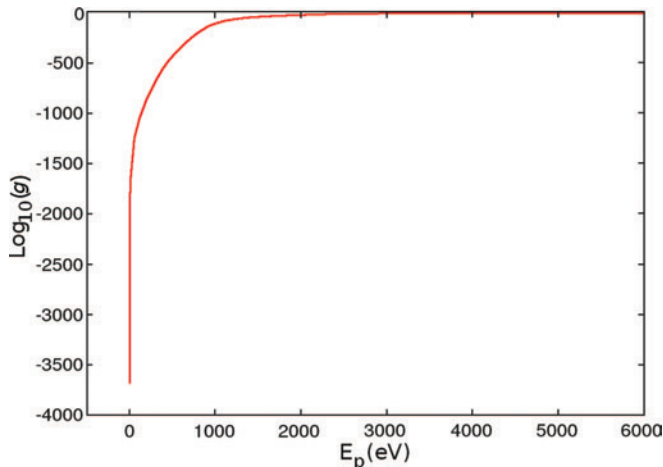


FIG. 4. Density of states $g(E_p)$ plotted for a 432 atom system from 0.01 to 6000 eV in Log_{10} scale.

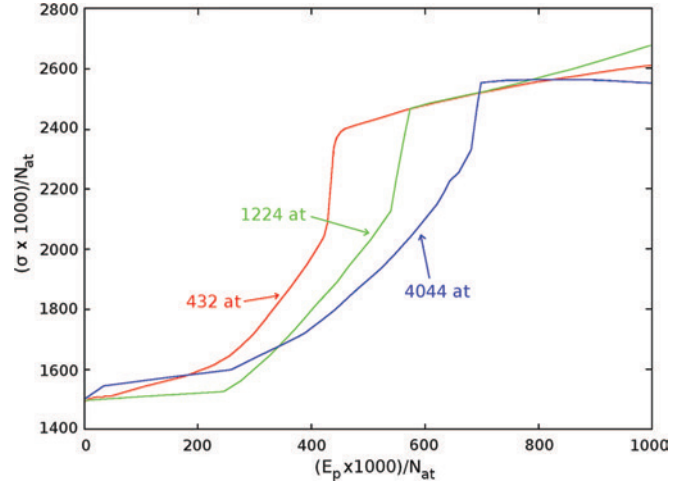


FIG. 5. Normalized curves: $\sigma \times 1000/N_{at}$ as a function of $E_p \times 1000/N_{at}$ (eV) for 432, 1224, and 4044 atoms.

sition region. This was the case even for the 1224 atom cluster in 30×10^9 steps, and a fortiori for 4044 atoms. Moreover, we observed that a MC configuration placed near the phase transition but in the solid zone was quickly attracted to the liquid zone with no return. This is the sign that the kinetic barrier from the liquid to the solid phase increases with the number of atoms. However, when the interval was completely inside the solid zone or the fluid zone, the MC point could make many round trips between the extremities, whatever the number of atoms in the cluster. It is stressed that, in spite of the impossibility to transit between the solid and the liquid zones, it was possible to accurately determine $\sigma(E_p)$ for such large clusters. This was achieved by exploring an interval ending just below the phase transition zone (within a few eV), and an interval starting just above. That way, a spline interpolation in the few eV range of this phase transition zone leaves only a small uncertainty.

The resulting $\sigma(E_p)$ curves are represented in Fig. 5, in units per atom, and show the same general shape as for the 432 atom cluster. The most obvious difference is that the energy of the triple intersection is shifted to higher values with the increase of the number of atoms.

IV. CANONICAL QUANTITIES AND PHASE TRANSITION CHARACTERIZATION

A. Caloric curve and heat capacity

Taking kinetic energy into account, canonical quantities, such as the direct caloric curve, can be computed as $U(\beta) = \langle E_k \rangle + \langle E_p \rangle$, with $\langle E_k \rangle = N_d/2\beta$,

$$U_p = \langle E_p \rangle = \frac{1}{Z_p} \int_{-\infty}^{+\infty} dE_p E_p e^{\ln[g(E_p)] - \beta E_p}, \quad (16)$$

Z_p being the configurational partition function (see Eq. (1)), and β the constant defining the inverse temperature (which must not be confused with the function $\beta(E_p)$).

The previous equation allows a discussion of the curves shape reported in Figs. 3 and 5 and clarifies the meaning of the mapping function $\sigma(E_p)$. When the equation $\beta = \beta_p(E_p)$,

or equivalently $\sigma(E_p) = \beta \cdot E_p$, has a single intersection at the energy E_p^β , the integrant of Eq. (16) has one maximum at this point. This situation occurs when $\sigma' < \beta$, for which Eq. (16) is very well approximated by a Gaussian expansion analogous to that of Eq. (11b). It results that $U_p \simeq E_p^\beta$ and the configurational heat capacity $C_{pV}(T) = (\partial U_p / \partial T)_V$ is simply related to σ by

$$\sigma \simeq \frac{C_{pV}}{k_B} \left[1 - \frac{\sigma'}{\beta} \right]. \quad (17)$$

σ follows consequently the variation of C_{pV} since the bracket is positive. When the equation $\beta = \beta_p(E_p)$ has a triple intersection, the integrant has two maxima and one intermediate minimum.^{10–12,17–19} This situation occurs when $\sigma' > \beta$ for the central intersection, and reveals a phase transition at the equality of these two maxima. The phase transition zone corresponds, therefore, to a high slope of σ . Then, $U(\beta)$ jumps from the first maximum to the second, and $(\partial U / \partial \beta)_V$ diverges to high values, implying the presence of a peak in the heat capacity. It results that σ does not follow anymore the variation of the heat capacity in the phase transition zone, but is rather related to the jump of U . To show evidence of the phase transition zone, the heat capacity is computed in the following by the Einstein relation on the variances:

$$C_V(T) = (\partial U / \partial T)_V \quad (18a)$$

$$= k_B \beta^2 [\langle \Delta E_k^2 \rangle + \langle \Delta E_p^2 \rangle]. \quad (18b)$$

$\langle E_p \rangle$ and $\langle E_p^2 \rangle$ are computed by numerical integration over E_p , knowing $g(E_p)$. The kinetic part is $N_d/2k_B$.

The resulting values for $U(T)$ and $C_V(T)$ are depicted in Fig. 6. In the direct caloric curve $U(T)$, a large slope is observed around the phase transition. The leap of 52 eV is associated with the latent heat of liquefaction. The phase transition is more clearly localized on the curve $C_V(T)$, with the peak centered at $T_{ph} = 2275$ K ($\beta = 5.10$ eV⁻¹), having a FWHM of 56 K.

For the clusters with 1224 and 4044 atoms, the phase transition is found at, respectively, $T_{ph} = 2805$ and 3271 K.

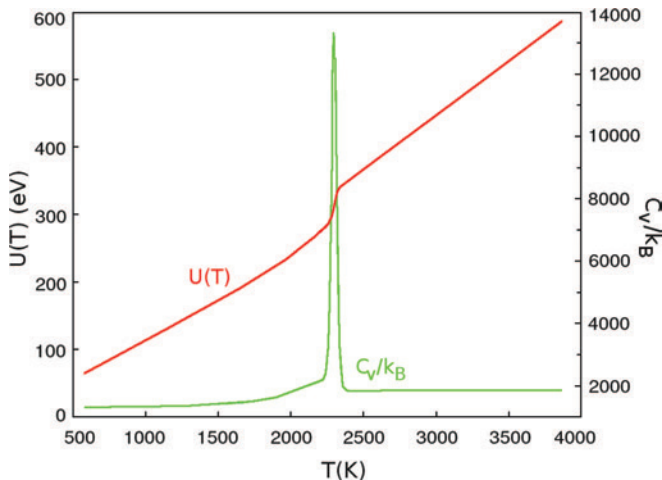


FIG. 6. Caloric curve $U(T)$ (red line, eV), and heat capacity $C_V(T)$ (green line, in k_B units) for a 432 atom system.

The temperatures T_{ph} of the three clusters are reported in Fig. 7 in terms of $N_{at}^{-1/3}$, which is proportional to the inverse of the cluster radius. The three points form a straight line, as expected from previous studies,^{29–32} and the bulk value can be extrapolated to $T_{ph} = 4150 \pm 50$ K. This high value is characteristic of the Rydberg pair potential which is not a realistic representation of metallic compounds because of the lack of an environment dependence²⁶ (for comparison, $T_{ph} = 1811$ K experimentally for bulk iron).

B. Lindemann-Berry analysis

Phase transitions can be probed at the atomic scale using the Lindemann-Berry order parameter,^{33,34} called L_B . Each MC computation is localized in a narrow energy range ΔE using the σ -linear bias at the step 1 of the method. A spatially restricted version $L_B(R_1, R_2)$ is used here, to monitor the spherical shell Sh between the radius R_1 and R_2 , and is defined by

$$L_B(Sh, \Delta E) = \frac{2}{N_{\text{pair}}} \sum_{i,j} N_{ij}^{\text{MC}} \left[\frac{\langle r_{ij}^2 \rangle}{\langle r_{ij} \rangle^2} - 1 \right]^{1/2}, \quad (19a)$$

where N_{ij}^{MC} is the number of MC steps for which both atoms ij of a pair are present in the shell Sh . N_{pair} is the sum over all pairs ij of N_{ij}^{MC} . The means are summed over the MC steps α

$$\langle r_{ij}^2 \rangle = \frac{1}{N_{ij}^{\text{MC}}} \sum_{\alpha} (r_{ij}^{(\alpha)})^2, \quad (19b)$$

$$\langle r_{ij} \rangle = \frac{1}{N_{ij}^{\text{MC}}} \sum_{\alpha} r_{ij}^{(\alpha)}. \quad (19c)$$

It is important to notice that the L_B coefficient is a dynamical one, either in MC or MD computations, and it depends on the number of steps performed. For an infinite number of steps, the full configurational space would be explored. L_B would necessarily characterize a diffusion of the pairs in the full accessible volume, with the limit value 0.375 whatever the potential energy. On the contrary, for a finite number of

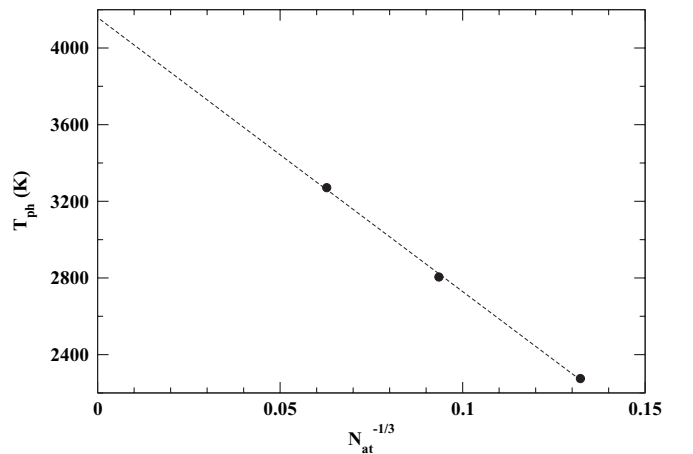


FIG. 7. Phase transition temperature T_{ph} (K), in terms of $N_{at}^{-1/3}$.

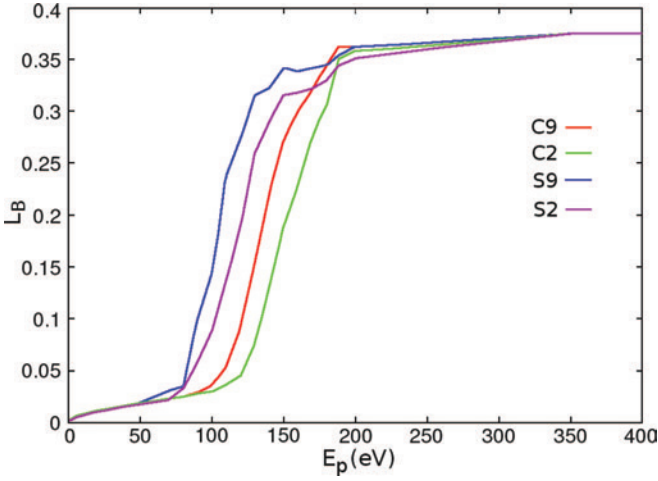


FIG. 8. Lindemann index L_B of a 432 atom cluster, for the core C (0 to 9 Å) and the surface shell S (9 to 12 Å), at 2×10^9 and 9×10^9 MC steps (indicated by: C2, C9, S2, and S9).

steps, a low value of L_B is expected in the solid phase where the movements remain local.

Then, the L_B coefficient calculated for the core (0 to 9 Å) and the surface shell (9 to 12 Å), after either 2×10^9 or 9×10^9 MC steps, are shown in Fig. 8. An increase of the number of MC steps simply translates the curves to lower energies of typically 20 eV, the diffusion being more extensive. Below 80 eV, the L_B coefficient is characteristic of local moves in a solid phase (≈ 0.05), both for the core and the surface atoms. At high energies (>200 eV), L_B converges to the value 0.375, which characterizes the full diffusion of a fluid phase in all the accessible volume. Between these energy ranges, the L_B curves indicate markedly that the surface shell melts more easily than the core. Indeed, the surface melting extends from 80 to 200 eV, while the core melting begins later, near 115 eV. The energy range between 80 and 200 eV, where this behavior is observed, matches the phase transition zone revealed above by the functions $\sigma(E_p)$, $U(T)$, or $C_V(T)$. In fact, the point where the σ curve increases steeply lies near the end of this range, at 185 eV. Hence, the present Lindemann-Berry analysis identifies the solid-liquid transition. Nevertheless, as the wall is located about 4 Å far from the $T = 0$ K cluster surface, the presence of a gas around the cluster is expected at this phase transition. Hence, three phases would be coexisting, the present calculation rather corresponding to a triple point. Sec. IV C states this hypothesis.

C. Volumic density analysis

In view of clarifying the presence of gas around the condensed cluster, the atomic density $n(R_1, R_2)$ was also monitored in successive shells between the radius R_1 and R_2 , to get a view of the accessible volume. The resulting density $n(r)$ is given in Fig. 9 at different energies. As for the Lindemann-Berry analysis the calculation is made at step 1 of the method in a narrow energy range, defined by a σ -linear bias.

In the solid phase (80 and 135 eV), or in the liquid phase ($E_p \geq 235$ eV), the presence of a gas of increasing density

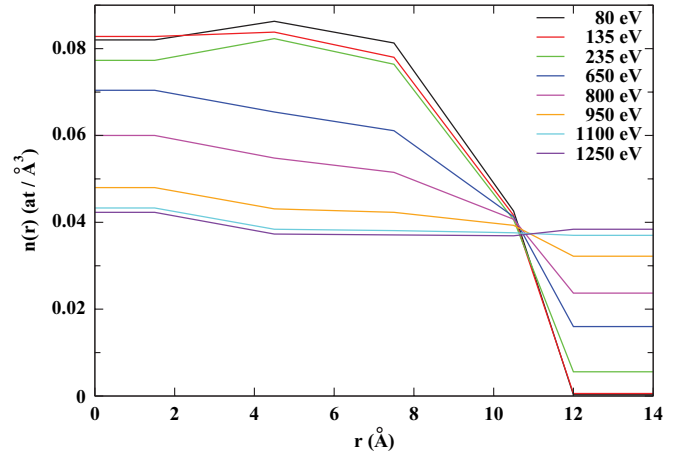


FIG. 9. Volumic density $n(r)$ of a 432 atom cluster at different energies: 80, 135, 235, 650, 800, 950, 1100, and 1250 eV.

(in terms of energy) can be observed between 12 and 14 Å. The surface defining the condensed cluster extends from 8 to 12 Å. Then, at energies corresponding to the solid-liquid phase transition zone, from 135 to 235 eV, a gas is actually present around the condensed cluster. The average number of atoms between 12 and 14 Å increases with energy, from 2.5 atoms (135 eV) to 24 atoms (235 eV). Hence, the solid-liquid phase transition discussed in Sec. IV B occurs at the triple point.

It is interesting to discuss the role of the wall radius choice, R_W , on this triple point. When R_W is smaller than a value $R_W^{TR1} \simeq 12$ Å, the cluster is expanded up to the wall when energy increases, expelling the gas phase before liquefaction can occur. Then, the solid phase is compressed without presence of a gas phase and the triple point is skirted round to upper pressures. In the opposite case where R_W is greater than a value R_W^{TR2} , the cluster vaporizes in the initial void next to the condensed phase, when energy increases, before liquefaction can occur, and the triple point is skirted round to lower pressures. The condition of the previous computations corresponds clearly to the intermediate case $R_W^{TR1} < R_W < R_W^{TR2}$ where the triple point is reached. The exact value of R_W and the total number of atoms determine the size of the cluster at this triple point and its precise characteristics.

A striking fact appears in Fig. 9. The position of the density fall, defining the surface of the condensed cluster, plotted in terms of energy, is a fixed point at 10.5 Å up to $E_p \simeq 1100$ eV, where the density becomes fully homogeneous inside the accessible sphere. This property can be interpreted by the fact that the liquid-gas equilibrium is kept until the critical point in pressure is reached at energies near 1100 eV. This situation can be explained following a reasoning similar to that developed above for the triple point. Indeed, when R_W is smaller than a given value R_W^{CR} , necessarily greater than R_W^{TR1} , the dilatation of the condensed phase is dominant over vaporization and the liquid phase reaches the wall, eliminating the gas phase. In the opposite case, $R_W > R_W^{CR}$, the vaporization is complete and the cluster surface moves towards the center of the sphere ($r = 0$), R_W^{CR} being necessarily smaller than R_W^{TR2} . The situation of Fig. 9 corresponds to $R_W \simeq R_W^{CR}$

and happens accidentally because of our combined choice of 432 atoms and of a volume $V = 11\,494 \text{ \AA}^3$. The latitude on the choice of the volume to remain in this particular situation is related to the fluctuation of the atom number in the external gaseous shell ($r \in [10.5-14 \text{ \AA}]$). This statistical fluctuation is estimated to be ± 16 atoms, corresponding to a latitude of about 4% for the total volume. The precise characteristics of this critical point depend now only on the total number of atoms.

D. Pressure study

Although the condition of a fixed wall, infinitely repulsive, is not well suited for a pressure study, the previous statements that the triple (respectively critical) point has been reached at energies around 185 eV (respectively 1100 eV) suggest to study the pressure curve in terms of temperature. The pressure is calculated for energies $E_p > 200$ eV by using the virial theorem

$$PV = N_{at}k_B T + \frac{1}{3}V_{ir,int}, \quad (20a)$$

with $V_{ir,int}$ being the internal virial term for N_{at} atoms in a volume V , expressed as

$$V_{ir,int} = \sum_i \vec{r}_i \cdot \vec{F}_i. \quad (20b)$$

The virial is computed by MC simulations with a canonical bias. For $E_p < 200$ eV, the accuracy of this method is completely insufficient, owing to large fluctuations of the virial. Then, the atomic density n_W near the wall has to be calculated to estimate the gas pressure using the perfect gas law. The correspondence between the two methods is very good in the range [200–400 eV], and the perfect gas law becomes more accurate at low temperatures (hence energies) for which densities are smaller.

In Fig. 10, the resulting pressure curve $P(T)$ is plotted in semi-logarithmic scale, at constant N_{at} and V . In the left part, the solid-gas equilibrium at temperatures $T < 2275$ K is observed, while for $T > 2275$ K the liquid-gas equilibrium is obtained. $T = 2275$ K is the triple point, with a pressure of 145 bars, corresponding to an average of 23 atoms vaporized in the shell between 12 and 14 \AA . The right part provides the pressure curve dilated up to 120 000 K, where a homogeneous fluid phase is completely settled with a density $n = 3.75 \times 10^{-2}$ atoms/ \AA^3 . Then, the gas obeys a nearly perfect gas law (the virial term being only 10% of the kinetic term). The homogeneity of the density is reached at an energy of about 1100 eV, corresponding to a temperature $T_c = 25\,200$ K, and a pressure $P_c = 156\,000$ bars. This is the critical point for the 432 atom system as discussed in Sec. IV C. For comparison, bulk iron has its critical temperature in the range 9250–9600 K, and its critical pressure in the range 8250–10 300 bars,³⁵ but we stress again that the Rydberg pair potential is not suited for a metal.

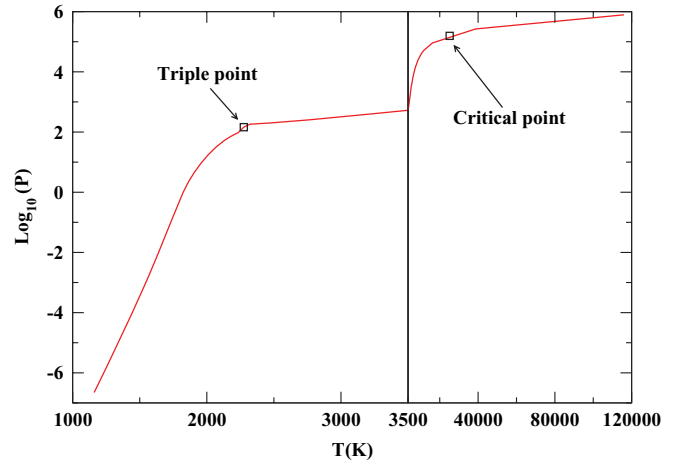


FIG. 10. Pressure curve $P(T)$ of a 432 atom cluster in semi-logarithmic scale. The left part depicts the zone of the triple point from 1000 to 3500 K. The right part is expanded up from 3500 to 120 000 K and shows the critical point.

V. DISCUSSION RELATIVE TO $\sigma(E_p)$ AND $g(E_p)$

A. Discussion of the $\sigma(E_p)$ variation

A quantitative discussion of the $\sigma(E_p)$ curve shape (Fig. 3) is outside the scope of this work, but we can qualitatively sketch the following process. At small energies, the configurational point is in the harmonic well of the minimum energy basin, so that σ is close to 644 (Eq. (7)). At higher energies, the negative anharmonicity (sub-harmonic potential, $n_i < 2$) and the appearance of excited basins, corresponding to local rearrangements of surface atoms, produce an increase of $\sigma(E_p)$. This increase of $\sigma(E_p)$ can be understood qualitatively assuming that each atom coordinate moves at the energy E_p in a mean potential roughly represented by the expression of Eq. (6a). The mean potential becoming sub-harmonic, n_i decreases below 2 and Eq. (6b) explains the increase of $\sigma(E_p)$. The phase transition zone is associated with the opening of a completely new basin corresponding to a global rearrangement, with a higher local energy minimum, but looser (i.e., with smaller frequency vibrations and more negative anharmonicities). This results in a sharp increase of $\sigma(E_p)$ (cf. Fig. 3(a)). The system can transit between the two types of basins thanks to the strong variation of a collective variable acting as an order parameter.

Then, a maximum is reached ($\sigma \simeq 1140$ for the 432 atom cluster, cf. Fig. 3(b)) and the subsequent decrease can be explained by the following simple and semi-quantitative model. At high energies, the system behaves as a gas of high volumic density, for which the attractive part of the potential becomes negligible and the repulsive part is dominant. Such a gas can be described again in the approximation of the mean potential perceived by each atom. Within these approximations, the virial term can be expressed through an excluded covolume $V_c(E_p)$ perceived by each atom in the presence of all the $N_{at} - 1$ other atoms, which corresponds to a van der Waals analysis.³⁶ At a qualitative level, Eq. (6a) can be evoked again: a stiff repulsive part corresponds to a strongly increasing n_i , much greater than 2, and to the fall of $\sigma(E_p)$.

At a more quantitative level, the covolume is modeled in the frame of the following argument. The repulsive part of the mean potential is assumed to be proportional to $1/(r - r_0)^{1/q}$, where r_0 is an atomic hard sphere radius, defining a covolume limit V_c^0 reached at infinite energy, and $1/q$ is an exponent linked to the stiffness of the potential repulsive part. The expression of the excluded covolume is easily deduced, once a reference energy E_p^0 (fixed at 1500 eV) has been chosen

$$V_c(E_p) = V_c^0 + \delta V_c^0 (E_p^0/E_p)^q. \quad (21)$$

$V_c^0 + \delta V_c^0$ is the covolume at the reference energy E_p^0 . The excluded covolume $V_c(E_p)$ is a decreasing function in terms of E_p . V_c^0 , δV_c^0 , and the exponent q are three parameters which have to be determined. The integrated DOS, $G(E_p)$, is the total accessible configurational space volume, proportional to $(V - V_c(E_p))^{N_{at}}$. The function $\sigma(E_p)$ can then be derived in this energy zone from the definition of Eqs. (5a) and (5c)

$$\sigma(E_p) = q(N_{at} - 1) \frac{\delta V_c^0}{V - V_c(E_p)} \left[\frac{E_p^0}{E_p} \right]^q - (q + 1), \quad (22)$$

where $V = 11\,494 \text{ \AA}^3$ is the total volume of the sphere. V_c^0 is estimated to be $0.161 \times V$ by assuming a cubic gathering of all the other $N_{at} - 1$ atoms at a distance defined by $a^* = -2$ (Eq. (15b)), i.e., in the highly repulsive part of the potential. It is now possible to reproduce the curve $\sigma(E_p)$ between 1500 and 6000 eV by fitting the two coefficients: $\delta V_c^0 = 0.217 \times V$ and $q = 1.46$. Then, the decrease of $\sigma(E_p)$ in this large energy range can be interpreted as due to the limitation of the accessible volume by the highly repulsive part of the potential, and its energy dependence.

B. Comparison of the σ -linear bias method with the Wang-Landau algorithm

In this section, a comparison of the σ -linear bias method with the WL algorithm is performed, focusing on the analysis of convergence properties, either speed or accuracy. Note that this comparison is relevant only for systems with continuous variables. Three intervals are investigated, ranging from the solid zone ([81, 171 eV], with a ratio $R_\alpha = 2.11$) to the liquid phase ([231, 1530 eV], $R_\alpha = 6.62$), through a phase transition zone ([150, 210 eV], $R_\alpha = 1.40$).

The WL algorithm is applied in the standard way,^{8,9} the modification factor decreasing according to the law $f \rightarrow \sqrt{f}$, from $f = 4$. The process is stopped when $f < 1 \times 10^{-6}$. At a given value f , the MC exploration starts from the value $g^{\text{WL}}(E)$ obtained at the end of the previous sequence. The first criterion of stopping is that the minimum value (H_{\min}) of the visiting energy histogram ($H^{\text{WL}}(E)$) becomes greater than $(1 - \epsilon_{\text{WL}}) \times \langle H^{\text{WL}}(E) \rangle$, with the common choice $1 - \epsilon_{\text{WL}} = 80\%$. According to the rule suggested by Zhou and Bhatt,³⁷ a second stopping criterion has to be simultaneously verified: $H_{\min} > K/\sqrt{\ln f}$ ($K = 10$ is chosen). This criterion guaranties that the ultimate fluctuations limit of $H^{\text{WL}}(E)$, $\simeq 1/\sqrt{\ln f}$, is smaller than H_{\min} . It turns out that this second condition was always checked before the first, and is, therefore, not restrictive. Finally, the number of MC steps,

N_f , is recorded for each f -sequence, as well as the number of round trips over the interval (ν_{RT}). At the end of the WL procedure, a simple Metropolis sampling is performed to check the convergence, as made in Ref. 10, the final $g^{\text{WL}}(E)$ being used as a bias $b(E)$.

It is observed that the first f -sequences give an estimated density $\ln[g_f^{\text{WL}}(E)]$ strongly biased compared to the final estimate, the discrepancy being progressively reduced when f approaches 1, as usual for the WL methods.³⁸ The use of a decorrelating parameter S , as proposed by Zhou and Bhatt,³⁷ allows to strongly diminish this bias, but at the price of an important slowing down of the WL process. For the phase transition range, MC samplings were tested with the decorrelating parameters $S = 1, 100$, and 1000 , the histogram being implemented only after S steps. The initial discrepancy of $\ln[g^{\text{WL}}(E)]$ to the converged density was strongly diminished with the increasing of S , at the beginning of the WL procedure. Nevertheless, the total number of steps of the f -cascade was heavily lengthened from 2.6×10^{10} for $S = 1$, to 9.4×10^{10} for $S = 100$, and 3×10^{11} for $S = 1000$. The decorrelating parameter is, therefore, practically useless and only the $S = 1$ value is discussed in the following.

The histogram resulting from the last Metropolis sequence (WL⁺), in the phase transition zone, is plotted in Fig. 11, compared to that of our algorithm (identical to curve (3) in Fig. 2(b)). It appears clearly that the last f -sequence of the WL algorithm has not converged properly, the population of the energy range above 185 eV being only 4.1×10^{-2} (corresponding to a logarithmic variation of 3.2 between the two half-zones).

The WL accuracy predicted by Zhou and Bhatt³⁷ or Lee *et al.*⁹ on $\ln[g_f^{\text{WL}}(E)]$ is $\simeq \sqrt{\ln f}$, i.e., $\simeq 10^{-3}$ for the histogram of Fig. 11. It is obvious that the convergence obtained is far from this theoretical limit and it is necessary to explain this bad convergence of the WL algorithm in the phase transition zone. Indeed, the previous estimate of the accuracy is a mathematical limit, reached only at an infinite number

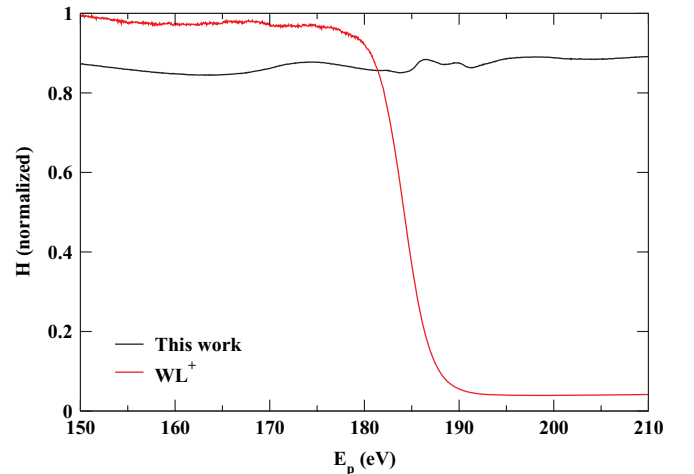


FIG. 11. Histograms obtained in the phase transition zone [150–210 eV] for a 432 atom system. “This work” refers to the curve already presented in Fig. 2(b) with 50×10^9 MC steps. “WL+” refers to the result of a Metropolis calculation (100×10^9 MC steps) following a Wang-Landau process (27×10^9 MC steps).

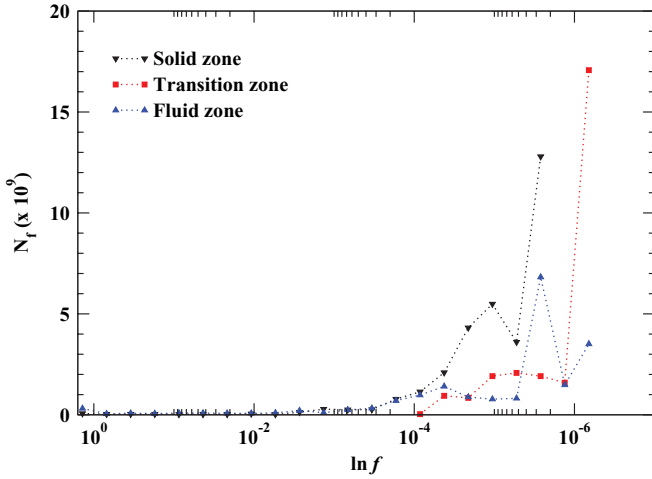


FIG. 12. Evolution of the number of MC steps with $\ln f$ for a 432 atom system. Three energy ranges are investigated: [81, 171 eV] (solid), [150, 210 eV] (transition zone), and [231, 1530 eV] (liquid).

of steps and is related to the fluctuations of the continuous self adaptation of the bias. In fact, the practical accuracy in the f -sequences turned out to be also limited by the number of round trips ν_{RT} , as in our method. The importance of the number of round trips has been discussed only by few authors,^{27,28} but not in the context of the WL algorithm. This role of ν_{RT} on the accuracy is observed strikingly in the phase transition zone ([150–210 eV]), and deserves now a careful analysis.

The result of Fig. 11 can be explained by studying the dynamical evolution of the WL histogram $H^{\text{WL}}(E)$ and of ν_{RT} . It is observed that the MC point wanders rapidly inside each half-range below and above the critical value (185 eV), but the crossing of the kinetic barrier needs several billion steps. This is the same phenomenon as observed in step 2 of our method. The number of MC steps necessary to make one round trip, N_{RT} , must therefore be compared to the number of steps of each f -sequence, N_f .

The evolution of N_f in term of $\ln f$ during the WL process is reported in Fig. 12 for all studied intervals. N_f increases strongly when f decreases towards 1 over all energy ranges, with successive stages, starting from $N_f \simeq 10^8$ steps at $f = 4$. For the phase transition zone, N_f increases particularly strongly for the last f -sequence, for which three round trips were recorded. This leads to an average number N_{RT} of 6×10^9 steps for a single round trip. For this reason, except for the last f -sequence, a single crossing of the kinetic barrier occurs.

Consequently, in all but the last f -sequence: (i) only one-half of the full zone is visited in a first part of the WL exploration; (ii) then, the second one is populated after the crossing; (iii) finally, the criteria for stopping are achieved, but no back turn to the first zone has occurred. Hence, no statistical equilibration is settled between the two half-zones. This means that a third necessary condition to stop an f -sequence is needed. The answer would be to enforce a sufficient number of round trips between the extremities of an interval.

The last f -sequence of 17×10^9 steps with a limited equilibration of three round trips deserves a particular analysis. A characteristic of the WL algorithm is the following logarithmic

relation between the density of states and the histogram for a bin i :³⁷

$$\ln [g_i^{\text{WL}}] = \ln(f) \cdot H_i^{\text{WL}} - \ln[\theta_i], \quad (23)$$

θ_i being the initial bias of the f -sequence. It results that the fluctuations of the histogram are amplified exponentially on $g^{\text{WL}}(E)$.

An estimate of these fluctuations between the two half-zones, in terms of the number of round trips is outside the scope of this work. However, at the end of the first round trip, it can be estimated that the difference of the accumulated histogram of visits between the two half-zones, $\delta \overline{H}_i^{\text{WL}}$, is about $2\epsilon_{\text{WL}} N_{RT}/N_b$, N_b being the number of bins in the interval. From the above equation, this implies a resulting discrepancy on the density of states, for $\nu_{RT} = 1$

$$\Delta \ln[\overline{g}^{\text{WL}}] = \ln(f) \cdot 2\epsilon_{\text{WL}} \frac{N_{RT}}{N_b}. \quad (24)$$

For the subsequent evolution of the fluctuations, it can be guessed that they diminish in terms of the number of round trips until their ultimate limit $\simeq 1/\sqrt{\ln f}$ is reached. Equation (24) can be applied to the case of the phase transition zone with $\ln f = 10^{-6}$, $\epsilon_{\text{WL}}=0.2$, $N_b = 1200$, and $N_{RT} = 6 \times 10^9$ steps. The predicted fluctuation is $\Delta \ln[\overline{g}^{\text{WL}}] = 2.0$, which must be compared to the observed variation of the logarithm in Fig. 11, i.e., 3.2. Equation (7) predicts, therefore, the correct order of magnitude of the fluctuations for a small number of round trips. This result is in contrast with the standard Metropolis approach we applied, for which the linear relation of Eq. (4) is available, leading to the following expression: $\Delta \ln[\overline{g}] \propto 1/\sqrt{\nu_{RT}}$, which is always smaller than 1 and decreases in terms of the number of round trips.

On the contrary, many round trips are recorded in the two other intervals (solid, liquid) allowing a much better convergence. Indeed, the histograms obtained during the last Metropolis stage (WL^+) are more flat. For example, in the fluid zone ([231, 1530 eV]), the histogram WL^+ oscillates between 0.7 and 1. Nevertheless, it indicates that a last correction has to be made according to Eq. (4) to get a flatness similar to that obtained with our method (cf. Fig. 2(a)). Consequently, this last Metropolis exploration is needed, even in these intervals, to improve convergence. It is interesting to apply Eq. (24) to the fluid zone with $\ln f = 10^{-6}$, $N_b = 13000$, and $N_{RT} = 3 \times 10^8$ steps, which gives $\Delta \ln[\overline{g}^{\text{WL}}] = 2.3 \times 10^{-2}$ at the end of the first round trip. This is much smaller than the discrepancy of $\simeq 0.3$ which has been observed after $H_T = 3.5 \times 10^9$ steps, corresponding to 12 round trips. Equation (24) applied with H_T instead of N_{RT} would give the correct order of magnitude $\Delta \ln[\overline{g}^{\text{WL}}] = 0.276$. This is probably the indication that the amplitude of fluctuations does not decrease after a small number of round trips in the WL process. It can be concluded that the convergence of the WL algorithm towards the ultimate limit $\simeq 1/\sqrt{\ln f}$ is very slow.

As can be noticed for the three intervals in Fig. 12, the first 15 sequences of the WL f -cascade are fairly fast, and give a rough approximation of $\ln [g(E)]$. The comparison with our method can be detailed for the fluid zone. These first 15 sequences totalize 3.5×10^9 steps. Unlike this, the σ -linear bias (step 1) provides directly a better guess with a similar

computational effort: 12 Gaussians converge in 5×10^8 steps each one. The last seven sequences of the WL f -cascade are much slower (14×10^9 steps, i.e., 80% of the calculation time). Only the last sequence of 3.5×10^9 steps determine the final accuracy of $\simeq 30\%$ with its 12 round trips, so that the six preceding sequences waste the computational time. In contrast to that, a single Metropolis sampling with the bias $b^{(1)}$ (step 2) of our method gives, in the same computational time of the last f -sequence, a corrected and converged DOS with an accuracy $\simeq 8\%$, although in a more pedestrian way.

VI. CONCLUSION AND PERSPECTIVES

A novel method is described in the present paper, providing the density of states of assemblies of several hundreds to several thousands of atoms. A local probing by a σ -linear bias has been shown to give a very good starting point, at a low computational cost, to obtain at the next steps flat histograms over a small number of large intervals, with only two subsequent iterations. Large clusters up to 4044 atoms, i.e., of a 4.2 nm diameter, have been treated. The present article reports a detailed analysis of the convergence, accuracy, and optimization of the method, using a fairly simple potential in order to be able to perform extensive MC sampling. This analysis is explained in Appendix A and gives informations of general interest for flat histogram methods. Section IV shows how this method, combined with various indicators, provides a powerful tool to analyze the thermodynamic state of a cluster. The curve $\sigma(E_p)$ obtained in such a way presents a characteristic shape linked to the physical and thermodynamical state of the cluster, particularly in the phase transition zone. It would deserve to be analyzed in the framework of a model valid from the solid to the fluid zone, more sophisticated than that of Sec. V A.

It remains nevertheless a difficulty of convergence below 100 eV, as mentioned at the end of Sec. III, due to an incomplete ergodicity in the solid phase. This has occurred in spite of the exploration by a flat histogram method of large intervals extended up to the phase transition zone at 185 eV. An elegant way of bypassing this problem, at no computer time expenses, would be to combine the present strategy with the parallel tempering/exchange replica approach,^{25,39,40} i.e., by exchanging the MC points between overlapping ranges. This combination would bring our method in kinship with the recent work of Kim *et al.*,²⁵ who have made parallel tempering between Gaussian Metropolis explorations.

A second problem of convergence has been described in Sec. II: the difficulty to cross the barrier of the phase transition, particularly from liquid to solid. This transition involves the transformation of a collective order parameter. One way to enhance the efficiency of this transition would be to use precisely such collective and concerted motions in the MC exploration of the configurational space,^{41,42} in view of simulating the real evolution of the system.

Finally, it has to be mentioned that we plan to explore the Fe system with the much more realistic MEAM potential, which takes into account all the environment of each atom. This potential being time consuming (three orders of magni-

tude slower than the Rydberg one), an EAM type potential is under development. This approximate potential, only three times slower than the Rydberg one, will be used in the nested potential approach,^{43,44} in which the fast EAM potential will alternate with the slower but much more sophisticated MEAM potential.

APPENDIX A: ANALYSIS OF THE LINEAR BIAS METHOD

Equation (11c) is assumed to be checked so that the histogram has a maximum E_{\max} . To optimize the determination of $\sigma(E_p)$, the parameters of the bias must be chosen so that the Gaussian histogram H is approximately centered at the middle of the interval, at $E_\alpha + \Delta_\alpha$. The index α characterizing the interval for quantities such as the slope B are omitted below.

The condition that the Gaussian is well drawn on the interval means that

$$\Delta = \frac{K_\Delta}{A^{1/2}} = K_\Delta \left[\frac{2E}{B - \sigma'} \right]^{1/2}, \quad (\text{A1})$$

with the parameter $K_\Delta \geq 1$ (typically around 3–4). Furthermore, we constrain Δ to be smaller than the value E of the interval center by a factor L_Δ , which is necessarily greater than 1, and must localize the histogram in a sufficiently well-defined zone of the σ curve, so that typically $L_\Delta \simeq 5 - 10$. These conditions give a lower bound for B

$$B - \sigma' = \frac{2K_\Delta^2 L_\Delta^2}{E} \quad (\text{A2a})$$

$$= \frac{2K_\Delta^2 L_\Delta^2}{\sigma} B_T \quad (\text{A2b})$$

$$\geq \frac{2L_\Delta^2}{\sigma} B_T, \quad (\text{A2c})$$

which is generally much smaller than the canonical value B_T owing to the denominator σ . Then, this condition allows a much larger choice than the canonical bias.

Now, the estimate of a maximum is a difficult task for a noisy curve. Then, it is made by cutting an arc at the top of the histogram, above $\exp(-K_M^2)$, with K_M being typically $1/2 - 1/3$. We develop $\ln[g/b^L]$ up to third order in energy around the maximum, taking into account a small asymmetry

$$H(\epsilon) = H_{\max} \times \exp \left[-A\epsilon^2 + \frac{D}{6}\epsilon^3 \right]. \quad (\text{A3a})$$

$$D = 2 \frac{B - \sigma'}{E^2} + \frac{\sigma''}{E}. \quad (\text{A3b})$$

D is the third-order coefficient. This provides the expansion

$$H(\epsilon) = H_{\max} \times \left(1 - A\epsilon^2 + \frac{D}{6}\epsilon^3 + \dots \right). \quad (\text{A3c})$$

The contribution of the third-order D term will be shown as not negligible. The average of the energy over the delimited arc, $\langle E \rangle_{\text{arc}}$, is computed (this quantity being the least noised).

Finally, keeping only the lower terms in K_M^2 , the maximum E_{\max} is expressed by

$$E_{\max} = \langle E \rangle_{\text{arc}} - D \frac{K_M^2}{12A^2}. \quad (\text{A4})$$

Then, the corresponding value of the bias at this maximum is (A being expressed by Eq. (11b))

$$\sigma_b^L(E_{\max}) = \sigma_b^L(\langle E \rangle_{\text{arc}}) + \sigma_{c1} + \sigma_{c2}, \quad (\text{A5a})$$

with the two contributions due to the third order,

$$\sigma_{c1} = -\frac{2K_M^2}{3} \frac{B}{B - \sigma'}, \quad (\text{A5b})$$

$$\sigma_{c2} = -\frac{K_M^2}{3} E \sigma'' \frac{B}{(B - \sigma')^2}. \quad (\text{A5c})$$

When B is sufficiently greater than σ' , σ_{c1} depends only on K_M . We get $-1/6$ for $K_M = 1/2$. The correction σ_{c2} is more involved and proportional to σ'' . The asymmetry D can be measured directly on the histogram arc by its third-order moment, but it is more accurate to compute an uncorrected σ curve, and from it to deduce the correction using Eqs. (A5b) and (A5c). We find that in some ranges, the correction σ_{c2} can reach 0.5–2, although it is generally much smaller.

Finally, an estimate of the error on $\langle E \rangle_{\text{arc}}$, due to the statistical fluctuations δH_i of the histogram number H_i , can be given. It will be made for simplicity linear least-square fits of the two sides of the arc defined by two values K_{Ma} and K_{Mb} . The accuracy can be expressed by

$$\delta(\langle E \rangle_{\text{arc}}) = \frac{2}{N_{\text{arc}} \frac{dH}{dE}} \sum_i \delta H_i, \quad (\text{A6})$$

where N_{arc} is the number of bins of the arc, and dH/dE is the slope of the histogram at the limit of the arc. The evaluation of δH_i can be made by remarking that the MC point wanders in a diffusion process from one side of the histogram to the other. Then, the counts accumulate around a bin i , leaves this region and returns to it after an average number of counts N_{RT} . We assume that, for a total number of MC steps H_T , δH_i is proportional to the square root of the number of round-trips ($\nu_{RT} = H_T/N_{RT}$, N_{RT} being the number of steps needed to go from one side to the other), and to the number of counts in the bin i during a round-trip ($N_{RT} H_i / H_T$)

$$\delta H_i = \left(\frac{N_{RT}}{H_T} \right)^{1/2} H_i, \quad (\text{A7})$$

Provided the system were out of the phase transition zone, it has been checked on various intervals Δ_E that N_{RT} obeys approximately a diffusion process with a diffusion constant D_E on the energy scale

$$(\Delta_E)^2 = D_E N_{RT}. \quad (\text{A8})$$

Choosing $(\Delta_E)^2 = 4/A$, assuming in Eq. (A6) that the δH_i are randomly distributed, and writing in Eq. (A7) that $H_i = H_T \cdot K_M / N_{\text{arc}}$, the slope dH/dE at the limit of the arc is expressed as

$$\frac{dH}{dE} = \frac{2}{\sqrt{\pi}} \delta E H_T K_M A,$$

δE being the width of the energy bin. Thus,

$$\delta(\langle E \rangle_{\text{arc}}) = 2 \left[\frac{\pi}{D_E H_T} \right]^{1/2} \frac{1}{\delta E (N_{\text{arc}} A)^{3/2}}. \quad (\text{A9})$$

Finally, $N_{\text{arc}} = 2K_M/\delta E(A)^{1/2}$ is eliminated and multiplied by B to obtain

$$\delta\sigma(\langle E \rangle_{\text{arc}}) = \left[\frac{\pi \delta E}{2D_E H_T K_M^3} \right]^{1/2} [2E]^{3/4} \frac{B}{[B - \sigma']^{3/4}}. \quad (\text{A10})$$

This result shows that $\delta\sigma$ is minimized for $B = 4\sigma'$, which is inside the condition of Eq. (11c). Reducing δE is also a way to improve the accuracy of $\delta\sigma$. Equation (A10) applied to the case $E = 100$ eV, $\delta E = 0.1$ eV, $D_E = 5 \times 10^{-5}$ eV², $K_M = 1/2$, $H_T = 10^9$ counts, and $B = 10$ eV⁻¹ gives $\delta\sigma = 0.8$, which is the correct magnitude of the observed fluctuations.

APPENDIX B: CORRECTION σ_{cd} DUE TO DISCRETIZATION

For a bin of extension δE , centered in E , the histogram H is populated according to the ratio

$$H = \frac{\delta G}{\delta G_b}, \quad (\text{B1})$$

where $\delta G = G(E + \delta E/2) - G(E - \delta E/2)$ is the number of states in the bin for the exact density of states (respectively δG_b for the bias number of states). The correcting factor q_g due to the non-zero extension δE can be defined as

$$\delta G = \delta E \cdot g \cdot q_g, \quad (\text{B2})$$

(respectively b and q_b for the bias). Obviously, if $\delta E \rightarrow 0$, $q_g \rightarrow 1$. With the previous two relations, the correction of σ due to the discretization can be derived

$$\sigma_{cd} = -E \frac{d}{dE} \ln \left(\frac{q_g}{q_b} \right). \quad (\text{B3})$$

δG (respectively δG_b) can be expressed as

$$\delta G = g(E) \int_{-\frac{\delta E}{2}}^{+\frac{\delta E}{2}} du \exp \left[\int_E^{E+u} dv \frac{\sigma_g(v)}{v} \right]. \quad (\text{B4})$$

Approximating the exponential in terms of u by a constant integrand in terms of v , we get the result for q_g in terms of the parameter x_g

$$q_g = \frac{sh(x_g)}{x_g}, \quad (\text{B5a})$$

$$x_g = \sigma_g(E) \frac{\delta E}{2E}, \quad (\text{B5b})$$

(respectively q_b , x_b , $b(E)$). This expression neglects higher order terms in $1/\sigma_g$. To avoid a too important correction, $x \lesssim 1$ must be chosen, and in this case

$$q = \frac{sh(x)}{x} = 1 + \frac{x^2}{6} + \dots$$

In this condition, using Eq. (13), the correction is expressed as

$$\sigma_{cd} \simeq \frac{2\sigma_H(E)}{3b(E)} x_b^2. \quad (\text{B6})$$

This quantity becomes negligibly small at the extremum of an histogram or for a quasi-flat histogram ($\sigma_H \rightarrow 0$).

- ¹J. R. Barker, J. Phys. Chem. **91**, 3849 (1987).
- ²L. Bonnet and J. C. Rayez, Chem. Phys. Lett. **338**, 385 (2001).
- ³F. Calvo, P. Parneix, and F. X. Gadéa, J. Phys. Chem. A **110**, 1561 (2006).
- ⁴F. Calvo and P. Parneix, J. Phys. Chem. A **113**, 14352 (2009).
- ⁵B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
- ⁶J. Lee, Phys. Rev. Lett. **71**, 211 (1993).
- ⁷P. M. C. de Oliveira, T. J. P. Penna, and H. G. Herrmann, Eur. Phys. J. B **1**, 205 (1998).
- ⁸F. G. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
- ⁹H. K. Lee, Y. Okabe, and D. P. Landau, Comput. Phys. Commun. **175**, 36 (2006).
- ¹⁰R. Faller and J. J. de Pablo, J. Chem. Phys. **119**, 4405 (2003).
- ¹¹P. Poulain, F. Calvo, R. Antoine, M. Broyer, and Ph. Dugourd, Phys. Rev. E **73**, 056704 (2006).
- ¹²G. Adjanor, M. Athènes, and F. Calvo, Eur. Phys. J. B **53**, 47 (2006).
- ¹³C. Desgranges and J. Delhommelle, J. Chem. Phys. **130**, 244109 (2009).
- ¹⁴M. Basire, P. Parneix, and C. Calvo, J. Chem. Phys. **129**, 081101 (2008).
- ¹⁵M. Basire, P. Parneix, C. Calvo, T. Pino, and Ph. Bréchnignac, J. Phys. Chem. A **113**, 6947 (2009).
- ¹⁶T. L. Hill, Nano Lett. **1**, 111 and 273 (2001).
- ¹⁷R. S. Berry and B. M. Smirnov, Phys. Usp. **52**, 137 (2009).
- ¹⁸M. Eryürek and M. H. Güven, Eur. Phys. J. D **48**, 221 (2008).
- ¹⁹A. Campa, T. Dauxois, and S. Ruffo, Phys. Rep. **480**, 57 (2009).
- ²⁰V. Heresanu, C. Castro, J. Cambedouzou, M. Pinault, O. Stephan, C. Reynaud, M. Mayne-L’Hermite, and P. Launois, J. Phys. Chem. C **112**, 7371 (2008).
- ²¹J. Kim, J. E. Straub, and T. Keyes, Phys. Rev. Lett. **97**, 050601 (2006).
- ²²B. J. Lee, M. I. Baskes, H. Kim, and Y. K. Cho, Phys. Rev. B **64**, 184102 (2001).
- ²³B. J. Lee, Acta Mater. **54**, 701 (2006).
- ²⁴D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).
- ²⁵J. Kim, T. Keyes, and J. E. Straub, J. Chem. Phys. **132**, 224107 (2010).
- ²⁶M. S. Daw and M. I. Baskes, Phys. Rev. B **29**, 6443 (1984).
- ²⁷S. Trebst, D. A. Huse, and M. Troyer, Phys. Rev. E **70**, 046701 (2004).
- ²⁸C. R. A. Abreu, J. Chem. Phys. **131**, 154113 (2009).
- ²⁹M. Wautelet, J. Phys. D: Appl. Phys. **24**, 343 (1991).
- ³⁰K. K. Nanda, S. N. Sahu, and S. N. Behera, Phys. Rev. A **66**, 013208 (2002).
- ³¹A. Safaei, M. A. Shandiz, S. Sanjabi, and Z. H. Barber, J. Phys. Condens. Matter **19**, 216216 (2007).
- ³²H. M. Lu, P. Y. Li, Z. H. Cao, and X. K. Meng, J. Phys. Chem. C **113**, 7598 (2009).
- ³³Y. Q. Zhou, M. Karplus, K. D. Ball, and R. S. Berry, J. Chem. Phys. **116**, 2323 (2002).
- ³⁴J. G. O. Ojwang, R. van Santen, G. J. Kramer, A. C. T. van Duin, and W. A. Goddard, J. Chem. Phys. **129**, 244506 (2008).
- ³⁵M. Beutl, G. Pottlacher, and H. Jäger, Int. J. Thermophys. **15**, 1323 (1994).
- ³⁶B. C. Eu, Phys. Chem. Chem. Phys. **9**, 6171 (2007).
- ³⁷C. Zhou and R. N. Bhatt, Phys. Rev. E **72**, 025701 (2005).
- ³⁸A. D. Swetnam and M. P. Allen, Phys. Chem. Chem. Phys. **11**, 2046 (2009).
- ³⁹U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
- ⁴⁰D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).
- ⁴¹J. J. de Pablo, M. Laso, J. I. Siepmann, and U. W. Suter, Mol. Phys. **80**, 55 (1993).
- ⁴²C. A. Hixson, J. P. Benigni, and D. J. Earl, Phys. Chem. Chem. Phys. **11**, 6335 (2009).
- ⁴³L. D. Gelb, J. Chem. Phys. **118**, 7747 (2003).
- ⁴⁴J. D. Coe, T. D. Sewell, and M. S. Shaw, J. Chem. Phys. **130**, 164104 (2009).